

Computing and Artificial Intelligence

<https://ojs.acad-pub.com/index.php/CAI>



2024 VOLUME 2 ISSUE 1
ISSN: 3029-2786 (Online)



1



Editorial Team

Editor-in-Chief

Shaohua Wan

University of Electronic Science and Technology of China
China

Associate Editor

Vijayakumar Varadarajan

La Trobe University
Australia

Editorial Board Members

Syed Muzamil Basha

REVA University
India

Junyu Xuan

University of Technology Sydney
Australia

Abdullah Ayub Khan

Bahria University
Pakistan

Ata Jahangir Moshayedi

Jiangxi University of Science and
Technology
China

Mingchu Li

Jiangxi Normal University
China

Jie Zhang

Xi'an Jiaotong-Liverpool University
China

Parikshit N Mahalle

Vishwakarma Institute of Technology
India

Anjum Razzaque

Western Illinois University
United States

William Cheng-Chung Chu

Fuyao Institute of Science and Technology
Taiwan

Umesh C. Pati

National Institute of Technology
India

Jun Ye

Ningbo University
China

Jiakai Wang

Zhongguancun Laboratory
China

Grigorios N. Beligiannis

University of Patras
Greece

Chin-Shiuh Shieh

National Kaohsiung University of Science
and Technology
Taiwan

Janusz Kacprzyk

Systems Research Institute of the Polish
Academy of Sciences
Poland

Esma Aïmeur

University of Montreal
Canada

Marcin Paprzycki

Polish Academy of Sciences
Poland

Inam Ullah

Gachon University
Korea

Shashi Kant Gupta

Eudoxia Research University
United States

Yi-Chung Hu

Chung Yuan Christian University
Taiwan

Neelamadhab Padhy

Gandhi Institute of Engineering and
Technology
India

Omar Cheikhrouhou

University of Sfax
Tunisia

Kibum Kim

Hanyang University
Korea

Maki K. Habib

The American University in Cairo
Egypt

Yousef Daradkeh

Prince Sattam bin Abdulaziz University
Saudi Arabia

Hao Ying

Wayne State University
United States

Pradeep Kumar Mallick

Kalinga Institute of Industrial Technology
India

Salah Bourennane

Aix Marseille Université
France

Ibrahim A Hameed

Norges Teknisk-Naturvitenskapelige
Universitet
Norway

Alberto Gotta

Consiglio Nazionale delle Ricerche
Italy

Shuo-Tsung Chen

Tunghai University
Taiwan

Saeed Alsamhi

Ibb University
Yemen

Chao Zhang

Shanxi University
China

Mohammad Khishe

Imam Khomeini Marine Science University
Iran

Pushpendu Kar

University of Nottingham Ningbo China
China

Sudan Jha

Kathmandu University
Nepal

Mohammed Baz

Taif University
Saudi Arabia

Xin Liu

Lappeenranta-Lahti University of
Technology
Finland

Ray-I Chang

National Taiwan University

Taiwan

Cihan Karakuzu

Bilecik Şeyh Edebali University

Turkey

Volume 2 Issue 1 • 2024

Computing and Artificial Intelligence

Editor-in-Chief

Prof. Shaohua Wan

University of Electronic Science and Technology of China, China



Computing and Artificial Intelligence

<https://ojs.acad-pub.com/index.php/CAI>

Contents

Articles

- 1 Medical image classification using a quantified hazard ratio and a multilayer fuzzy approach**
Kishore Kumar Akula, Monica Akula, Alexander Gegov
- 16 Identifying voices using convolution neural network models AlexNet and ResNet**
Abdulaziz Alhowaish Luluh, Muniasamy Anandhavalli
- 31 A comparison of cepstral and spectral features using recurrent neural network for spoken language identification**
Irshad Ahmad Thukroo, Rumaan Bashir, Kaiser Javeed Giri
- 45 Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse software learning needs with a single document knowledge base**
Yuchen Wang, Yin-Shan Lin, Ruixin Huang, Jinyin Wang, Sensen Liu
- 70 The AI spectrum under the doctrine of necessity: Towards the flexibility of the future legal praxis**
Lambrini Seremeti, Ioannis Kougias
- 77 Utilizing emotion recognition technology to enhance user experience in real-time**
Yuanyuan Xu, Yin-Shan Lin, Xiaofan Zhou, Xinyang Shan

- 90 The electronic knowledge publishing based on intellect modeling**
Konstantin M. Golubev
- 104 Is it possible to detect cribriform adverse pathology in prostate cancer with magnetic resonance imaging machine learning-based radiomics?**
Hüseyin Bıçakçioğlu, Sedat Soyupek, Onur Ertunç, Avni Görkem Özkan, Şehnaz Evirmeler, Tekin Ahmet Serel
- 105 From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI**
Zarif Bin Akhtar
- 136 Ethical considerations of AI in financial decision**
Yifei Wang

Reviews

- 144 Review of quantum algorithms for prediction of hazardous asteroids**
Priya Pareshbhai Bhagwakar, Chirag Suryakant Thaker, Hetal A. Joshiara
- 153 Revolutionizing Neurosurgery and Neurology: The transformative impact of artificial intelligence in healthcare**
Habib Hamam

Medical image classification using a quantified hazard ratio and a multilayer fuzzy approach

Kishore Kumar Akula^{1,*}, Monica Akula², Alexander Gegov^{3,4}

¹ Statistics eTeachers Group, Royal Statistical Society, North York, Ontario M3C 2Z1, Canada

² Department of Neuroscience, McMaster University, Hamilton, Ontario L8S 4L8, Canada

³ School of Computing, University of Portsmouth, Portsmouth PO1 3HE, United Kingdom

⁴ English Faculty of Engineering, Technical University of Sofia, 1756 Sofia, Bulgaria

* **Corresponding author:** Kishore Kumar Akula, kishorekumarakula@gmail.com

CITATION

Akula KK, Akula M, Gegov A.
Medical image classification using a
quantified hazard ratio and a
multilayer fuzzy approach.
Computing and Artificial
Intelligence. 2024; 2(1): 450.
<https://doi.org/10.59400/cai.v2i1.450>

ARTICLE INFO

Received: 2 January 2024

Accepted: 29 January 2024

Available online: 18 February 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence
is published by Academic Publishing
Pte. Ltd. This work is licensed under
the Creative Commons Attribution
(CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: We previously developed two AI-based medical automatic image classification tools using a multi-layer fuzzy approach (MFA and MCM) to convert image-based abnormality into a quantity. However, there is currently limited research on using diagnostic image assessment tools to statistically predict the hazard due to the disease. The present study introduces a novel approach that addresses a substantial research gap in the identification of hazard or risk associated with a disease using an automatically quantified image-based abnormality. The method employed to ascertain hazard in an image-based quantified abnormality was the cox proportional hazard (PH) model, a unique tool in medical research for identifying hazard related to covariates. MFA was first used to quantify the abnormality in CT scan images, and hazard plots were utilized to visually represent the hazard risk over time. Hazards corresponding to image-based abnormality were then computed for the variables, ‘gender,’ ‘age,’ and ‘smoking-status’. This integrated framework potentially minimizes false negatives, identifies patients with the highest mortality risk and facilitates timely initiation of treatment. By utilizing pre-existing patient images, this method could reduce the considerable costs associated with public health research and clinical trials. Furthermore, understanding the hazard posed by widespread global diseases like COVID-19 aids medical researchers in prompt decision-making regarding treatment and preventive measures.

Keywords: cox proportional hazards model; hazard ratio; CT scans; fuzzy system; survival analysis

1. Introduction

Physicians rely heavily on computed tomography (CT) scans to detect diseases like cancers. They spend much of their time observing one image to decide if the patient has cancer, and if so, the stage of the cancer. Qualitative analysis of an abnormality in a diagnostic image is not always a robust method of analyzing an abnormality and may be contribute to reduced inter-rater reliability. However, when the abnormality can be quantified as a percentage, this both improves consistency in the assessments performed by different healthcare practitioners, while also providing a more accurate method of evaluating the aberration.

Our previous studies described methods for quantification of the abnormality in images [1]. A novel method, an AI-based medical image classification tool using a multi-layer fuzzy approach (MFA), was introduced in the first study. The second study focused on developing a more robust version of the MFA method using the many-to-many relation [2] (MCM) (manuscript under review) to find the abnormality in the

images, which is the disease present in the objects in the images. In the first study, a simple process from cognitive science known as assessment was used, which involves comparison of a normal image with an abnormal image. In this process of comparison, the structural similarity index (SSI) between two images is the similarity among the images [3]. When this similarity in percentage form is subtracted from 100, the calculation quantifies the abnormality in the abnormal image. In our second study using MCM, the comparison was made more robust, such that instead of considering one image, multiple normal images were compared with a single abnormal image, and all the similarity scores were averaged to obtain the abnormality in the image. This process was continued for all of the abnormal images, which was a more robust and accurate method of analyzing the images. The images were then classified, which was performed using multilayer fuzzy systems, computational intelligence rules, computer vision from AI and manual testing tools.

A literature search was conducted using the key words for any studies like MFA or MCM that is to convert the image based abnormality to quantity or finding the hazard ratio corresponding to the image based abnormality, but no similar studies were found [4,5].

1.1. The rationale of the study

Currently, there is limited research on using diagnostic image assessment tools, particularly methods that can be used for small and large data sets like the MFA and MCM methods, for prognostic applications, such as prediction of mortality risk. Moreover, following the inspection of patient images, such as CT scans, the images are not typically used for subsequent research endeavours. In the present study, a novel method was developed using conversion of image-based disease severity quantity into the hazard ratio through the use of an Artificial Intelligence (AI) medical image classification-based multilayer fuzzy approach (MHM). In this method, the hazard ratio corresponding to an image-based disease severity quantity was found using an AI medical image classification-based multilayer fuzzy approach, and survival analysis, a domain of biostatistics, as well as concepts from MFA [1], and MCM, a modified version of MFA [2], were used. This study posits that quantifying abnormalities in these images and determining the associated hazard within a group of individuals could substantially contribute to public health efforts and research initiatives to optimize resource allocation, substantially improve treatment outcomes that increase survival and reduce the expenses incurred during the conduct of clinical trials involving patient studies. Furthermore, the simplicity of incorporating basic software code in this study renders it easily applicable with existing technologies. Ultimately, the utilization of images acquired for the individualized treatment of patients can extend the scope of the study to encompass the health of larger populations, countries, or even the entire global patient demographic. This extension arises from the study's classification of patients based on the hazard or risk associated with abnormalities resulting from various diseases.

1.2. Aims of the study

1.2.1. Primary aim

The primary aim of the study is the following:

- a) To quantify the abnormality in the form of images present in the objects of images.
- b) Finding the hazard ratio corresponding to this quantified abnormalities in a set of images as well as to classify the data into some sub-groups and to check the effect of the hazard due to the abnormality.
- c) To classify the abnormality based on the hazard or risk due to the abnormality.

1.2.2. Secondary aim

The secondary aim of the study is to apply the concepts of the primary aim to a medical CT scan image data set taken to diagnose the lung cancer, as described in section 1.1. The specific components of the second aim are as follows:

- a) To quantify the cancer or abnormality present in image form in a CT scans data set.
- b) To find the hazard ratio corresponding to the above-quantified abnormality of the data set, as well as to classify the data into some sub-groups and to check the effect of the hazard present in the images due to abnormality.
- c) To classify the abnormality based on the hazard ratio due to the abnormality present in the CT scans of the data.

2. Materials and methods

2.1. Materials

2.1.1. Participants

In the dataset [6] we considered, nothing was known about the age, gender, participants and ethnicity of the patient. The only information available was a confirmed lung cancer, and did not include the time of event, status or smoking status. To find the hazard ratios using the Cox PH model, few more variables were needed. That is in order to make the dataset a survival data set, some variables such as ‘time’, ‘status’ were simulated and for the subgroup analysis ‘smoking-status’, ‘gender’, and ‘age’ were simulated.

2.1.2. Data set and data dictionary used in the current MHM study

The number of CT scans in the data set used in this study was 67 [6]. Among the CT scans, images of the right lungs were extracted from the CT scan. The right lung was chosen at random for study due to the noise in the images. Additionally, the normal image data set was a small data set with 20 images, similar to the MCM study [2].

2.1.3. Variable description

First, after finding the abnormality in the different images, they were classified as per the stages of abnormality (**Table 1**). Stage 1 has less abnormality than the other stages and more normality, and stage 4 has greater abnormality and less normality, such that from stages 1 to 4, the normality decreases or the abnormality decreases. Furthermore, the elements at each stage represent the normality percentage in the images. In addition, the variable ‘time’ is the time to event, ‘status’ is the occurrence of disease or death, ‘gender’ is whether the patient is male or female, and ‘smoking’

is the smoking-status.

2.1.4. Data types

The variables ‘stage 1 to 4’ and time were continuous variables. Stages were in percentages and time was in time units, which can be days, months or years, while the rest were categorical variables. ‘Status’ can be categorized as live or dead, ‘gender’ was female (1) or male (0), and similarly for non-smoking or smoking- status.

2.1.5. Survival analysis time data

The survival analysis [7,8] data only has information on abnormalities in the form of CT scans. However, to find the hazard due to the cancer in that particular study, area or group, we need information on a few more variables, so these variables were simulated, because in order to find the hazard ratios, these variables are needed. Here, we included gender, age and smoking-status, and by using these variables the method was developed.

Survival time is defined as time until retrieval or time until an end point whether it is medical or industrial [9], and can also be called the ‘lifetime’. In the current data set, it is the time to disease (cancer), disorder occurrence or death. The units of the time in this study can be days, months or years.

The variable, ‘status,’ is the status of the patient, which is the death of the patient. It is a binary variable, such that ‘0’ denotes being alive or not affected, and ‘1’ denotes being affected by abnormality or cancer or death. The next variable added was ‘gender’, which is male or female, a binary variable. Lastly another variable ‘smoking-status’ was added, which was also a binary variable.

2.2. Methods

2.2.1. Test statistic

The test statistic used for the Cox PH model is the chi-square test for the p -value assessment of each variable or covariate used in the model.

2.2.2. Finding the risk or hazard due to the quantified abnormality or disease

A hazard represents any factor with the capability to cause damage, harm, or negative health consequences to an object or individual. The hazard ratio is also known as the relative risk [10]. The ratio compares the risk of disease or death between the exposed and unexposed populations, like, for example, the ratio of the number of patients in a country with lung cancer and the population of that country. The baseline hazard is the hazard when all the covariates are zero or at their reference levels. In medical sciences and epidemiology, the hazard ratio plays a very important role in helping assess potential patient outcomes and classify patients on the basis of the hazard. It explains how much risk is associated with a certain disease, drug or a habit like smoking. To find the hazard or risk due to the quantified abnormality or disease, the cox PH was used, which falls under the domain of survival analysis, a branch of statistics. The primary emphasis of survival analysis is on the time until an event, which is death, or until the disease occurs. This could be the time until a patient experiences a relapse or the time until a machine fails, depending on the application of the study.

2.2.3. Cox proportional hazard (PH) model to find the hazard ratios of variables

The cox PH semi-parametric regression model has been widely used in the medical and clinical fields, as well as the industry [11,12]. Since the Cox PH model is a statistical model, certain procedures used with statistical models, like hypothesis testing, need to be implemented when calculating the coefficients obtained when Cox PH model is used.

Normally, in clinical/medical, epidemiology, and industry studies, the Cox PH model is used in the context of treatment or intervention, age, gender, disease stage or severity, socioeconomic status, smoking-status, and other important factors. In addition to the mandatory variables to use or to calculate hazard ratios are time to event, and status of the disease (please clarify this sentence). In the current study, the covariates of the Cox PH model are the abnormalities in the images converted into percentages using MCM, a novel approach to find the hazard ratio, that provides information on mortality risk, corresponding to the abnormality obtained using MFA and MCM.

2.2.4. Steps in using Cox PH model

The steps in using Cox PH model to find the hazard due to image based quantified abnormality are the following:

Step 1. The data was prepared to use with the Cox PH model, using the CRAN-R software and with the proper libraries, which yielded the hazard ratios, p -values, and confidence intervals for hazard ratios.

Step 2. Next, the PH model was tested for accuracy, which was done by checking the statistical significance of the coefficients. The p -value associated with each beta coefficient helps assess the statistical significance of the corresponding covariate. A small p -value (typically less than a chosen significance level, e.g., 0.05) indicates that the covariate is statistically significant in predicting the hazard, suggesting that it is likely not due to random chance.

Step 3. The proportional hazards assumptions were checked. The Cox PH model was based on some assumptions [10–14], and when the model is used, the assumptions are to be checked. If the assumptions are not satisfied, then it means that the mode was not fit properly. At the same time, for all data sets, there is no need that all assumptions are to be checked. In this study, the only key assumptions that were checked were:

1) The hazard ratio is presumed to be constant over time, which means that the ratio of the hazards for any two persons or patients or commodities is constant over time.

2) The independence assumption, which states that the observations in the dataset are expected to be independent. This means that the incidence or non-incidence of an event for one patient, person or commodity does not provide information about the occurrence or non-occurrence of an event for any other subject.

2.3. Statistical procedures used in MHM

2.3.1. The software used

The software used in the current study were Python with Anaconda as the backend and spider as the frontend. Python was utilized to compare images and to acquire the similarity indices as described in our previous MCM study [2] using

OpenCV from AI. Python was also used to plot some graphs. Next, CRAN-R was used to calculate the hazard ratios using Cox PH model and to plot some additional graphs.

2.3.2. Data cleaning

The most important step in the MHM method was ensuring that when two images were compared, only the object, lungs, were compared. Parts of the image other than lungs were avoided, so that only the normal lung was compared with the abnormal lung in order to avoid any noise when obtaining the similarity index.

2.3.3. Statistical tests and tools used

As the secondary aim involves the application of the current study MHM to CT scans of lung cancer, in order to find the hazard due to the abnormality in CT scans, the Cox PH model was used. The hazard ratios were obtained by using the COX PH model on the abnormality data. First, the global model was tested for statistical consistency followed by the coefficients. Lastly, the confidence intervals were examined to check the model for goodness of fit or not.

2.3.4. Method to establish the primary aim

All the steps in the primary aim are shown in the schema in **Figure 1**. These steps are the same as the MFA or MCM method [1,2] for obtaining the fuzzy set. The general form of the fuzzy set obtained is {Patients’ ID, abnormality score}. Next, the Cox PH model was used together with some basic statistics to get hazard ratios and the survival probabilities on graphs for the variables. The rest of the method from the fuzzy logic steps until the end of the method was similar to the MCM method.

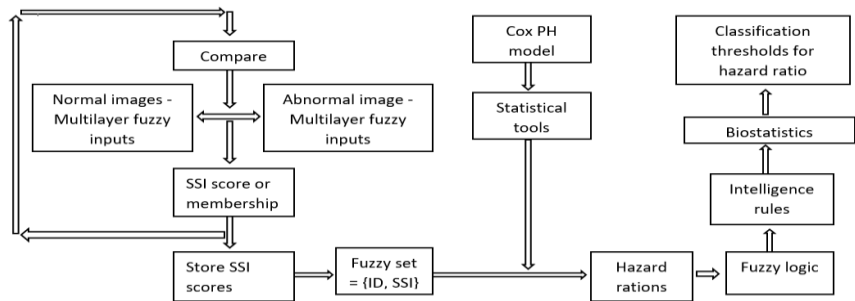


Figure 1. The schema for the study MHM.

2.3.5. Method to establish the primary aim 1.2.1 (a)

To quantify the abnormality in the form of images, MCM [2] was used. That is, a few normal images were compared with the abnormal image to get the SS followed by taking the mean of all the SSs to get the final SS of the abnormal image. The abnormality was obtained by subtracting the normality in percentage form from 100, which was the process used to quantify the abnormality in the form of an image.

2.3.6. Method to establish the primary aim 1.2.1 (b)

The data classified in the above section was used with the Cox PH model to calculate the hazard ratios as discussed in section 1.4.4, and three steps were used with the Cox PH model.

Step 1: The model is $h(t/X) = h_0(t)\exp(\beta_1\text{Stage1} + \beta_2\text{Stage2} + \dots + \beta_k\text{Stagek})$, where the required hazard ratios are $\exp(\beta_k)$. To get the specific β_k the data has to be used, which was carried out in the subsequent sections using CRAN-R software.

Step 2: The global statistical tests, the likelihood ratio test, Wald test, and logrank tests were used to check for the significant overall association between the variables. The p -value for each variable was used to check if the coefficients occurred by chance or if there is statistical consistency depending on the obtained p -value being less than or greater than a standard value (0.05).

Step 3: The Cox PH model’s assumptions were checked as mentioned in section 1.4.7 (b) above. The first assumption is the constant hazard ratio (proportionality), which states that the hazard ratio should remain constant over time. A rule of thumb to prove the above is that the hazard curves for the groups should be proportional and cannot cross each other [9]. The second assumption is that the observations in the dataset are expected to be independent, and in the currently used data set, neither the patients nor their data are related to each other. Confidence intervals are discussed in detail numerically while discussing the secondary aim.

2.3.7. Methods to establish the primary aim 1.2.1 (c)

In the previous section, the hazard ratio was obtained as $\exp(\beta_i)$, the hazard ratio or the relative risk based classification is as follows [8–11]:

If $\exp(\beta_i) < 1$, then the hazard or risk is reduced.

If $\exp(\beta_i) = 1$, then there is no effect on hazard or risk.

If $\exp(\beta_i) > 1$, then there is increase in hazard, where ‘exp’ stands for the exponentiation of β_i .

3. Results

3.1. Methods to establish the secondary aim 1.2.2 (a)

The secondary aim is the application of the primary aim to a data set of CT scans taken for the diagnosis of the lung cancer. To quantify the abnormality in the form of CT scans or images of the data set, that is, to quantify the abnormality or cancer in the form of images present in the CT scans, the abnormality present in the lung in the CT scan images of the data set was quantified by using the MCM method [2] and the method presented in the primary aim 1.2.1 (a).

In addition, a few more variables were simulated for the full application of the primary aims. The new sample data set for stage 4 lung cancer is show in **Table 1**. The only variable that was not simulated was ‘stages’. This variable consisted of the abnormalities obtained by using MCM to get SSI and subtracting the SSI from 100. The Cox PH model can be used on the data set to find the hazard ratios (**Table 1**).

Table 1. A sample of the data with simulation for variables and abnormality classified as stage 4.

Patient	Time	Status	Stage 4	Age	Gender	Smoking-status
1	6.6	0	48.80	50	Female	Smoker
2	6.05	1	46.69	50.05	Female	Non-Smoker
3	2.75	0	52.40	50.12	Male	Non-Smoker
4	49.67	1	36.76	50.04	Female	Non-Smoker

Table 2. The p -values for global statistical tests for stage 3.

Test	Calculated value	Degrees of freedom	p -value
Likelihood ratio test	13.98	4	0.0007
Wald test	10.7	4	0.03
Score (Logrank)	13.36	4	0.01

3.2. Methods to establish the secondary aim 1.2.2 (b)

The Cox PH model has to be used in three steps to find the hazard ratio corresponding to the above-quantified abnormality of the dataset, which is described below for the current data set:

Step 1. The hazard ratios, and p -values attached with the covariates ‘stage1–4’, ‘gender’, and ‘smoking-status’ were calculated by applying the Cox PH model using the CRAN-R packages. Since this model is a statistical model, basic statistical concepts like global study’s p -values, and the p -values for the covariates were also obtained to check the significance of the model fit to the data. The results were tabulated below. In these tables all the hazard ratios were related to the covariates or variables.

Step 2. The statistical significance of each of the coefficients in the table and whether they formed due to chance or not was checked. The global statistical significance was checked in **Table 2**. Subsequently, the statistical significance of the variables, and the confidence intervals for the good fit of the parameters were also checked.

Firstly, the tables for all stages and the statistical analysis of the global model with p -values, for example, are given in **Table 2**.

The p -values for the three overall tests (likelihood, Wald, and score) for stages 1–4, showed significance, suggesting that the model holds importance. These tests assess the general null hypothesis that all beta coefficients (β) equal 0. In the given instance, the test statistics closely aligned, leading to a firm rejection of the general null hypothesis. That is, the coefficients were existing non zeros.

Secondly, the statistical significance of each covariate in the partially simulated data was checked, which in the below discussion are the hazard ratios or the values under the column with the title ‘exp(coefficients)’, ‘exp(β)’ or ‘ e^β ’ (**Tables 3–6**).

Table 3. Full Cox PH model-hazard ratios and p -values for variables of stage 1 lung cancer data.

Covariate	Coefficient	Exp (coefficient)	Standard Err. (coefficient)	Z	Pr (> Z)
Stage 1	-0.85790	0.42405	0.38745	-2.342	0.0192
Age	-0.06678	0.93540	0.01762	-1.861	0.0627
Gender male	-1.57594	0.20681	0.32392	-2.051	0.0403
Smoker	1.69714	5.45834	0.32379	2.190	0.0285

Table 4. Full Cox PH model—hazard ratios and *p*-values for variables of stage 2 lung cancer data.

Covariate	Coefficient	Exp (coefficient)	Standard Err. (coefficient)	Z	Pr (> Z)
Stage 2	0.89030	2.43585	0.3874	2.298	0.0216
Age	0.01564	1.01576	0.0176	0.888	0.3746
Gender male	0.19913	1.22034	0.3239	0.615	0.5387
Smoker	0.72066	2.05579	0.3237	2.226	0.0260

Table 5. Full Cox PH model—hazard ratios and *p*-values for variables of stage 3 lung cancer data.

Covariate	Coefficient	Exp (coefficient)	Standard Err. (coefficient)	Z	Pr (> Z)
Stage 3	0.04527	1.04631	0.01990	2.275	0.0229
Age	0.20473	1.22762	1.6417	0.125	0.9008
Gender male	-0.27398	0.76035	0.30131	-0.909	0.3632
Smoker	-0.53360	0.58647	0.30069	0.30069	0.0760

Table 6. Full Cox PH model—hazard ratios and *p*-values for variables of stage 4 lung cancer data.

Covariate	Coefficient	Exp (coefficient)	Standard Err. (coefficient)	Z	Pr (> Z)
Stage 4	-0.16303	0.84857	0.06953	-2.234	0.0190
Age	0.09225	1.096642	0.04497	2.051	0.0402
Gender male	-0.08372	0.91969	0.56618	-0.148	0.8824
Smoker	-0.39789	0.67174	0.61522	-0.647	0.5178

*where Err is the error.

Table 3 shows that the variables ‘stage 1’ and ‘smoking smoker’ (wherein patient is a smoker) had *p*-values less than 0.05, that is, these values did not exist by chance. The hazard ratio for the variable stage 1 was $\exp(\beta) = 2.43585$ and for the variable, ‘smoking smoker’, this value was 2.05579, and the *p*-values for other variables were less than 0.05. We can also consider the hazard ratios of those variables for which the *p*-value is greater than 0.05, but they could have occurred by chance, that is it suggests that the observed data is not consistent with the null hypothesis.

Similarly, from **Table 4**, the variables ‘stage 2’, ‘gender male’, and ‘smoking smoker’ had *p*-values less than 0.05, which is to say that they were statistically significant, and in **Table 5**, the variables ‘stage 3’ and ‘age’ had *p*-values less than 0.05, while **Table 6** shows that the variable ‘stage 4’ had a *p*-value less than 0.05.

To conclude, the spread of the cancer as seen on CT scans were converted to quantities and the hazards corresponding to the cancer were calculated. These models can be used to predict the trends in hazard or risk due to lung cancer and the other variables. Some variables had statistical significance and others did not. The variables that were not statistically significant were dropped from the Cox PH model.

The only specific models with statistical significance for the stages of lung cancer were as follows:

$$\text{For stage 1: } h_1(t/X)/h_0(t) = e^{(0.89030\text{Stage 1} + 0.01564\text{Age} + 0.72066\text{Smoking smoker})}$$

where $h_1(t/X)/h_0(t)$ can be interpreted as the ratios of infected to non-infected, smoking

to non-smoking, and female to male.

Step 3. The 3rd step in finding the hazard ratios is using the confidence intervals to check if the parameters, that is, hazard ratios are effective for prediction using the above models for all stages of lung cancer. The following rules of thumb apply only to parameters or hazard ratios that are in the model, which were the variables that were statistically significant. Firstly, the parameters should lie inside the confidence intervals, and we can observe from **Table 7** that the parameters are lying in the corresponding confidence intervals. Secondly, the width of the confidence intervals should be very narrow, and this can also be observed.

Table 7. Full Cox PH model—confidence intervals for stage 4.

Covariate	Exp (coefficient)	Lower limit 0.95	Upper limit 0.95
Stage 4	0.8496	0.7413	0.9736
Age	1.0966	1.0041	1.1977
Gender male	0.9198	0.3032	2.7898
Smoker	0.6717	0.2012	2.2432

The hazard gradually increased as time passes as seen in **Figure 2**. At 45 units of time, the hazard increases very steeply, leading to the next stage of cancer. Here, if the advancement of the spread of lung cancer or abnormality is increased, then the hazard increases with time.

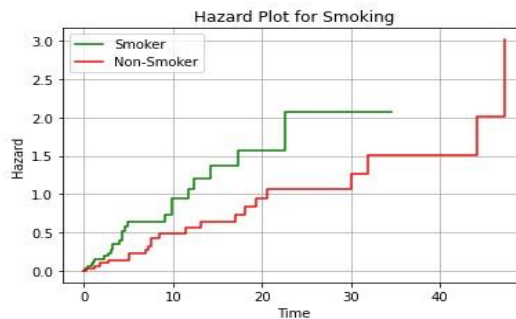


Figure 2. Hazard for smoking status among stage 1 lung cancer patients.

In **Figure 2**, the data for the variable, smoking-status, is available until 35 units of time (what is the specific unit here?). The fluctuation in hazard in males and females can be observed in **Figure 3**. Males have a lower hazard than females. The other variable that had a p -value < 0.05 was ‘gender male’ which had a hazard of 0.2068, and this indicates that males had a decrease of lung cancer by 0.2068 compared to females. Similarly the pattern in hazard can be noticed for the variables in **Figures 2–4**.

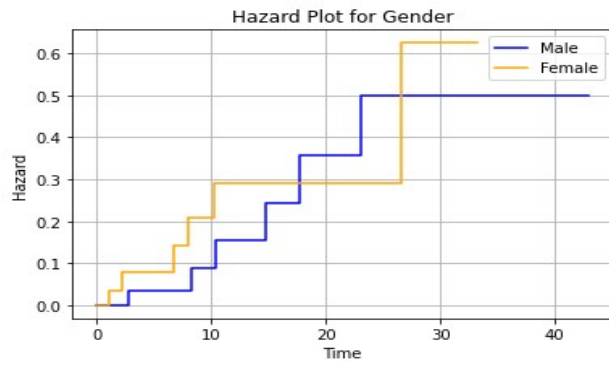


Figure 3. Hazard for the variable, gender.

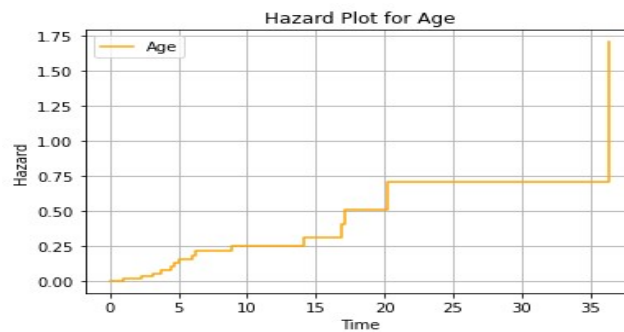


Figure 4. Hazard for the variable age for stage 3 of lung cancer.

3.3. Methods to establish the secondary aim 1.2.2 (c)

Firstly, stage 1 of lung cancer was considered (Table 3). The hazard ratio for the variable ‘stage 1’ was $2.4359 > 1$. This implies that for every unit increase in the abnormality, the hazard increases by 2.4359. In addition, this decrease was not by chance as the variable, ‘stage 1’, was statistically significant. Next, at this stage 1, the other variable, smoking, a categorical variable, had a p -value < 0.05 with a hazard ratio of $2.0558 > 1$. This also indicates that being a smoker has 2.0558 times greater risk or hazard than being a non-smoker.

Secondly, the variables ‘stage 2’ (Table 5), ‘gender male’ and ‘smoking smoker’ had p -values < 0.05 , and hence, were statistically significant. The hazard ratios and the classification for these variables were as follows:

Variable ‘stage 2’ had a hazard ratio of $0.4241 < 1$, which indicates that the hazard decreases by 0.4241 for every unit of abnormality. Normally, stage 2 should have a hazard greater than 1. This decrease in hazard may be attributed to treatment used if any. The last variable considered in the stage 2 category of lung cancer was the variable, smoking. Table 4 shows the hazard for the variable, ‘smoking smoker’, indicating that a person who smokes has a hazard ratio 5.4583 times greater than that of non-smokers at stage 2 as depicted in Figure 5.

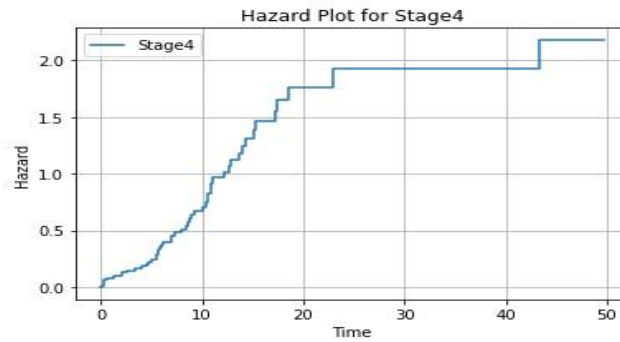


Figure 5. Hazard for the variable, ‘stage 4’.

Thirdly, at stage 3 of lung cancer (**Table 5**), the variables for which the p -value < 0.05 were for the variables, ‘stage 3’ and age. The hazard for the variable stage 3 was 0.8496, which means that the hazard is decreased in the patients who were affected. This decrease might be attributed to treatment for cancer. Next, the variable, age, had a hazard of 1.0966, indicating that the hazard was slightly increasing with the age of the patient.

Lastly, the only statistically significant variable is stage 4, for which the p -value is < 0.05 , with a hazard ratio of 1.0463. This signifies that there is increase in hazard due to cancer.

In this section, lung cancer in CT scans was classified on the basis of the hazard due to the lung cancer as follows:

A hazard ratio of < 1 implies a decrease in the hazard due to cancer.

A hazard ratio > 1 implies an increase in hazard due to lung cancer.

The hazard ratio for stage 1 > 1 , and the hazard ratio for smokers at stage 1 lung cancer > 1 .

The hazard for stage 2 < 1 , and the hazard ratio for gender male of stage 2 < 1 .

The hazard ratio for smokers with stage 2 lung cancer > 1 .

The hazard ratio for stage 3 < 1 , and the hazard ratio for age at stage 3 lung cancer > 1 .

The hazard ratio for stage 4 (**Table 6**) lung cancer > 1 .

The assumptions of the PH model were checked. For the first assumption, the hazard curves were not intersecting each other for the stage 1 variable. In addition, this has been verified for other stages, and the PH assumptions were met (**Figures 6 and 7**). This indicates that the Cox PH model fit was successful.

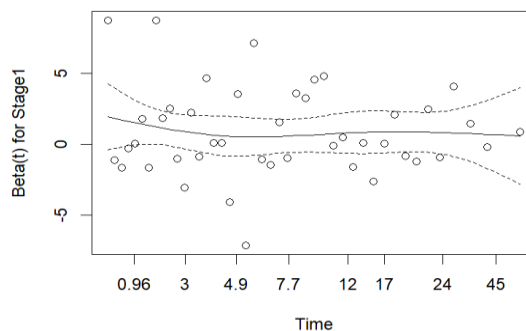


Figure 6. The hazard curves for the variable, ‘stage 1’.

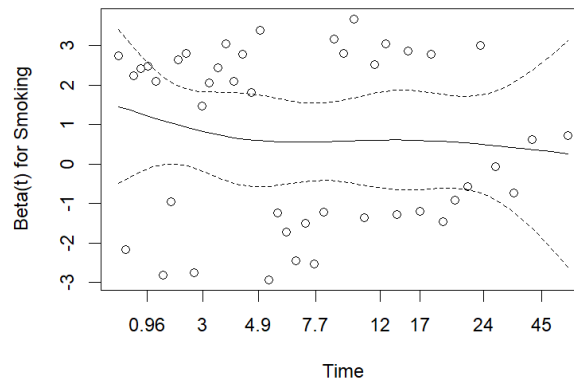


Figure 7. The hazard curves for the variable, ‘smoking’.

For the second assumption, the patients were independent of each other, so the second PH assumption was also met.

4. Discussion

The hazard ratio is an important indicator of mortality risk that provides information on disease prognosis and helps inform decisions about treatment made by healthcare professionals. There is limited literature on using biostatistics in combination with AI-based automated image analysis methods, particularly methods that can be used to analyse small data sets without the need for a training data set. The purpose of the current study was to find the hazard or risk due to the abnormality present in the images of a group of people or commodities using a quantified abnormality score calculated from the previously developed MFA and MCM methods [1,2]. This is a novel approach to using AI-based automated image analysis methods to determine the mortality risk associated with disease and was used for the first time.

The MFA or MCM method was used successfully to first find the cancer in the CT scans together with some simulated variables and to classify them on the basis of lung cancer (**Tables 1–4**), which would give the physician information on the cancer spread in the form of a numerical quantity, leading to a better understanding of a patient’s disease progression. Secondly, for the group of patients at the different stages of cancer, the hazard ratio was found using techniques from survival analysis in a novel approach. Furthermore, the hazard was studied within subgroups of patients, like, for example, on the basis of smoking. Thirdly, the groups of patients with lung cancer were classified on the basis of hazard due to the cancer affecting them, or any association with their gender or smoking status. There is no other research in the literature like the current study that calculates the hazard corresponding to the converted cancer in image form to a quantified hazard. Moreover, none of the results contradict the hypotheses of this MHM study.

The clinical significance of the current study is that typically in medicine, the hazard due to cancer is attributed to deaths, and here, for the first time, the hazard due to cancer has been estimated using information on cancer in image form before the death of the patients. With this, the physician not only has information about the cancer in image form, but also new information on the numerical hazard due to the cancer, and this could potentially lead to improved treatment. Another significant finding of this study is that the CT scans taken to study the particular patient can be useful to

study the public health of the group. This can impact medicine by allowing healthcare professionals to provide better, earlier treatment and save more lives. Furthermore, the MHM method not only saves a lot of physicians' time, but it can also provide a means of increasing inter-rater reliability. In addition, normally when cancer is studied, other variables or symptoms are mixed with cancer and increase the number of confounding variables. This method does not involve any confounding with other variables, because in the CT scans, the cancer is visible and is converted to number, and then its corresponding hazard is calculated. Hence, there are no confounding variable interactions in this study of cancer.

One limitation of this study is that it needs a considerable amount of data; however, it does not require data sets of a thousand or more, because the Cox PH model only needs enough data points for its assumptions to be met. In addition, the idea of the study has greater generalizability to other fields, but if the images are from a different area of science then the thresholds of classification must be found again. However, the thresholds obtained for cancer in the current study can be used for other CT scans of cancer or other populations with lung cancer. The recommendations for future research are to use the same CT scans or original format of the images instead of conversion to other formats. This will enhance the study and remove extra fuzziness due to the conversion of images.

The key findings of this MHM study are the better understanding of the cancer or abnormality because of its ability to find the hazard ratio corresponding to the cancer. There is also no need to devote extra funding for the study of cancer, as existing CT scans were used in this study. Moreover, the method is simple to write using a CRAN-R package. Overall, MHM is based on the simple idea of converting abnormality in image form to a number, and in turn, finding the hazard due to the abnormality.

5. Conclusion and future directions

To conclude, this is a cross application study where AI, fuzzy systems, and computational intelligence techniques were used to convert abnormality or cancer in the form of an image to a quantity, and subsequently find the hazard resulting from the abnormality or cancer using the Cox PH model of survival analysis, after which the hazard was classified into categories. This was a novel approach that was used for the first time in the literature. Moreover, as very little software was used along with existing images, this study has practical applications. Future research will focus on classifying abnormality in images into survival probabilities.

Author contributions: Conceptualization, KKA; methodology, KKA; software, KKA; validation, KKA, AG and MA; formal analysis, KKA; investigation, KKA; resources, KKA; data curation, KKA; writing—original draft preparation, KKA; writing—review and editing, KKA and MA; visualization, KKA; supervision, KKA, MA, and AG; project administration, KKA. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Akula KK, Gegov A, Arabikhan F. Artificial Intelligence-Based Medical Image Classification using a Multilayer Fuzzy Approach. *Wseas Transactions on Computers*. 2023; 22: 206-217. doi: 10.37394/23205.2023.22.24
2. Akula KK, Gegov A, Arabikhan F, et al. Medical Image Classification using a Many to Many Relation and a Multilayer Fuzzy Approach. *Electronics MDPI* 2024 (under review).
3. Wang Z. Multi-scale structural similarity for image quality assessment. In: *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2003.
4. Wulczyn E, Steiner DF, Xu Z, et al. Deep learning-based survival prediction for multiple cancer types using histopathology images. Hsieh JCH, ed. *PLOS ONE*. 2020; 15(6): e0233678. doi: 10.1371/journal.pone.0233678
5. Furrer MA, Sathianathen N, Gahl B, et al. Oncological outcomes after attempted nerve-sparing radical prostatectomy (NSRP) in patients with high-risk prostate cancer are comparable to standard non-NSRP: A longitudinal long-term propensity-matched single-centre study. *BJU International*. 2023; 133(1): 53-62. doi: 10.1111/bju.16126
6. National Cancer Institute. Available online: <https://www.cancerimagingarchive.net/> (accessed on 23 October 2023).
7. John M. Last, *A dictionary of epidemiology*, Oxford University Press, 2001.
8. Bradburn MJ, Clark TG, Love SB, et al. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer*. 2003; 89(3): 431-436. doi: 10.1038/sj.bjc.6601119
9. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972; 34(2): 187-202. doi: 10.1111/j.2517-6161.1972.tb00899.x
10. Balakrishnan N, Rao CR. *Handbook of Statistics 23: Advances in Survival Analysis*, North Holland. 2004.
11. Bradburn MJ, Clark TG, Love SB, et al. Survival Analysis Part II: Multivariate data analysis—an introduction to concepts and methods. *British Journal of Cancer*. 2003; 89(3): 431-436. doi: 10.1038/sj.bjc.6601119
12. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika*. 1982; 69(1): 239-241. doi: 10.1093/biomet/69.1.239
13. Kleinbaum DG, Klein M. *Survival Analysis*. Springer New York; 2012. doi: 10.1007/978-1-4419-6646-9
14. Deo SV, Deo V, Sundaram V. Survival analysis—part 2: Cox proportional hazards model. *Indian Journal of Thoracic and Cardiovascular Surgery*. 2021; 37(2): 229-233. doi: 10.1007/s12055-020-01108-7

Article

Identifying voices using convolution neural network models AlexNet and ResNet

Abdulaziz Alhowaish Luluh, Muniasamy Anandhavalli*

College of Computer Science, King Khalid University, Abha 62529, Kingdom of Saudi Arabia

* **Corresponding author:** Anandhavalli Muniasamy, anandhavalli.dr@gmail.com

CITATION

Luluh AA, Anandhavalli M.
Identifying voices using convolution
neural network models AlexNet and
ResNet. *Computing and Artificial
Intelligence*. 2024; 2(1): 441.
<https://doi.org/10.59400/cai.v2i1.441>

ARTICLE INFO

Received: 2 January 2024
Accepted: 30 January 2024
Available online: 12 February 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence
is published by Academic Publishing
Pte. Ltd. This work is licensed under
the Creative Commons Attribution
(CC BY) license.
[https://creativecommons.org/licenses/
by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: Deep learning (DL) techniques which implement deep neural networks became popular due to the increase of high-performance computing facilities. DL achieves higher power and flexibility due to its ability to process many features when it deals with unstructured data. DL algorithm passes the data through several layers; each layer is capable of extracting features progressively and passes it to the next layer. Initial layers extract low-level features, and succeeding layers combine features to form a complete representation. This research attempts to utilize DL techniques for identifying sounds. The development in DL models has extensively covered classification and verification of objects through images. However, there have not been any notable findings concerning identification and verification of the voice of an individual from different other individuals using DL techniques. Hence, the proposed research aims to develop DL techniques capable of isolating the voice of an individual from a group of other sounds and classify them based on the use of convolutional neural networks models AlexNet and ResNet, that are used in voice identification. We achieved the classification accuracy of ResNet and AlexNet model for the problem of voice identification is 97.2039 % and 65.95% respectively, in which ResNet model achieves the best result.

Keywords: deep learning (DL); voice identification; convolutional neural network (CNN); AlexNet; ResNet

1. Introduction

Due to the continual advancement of artificial intelligence (AI) technology, deep learning (DL) has become increasingly important in the last decade. DL is an artificial intelligence (AI) structure that impersonates the activities of the human mind in handling information for use in identifying objects, perceiving speech, interpreting dialects, and making decisions based on the data collected [1]. Deep learning refers to a machine-learning algorithm that can locate objects, identify speech, translate, and interpret languages, and make decisions without the need for human supervision. So, deep learning has the advantage that the software builds the feature set independently and without supervision. While deep learning techniques have been successfully implemented in recognizing characters and objects, its development in identifying voices and interpreting them is still under process. Deep learning achieves this by employing a multi-layered calculation design known as neural networks. The design of the human brain influences the neural organizing plan. In the same way that humans think carefully to discern designs and order various types of data, neural networks may be trained to do the same thing with data. The individual layers of neural organizations can likewise be considered as such a channel that works from gross to unpretentious, improving the probability of identifying and yielding a right outcome [2]. While the AI's performance in voice recognition has been widely praised, it is also necessary to

address the AI's issues in distinguishing the voices of individuals. As voice-based AI systems are now included in personal security and essential surveillance technologies, the requirement for distinguishing voices is higher than ever. The current research used deep learning techniques to attempt to recognize voices in this regard. It focuses primarily on the usage of convolutional neural networks, which provide non-linearity toward an audio-based set, making it easier to operate. Moreover, according to Deng and Liu [3], the relevance of identifying and interpreting voices has grown in recent years. The increasing importance is due to an increase in demand for voice-controlled AI systems in the technology.

The human brain works in the same way with each point involving humans accessing new data, the brain attempts to contrast it to other known items. A similar idea is additionally utilized by deep neural networks where the systems use the data collected to be compared to already existing information [4]. Deep learning is notable for its relevance in image recognition. However, another critical utilization of the innovation is in speech recognition utilized in digital assistances such as Amazon's Alexa or messaging through voice recognition [5]. The benefit of deep learning in voice identification comes from the adaptability and flexibility that comes from deep neural networks that lately become more accessible. Meanwhile, A critical challenge in identifying the right words involves the variability of the voice produced for the same word given in different accents such as Hello vs. Hellooo. When issued with an audible sentence the voice recognition starts by altering the voice waves using Fast Fourier Transformation and concatenating structures from adjacent windows in order to come up with a spectrogram [6]. The main aim is to take down the dimensionality of the univariate voice data enabling prediction of the letters and words that are coming next. Computer oriented processing and the recognition of human voices is referred to as speech recognition. It can be used to validate user credentials in certain systems as well as give instructions to digital assistants such as Siri, Cortana or Google Assistant. These systems work through storage of human voices which then train the voice identification system to listen in for vocabulary and patterns within the speech produced.

Another vital test in voice identification is the issue of redundancy; to decipher progressively, the model should foresee words accurately without the entire sentence. A portion of the deep learning models like bi-directional repetitive neural networks advantage exceptionally from utilizing the entire sentence because of the additional unique [7]. The arrangement in diminishing inactivity is to remember restricted setting for the model construction by permitting the neural organization to approach a short measure of data after a particular word [8]. At last, deep learning is still genuinely in its outset, however, is rapidly moving toward a cutting-edge capacity in voice identification. Studies indicate that there is still a great deal of opportunity to get better in model designing to diminish inertness and increment precision [9].

Though object verification and classification of images have benefited greatly from the widespread use of deep learning, the identification and verification of individual voices remains conspicuously lacking. Many DL developments to far have gone toward translating and understanding voice instructions, which has left a gap in the capacity to identify and authenticate individual voices within a group. In order to close this gap, this article uses deep learning methods for voice identification, namely

convolutional neural networks models like AlexNet and ResNet. Voice recognition for individuals involves a machine consuming voices, speeches, utterances, and phrases from an individual so that each time the individual speaks, the speech is converted to voice waves and compared to previously stored datasets in order to recognize the specific person. It is important to note that the identification of voices through deep learning improves with time and the measure of data fed into the system. The architecture of deep learning, such as AlexNet and ResNet models and the network of the adversarial generative and recurrent neural networks for memory and long-term as obstetric non-supervised models. Neural Networks sets the background of the AlexNet and ResNet as models discriminatory moderated, presented and developed it. The application of this structure on a large scale to many applications such as automatic speech, natural language processing, and the establishment of voice and detection of objects in three-dimensional where it is made proven that it produces results developed for different tasks.

In summary, this paper makes the following contributions:

- Apply convolutional neural network models and evaluate the performance in identifying and verifying the voice of a particular speaker.
- Explore ResNet and AlexNet models in identifying the voice.
- Evaluate the model’s results and compare them with other deep learning or machine learning models.

2. Literature review

Khalil et al. [10] is presenting different deep learning techniques for speech emotion recognition (SER) as it is known that the recognition of emotion is a very challenging part of the human computer interaction (HCI). Extraction of the emotion from the voice signal is being done using many techniques, which is classification techniques and a well-established analysis of the speech. All the traditional techniques now have been replaced by the deep learning technique for speech emotion recognition. A convolution neural network has been used for the classification feed-forward architecture (Figure 1).

In Figure 1, the basic layer-wise architecture of the convolution neural network is presented.

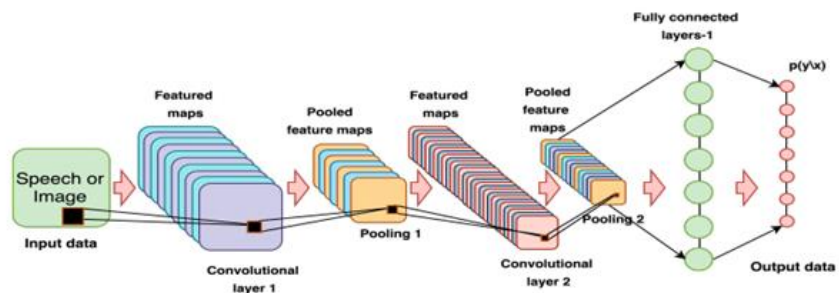


Figure 1. Architecture of CNN layer-wise.

Recognition of the pattern is done with the help of a convolution neural network, and classification of the data in a better form is provided by the CNN. In each layer small size of a neuron are present in these types of networks, the layer is of architecture

which is designed model which is used for the processing receptive field which is a type of input data [10].

On the other hand, Zhang et al. [11] point to the use of convolutional neural network as an effective method to be involved with deep learning techniques for recognizing sounds. They believe that since convolutional neural networks have multiple hidden layers, the users could set more parameters for refining the sounds that pass through the AI learning process. As in agreement, Deng and Liu [3] says that by representing the audio input using the spectrogram image, it is possible to identify the minute the changes in the hertz and decibels of a sound. In this regard, Deng and Liu [3] carried on an analysis that involved isolation of specific sounds from multiple sounds using the convolutional neural network. By identifying the spectrogram plots, the network was successful in isolating the specific sound. However, Lemley et al. [12] argue that though the approach would yield expected results in isolating environment sounds, it would not be able to distinguish between the words spoken by a person. The reason is that, regarding human speech, the spectrogram plots are hard to identify, as they look similar to one another. Even so, Parcollet et al. [13] stated in their study that convolutional neural network when used with deep learning techniques would enable an AI to improve its classification parameters and identify sounds of individuals, given that a specific speech input has already been made.

In Zhang et al. [14] they focus on the details of the Adam optimization algorithm, which is utilized to solve the objective function iteratively. As a result, by constantly modifying the first order and second-order momentum with an average of gradients over iteration, the Adam algorithm has great strength in resisting noise pollution.

In ShabanAl-Ani [15], fingerprint recognition is one of the most famous biometrics. Due to its excellence and consistency over time, fingerprints have been used for identification for more than a century. In this study, discrete cosine transform (DCT) is applied to the entire image. DCT communicates a limited succession of items as far as an aggregate of cosine capacities swaying at various frequencies. This gives DCT transactions, which are universal fingerprint recognition policy features based on separate pocket conversion technology, suggested. This algorithm relies on the matching process through the link in the images, and then uses DCT technology to extract important features. This study faces many problems such as the low quality of the fingerprint input image.

According to Krizhevsky et al. [16], AlexNet famously won the 2012 ImageNet LSVRC-2012 competition by a large margin (15.3% vs. 26.2% (second place) error rates). The network had a very similar architecture to LeNet but was deeper, with more filters per layer, and with stacked convolutional layers.

In their study on sound classification using AI, Andrew [17] figured out that computers could read a sound better if it were to be given an image representation of those sounds. As such, Andrew [17] suggests carrying on a visual inspection of the audio files and then differentiating them based on the difference in their graphs. Malik et al. [18] argue that using the image representation of sound waves could offer inaccurate results in classifying sounds. In their study carried out urban sound classification, the study realized that, for certain sounds, the visual representation remains almost identical to one another. In such cases, it becomes hard to differentiate one sound from another.

In another study, Pons et al. [19] conducted an experiment on isolating environmental sounds using deep learning techniques. The study made use of a neural network but with a single hidden layer. The neural network was placed as a middle layer to filter the sounds based on given parameters. The experiment was successful in isolating the sounds in the environment. However, as it was stated by Bunrit et al. [20], using a single hidden layer limits the number of parameters that the user could set in recognizing the sounds. While it may be successful in differentiating sounds with notable level of frequency, it cannot isolate mild sounds that the environment may also contain.

According to the author Heaven [21], the categorization of speaker verification and speaker identification can be done in the speaker recognition approach. A spectrogram image is being used for the training of the CNN method used in the identification and verification of the voice and is best than the other system used. 95.83% was the average classification result of the method used for the testing. In the case of the MFCC method, the average was 91.26% and 49.77% for the raw signal as an input to the CNN model. So, when the shot voice is used as an input, then the method is very efficient.

3. Methodology

This research aims to use convolutional neural networks models, AlexNet and ResNet, to identify the voices of the individual in an environment and classify voice dataset to achieve the best accuracy of the voice identification. In this section we explained the methodology of our work. **Figure 2** shows the proposed methodology.

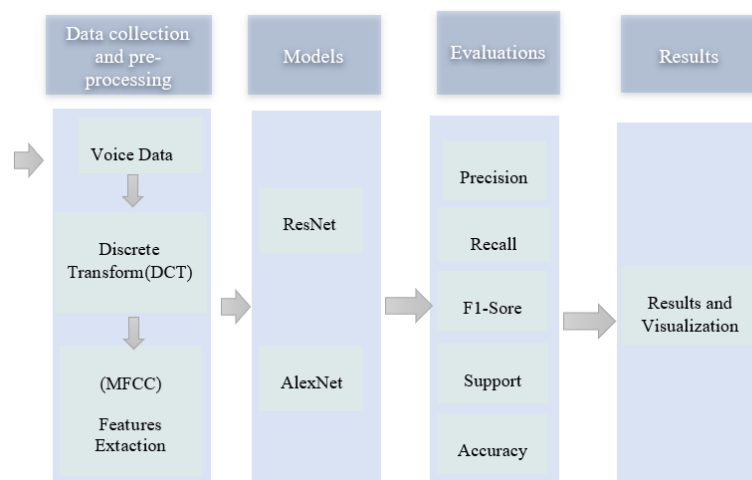


Figure 2. Proposed methodology structure.

Qualitative research methodology uses non-numerical data for understanding the underlying reasons behind selected subject [22]. Meanwhile, quantitative research methodology encourages the use of numerical data, it can help in identifying patterns and making predictions [23]. Our research requires the use of numerical data and recognition of patterns in the changes with regard to voices and their classification. Hence, we used quantitative research methodology. The data required for carrying on the analysis is an existing voice dataset from OpenSLR website. We used MFCC to extract features from voices, and then we applied the ResNet and AlexNet models for

the training and testing process. The varying levels of accuracy in identifying the specific set would enable the current research to understand the level of applicability of CNN in identifying voices.

The following are the steps to implement the proposed methodology of voice identification of the speaker:

Step 1: Collection of voice datasets contains 2937 audio files.

Step 2: Feature extraction using MFCC.

Step 3: Divide the dataset into train and test sets as a ratio of 9:1.

Step 4: Set labels IDs for recognition.

Step 5: Training the ResNet and AlexNet models for about 50 epochs for better prediction.

Step 6: Save the accurate trained models to test voices.

3.1. Dataset description

We used raw human voice dataset from LibriSpeech ASR corpus [24]. LibriSpeech is a corpus of approximately 1000 h and 16 kHz of English speeches, prepared by Vassil Panayotov with the assistance of Daniel Povey.

ID	SEX	SUBSET	MINUTES	NAME
14	F	train-clean-360	25.03	Kristin LeMoine
16	F	train-clean-360	25.11	Alys Attewater
17	M	train-clean-360	25.04	Gord Mackenzie
19	F	train-clean-100	25.19	Kara Shallenberg

Figure 3. Dataset description.

The data is derived from read audiobooks from the LibriVox project and has been carefully segmented and aligned. The used dataset is “dev-clean.tar.gz”, that contains 2937 audio of people and is considered as voice classification of 40 classes, and the samples are shown in Figure 3. The samples of shape of audio signals are shown in Figure 4.

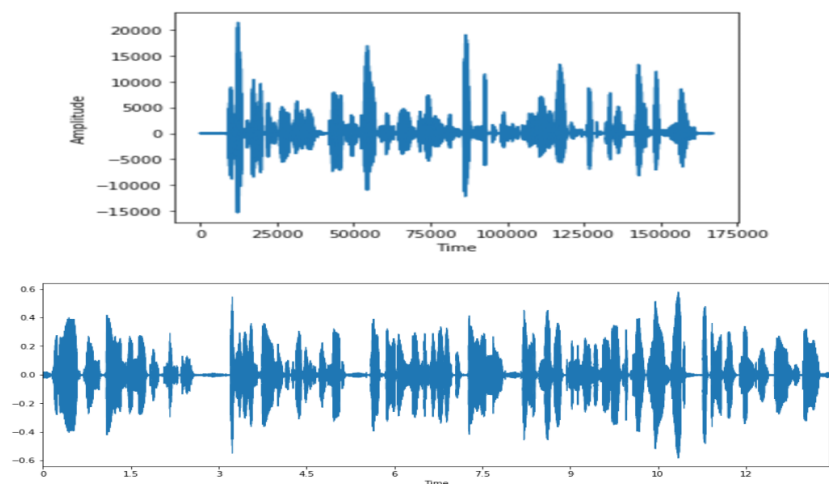


Figure 4. Sample of shape of original audio signals.

3.2. MFCC (Mel-Frequency Cepstral Coefficients)

The main process in voice identification is speech pre-processing. We used MFCC (Mel-Frequency Cepstral Coefficients) as a tool for extracting voice dynamics functions as acoustic features (such as MFCCs) as inputs and speaker IDs as target variable (**Figure 5**).

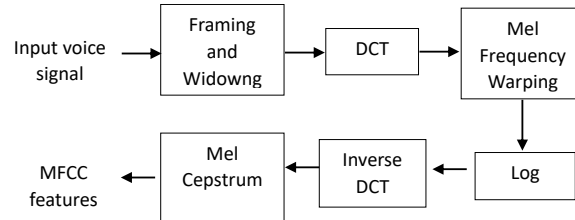


Figure 5. MFCC process flow diagram.

Voice signals in the time domain change very quickly, so we convert speech signals from the time domain to the frequency domain, then the corresponding spectrum can be clearly defined. We separate the signals into frames and call the window function to increase the continuity of voice signals in the frame. Discrete cosine transform (DCT) is being used for quantitative evaluation of spectral energy data into data units that can be analyzed by MFCC [25]. The famous audio processing library “librosa 0.7.2” has been used to compute the MFCC features of audios (**Figure 6**). The sampling rate is set to 16,000 and the number of MFCCs is set to 24.

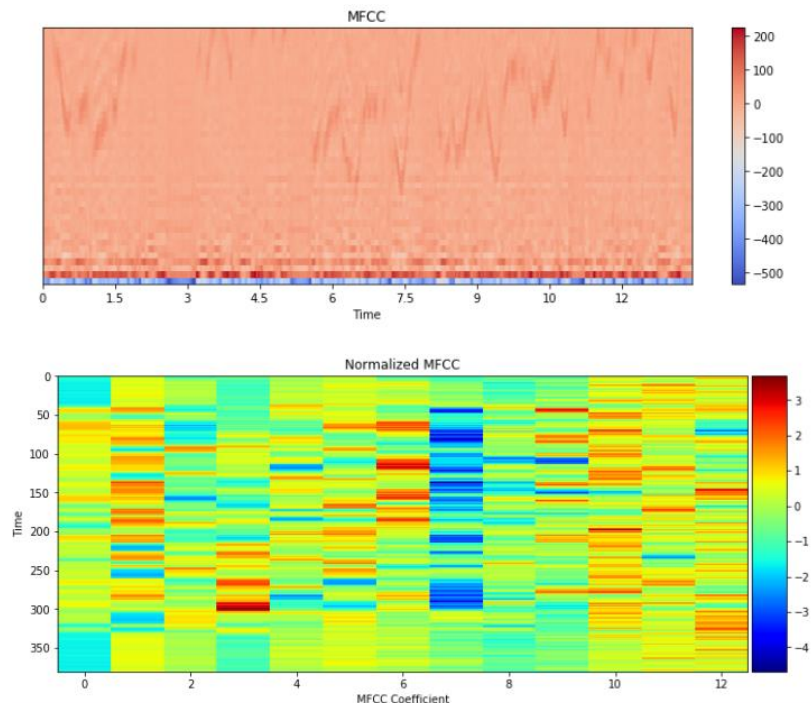


Figure 6. Sample of MFCC extracted features of audio.

We can observe that more no. of filters in low frequency region, and lesser no. of filters in high frequency region. Finally, these extracted features that we used in training and voice identification model using ResNet and AlexNet.

3.3. CNN models for classification

The ResNet (residual neural network) consists of many CNN layers, batch norm layers, max pooling layers, and ReLu layers. The core idea of ResNet is introducing a so-called “identity shortcut connection” that skips one or more layers, so it’s very good at classification tasks and its block model is shown in **Figure 7**.

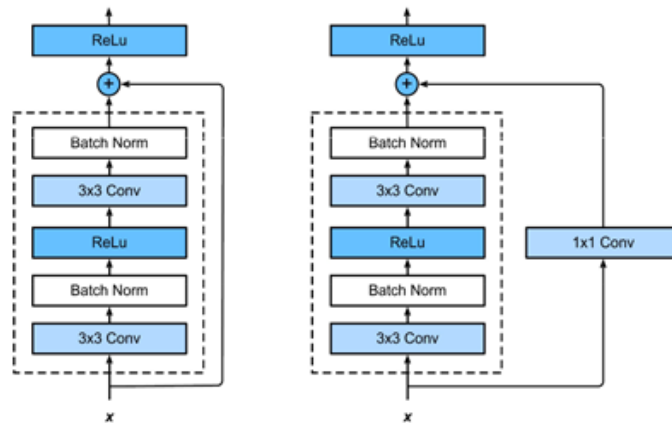


Figure 7. ResNet block model.

Figure 7 shows the ResNet consists of many CNN layers, batch norm layers, max pooling layers, and ReLu layers. The core idea of ResNet is introducing a so-called “identity shortcut connection” that skips one or more layers, so it’s very good at classification tasks. ResNet model consists of 11 ResNet blocks, and a ResNet block consists of 3 CNN layers and batch norm and max pooling layers have been applied for this research.

AlexNet (**Figure 8**) is a convolutional neural network which has had a large impact on the field of machine learning, specifically in the application of deep learning to machine vision.

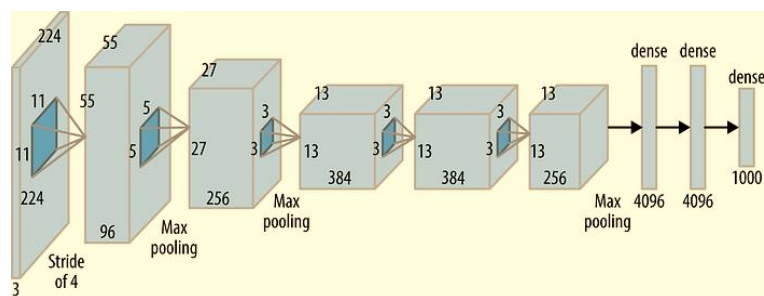


Figure 8. Architecture of AlexNet.

AlexNet consists of 11×11 , 5×5 , 3×3 , convolutions, max pooling, dropout, data augmentation, ReLU activations, SGD with momentum. It attached ReLU activations after every convolutional and fully connected layer. AlexNet was trained simultaneously on two Nvidia Geforce GTX 580 GPUs which is the reason why their network is split into two pipelines.

3.4. Evaluations

Classification evaluations metrics (**Figure 9**) includes accuracy, precision, and

recall. Precision is measured the positive patterns that are correctly predicted from the total predicted patterns in a positive class. Recall is used to measure the fraction of positive patterns that are correctly classified, and the accuracy metric measures the ratio of correct predictions over the total number of instances evaluated. F-score derived from two measures which are precision and recall, this metric represents the harmonic mean between recall and precision values [26].

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Figure 9. Description of metrics.

True positive (TP) means the patient has the condition and the test was positive. True negative (TN) indicates that the patient is healthy with a negative test result. A false negative (FN) indicates when negative samples are incorrectly expected to be positive. False positive (FP) refers to when positive patient samples are incorrectly anticipated to be negative.

4. Results and discussion

In this section, a description of the experimental results and performance evaluation, which shows the effectiveness of models that extract results of experiment and comparison between results with graphs and tables has been discussed. The proposed model was built with python deep learning framework: “PyTorch 1.2”, the code runs in google colab.

Training parameters:

The dataset is split into training and test sets at a ratio of 9:1, which is the most effective split ratio for our learning process. In this research we used the loss function which is the Adam optimization. Adam optimization algorithm is an extension to stochastic gradient descent that has recently seen broader adoption for deep learning applications. **Table 1** shows Hyperparameters values that control the learning process.

Table 1. Hyperparameters for data training.

No.	Hyperparameter	Value
1	Learning rate	0.0005
2	Batch size	16
3	Num of epochs	50
4	Loss function	Adam

The performance of deep learning is affected by hyperparameters. Also, they are critical for the efficacy of the optimization and model-fitting process [27]. So, we used the appropriate values to train our CNN models that gave the best performance. The chosen hyperparameters are the Learning rate, the ratio into training and validation is

0.0005, batch size is 16, the size of a batch must be more than or equal to one and less than or equal to the number of samples in the training dataset. Also, the number of epochs is 50, the number of complete passes through the training dataset.

Figure 10 shows the training (loss, accuracy) for ResNet model with 50 iterations from 1 to 49. Epochs is the number of complete passes through the training dataset, each iteration show loss, accuracy and val-accuracy for make improvement in ResNet model to arrive the greatest accuracy and lowest loss in our ResNet model with loss 0.0669% and higher accuracy 97.2039% and val-accuracy is 88.970% respectively.

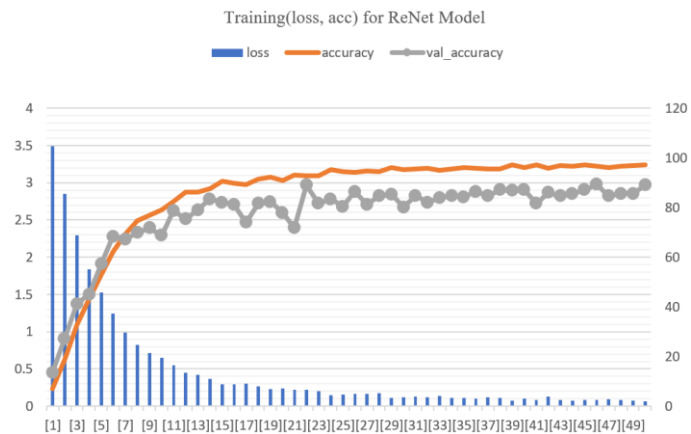


Figure 10. Training (loss, acc) for ResNet model.

Figure 11 shows the information about the accuracy of the ResNet model in each epoch (50 epochs) or iteration along with the loss is in the graph, by increased epoch number lowered loss value. The X-axis of the graph is the epochs, and the Y-axis is the result percentage. The minimum loss value is 0.0259, best training accuracy is 98.64%, and best validation accuracy is 87.86%. Also, we show the relation between training loss and accuracy for data using the ResNet model. Training loss value in data reduced and the validation accuracy is still lower than training accuracy then this model is excellent for our system, this means our model is fitted.

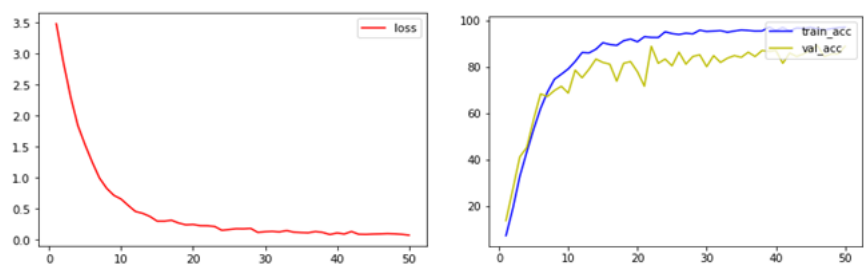


Figure 11. Line plots for (Training loss, Train-accuracy and Val-accuracy) of the ResNet model.

Classification report of the ResNet model is given in **Figure 12**. Report shows the results of 40 number of data voices from dataset to identify this voice belonged to whom. So, the test classification report that belongs to the speaker with an ID 84, ResNet model shows 90% accuracy, precision value 83%, recall value 100% and F1-score is 91%.

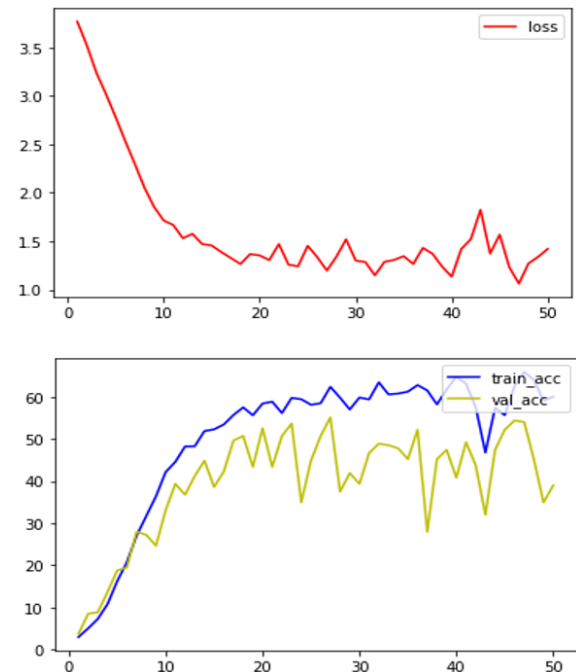


Figure 14. Line plots for (Training loss, Train-accuracy and Val-accuracy) of the AlexNet model.

Figure 14 shows the information about the accuracy of the AlexNet model in each epoch (50 epochs) or iteration along with the loss is in the graph. The x -axis of the graph is the epochs and the y -axis is the result percentage. The minimum loss value is 1.064%, best training accuracy is 65.95%, and best validation accuracy is 54.04%. The validation accuracy is still lower than training accuracy. So, from achieved results of the two models, ResNet model gives better accuracy than AlexNet model.

Classification report of the AlexNet model is given in **Figure 15**. Report shows the results of data voices from dataset to identify this voice belonged to whom. So, the test classification report that belongs to the speaker with an ID 1988, AlexNet model shows 39% accuracy.

Figure 16 shows the average of classification metrics of the ResNet and AlexNet Models. Based on the results, the ResNet model produces a higher average by precision 91.3% and recall 89.9 % and F1-score 89.7 %. We found that the results of the AlexNet model with classification metrics values of precision 42.5% and recall 41.2 % and F1-score 37.1% respectively.

the results of precision 42.5 %, recall 41.3 %, and F1-score 37.1%, respectively.

Table 2. Comparison between the proposed research and related work.

Study	Classification models	Accuracy
Proposed work	ResNet	97.2039%
	AlexNet	65.95%
Mamyrbayev O et al. [28]	Support vector method	83%
	Robust scaler method	90%
Bunrit S et al. [20]	CNN model	95.83%
	MFCC model	91.26%
Khalil R et al. [10]	Deep recurrent neural network	92.3%
Pons J et al. [19]	CNN model	88.95%

5. Conclusion

This study offers deep learning-based based voice recognition models. Convolutional neural networks have been used for identifying and verifying the speakers by their voices, and we tested and identified the voice of a particular speaker. In the initial step of data preparation, we employed the MFCC algorithm in the voice preprocessing process. ResNet model and AlexNet, models of convolutional neural network have been applied for voice identification. Classification accuracy of ResNet and AlexNet model for the problem of voice identification is 97.2039 % and 65.95% respectively. In the future work, we planned to incorporate voice verification technology into a multi-level and hybrid authentication approach to improve authentication reliability.

Author contributions: Conceptualization, AAL and MA; methodology, AAL; software, AAL; validation, AAL, and MA; formal analysis, MA; investigation, AAL; resources, AAL; data curation, AAL; writing—original draft preparation, AAL; writing—review and editing, MA; visualization, AAL; supervision, MA; project administration, MA. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Bae HS, Lee HJ, Lee SG. Voice recognition based on adaptive MFCC and deep learning. In 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA). 2016. pp. 1542-1546. doi: 10.1109/ICIEA.2016.7603830
2. Zhou R, Liu F, Gravelle CW. Deep Learning for Modulation Recognition: A Survey With a Demonstration. IEEE Access. 2020, 8: 67366-67376. doi: 10.1109/access.2020.2986330
3. Deng L, Liu Y. Deep Learning in Natural Language Processing. Springer Singapore, 2018. doi: 10.1007/978-981-10-5209-5
4. Deng L, Hinton G, Kingsbury B. New types of deep neural network learning for speech recognition and related applications: an overview. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Published online May 2013. doi: 10.1109/icassp.2013.6639344
5. Terzopoulos G, Satratzemi M. Voice Assistants and Smart Speakers in Everyday Life and in Education. Informatics in Education. Published online September 16, 2020: 473-490. doi: 10.15388/infedu.2020.21
6. Nguyen G, Dlugolinsky S, Bobák M, et al. Machine Learning and Deep Learning frameworks and libraries for large-scale

- data mining: a survey. *Artificial Intelligence Review*. 2019, 52(1): 77-124. doi: 10.1007/s10462-018-09679-z
7. Lee JG, Jun S, Cho YW, et al. Deep Learning in Medical Imaging: General Overview. *Korean Journal of Radiology*. 2017, 18(4): 570. doi: 10.3348/kjr.2017.18.4.570
 8. Fayek HM, Lech M, Cavedon L. Evaluating deep learning architectures for Speech Emotion Recognition. *Neural Networks*. 2017, 92: 60-68. doi: 10.1016/j.neunet.2017.02.013
 9. Bashar A. Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence and Capsule Networks*. 2019, 2019(2): 73-82. doi: 10.36548/jaicn.2019.2.003
 10. Khalil RA, Jones E, Babar MI, et al. Speech Emotion Recognition Using Deep Learning Techniques: A Review. *IEEE Access*. 2019, 7: 117327-117345. doi: 10.1109/access.2019.2936124
 11. Zhang Y, Pezeshki M, Brakel P, et al. Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks. *Interspeech 2016*. Published online September 8, 2016. doi: 10.21437/interspeech.2016-1446
 12. Lemley J, Bazrafkan S, Corcoran P. Deep Learning for Consumer Devices and Services: Pushing the limits for machine learning, artificial intelligence, and computer vision. *IEEE Consumer Electronics Magazine*. 2017, 6(2): 48-56. doi: 10.1109/mce.2016.2640698
 13. Parcollet T, Zhang Y, Morchid M, et al. Quaternion Convolutional Neural Networks for End-to-End Automatic Speech Recognition. *Interspeech 2018*. Published online September 2, 2018. doi: 10.21437/interspeech.2018-1898
 14. Zhang Z. Improved Adam Optimizer for Deep Neural Networks. 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Published online June 2018. doi: 10.1109/iwqos.2018.8624183
 15. ShabanAl-Ani M, M. Al-Aloosi W. Biometrics Fingerprint Recognition using Discrete Cosine Transform (DCT). *International Journal of Computer Applications*. 2013, 69(6): 44-48. doi: 10.5120/11849-7598
 16. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 2017, 60(6): 84-90. doi: 10.1145/3065386
 17. Andrew. Nuts and bolts of building AI applications using Deep Learning. NIPS Keynote Talk. 2016.
 18. Malik M, Adavanne S, Drossos K, et al. Stacked convolutional and recurrent neural networks for music emotion recognition. 2017.
 19. Pons J, Slizovskaia O, Gong R, et al. Timbre analysis of music audio signals with convolutional neural networks. In: *Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO) IEEE*. pp. 2744-2748.
 20. Bunrit S, Inkian T, Kerdprasop N, et al. Text-Independent Speaker Identification Using Deep Learning Model of Convolution Neural Network. *International Journal of Machine Learning and Computing*. 2019, 9(2): 143-148. doi: 10.18178/ijmlc.2019.9.2.778
 21. Heaven D. Why deep-learning AIs are so easy to fool. *Nature*. 2019, 574(7777): 163-166. doi: 10.1038/d41586-019-03013-5
 22. Taylor GR. Integrating quantitative and qualitative methods in research. University press of America. 2005, doi: 10.4236/psych.2020.115053
 23. McEvoy P, Richards D. A critical realist rationale for using a combination of quantitative and qualitative methods. *Journal of Research in Nursing*. 2006, 11(1): 66-78. doi: 10.1177/1744987106060192
 24. Open Speech and Language Resources. OpenSLR (Version 1) [Dev-clean]. Available online: <http://www.openslr.org/12/> (accessed on 7 September 2023).
 25. Nasr MA, Abd-Elnaby M, El-Fishawy AS, et al. Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients. *International Journal of Speech Technology*. 2018, 21(4): 941-951. doi: 10.1007/s10772-18-9524-7
 26. Hossin M, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 2015, 5(2): 3-5. doi: 10.5121/ijdkp.2015.5201
 27. Han J, Gondro C, Reid K, et al. Heuristic hyperparameter optimization of deep learning models for genomic prediction. De Koning DJ, ed. *G3 Genes|Genomes|Genetics*. 2021, 11(7). doi: 10.1093/g3journal/jkab032
 28. Mamyrbayev O, Mekebayev N, Turdalyuly M, et al. Voice Identification Using Classification Algorithms. *Intelligent System and Computing*. Published online April 29, 2020. doi: 10.5772/intechopen.88239

A comparison of cepstral and spectral features using recurrent neural network for spoken language identification

Irshad Ahmad Thukroo*, Rumaan Bashir, Kaiser Javeed Giri

Department of Computer Science, Islamic University of Science and Technology, Kashmir 192122, India

* Corresponding author: Irshad Ahmad Thukroo, thukrooirshad@gmail.com

CITATION

Thukroo IA, Bashir R, Giri KJ. A comparison of cepstral and spectral features using recurrent neural network for spoken language identification. *Computing and Artificial Intelligence*. 2024; 2(1): 440.
<https://doi.org/10.59400/cai.v2i1.440>

ARTICLE INFO

Received: 22 December 2023
Accepted: 22 January 2024
Available online: 30 January 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: Spoken language identification is the process of confirming labels regarding the language of an audio slice regardless of various features such as length, ambiance, duration, topic or message, age, gender, region, emotions, etc. Language identification systems are of great significance in the domain of natural language processing, more specifically multi-lingual machine translation, language recognition, and automatic routing of voice calls to particular nodes speaking or knowing a particular language. In his paper, we are comparing results based on various cepstral and spectral feature techniques such as Mel-frequency Cepstral Coefficients (MFCC), Relative spectral-perceptual linear prediction coefficients (RASTA-PLP), and spectral features (roll-off, flatness, centroid, bandwidth, and contrast) in the process of spoken language identification using Recurrent Neural Network-Long Short Term Memory (RNN-LSTM) as a procedure of sequence learning. The system or model has been implemented in six different languages, which contain Ladakhi and the five official languages of Jammu and Kashmir (Union Territory). The dataset used in experimentation consists of TV audio recordings for Kashmiri, Urdu, Dogri, and Ladakhi languages. It also consists of standard corpora IIIT-H and VoxForge containing English and Hindi audio data. Pre-processing of the dataset is done by slicing different types of noise with the use of the Spectral Noise Gate (SNG) and then slicing into audio bursts of 5 seconds duration. The performance is evaluated using standard metrics like F1 score, recall, precision, and accuracy. The experimental results showed that using spectral features, MFCC and RASTA-PLP achieved an average accuracy of 76%, 83%, and 78%, respectively. Therefore, MFCC proved to be the most convenient feature to be exploited in language identification using a recurrent neural network long short-term memory classifier.

Keywords: MFCC; RASTA-PLP; spectral features; RNN-LSTM; SNG

1. Introduction

The technique of recognising the language of a speech segment irrespective of the speaker gender, speaker emotions, speaker tune, and speaker age is known as spoken language identification (LID). Artificial intelligence has achieved new goals in developing intelligent algorithms for processing languages (both text and audio), making humans more interactive with the systems. Many languages have been computationally modelled with a lot of effort [1–3]. In the early days of voice recognition and speaker recognition, speech signals were portrayed as a normal (smooth input without first- and second-order derivatives) input to systems. On a daily basis, LID is utilised in a range of applications, including emergency call routing and multilingual translation systems [4,5], where the response time of the operator is crucial. Despite the use of high-level techniques such as phonotactic and prosody as major sources of information in today's state-of-the-art recognition systems, acoustic

modelling is still employed. The feature extraction procedure is the most effective and critical phase in any pattern recognition activity and depends on the selection and calculation of features. The most often utilised LID characteristics are MFCC features [6–9] which have achieved great accuracy. It also depends on various types of parameters, like demand, computer resources available, and the amount of language data available, all of which impact the reason for picking appropriate language recognition capabilities. Despite significant advances in state-of-the-art voice recognition systems, there is a massive and challenging challenge ahead of us, particularly for languages with limited resources. There has been relatively little study done on the languages spoken in Jammu & Kashmir and Ladakh. To our knowledge, no language identification has been made for the Ladakhi language. In this work, we have worked on six languages, which include five languages of JK and one language of Ladakh. We used two types of features: cepstral and spectral feature vectors. The cepstral features used are MFCC and RASTA-PLP, and the spectral features used are spectral roll-off, spectral flatness, spectral bandwidth, spectral contrast, and spectral centroid individually. By using RNN-LSTM as a backend classifier, it was found that MFCC features perform well in comparison to RASTA-PLP and spectral features on all performance metrics.

The major contribution of the proposed spoken language feature selection and detection model is given below:

- To develop a novel speech corpus for Ladakhi and languages spoken in UT (Jammu & Kashmir).
- To determine various cepstral and spectral features used for spoken language identification.
- To build RNN-LSTM models using three feature vectors.
- To validate the efficiency of the proposed language model on trained datasets using different performance metrics.

The remaining sections of this paper are given here. In Section II, a state-of-the-art literature survey is presented. The III section provides a detailed description of the datasets used in the experimentation. In Section IV, a brief introduction to pre-processing (noise reduction) and feature vectors used in the proposed model is presented. Experiments and results are put forward in Section V. Section VI presents a conclusion and recommendations for future work.

2. Literature survey

To remedy the prior limitations, similar advancements are being made to propose better LID systems. In LID, the most important activity is the generation of feature maps from a raw speech signal. The feature extraction techniques that have achieved significant accuracy are Linear Prediction Coefficients (LPC) [10,11], cepstral coefficients generated from LPC (LPCC) [12,13], PLP [14,15], RASTA [16], RASTA-PLP [17–19], independent component analysis [20], MFCC [21], and kernel-based approaches [22]. MFCC has made progress in different discourse applications, particularly in LID [23–26], and it achieves high exactness. Notwithstanding, as far as the all-out inclusion aspect (the quantity of MFCC highlights partitioned by the quantity of casings), MFCC actually has significant constraints, bringing about high

time utilization and asset exhaustion. Many examinations [27–31] have taken a gander at the MFCC approach to further develop distinguishing proof exactness and lessen the absolute element aspect to conquer the intricacy, time responsibility, and restricted asset issues.

Mukherjee et al. [32] created a novel version of MFCC named MFCC-2 (second-level MFCC) to govern the MFCC's wide and uneven dimensionality, which has been used to categorize English, Bangla, and Hindi languages. The 399-value feature set of the original MFCC was used to create the MFCC-2 features. The 19 MFCC coefficients, each of which defines a Mel-scale frequency spectrum, are referred to as the number of bands. Between the global extremes of each band, 18 equally spaced classes were designated for each clip, and the number of occurrences of energy values in each such class was recorded. As a consequence, the size of the function stayed constant at $19 \times 18 = 342$ values. In addition, for each clip from the MFCC feature set that was originally produced, the number of energies, mean, and standard deviation for each of the 19 bands were measured. As a consequence, $342 + 3 \times 19 = 399$ values were produced as a feature set. Principal Component Analysis (PCA) was used to evaluate the system's performance with lower-dimensional features. Additional feature sets were assessed, and features with 0% separation functionality were removed.

Many research scientists have helped to hybridize feature extraction methodologies, especially in speech processing applications including native speaker identification, background music identification, and voice recognition. In the case of spoken language identification, however, this domain has yet to be examined. Boussaid and Hassian [33] used PLP, RASTA-PLP, and MFCC to build a speaker recognition system based on 11 Arabic words (along with their first order derivatives). PCA was employed to reduce dimensionality. Hybrid features were employed on the front end, while the FFBPNN classifier was used on the back end. Two gradient learning methods, scaled conjugate (SC) and Levenberg-Marquardt (LM), were used to improve the performance of Feed forward Back Propagation Neural Network (FFBPNN). The TrainSC learning algorithm beats the TrainLM learning technique, according to the findings of the experiments.

Samarpan et al. [34] fostered a hybrid feature selection method based on Harmony Search (HS) and Naked Mole-Rat (NMR) algorithms for SLID. MFCC and RASTA-PLP attributes were removed from a bunch of single-speaker speech datasets for ten dialects (VoxForge, IIT-Madras). The versatile HS procedure has been converged with another nature-roused approach NMR calculation to construct another cross-breed Feature Selection (FS) calculation that picks the best subset of elements while likewise lessening model intricacy to assist it with preparing quicker. To get to the exhibition of this half-component determination method utilizing regular measurements, five classifiers (Random Forest, Multi-layer Perceptron, k-Nearest Neighbour, Nave Bayes, and Support Vector Machine) have been picked in the backend. In contrast with different classifiers, Random Forest acquires the most noteworthy exactness of 99.89% on CSS10, 98.22% on VoxForge, and 99.75% on the IIT-Madras discourse corpus.

Bashir and Quadri [35] proposed a bilingual script identification system for Kashmiri and English machine-generated languages using two features: horizontal

profile coefficients (peaks) and valleys. Experiments were done using machine-generated scripts, and it was found that for the Kashmiri language, lower peaks were achieved as compared to the English language. After working on 500 text lines, it was found that 481 text lines were identified correctly, thus achieving an accuracy of 96.2%.

Thukroo and Bashir [36] proposed a mel-spectrogram-based approach using convolutional neural networks for the officially spoken languages of Jammu and Kashmir and Ladakhi. The dataset contains six languages, i.e., Kashmiri, Ladakhi, Urdu, English, Hindi, and Dogri. Initially, speech segments were converted into mel-spectrograms by using the inverse Fourier transformation to log the Fourier transformation of a time-domain signal, and at the backend, CNN serves as a classifier. Experiments were conducted on recorded speech, IIIT-H, and VoxForge. It was found that while training the model at 100 epochs, an average accuracy of 100% was achieved. One of the main drawbacks of this model is that speech segments were converted into image domain; therefore, the focus has been shifted to image domain rather than linguistic characters such as syntax and semantics of the language. Second, testing was done by using speaker-dependent samples instead of speaker-independent samples. Third, the effect of noise has not been tested in a real domain.

From the above literature survey, it was found that very little work has been done related to low-resource languages like Kashmiri and Ladakhi, both in terms of written and spoken the main challenge of any low-resource language is the creation of a speech and written corpus, as there is no freely available corpus of the above-mentioned languages. Second, whatever work has been done related to the Kashmiri language is either done in the image domain or by using cepstral features, i.e., MFCC; the effect of spectral and temporal features has not been done yet. Third, to check which linguistic features perform better in terms of spoken language has not been performed yet. In this paper, we will focus on which spectral and cepstral features are used in language identification and perform a comparative study to check which feature vector performs better in terms of spoken language identification.

3. Dataset

Most Asian and African languages are resource-poor, which poses a great challenge for processing such languages in the automation domain. As the paper is related to the processing of resource-scarce languages, one of the main issues was the creation of a sizeable dataset that is not readily available. After the data collection, we trained and tested the model on six phonetically similar languages, which are the official languages of Jammu and Kashmir and the Ladakhi language (which too has official status in the Ladakh region and is also spoken in Tibet and Gilgit-Baltistan). The languages in this dataset belong to four groups, i.e., Tibetan, Indo-Aryan, Germanic, and European. We created our own dataset for four languages spoken in Jammu and Kashmir, i.e., Dogri, Urdu, Kashmiri, and Ladakhi. The dataset is recorded with standard recording devices from different regional radio and TV channels, containing telephonic conversations, formal communication, group discussions, and formal interviews. Recording was done at a sample rate of 16 kHz with a 16-bit resolution. For the other two languages, i.e., Hindi and English, we used VoxForge

and the IIT-H corpus. A total of 3000 audio files have been used, of which 1800 samples are used for model training and 600 samples are used for testing and validation. All files are in Wav. format with a duration of 5 seconds each. This dataset is intended to support research in the field of speech processing, particularly in the context of low-resource languages.

3.1. Dataset composition

Language Diversity: The dataset is meticulously curated to include a balanced representation of the four languages—Kashmiri, Dogri, Urdu, and Ladakhi. This diversity allows for a comprehensive analysis of the acoustic characteristics unique to each language.

Speaker Demographics: To ensure the dataset’s richness and diversity, speakers were selected from various age groups, genders, and socio-economic backgrounds, reflecting the natural variation in spoken language.

Recording Conditions: Recordings were captured in diverse environments, including indoor and outdoor settings, to simulate real-world scenarios and capture the variability in acoustic conditions.

3.2. Data collection process

Native Speakers: Native speakers were recruited to ensure authentic pronunciation and natural speech patterns for each language.

Script Variety: The dataset includes a variety of scripted and spontaneous speech to capture both formal and informal language use.

Recording Equipment: High-quality recording equipment was used to minimize noise interference and maintain the fidelity of the audio recordings.

3.3. Dataset statistics

Size: The dataset includes a substantial number of hours of audio recordings for each language, totaling several hundred hours, providing sufficient data for training and evaluating speech processing models.

Accent and Dialect Variation: The dataset accounts for accent and dialectal variations within each language, enhancing its representativeness.

Speaker Distribution: The dataset ensures a balanced distribution of speakers across different demographic categories, preventing biases in model training.

In conclusion, this audio dataset and its accompanying analysis provide valuable resources for advancing research in speech processing, particularly for languages with limited linguistic resources. The comprehensive nature of the dataset ensures its applicability in various research domains, from automatic speech recognition to linguistic studies of underrepresented languages.

4. Proposed model

Figure 1 depicts the basic architecture of our model, consisting of mainly three components, i.e., pre-processing, feature extraction, and classification. Pre-processing is done to remove high noise and silence using the Spectral Noise Gate (SNG) [37]. The pre-processed signal is then forwarded for feature extraction (using MFCC,

RASTA-PLP, and spectral features), and at the end, RNN-LSTM classifiers are used for classification. The raw input speech signals have some unwanted distortions or noise. Therefore, it cannot be processed directly through the LID model. Hence, the input speech signal should undergo a pre-processing method before the feature extraction process. The pre-processing method analyses the input speech signals to see if some background noise corrupts them. In this paper, the pre-processing of input speech signals is done using the SNG technique.

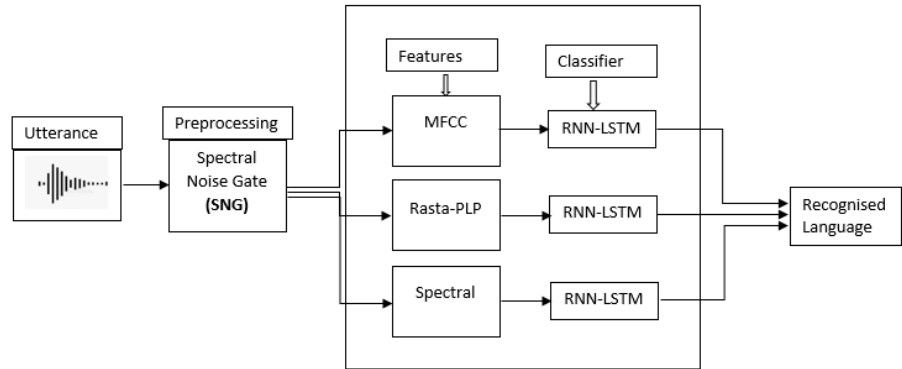


Figure 1. Architecture of model.

4.1. Spectral noise gate (SNG)

It is a common technique used for manipulation and resonant mix, which attenuates a signal corresponding to a certain threshold. This technique is also used in the general noise removal process. The sound spectrum has undergone some attenuation operations. The following steps follow the noise removal processing using SNG:

- 1) The noise audio clip is processed to calculate Fast Fourier Transform (FFT).
- 2) Statistics are calculated on the FFT of the noise in terms of frequency.
- 3) By using the calculated statistics, a threshold value is calculated.
- 4) FFT is calculated on the input signal.
- 5) The signal FFT and the threshold are compared to determine a mask value.
- 6) The determined mask is smoothed through a filter over time and frequency.
- 7) The smoothed mask is applied to the already calculated FFT of the input signal, and finally, the pre-processed signals are attained by applying inverse FFT.

Therefore, the SNG technique produced signal that is used for further feature extraction.

4.2. Feature extraction techniques

The pre-processed speech signals S_{kl}^{pr} are further given as input to the feature extraction process. The feature extraction is a process of deriving appropriate information from speech utterances. This paper uses two types of feature extraction methods such as cepstral features and spectral features. The cepstral features include MFCC, RASTA-PLP, and spectral features such as spectral roll-off, spectral flatness, spectral centroid, spectral bandwidth, and spectral contrast.

4.2.1. Cepstral features

This technique is used to separate the speech signals into their source and its system components. The features are less correlated in the cepstral domain, and it is invariant to amplitude and transition changes.

MFCC [38]: It is a predominant feature extraction method for signifying the information of speech utterances. The continuous speech utterances are recognized by matching the input signal $S_{kl}^{pr}(c)$ with a group of words or sentences. The first step is called parameterization. The input signal is transformed to the parameters to reduce the amount of redundant data. The familiar parameters in the recognition system are the MFCC. The MFCC is calculated by applying inverse Discrete Fourier transformation to the log of Fourier transformation of a time-domain signal as described below.

- 1) Separation of window: We know that audio signal changes continuously with time, as it very much difficult to deal with continues signals. We do segmentation to make continuous speech signals static for a particular time in order to calculate various cepstral and spectral feature various. A window size of 2048 samples with hope length of 512 has been used.
- 2) Periodogram Generation: In order to determine various frequencies present in speech segment Fast Fourier Transformation (FFT) is applied to each windowed frame. This is done as analogue with human ear which contains cochlear membrane that vibrates at different places with respect to incoming speech signal frequency. The output after applying FFT on a windowed frame is called periodogram.
- 3) Filterbanks application: In order to know how much energy is present in a specific frequency range, filterbanks were applied to bundle periodogram bins. This is done by applying mel-filterbanks and the formula for converting hertz to mel scale is given as

$$m = 2595 \log_{10} (1 + f/700)$$

- 4) Log application: As human ear is more concerned about the lower frequencies and linear changes, because we don't hear loud volumes linearly. To make it compatible for further we apply logarithms for two reasons, one for getting smaller values and other for channel normalization.
- 5) Applying Discrete Cosine Transformation: Applying discrete cosine transformation (DCT) on the log filterbank energies is the final step in computing MFCC features. Because our filterbanks are all overlap, and the energies of the filterbanks are highly connected. The DCT de-correlates the energies, allowing for the application of diagonal covariance metrics to represent the features.

The MFCC is calculated in Equation (1).

$$C(x(t)) = F^{-1} \log (F[x(t)]) \quad (1)$$

where $x(t)$ is the normal signal in the time domain, F is the discrete Fourier transformation, F^{-1} is the inverse discrete Fourier transformation (Discrete Cosine Transformation) and C represented the cepstrum. The number of MFCC coefficients used here are 128.

RASTA-PLP [36]: The short-term speech signals are represented by using Perceptual Linear Prediction (PLP) feature extraction method. RASTA-PLP is the

enhanced form of PLP method, which overcomes the limitations of PLP technique. This improved technique suppresses the adverse frequencies and increases the robustness of PLP in terms of noise. In RASTA-PLP method, critical band spectral resolution is applied for audible spectrum analysis, and a band-pass filter is employed for smoothening spectral variations, which are performed in each-frequency sub-band. This process leads to deriving a new spectral estimation, which is less prone to such variations. Then, the non-linear transformation process is done on the filtered speech signal spectral representation. The RASTA-PLP method tries to include the noise cancellation feature of the human auditory system and it is considered as the main advantage of this feature extraction method for the SLID system. The transformation function calculation of RASTA-PLP is formulated in Equation (2) using IIR filter as:

$$H(z) = 0.1 \times \frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{z^{-4}(1 - 0.98z^{-1})} \quad (2)$$

The fastest spectral change of the log spectrum is determined by the low cut-off frequency. The fastest spectral change is preserved in the output parameters, which is denoted by the high cut-off frequency. The higher values of the band-pass filter attenuate the convolution noise. The previous outputs are stored in the memory of RASTA filter. The current analysis results depend on these stored output values.

Thus, the RASTA-PLP extracted features are represented as FR_{dr}^{RASTA} , where the total number of RASTA-PLP extracted features are attained as 27.

4.2.2. Spectral features

The spectral features are a kind of frequency-based feature. They are commonly used to classify speech and audio signals. It is obtained by transforming the time-based signal into the frequency domain using FFT.

Spectral roll-off [39]: It is defined as the signal's kl^{th} percentile of spectral distribution. The percentage varies from 80 to 90 percent. It is the frequency obtained below which the magnitude distribution's percentile kl^{th} is concentrated. Spectral roll-off calculation for the given input signal S_{kl}^{pr} is given in Equation (3).

$$SRF = \sum_{qr=0}^{mn_r} |S_{kl}^{pr}(qr)| = \frac{CR}{100} \sum_{qr=0}^{MN-1} |S_{kl}^{pr}(qr)| \quad (3)$$

Here CR is the range, qr ranges from two band edges (0, MN-1). Hence, the spectral roll-off extracted features are represented as FR_{dr}^{SPR} .

Spectral flatness [39]: The spectral flatness finds the differentiation between noise and harmonic-like sounds. The spectral flatness is nearly zero for harmonic sounds and around one for noise-like sounds. In power spectrum, it measures the uniformity in the frequency distribution. It is computed as the ratio of the geometric mean to the arithmetic mean.

$$SFT = \frac{\prod_{qr=0}^{MC-1} |S_{kl}^{pr}(qr)|^{\frac{1}{MC}}}{\frac{1}{MC} \sum_{qr=0}^{MC-1} |S_{kl}^{pr}(qr)|} \quad (4)$$

Therefore, the spectral flatness extracted features are represented as FR_{dr}^{flat} .

Spectral centroid [39]: It is a spectral characterization metric used in Digital Signal Processing (DSP). It specifies where the spectrum's center of mass is located. It's calculated as the weighted mean of the signal's accessible frequencies, and it's defined by a Fourier transform using their magnitudes as weights, as in Equation (5).

$$SCD = \frac{\sum_{qr=0}^{MC-1} fe(qr)mg(qr)}{\sum_{qr=0}^{MC-1} mg(qr)} \quad (5)$$

Here, the weighted frequency value or the magnitude is denoted by $mg(qr)$, and its central frequency is denoted by $fe(qr)$ ranging from 0 to MC-1 (band edges). Hence, the spectral centroid extracted features are represented as FR_{dr}^{cen} .

Spectral bandwidth [39]: In speech signal, the spectral bandwidth is defined as the bandwidth of signal at one-half the peak maximum used to determine the narrowness of a wave spectrum is used in music classification and environmental sound recognition.

Hence, the spectral centroid extracted features are represented as FR_{dr}^{cen} .

Spectral Contrast [39]: It is defined as the difference between the peak values and valley values of the spectrum, used in music and music mood classification. Hence, the spectral contrast extracted features are represented as FR_{dr}^{cont} .

5. Experiments and results

One of the most prominent architectures of RNN (recurrent neural network) is LSTM, which processes the entire sequence, such as audio and video. Instead of processing a single data point, such as an image, to cover long-term dependencies, As RNN is facing vanishing gradient problems when training, LSTM overcame this problem to provide better results for problems that require longer sequences of data. LSTM is insensitive to the gap length of input in comparison to RNN, Markov models, and other sequence learning methods. We use the RNN-LSTM architecture, which contains five layers, i.e., input and output layers, with three hidden layers.

The number of neurons in the input layer is 128 in order to match the shape of the input feature vector, i.e., (128,409). The input is followed by three hidden layers with nodes 128, 64, and 32, respectively, activated by the non-linear function ReLu (Rectified Linear Activation Function). A dropout of 20%, 40%, and 40% neurons has been carried out to prevent overfitting. The Softmax function is used as an output layer to determine the number of classes, i.e., six. As we have calculated 128 MFCC features for each speech segment, to make it compatible with the input layer, we use 128 nodes. Then we gradually decrease the nodes to 64 and 32 to make it suitable for the output layer. The model is optimised by RMSprop (Root Mean Squared Propagation) using sparse categorical cross-entropy as a loss function, given by Equations (6) and (7). One of the reasons for using RMSprop as an optimizer is that it converges much faster to the local minima. In addition, it can overcome getting stuck in saddle points of any of the dimensions, which may be quite common in multi-dimensional functions, and works well with sparse problems by providing easy tuning parameters in comparison to stochastic gradient descent (SGD).

$$S(x)_i = \frac{\exp(x)_i}{\sum_{j=0}^5 \exp(x_j)} \quad (6)$$

From the above equation, S represents Softmax function, the input and output is represented by x_i and x_j respectively. The $\exp()$ represents exponentiation for both input and output vectors. Value of J ranges from 0 to 5 to determine number of categories in which an audio sample is to be classified.

$$\text{CCE} = - \sum_{j=0}^5 t_j \log(S(x)_j) \quad (7)$$

In the above equation $S(x)_j$ defines Softmax probability for the j -th class and t_j represents the truth label.

The distribution of data for our model is 60%, and 20% of the data is allocated for training and testing, containing 1800 and 600 files, respectively, and the remaining 20% of the data is used for validating the model's performance, containing 600 audio files. We train our model by choosing the optimal batch size and number of epochs as 64 and 100, respectively. The architecture of our model is shown in **Table 1**, which has MFCC trainable parameters of 272806. We did experiments on six languages, i.e., Hindi, Dogri, Kashmiri, Urdu, English, and Ladakhi. The data to be trained is first converted into a sequence of acoustic feature vectors like MFCC, RASTA-PLP, and spectral features (spectral roll-off, spectral flatness, spectral bandwidth, spectral centroid, and spectral contrast). The acoustic features serve as input to the RNN-LSTM classifier, which has RMSprop as an optimizer. The loss and accuracy of both training and testing data are represented by **Figures 2** and **3**, respectively.

Table 1. Model architecture of Recurrent Neural Network (Long Short Term Memory).

Model: "sequential_8"		
Layer (type)	Output shape	Param #
lstm_8 (LSTM)	(None, 128)	245760
dropout_24 (Dropout)	(None, 128)	0
dense_32 (Dense)	(None, 128)	16512
dense_33 (Dense)	(None, 64)	8256
dropout_25 (Dropout)	(None, 64)	0
dense_34 (Dense)	(None, 32)	2080
dropout_26 (Dropout)	(None, 32)	0
dense_35 (Dense)	(None, 6)	198
Total params: 272,806		
Trainable params: 272,806		
Non-trainable params: 0		

Table 2 shows the model's performance on validation data (20% of the data corpus) using standard performance metrics, i.e., recall, precision, F1 score, macro accuracy, weight accuracy, and accuracy, using three different feature vectors, i.e., MFCC, RASTA-PLP, and spectral. It was found that Kashmir shows the highest precision of 98% and Hindi shows the lowest precision of 61% by using MFCC features. By using RASTA-PLP features, English shows the highest precision of 96%, while Urdu shows the lowest precision of 47%. By using spectral features, English shows the highest precision of 98%, while Urdu shows the lowest precision of 50%. Overall, Kashmiri shows the highest average precision of 95%, while Urdu shows the lowest average precision of 64%. By analysing the various performance metrics in **Table 2**, it is found that MFCC performed well in comparison to RASTA-PLP and spectral features.

Table 2. Performance of different cepstral and spectral features using standard performance metrics.

	MFCC			RASTA_PLP			SPECTRAL		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score	Recall	Recall	F1-Score
Dogri	00.99	00.93	00.96	00.98	00.88	00.92	00.61	00.85	00.71
English	00.60	00.62	00.61	00.46	00.96	00.62	00.86	00.98	00.91
Hindi	00.61	00.61	00.61	00.86	00.65	00.74	00.96	00.88	00.92
Kashmiri	00.98	00.98	00.98	00.82	00.90	00.86	00.85	00.97	00.90
Ladakhi	00.90	00.90	00.90	00.78	00.82	00.65	00.60	00.75	00.67
Urdu	00.92	00.95	00.93	00.67	00.47	00.63	00.82	00.50	00.62
Macro average	00.83	00.83	00.83	00.76	00.78	00.73	00.78	00.83	00.79
Weighted average	00.83	00.83	00.83	00.76	00.78	00.73	00.78	00.83	00.79
Accuracy	00.83			00.76			00.78		

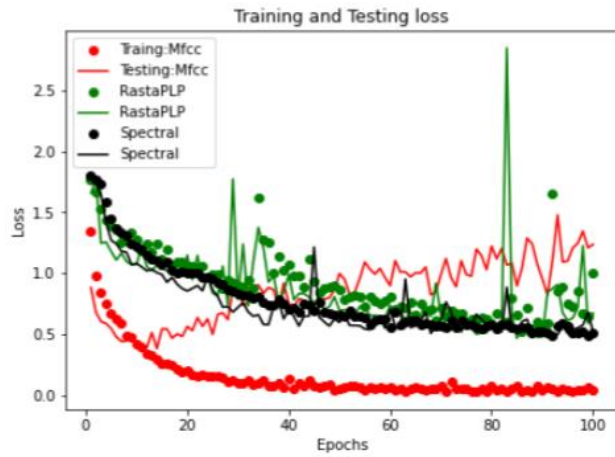


Figure 2. Cross-Category loss values of three feature vectors using training and testing data.

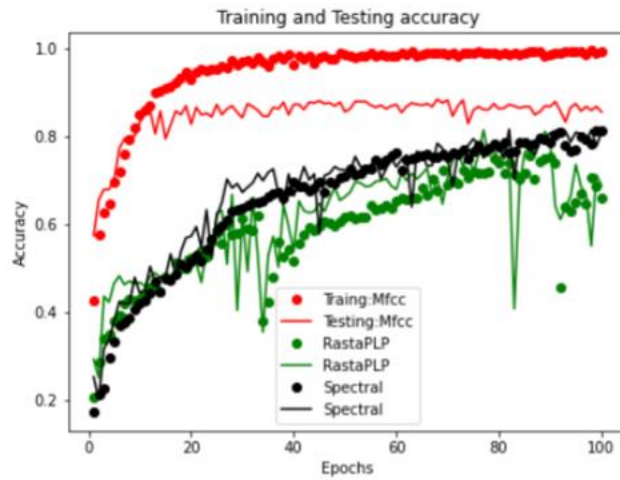


Figure 3. Accuracy of three feature vectors using training and testing data.

6. Conclusion and future work

Language identification of an audio signal is a basic task in multi-lingual speech processing systems, as it acts as a front end for various NLP tasks dealing with speech. The paper presents the comparison of various cepstral and spectral features in the process of spoken language identification using RNN-LSTM. The model has been

trained and tested on the official languages of Jammu and Kashmir (Union Territory) and Ladakhi. Various features of audio signals, such as cepstral and spectral, are used to model the language identification system. The dataset that is used in experimentation consists of some readily available portions for Hindi and English and self-created portions for the other four languages. The model has been trained on 100 epochs with RMSProp as an optimizer. We evaluate relative performance in terms of common performance criteria like precision, recall, f1-score, and accuracy. On average, MFCC, RASTA-PLP, and spectral characteristics are recognized at 83%, 76%, and 78%, respectively. The results indicate that utilising a recurrent neural network (LSTM) classifier, MFCC is the best feature for language identification. The model thus created is first of its nature related to the given set of languages, as it contains a good number of languages for which least work has been done and the Ladakhi language for which it is the first of its kind to map it with automatic processing. The main challenge was the dataset, which we handled gracefully. One of the issues with our dataset is that it lacks generality, i.e., sampling bias, limited context, temporal dynamics, feature representation, and domain shift, which lowers the accuracy of the model as the dataset is homogeneous, which needs to be considered for future work. Moreover, the model may be updated by using CNN, GRU, and attention mechanisms along with expanded datasets to make the system more relevant regarding the extraction of feature sets and subsequent linguistic categorization.

Author contributions: Conceptualization, IAT and RB; methodology, IAT; software, IAT; validation, IAT and RB; formal analysis, IAT; investigation, IAT; resources, IAT; data curation, IAT; writing—original draft preparation, IAT, RB and KJG; writing—review and editing, IAT, RB and KJG; visualization, IAT; supervision, RB and KJG. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. China Bhanja C, Laskar MA, Laskar RH. Modelling multi-level prosody and spectral features using deep neural network for an automatic tonal and non-tonal pre-classification-based Indian language identification system. *Language Resources and Evaluation*. 2021, 55(3): 689-730. doi: 10.1007/s10579-020-09527-z
2. Lee HS, Tsao Y, Jeng SK, et al. Subspace-Based Representation and Learning for Phonotactic Spoken Language Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020, 28: 3065-3079. doi: 10.1109/taslp.2020.3037457
3. Chandak C, Raesy Z, Rastrow A, et al. Streaming language identification using combination of acoustic representations and ASR hypotheses. *arXiv*. 2020. doi.org/10.48550/arXiv.2006.00703
4. Gemmeke JF, Van Hamme H, Cranen B, et al. Compressive Sensing for Missing Data Imputation in Noise Robust Speech Recognition. *IEEE Journal of Selected Topics in Signal Processing*. 2010, 4(2): 272-287. doi: 10.1109/jstsp.2009.2039171
5. Wang P, Tan K, Wang DL. Bridging the Gap Between Monaural Speech Enhancement and Recognition With Distortion-Independent Acoustic Modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020, 28: 39-48. doi: 10.1109/taslp.2019.2946789
6. Albadr MAA, Tiun S, Ayob M, et al. Mel-Frequency Cepstral Coefficient Features Based on Standard Deviation and Principal Component Analysis for Language Identification Systems. *Cognitive Computation*. 2021, 13(5): 1136-1153. doi: 10.1007/s12559-021-09914-w
7. Biswas M, Rahaman S, Kundu S, et al. Spoken Language Identification of Indian Languages Using MFCC Features. *Machine Learning for Intelligent Multimedia Analytics*. Published online 2021: 249-272. doi: 10.1007/978-981-15-9492-

- 2_12
8. Wicaksana VS, S.Kom AZ. Spoken Language Identification on Local Language using MFCC, Random Forest, KNN, and GMM. *International Journal of Advanced Computer Science and Applications*. 2021, 12(5). doi: 10.14569/ijacsa.2021.0120548
 9. Athiyaa N, Jacob G. Spoken language identification system using MFCC features and gaussian mixture model for Tamil and Telugu Languages. *International Research Journal of Engineering and Technology(IRJET)*. 2019, 6(4): 4243–4248.
 10. Das A, Guha S, Singh PK, et al. A Hybrid Meta-Heuristic Feature Selection Method for Identification of Indian Spoken Languages From Audio Signals. *IEEE Access*. 2020, 8: 181432-181449. doi: 10.1109/access.2020.3028241
 11. Das HS, Roy P. Bottleneck Feature-Based Hybrid Deep Autoencoder Approach for Indian Language Identification. *Arabian Journal for Science and Engineering*. 2020, 45(4): 3425-3436. doi: 10.1007/s13369-020-04430-9
 12. Qu D, Wang B, Wei X. Language identification using vector quantization. In: *Proceedings of the 6th International Conference on Signal Processing; 26–30 August 2002; Beijing, China*. 492–495. doi: 10.1109/ICOSP.2002.1181100
 13. Maity S, Kumar Vuppala A, Rao KS, et al. IITKGP-MLILSC speech database for language identification. 2012 National Conference on Communications (NCC). Published online February 2012. doi: 10.1109/ncc.2012.6176831
 14. Sarthak, Shukla S, Mittal G. Spoken Language Identification Using ConvNets. *Ambient Intelligence*. Published online 2019: 252-265. doi: 10.1007/978-3-030-34255-5_17
 15. Lopez-moreno I, Gonzalez-dominguez J, Plchot, D. Martinez O, et al. Google Inc ., New York, USA ATVS-Biometric Recognition Group, Universidad Autonoma de Madrid, Spain Brno University of Technology, Czech Republic Aragon Institute for Engineering Research (I3A), University of Zaragoza, Spain. 2014. pp. 0–4.
 16. Hermansky H, Morgan N. RASTA processing of speech. *IEEE Transactions on Speech and Audio Processing*. 1994, 2(4): 578-589. doi: 10.1109/89.326616
 17. Hermansky H, Morgan N, Bayya A, Kohn P. RASTA-PLP speech analysis technique. In: *Proceedings of the ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing; 23–26 March 1992; San Francisco, CA, USA*. pp. 121-124. doi: 10.1109/ICASSP.1992.225957
 18. Kingsbury BED, Morgan N. Recognizing reverberant speech with RASTA-PLP. In: *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing; 21–24 April 1997; Munich, Germany*. pp. 1259–1262. doi: 10.1109/ICASSP.1997.596174
 19. Razia Sulthana A, Mathur A. A State of Art of Machine Learning Algorithms Applied Over Language Identification and Speech Recognition Models. *International Virtual Conference on Industry 40*. Published online 2021: 123-132. doi: 10.1007/978-981-16-1244-2_10
 20. Ghanghor N, Krishnamurthy P, Thavareesan S, et al. IITK@DravidianLangTech-EACL2021: Offensive language identification and meme classification in Tamil, Malayalam and Kannada. In: *Chakravarthi B, Priyadharshini R, Kumar MA, et al., Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages; 20 April 2021; Kyiv, Ukraine. Association for Computational Linguistics; 2021*. pp. 222–229.
 21. Anusuya MA, Katti SK. Speech recognition by machine: A review. *International Journal of Computer Science and Information Security*. 2009. 6(3): 181–205.
 22. Schutte KT. *Parts-Based Models and Local Features for Automatic Speech Recognition [PhD thesis]*. Massachusetts Institute of Technology; 2009.
 23. Deshwal D, Sangwan P, Kumar D. Feature Extraction Methods in Language Identification: A Survey. *Wireless Personal Communications*. 2019, 107(4): 2071-2103. doi: 10.1007/s11277-019-06373-3
 24. Han W, Chan CF, Choy CS, Pun KP. An efficient MFCC extraction method in speech recognition. In: *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems (ISCAS); 21–24 May 2006; Kos, Greece*. pp. 145–148. doi: 10.1109/ISCAS.2006.1692543
 25. Dewi Renanti M, Buono A, Ananta Kusuma W. Infant cries identification by using codebook as feature matching, and MFCC as feature extraction. *Journal of Theoretical and Applied Information Technology*. 2013, 56(3): 437–442.
 26. Trang H, Tran Hoang Loc, Huynh Bui Hoang Nam. Proposed combination of PCA and MFCC feature extraction in speech recognition system. 2014 International Conference on Advanced Technologies for Communications (ATC 2014). Published online October 2014. doi: 10.1109/atc.2014.7043477
 27. Ahmed AI, Chiverton JP, Ndzi DL, et al. Speaker recognition using PCA-based feature transformation. *Speech Communication*. 2019, 110: 33-46. doi: 10.1016/j.specom.2019.04.001

28. Krishna SR, Rajeswara R. SVM based emotion recognition using spectral features and PCA. *International Journal of Pure and Applied Mathematics*. 2017, 114(9): 227–235.
29. Sabab MdN, Chowdhury MAR, Nirjhor SMMI, et al. Bangla Speech Recognition Using 1D-CNN and LSTM with Different Dimension Reduction Techniques. *Emerging Technologies in Computing*. Published online 2020: 158-169. doi: 10.1007/978-3-030-60036-5_11
30. Saleh MAM, Ibrahim NS, Ramli DA. Data reduction on MFCC features based on kernel PCA for speaker verification system. *WALIA Journal*. 2014, 30(S2): 56–62.
31. Winursito A, Hidayat R, Bejo A. Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition. *2018 International Conference on Information and Communications Technology (ICOIACT)*. Published online March 2018. doi: 10.1109/icoiact.2018.8350748
32. Mukherjee H, Obaidullah SM, Santosh KC, et al. A lazy learning-based language identification from speech using MFCC-2 features. *International Journal of Machine Learning and Cybernetics*. 2019, 11(1): 1-14. doi: 10.1007/s13042-019-00928-3
33. Boussaid L, Hassine M. Arabic isolated word recognition system using hybrid feature extraction techniques and neural network. *International Journal of Speech Technology*. 2017, 21(1): 29-37. doi: 10.1007/s10772-017-9480-7
34. Guha S, Das A, Singh PK, et al. Hybrid Feature Selection Method Based on Harmony Search and Naked Mole-Rat Algorithms for Spoken Language Identification From Audio Signals. *IEEE Access*. 2020, 8: 182868-182887. doi: 10.1109/access.2020.3028121
35. Bashir R, Quadri S. Identification of Kashmiri script in a bilingual document image. *2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*. Published online December 2013. doi: 10.1109/iciip.2013.6707658
36. Thukroo IA, Bashir R. Spoken Language Identification System for Kashmiri and Related Languages Using Mel-Spectrograms and Deep Learning Approach. *2021 7th International Conference on Signal Processing and Communication (ICSC)*. Published online November 25, 2021. doi: 10.1109/icsc53193.2021.9673212
37. van Keeken A. *Understanding Records. A Field Guide to Recording Practice*. Second Edition. By Jay Hodgson. New York: Bloomsbury, 2019. 233 pp. ISBN 978-1-5013-4237-0. *Popular Music*. 2021, 40(1): 172-174. doi: 10.1017/s0261143021000192
38. Deshwal D, Sangwan P, Kumar D. A Language Identification System using Hybrid Features and Back-Propagation Neural Network. *Applied Acoustics*. 2020, 164: 107289. doi: 10.1016/j.apacoust.2020.107289
39. Sharma G, Umopathy K, Krishnan S. Trends in audio signal feature extraction methods. *Applied Acoustics*. 2020, 158: 107020. doi: 10.1016/j.apacoust.2019.107020

Article

Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse software learning needs with a single document knowledge base

Yuchen Wang¹, Yin-Shan Lin^{2,*}, Ruixin Huang³, Jinyin Wang⁴, Sensen Liu⁵

¹ School of Architecture, University of Hawaii at Manoa, 2410 Campus Road, Honolulu, HI 96822, United States

² Khoury College of Computer Science, Northeastern University, 360 Huntington Ave, Boston, MA 02115, United States

³ Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, United States

⁴ Department of Computer Science, Stony Brook University, 100 Nicolls Rd, Stony Brook, NY 11794, United States

⁵ Department of Electrical and Systems Engineering, Washington University in St. Louis, 1 Brookings Dr, St. Louis, MO 63130, United States

* Corresponding author: Yin-Shan Lin, lilianlin003@163.com

CITATION

Wang Y, Lin Y, Huang R, et al.
Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse software learning needs with a single document knowledge base.
Computing and Artificial Intelligence. 2024; 2(1): 535.
<https://doi.org/10.59400/cai.v2i1.535>

ARTICLE INFO

Received: 2 February 2024

Accepted: 1 April 2024

Available online: 19 April 2024

COPYRIGHT



Copyright © 2024 by author(s).

Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.

<https://creativecommons.org/licenses/by/4.0/>

Abstract: This paper begins with a theoretical exploration of the rise of large language models (LLMs) in Human-Computer Interaction (HCI), their impact on user experience (HX) and related challenges. It then discusses the benefits of Human-Centered Design (HCD) principles and the possibility of their application within LLMs, subsequently deriving six specific HCD guidelines for LLMs. Following this, a preliminary experiment is presented as an example to demonstrate how HCD principles can be employed to enhance user experience within GPT by using a single document input to GPT's Knowledge base as new knowledge resource to control the interactions between GPT and users, aiming to meet the diverse needs of hypothetical software learners as much as possible. The experimental results demonstrate the effect of different elements' forms and organizational methods in the document, as well as GPT's relevant configurations, on the interaction effectiveness between GPT and software learners. A series of trials are conducted to explore better methods to realize text and image displaying, and jump action. Two template documents are compared in the aspects of the performances of the four interaction modes. Through continuous optimization, an improved version of the document was obtained to serve as a template for future use and research.

Keywords: Large Language Models (LLMs); Human-Computer Interaction (HCI); User Experience (UX); Human-Centered Design (HCD); GPTs; knowledge base; user needs

1. Introduction

Since the emergence of Large Language Models (LLMs) and blowout development from 2022, their integration with Human-computer interaction (HCI) marks the beginning of a new chapter in this interaction. This shift heralds a shift away from traditional HCI, which primarily relied on graphical user interfaces and command-line inputs, toward more sophisticated AI-driven interfaces and models. As Gokul [1] points out, LLMs are reshaping the Artificial Intelligence (AI) landscape with their advanced capabilities in processing and generating human-like language. Their applications extend into various creative domains, including music, art, and storytelling. However, in the aspect of user experience (UX), the LLMs and their applications still present challenges.

2. The impact of LLMs on HCI and UX

2.1. Redefining UX with LLMs in HCI

LLMs have been pivotal in transforming UX.

One of the big transformations brought about by LLMs is personalization. These models analyze user data and learn from individual interaction patterns to tailor responses and suggestions.

The incorporation of advanced Natural Language Processing (NLP) capabilities in LLMs marks another stride forward. This development allows for more intuitive and human-like interactions.

Additionally, context-aware interactions signify a significant advancement in HCI, brought about by LLMs [2]. These models not only recognize words but also comprehend the context of user requests [3] and predict user's preference [4].

2.2. Challenges from UX

As we transition from the exploration of the positive advancements of LLMs in HCI, it becomes imperative to critically examine the multifaceted challenges that accompany this technological integration.

2.2.1. Ethical consideration

As for ethical consideration, while LLMs offer immense potential in HCI, they introduce complex ethical challenges that significantly impact user experience. Ethical challenges mainly come from two aspects: The technology inherent defects, such as specification gaming and side effects [5], pressure to deploy unsafe systems [6] and risks from advanced misaligned AI [7], and inappropriate use, such as Misinformation Harms and Malicious Uses [8].

2.2.2. Supportiveness of user needs

The integration of LLMs into HCI presents a range of technical complexities to meet user's advanced needs, such as the need for higher-speed content generation and more accuracy to the background context, which involves transformers, tokens, reinforcement learning from human feedback (RLHF) and natural language processing (NLP) [9].

Additionally, models often produce outright fabrications that may appear plausible [10]. It's widely acknowledged, through both research and anecdotal evidence, that LLMs often face a pervasive problem of hallucination, or "hallucinated" content.

Subramonyam et al. [11] focus on integrating user experience and needs into the AI development process, finding the problems such as low-level design and share information across expertise boundaries. Zhang et al. [12] use LLM to answer student questions classified into four types, and finds the system effectively ignores questions that it cannot address.

3. Introducing HCD to LLMs and their applications

3.1. Principles of HCD

HCD, or HCAI, introduced by Don [13], is a problem-solving approach with its core positioning real individuals at the center of the development process. This approach is focused on delivering equitable results and upholding the utmost respect for privacy, thereby aligning AI functionalities with human values and ethics [14]. The essence of HCD lies in consistently prioritizing the user's desires, challenges, and preferences throughout every stage of the design and development process [15].

Major principles of UCD includes early and active involvement of the user during the design process, clarification of user, user feedback is incorporated into the product's lifecycle and the product is improved using an iterative design process [16]. For instance, Jaimes et al. [17] emphasize the importance of mixed-initiative human-computer systems, highlighting how user input plays a crucial role in shaping the functionality and responsiveness of these systems. Similarly, Mack et al. discusses the criticality of including diverse user perspectives in research methods, ensuring that systems are accessible and meet varied user needs [18].

Some research explores the way to encourage public early participation in public decision making or affecting users' climate-controlling behavior by using new technology, such as augmented reality (AR) [19] and virtual reality (VR) [20], which are also applications of HCAI. Research by Seffah and Andreevskaia [21] developed a skill-oriented program towards developers and students based on analyzing UCD knowledge and techniques.

3.2. Previous attempts to reflect HCD in LLMs applications

This study mainly focusses on the methods of enhancing the supportiveness of user needs by applying HCD principles. HCD prioritizes the needs, preferences, and contexts of users [22], ensuring that LLM-driven interactions are not only efficient but also resonate with the users' expectations.

Petridis et al. [23] explore the possibility of incorporating prompt-based prototyping into designing functional user interface (UI) mock-ups, finding LLMs potentially reduce the time needed to create a functional prototype. Park and Choi [24] introduce LLMs into audience simulation for public speech and uses AudiLens to provide flexibility to the speaker. Di Fede et al. [25] introduce the Idea Machine combined with LLMs to empower people engaged in idea generation tasks.

Korbak et al. [26] explore alternative objectives for pretraining LMs (Language Model) to create text aligned with human preferences. Study by Rastogi et al. [27] find existing auditing tools use either or both humans and AI to find failures. They create the evaluation tool: AdaTest++, which is powered by GPT3 and Azure's sentiment analysis model.

3.3. Build HCD guidelines to enhance UX in LLMs

From the above discussion, the HCD principles related to LLMs can be concluded as the following (**Table 1**).

These six guidelines are crucial to LLMs like GPT in meeting user expectations and needs effectively. They also provide possible ways to optimize related design including AI agent, application, platform, user interface and the construction of knowledge base.

Table 1. HCD guidelines related to LLMs for enhancing UX.

Principles	Requirements
High efficiency	Fast response
Feedback consideration	Collect feedback; Update periodically or imperiodically
High supportiveness for diverse needs	Generality and specificity consideration; Personalization and Customization
Emotional consideration	Understanding the emotional of user conversations and providing more humanized interactions
High Simplicity	Easy input; Effortless expression; Multimodal input and output: images, text, voice, etc.
High Reliability	Authenticity; Accuracy

In the following section, a preliminary experiment is conducted to apply these HCD principles into the enhancement of LLM’s interaction capabilities.

4. Improving UX by optimizing a single document as principal knowledge in GPT: A preliminary experimental study

This preliminary experiment mainly focuses on the UX enhancement from the aspect of supportiveness for users’ diverse needs. Other HCD principles, such as simplicity and reliability, will also be taken into consideration in the experiment design. The experiment takes ChatGPT-4 as an example, exploring how to use a single document as the main material of knowledge base to construct a custom GPT.

4.1. Materials and methods

4.1.1. Virtual experimental environment: ChatGPT-4 and GPTs editor

The working environment for this study is set in ChatGPT-4 and GPTs Editor.

GPTs editor is a relatively new function as one part of ChatGPT-4. It’s a specialized environment for creating and tuning GPT models based on GPT editor’s preset configuration, including descriptions of this GPT, instructions, knowledge, starters and actions, allowing adjustments to the model’s responses, capabilities, and interaction style. In “Configure” interface, the “Instructions” area provides overall control rules for GPT to follow during interaction. “Conversation starters” allows users to start a conversation by just clicking corresponding buttons. “Knowledge” provides a preset knowledge base where editor can upload files as data in certain formats, such as docx, pdf or jpg.

After the new GPT being created, it will be imported in ChatGPT-4 automatically, which provides an environment for users to interact.

4.1.2. Principal objective: Meeting users’ diverse needs in software learning interaction

This study defines a goal as taking ChatGPT-4 as a software learning tool that provides knowledge and solutions for novices in learning a new software. This hypothetical scenario is designed to simulate how LLMs can synthesize newly inputted knowledge and utilize it in multiple ways, which can be considered as one of the typical applications which use LLMs to serve a specific group of people. UX in this study can be evaluated by the quality of dialogues during interaction.

Visual Scripting, a tool inside Unity software, is taken as the software in the optimization process. It allows for the creation of logic and game behaviors without

writing code directly. By using visual graphical nodes and connecting them with lines, Unity developers can construct complex game logic and interactions. The advantages are as follows:

- (1) GPT has less inherent knowledge about Visual Scripting itself, even if some related coding knowledge is trained into GPT. Therefore, the pre-existing knowledge will less interfere the evaluation.
- (2) The Visual Scripting Manual on official website can be used as a reliable source for constructing the knowledge for GPT.
- (3) The images of how to use Visual Scripting are easier to make and since it is an intuitive tool.

4.1.3. New knowledge resource: A single document input in knowledge base

A single document is used as the main new knowledge resource uploaded in GPTs Editor’s “Knowledge” area. It is a Microsoft Word document in docx format, serving as the new knowledge resource and control module. It is composed of a control part and a software knowledge part. The advantages of using a single document are as follows:

- (1) Simplicity and customization consideration: It is easy for a real creator to replace certain parts of the template document to make another GPT as a tutorial for learning other software.
- (2) Compatibility consideration: The docx document can contain the knowledge both in forms of natural languages or codes. The arrangement of content is also easy to be adjusted.
- (3) Variables control: To avoid black box effect which often exists in AI product, the single document can be easily optimized, which helps to explore a method of getting a relatively controllable result.

4.1.4. Users’ needs and requirements definition

Different groups of users may have different needs for the usage of a software learning GPT, while a single user may also have needs for multiple ways to use it. The following diagram shows the possible needs (Figure 1).

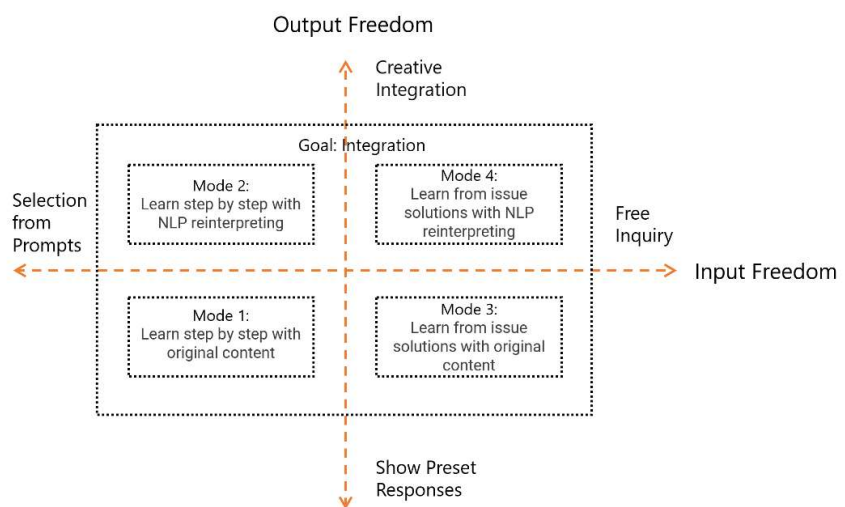


Figure 1. Software learners’ diverse needs in using this GPT.

It can be considered that different using modes of this GPT are based on users’

requirements for varying degrees of input and output freedom. The overall goal is to integrate these modes.

For the input freedom, inputting single a number or letter based on the given prompts is an alternative to select a desired action, such as to start the tutorial, or jump to a certain section of the tutorial, which requires less input freedom. Users also have the demands for inputting a complex issue and then getting solutions, which requires more input freedom.

For the output freedom, the alternative of strictly showing the original content from the knowledge part of the document is needed, meaning less output freedom, which can be applied in the scene that users hope to strictly obey the software guidance from a traceable source. In other cases, the output needs to display content in a creative way by using more natural and coherent language to rewrite and reorganize the knowledge, meaning more output freedom.

Therefore, four types of modes are supposed to be realized:

Mode 1: Learning step-by-step with original content, enabling users to learn from printed original content retrieved from the knowledge bases words by words.

Mode 2: Learning step-by-step, similar to the previous one, but use NLP to reinterpret original content.

Mode 3: Learning by issue solutions, allowing users to receive solutions for their issues while using Visual Scripting, and the solutions should print the original sentences of related knowledge.

Mode 4: Learning by issue solutions, similar to the previous one, but use NLP to reinterpret original content.

4.1.5. Expected outcomes

The overall optimization process can be illustrated as the following diagram (Figure 2).

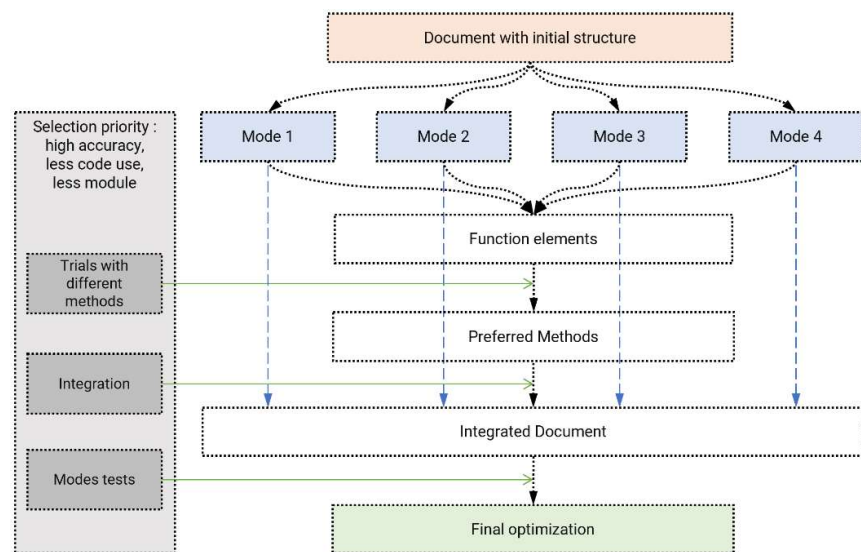


Figure 2. Document optimization process.

The process involves a series of examinations. Firstly, different functions to realize these modes will be analyzed. Different methods will be tested to within the document initial structure. Then, these methods will be filtered and selected based on

the results of tests. The document will be adjusted by using the preferred methods. The four modes will be tested within the new document to check the result of performance and finally, it will be optimized again based on the result and then be retested. The selection principles in each step include high accuracy, less code use and less module use.

The expected outcomes of this study are as follows:

- (1) To figure out how the elements inside the single knowledge document as well as the GPT configurations will affect interaction quality.
- (2) To explore the possibility to integrate this GPT's four modes.
- (3) If possible, through optimization process, a document that can better integrate these modes get be finally obtained as a template for future use, which can be seen as the application of "feedback consideration" HCD principle.

4.1.6. Methods of UX evaluation

The method of UX evaluation is to assess the accuracy of interaction rather than speed. To emphasize the ability of GPT comprehensively utilizing newly input knowledge, is supposed to control the variables in the input-output process. The following methods are employed to prevent GPT's directly using knowledge of Visual Scripting to interfere the evaluation:

- (1) Multiple forms of knowledge composition

The images of Visual Scripting nodes and connections are taken as knowledge together with text. Some of the original text and images from Unity Visual Scripting Manual 1.9.1 version [28] is extracted or rewritten and then be placed into the document.

- (2) Closing web browsing

Web browsing action in GPT may introduce original online resources, so closing it can isolate environment.

- (3) Methods of output control

The output is supposed to mix text and images from the knowledge, making it challenging to achieve user's goals.

The average accuracy of the results in each trial will be assessed through 5 times of repetitive complete chatting, the functional elements of the four modes and their evaluation criteria of the interaction result are as follows (**Table 2**).

Table 2. Criteria for approximate accuracy assessment.

Functional Elements and Criteria List for modes	
(A) Whether the jump action is successful and smooth;	
(B) Whether the output obey the sequences of the original steps;	
(C) Whether related images can be successfully displayed together with text.	
(D) Whether the output display original content in each step completely;	
(E) How much the reinterpreting using NLP deviates the original contents in the document, producing wrong content or "hallucinated" content (content that seems to be correct but has no relevance with the document original content);	
(F) How helpful the selected content is to the user's question (the designed questions are designed to be satisfied by some certain parts in the document);	
(G) Whether the output solutions cross enough necessary range of knowledge chapter in the document.	
Mode 1	Mode 2
(A), (B), (C), (D)	(A), (B), (C), (E)
Mode 3	Mode 4
(C), (D), (F), (G)	(C), (E), (F), (G)

For testing the function in Mode 3 and 4 that answer user’s questions, the question list is designed as follows (Table 3).

Table 3. Question list for Mode 3 and 4.

Question Number	Content
Q1	How can I use nodes to change the sprite of GameObject “A” when a time duration finishes?
Q2	How can I use nodes to sets the velocity of GameObject “B” to half of its original velocity when a “B” enters a trigger collider in 2D space?
Q3	How can I use nodes to add value 1 to the existing object variable named “C” and set back to “C”?
Q4	How can I make another Script Graph named “D” inside a Script Graph named “E” to receive a float from “Script E”, then returns true if the float is greater than 1 and less than 2, otherwise returns false to “Script E”?
Q5	If UI button “G” has a Script Graph named “H” and GameObject “J” has a Script Graph named “K”, how can I use nodes in “H” to trigger the event in “K” when clicking the button “G”?

These questions are with high complexity and less specificity, meaning to require crossing different sections of knowledge part in the document to find answers, and less mention any specific name of node and the phrase “Visual Scripting”. The intent is to make it easy to recognize whether it use the new knowledge (Figure 3).

You
In Visual Scripting, if I want to make a flow that called at every frame while the graph is active, what node can I use to start with?

ChatGPT
In visual scripting, particularly in game engines like Unity (with Bolt) or Unreal Engine (with Blueprints), if you're looking to execute a flow at every frame while the graph is active, you'd generally use a node that responds to the game's update cycle.

For Unity using Bolt, the node you're looking for is often called 'Update' or 'On Update'. This event is called once per frame and is the place where you can execute logic that needs to check or perform actions every frame.

(a)

You
How can I use nodes to add value 1 to the existing object variable named "C" and set back to "C"?

ChatGPT
Using nodes to manipulate an object's variable depends on the context you're working within. Let's consider a couple of common scenarios where nodes might be used: game development with a node-based engine like Godot, and a hypothetical scenario in a generic programming environment that organizes logic or operations in a "node-like" structure.

In a Game Engine (e.g., Godot)
If you're using a game engine that organizes entities as nodes (like Godot), and you want to increment a variable named 'C' in an object, you'll likely be working with scripts attached to those nodes. Assuming 'C' is a property of an object (node) that has a script component, here's how you could increment 'C':

(b)

Figure 3. GPT’s responses to different questions without new document input. (a) Question with less complexity and high specificity; (b) Question with high complexity and less specificity.

4.1.7. Structural design of the single document

The initial structure of the new document is designed as follows Figure 4.

The document includes two parts: Part 0 provides an introduction and response methods, serving as a general control part; Part 1 provides the knowledge of Visual Scripting, structured into chapters and sections based on the content. After each output, users can change interaction mode directly.

Here is a brief introduction of these parts:

Part 0 includes:

(1) “Overview of This GPT’s Rules” section: It provides general rules to control the interaction. It includes the following parts:

a) Descriptions of this GPT and the document.

b) The general rules, such as “Refer to Section 2 in Part 0 for initial dialogue rules” and “Interactive requests needing user responses are enclosed in braces {}, such as {Enter 1: Continue}”. Prohibited interaction ways area also provided, such as

“Refuse interactions that does not meet current interactive requests inside {}”.

(2) “Initial Response Method for Dialogues” section: It provides ways to go to different parts corresponding to the input Starters.

(3) “Start learning” section: It provides ways to process Part 1 section by section.

(4) “Finding Solution” section: It provides ways to provide solutions to user’s questions.

(5) “Introduction” section: It provides the basic information about how to use this GPT.

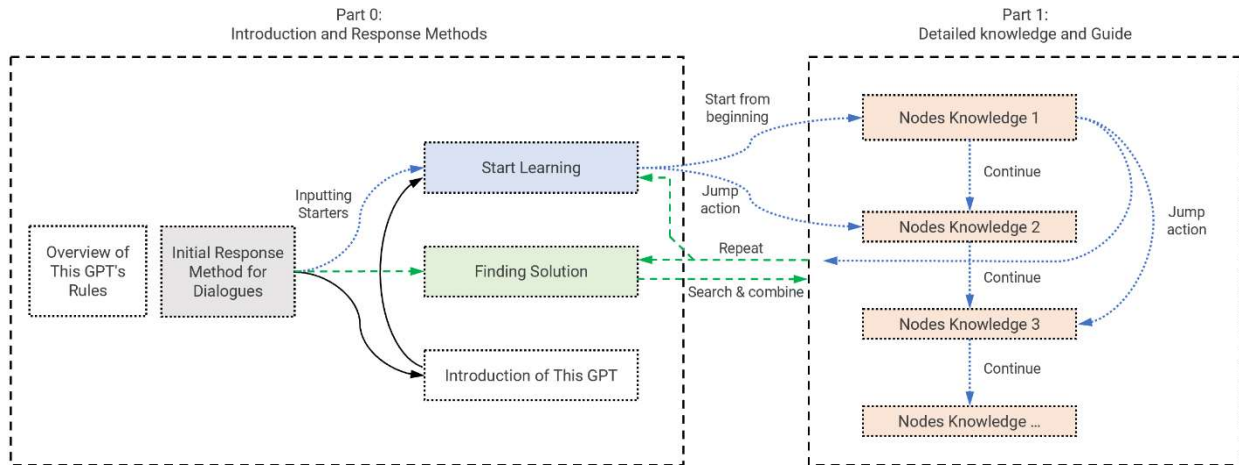


Figure 4. Initial structure of the single document.

Part 1 includes:

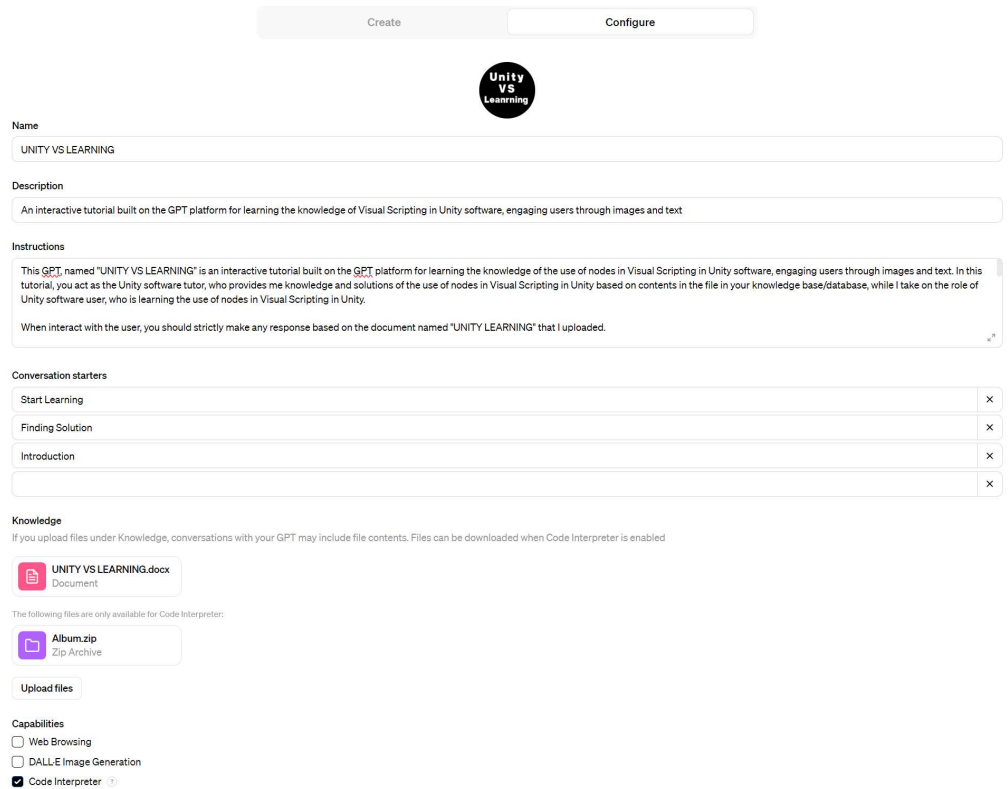


Figure 5. Configuration of GPT.

14 Chapters of detailed knowledge are provided as well as some guidance. There is a title of each chapter and each section, but no overall introduction in each Chapter.

It is found that the extraction of images directly from the Word file is not possible. Therefore, a zip file containing multiple images is uploaded as supplementary material. Each image is named in following format: “Chapter number” + “Section number” + the image’s sequence number, such as “060401”. Also, the “Instructions” area is filled with content from “Overview of This GPT’s Rules” section. In addition, the code interpreter is turned on for processing code in the document. The configuration of the new GPT is set as **Figure 5**.

4.2. Module and structure trials: Results and analysis

4.2.1. Original text display

For displaying the original text in the document, different ways are examined by several trials.

(1) Trial 1: Executing the “print ()” function when jumping from somewhere else

Here is an example. The instructions in Part 0’s “Overview of This GPT’s Rules” reads: “Text that needs to be directly printed will be with clear instructions such as ‘print (Hello)’, and will be enclosed in brackets marks ().” The instruction in “Initial Response Method for Dialogues” in Part 0 reads: “If ‘Start Learning’ is inputted by the user, go to Part0, Section 3A.” and in Section 3A, it uses “print ()” function in each step following a serial number, such as “1. Hi. Welcome...” and “2. The following is...”. The result shows it can proceed printing text step by step easily as following screenshot (**Figure 6**). The limitation is it has to proceed from the first step in a section.

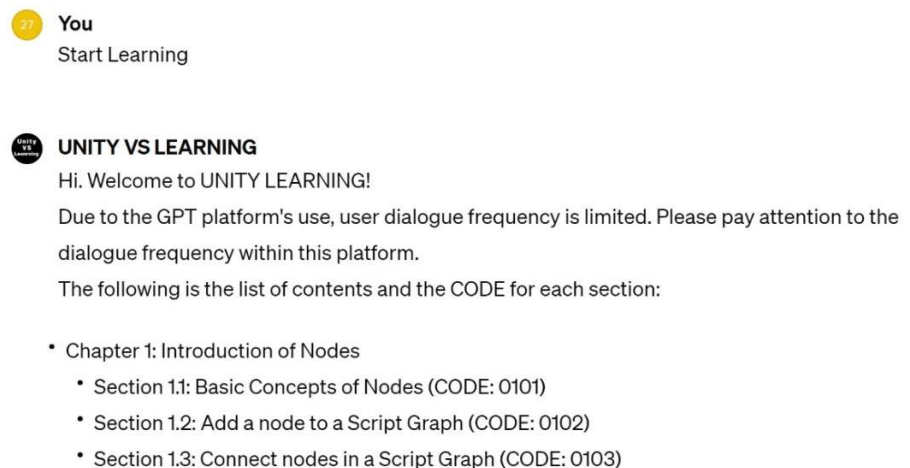


Figure 6. Result of text display Trial 1 (part of the whole result).

It is also examined whether it can work if removing the “print ()” function in Trial 1. It shows that when jumping from somewhere else, this approach cannot always keep the text printed in its original form.

(2) Trial 2: Using printing command to print text in another place

Here is an example. The instruction in “Initial Response Method for Dialogues” in Part 0 reads: “If ‘Start Learning’ is inputted by user, print Point 4 in Part 1, Chapter 1, Section 1.1, then print Point 4 in Part 1, Chapter 1, Section 1.3” or “If ‘Start

Learning' is inputted by user, proceed the following steps: Step 1. print Point 4 in Part 1, Chapter 1, Section 1.1; Step 2. print Point 4 in Part 1, Chapter 1, Section 1.3.”

Some variables here include: a) whether directly providing the text content in Section 3A, or with “printing ()” function; b) whether use single number at the beginning of the text like “2.”, or use serial number like “Step 2”, “Point 2”

All results with different combination of variables successfully printed the text, however, they have to continue printing until the end of the section. Also, if it is required to print sperate parts in Part 1, it can just finish the first one.

(3) Trial 3: Using command to execute the “print ()” function in somewhere else

Here is an example. “If ‘Start Learning’ is inputted by user, proceed Step 1 in Part 1, Chapter 3, Section 3.1” and in the corresponding section, it use “print ()” function.

With different variable forms, the results are similar to those in trial 2. However, it can display two separated parts of text in one time.

(4) Trial 4: Searching text to display

An approach is to use instructions in Part 0 to force GPT answer user’s question with original text content as follows: “If users ask you any question, please print any useful information in Part 1 that can answer user’s question. Please print the original text and do not rewrite them or add your own words. Please notice that the useful information in Part 1 can be over one place, so please find as much as possible.”

The results are as follows (**Figure 7**). The red square is to mark the original text. It can be seen that even under strong instructions, it still tries to rewrite the original text to make the content coherent. The reason might be it has a strong weight of using NPL since it is a LLMs.

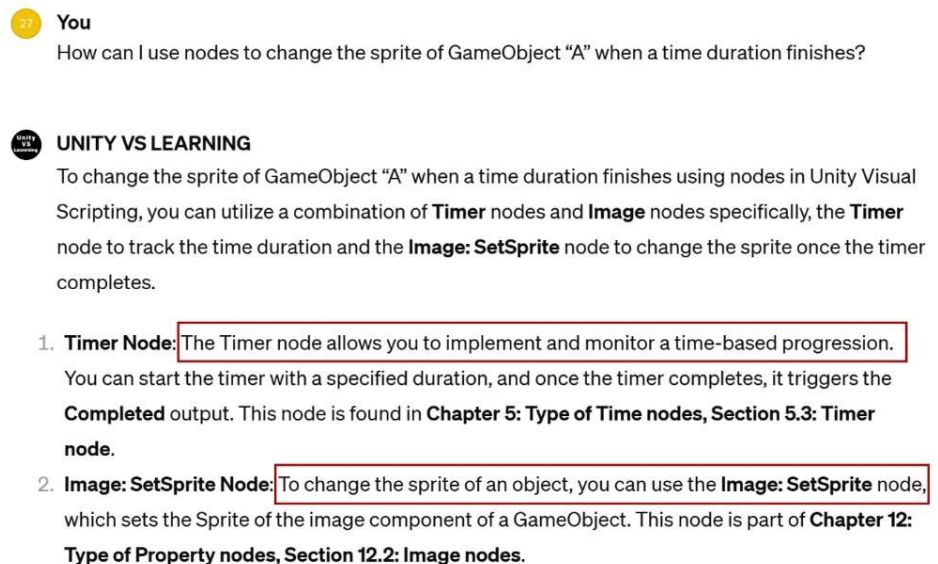


Figure 7. Result of text display Trial 4 with easy input (part of the whole result).

However, if the requirement of using original text is added into user’s input, GPT will largely increase the weight of using original text, as shown in **Figure 8**, which shows that the user’s input plays a decisive role during interaction.

Another test is to use an existing printed instruction as context to force GPT print the original text. It induces GPT to print out the command first as context and then the

user adds a signal from the printed instruction in new input. The result shows this method also does not work.

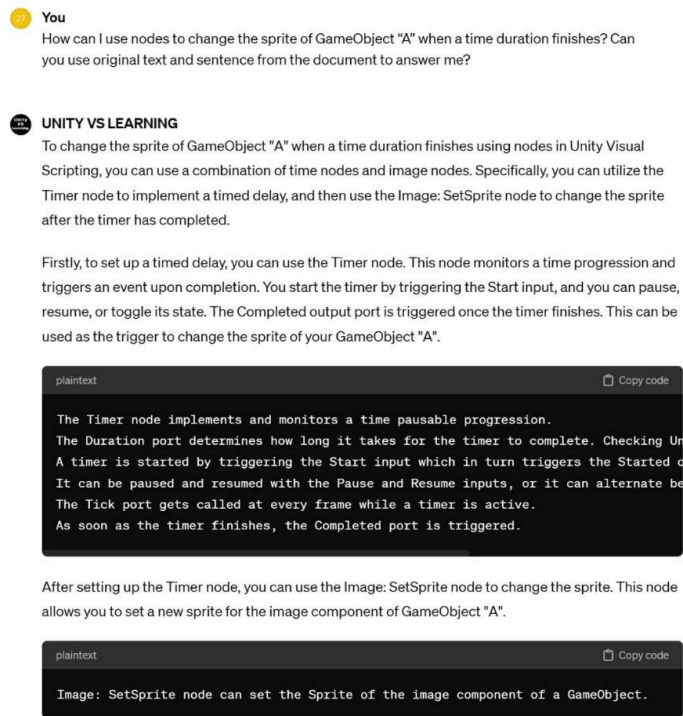


Figure 8. Result of text display Trial 4 with strong input (part of the whole result).

It can be inferred that without advanced code, it is difficult to force GPT print the original text when asked a question that is not preset, because this process may involve several steps with different ability. Therefore, it is preferred to design the Mode 3 to work with user's additional input instruction. The results also show that "print ()" function of every step in Part 1 rarely interfere the information searching process.

Move the conditions into "Instructions" area are also testes and the results show they sometime works.

(5) Trial 5: Reinterpreting text in natural language

Reinterpreting text in natural language is required by Mode 2 and Mode 4. Due to GPT's NLP characteristic, if there is no additional command, it is easy to realize the reinterpreting function. However, if there is command like "print ()" in the content, or instruction to print original text, it is needed to add a conversion mechanism. Considering the convenience for future editors of this document, it is not supposed to provide conditional statements with many options after each section in Part 1 for switching modes. It is found that when conditional statements are only placed in part 0, it is difficult to reinterpret the text in Mode 4. It works only if the user add prompt like "in natural language" into the input. However, when trying to move conditional statements into "Instructions" area, it works well.

4.2.2. Image display

Images are compressed into zip file named "Album" uploaded additionally. Four ways to display images are examined here.

Trial 1: Directly commanding to show images

When using command like “show the image named ‘010101.jpg’ from the zip file”, “display image named ‘010101’ from ‘/mnt/data/Album.zip’” or “extract the image named ‘010101.jpg’ from the zip file “Album.zip” in “/mnt/data/” and display”, it is sometimes failed and no image is shown.

Trial 2: Using half code to extract every time

When using sentence like “Please use code including ‘display.display(image)’ to display image named “010101.jpg” in path ‘/mnt/data/Album.zip’” to display code, it sometimes works, while in other time it fails. Sometimes, even if it generates the complete required information, the sequence of image and text is disordered and it always display an image at first. This may be because GPT takes this order as a natural-language description rather than a code, so it mixes the information with other to utilize.

Trial 3: Using prompt to call a shared piece of code

A shared piece of code to display images is put in the Part 0 (**Figure 9a**), for example, in Section 7. And in Part 1, it provides the following prompt in each place where needed: “Please use code in Part 0, Section 7 to display an image (image_name = ‘010101.jpg’).” This approach can show the image successfully. However, it also has the disadvantages of disorder, giving extra description (marked with red square in the figure) (**Figure 9b**) and sometimes turn into reinterpreting in the middle.

```

from PIL import Image
import IPython.display as display
import zipfile
import os

def extract_and_display_image(zip_path, image_name, extract_path='/mnt/data/extracted_images'):
    if not os.path.exists(extract_path):
        os.makedirs(extract_path)

    with zipfile.ZipFile(zip_path, 'r') as zip_ref:
        if image_name in zip_ref.namelist():
            zip_ref.extract(image_name, extract_path)
            full_path = os.path.join(extract_path, image_name)
            image = Image.open(full_path)
            display.display(image)
            return full_path
        else:
            print(f"Image '{image_name}' not found in the zip file.")
            return None

# Display the first image as per the instructions
zip_path = '/mnt/data/Album.zip'
image_name = '010201.jpg' # Adjust based on the requested image name
extracted_image_path = extract_and_display_image(zip_path, image_name)

if extracted_image_path:
    print(f"Image successfully extracted, stored at: {extracted_image_path}")
    
```

(a)

UNITY VS LEARNING
 Nodes are the most basic part of scripts in Visual Scripting. A node can listen for events, get the value of a variable, modify a component on a GameObject, and more. Nodes appear as blocks in the Graph Editor.


Please use code in Part 0 Section 7 to display an image (image_name = '010101.jpg')

You can arrange and connect these blocks with edges to create logic for an application. Add nodes with the fuzzy finder. Click and drag a node to move it in the Graph Editor. Visual Scripting highlights nodes in your current selection. All nodes have a header which displays the node's name. Some node headers might contain additional information or controls. Ports on the left side of a node are Input Ports. Ports on the right side of a node are Output Ports. An input port or output port can be a Control Port or a Data Port.

- Control Ports control the logical flow in a graph. They tell Visual Scripting what order to execute the nodes in a graph from left to right. The icon for a control port is always an arrow. These arrows display the direction of the flow of logic in a graph.
- Data Ports send and receive data such as number values or GameObjects between nodes. They have colors that correspond to the specific type they expect to receive as inputs or send as outputs. Their icons change based on their type.

(Enter 2: Continue next section) or (Enter the CODE of section to go)

Let's display the image as instructed.



Here's the visual representation of nodes in Visual Scripting within Unity. As described, nodes are depicted as blocks within the Graph Editor, where they can be arranged and connected to create the logic for an application.

(Enter 2: Continue next section) or (Enter the CODE of section to go) (-)

(b)

Figure 9. Code and Result of image display Trial 3. (a) Code of Trial 30; (b) Result of Trial 30.

Trial 4: Using complete code to extract one image every time

This approach is using code in (**Figure 9a**) in each place where needed. The code is executed every time when being proceed together with text printing. In most results, the image can be displayed in correct sequence. The disadvantage is that the document needs to repetitively provide the code.

Trial 5: Extracting all images from the beginning

This is to use a complete code (**Figure 10a**) to extract all images from the beginning, then use following request to show the extracted images where needed:

“Please use code <image_path = ‘/mnt/data/extracted_images/010101.jpg’ display.display(Image.open(image_path))> to display an image in path ‘/mnt/data/extracted_images’.”

The performance is relatively good with correct sequence and makes less errors (**Figure 10b**).

```
import zipfile
import os
from PIL import Image
import IPython.display as display

def extract_all_images(zip_path, extract_path='/mnt/data/extracted_images'):
    if not os.path.exists(extract_path):
        os.makedirs(extract_path)
    with zipfile.ZipFile(zip_path, 'r') as zip_ref:
        for file in zip_ref.namelist():
            if file.lower().endswith(('.png', '.jpg', '.jpeg', '.gif', '.bmp')):
                zip_ref.extract(file, extract_path)
                print(f"Extracted '{file}' to '{extract_path}'")
zip_path = '/mnt/data/Album.zip' # Path to your zip file
extract_all_images(zip_path)
```

```
STDOUT/STDERR
Extracted '010101.jpg' to '/mnt/data/extracted_images'
Extracted '010201.jpg' to '/mnt/data/extracted_images'
Extracted '010301.jpg' to '/mnt/data/extracted_images'
Extracted '010302.jpg' to '/mnt/data/extracted_images'
Extracted '010303.jpg' to '/mnt/data/extracted_images'
Extracted '020101.jpg' to '/mnt/data/extracted_images'
Extracted '020102.jpg' to '/mnt/data/extracted_images'
```

(a)

The screenshot shows a Unity VS Learning interface. It starts with a 'You Start Learning' section. This is followed by a 'UNITY VS LEARNING' section with a message: 'The images have been successfully extracted. You can now continue with the next part of the tutorial.' Below this is a prompt: '[Enter 1: Start from Chapter 1] (5)'. Another 'You' section follows with the number '1'. This is followed by another 'UNITY VS LEARNING' section with the text: 'Chapter 1: Introduction of Nodes' and 'Section 1.1: Basic Concepts of Nodes (CODE: 0107)'. A paragraph explains: 'Nodes are the most basic part of scripts in Visual Scripting. A node can listen for events, get the value of a variable, modify a component on a GameObject, and more. Nodes appear as blocks in the Graph Editor.' Below this is a sub-section: 'Let's take a look at an example image to better understand the concept of nodes.' This is followed by a small image of a graph editor showing a node. Below the image is a paragraph: 'You can arrange and connect these blocks with edges to create logic for an application. Add nodes with the fuzzy finder. Click and drag a node to move it in the Graph Editor. Visual Scripting highlights nodes in your current selection. All nodes have a header which displays the node's name. Some node headers might contain additional information or controls. Ports on the left side of a node are Input Ports. Ports on the right side of a node are Output Ports.'

(b)

Figure 10. Code and Result of image display Trial 5. (a) Code of Trial 5; (b) Result of Trial 5.

Trial 6: Display image by searching

It is also required to display an image when giving solutions to user’s question. Two ways are explored based on Trial 5’s method which extracts all images first.

One approach is to add additional instruction to all of the conditional statements in “Control Center” section (will be discussed later), “Overall Rules” section and “Instructions” area as follows:

“Please also execute the steps with code for displaying images that is very close to the information you find and in the same section of the information you find, which helps to illustrate the text information.”

The results show it works in printing the text it found and displaying the corresponding images. **Figure 11** shows two pieces of one result. There is one disadvantage that it sometimes put all images together, even when an addition instruction.

Another approach is similar to the previous one, but placing code in part 0 and in Part 1, telling GPT to execute this code to display an image. The results and disadvantages are almost the same as the first approach.

You
How can I use nodes to change the sprite of GameObject "A" when a time duration finishes?

UNITY VS LEARNING

To change the sprite of a GameObject "A" when a time duration finishes using nodes in Unity Visual Scripting, you would typically use a combination of time nodes, specifically a Timer node to track the duration, and Image nodes to change the sprite of the GameObject. Here's a general approach on how you could set this up:

1. **Timer Node:** This node is used to count down the time duration. Once the set duration is reached, it triggers an event indicating that the time has elapsed.
2. **Image: SetSprite Node:** This node is utilized to change the sprite of a GameObject. You would connect this node to be triggered after the Timer node completes, specifying the new sprite you want to apply to the GameObject.
3. **Start Node:** Often, the sequence is initiated by a Start node or any other event node that suits your game logic, to start the timer at the beginning or at a specific moment in your game.

Here are the steps you might follow, simplified for clarity:

- Use a **Start node** to initiate the sequence when your game starts or when the GameObject becomes active.
- Connect the Start node to a **Timer node** and set the duration to your desired time.
- Once the Timer node's duration finishes, use its **Completed output** to trigger an **Image: SetSprite node**.
- The **Image: SetSprite node** would be configured with the target GameObject "A" and the new Sprite that you want to apply to it.

This setup ensures that once the specified time duration is over, the sprite of GameObject "A" will change to the new sprite you've designated.

(a) Piece 1.

To visually represent this setup, I will display the Timer and Image: SetSprite nodes from the extracted images which closely relate to this explanation. Let's start with an illustration for the Timer node.



This image shows how a Timer node could be configured. In your case, you would set the duration for how long you want to wait before changing the sprite.

Next, let's visualize how an Image: SetSprite node might be represented. Since we don't have a direct example of an Image: SetSprite node in the extracted images, I'll show you an example of another closely related Image node to give you an idea. Let's look at an Image node example. [-]



This image is actually of a Transform node, but it serves to illustrate the concept of interacting with GameObject properties through nodes. For changing the sprite, you would use an "Image: SetSprite" node similarly, specifying the GameObject "A" and the new sprite as inputs.

(b) Piece 2.

Figure 11. Results of image display Trial 6.

4.2.3. Jump action

This part explores how to jump from one place to another in different ways. Using conditional statements is to add an additional part where needed with several "if" conditions corresponding to users' input, and they just exist in the document and are not be printed. Using interactive request is to provide a request enclosed in braces such as {Enter 1: go to Section 3} at the end of each section. Using section title is to place prompt in section title for positioning.

(1) Trial 1: About section title and "Initial Response Method for Dialogues"

* Data Ports send and receive data, such as number values or GameObjects, between nodes. They have colors that correspond to the specific type they expect to receive as inputs or send as outputs. Their icons change based on their type. [-]

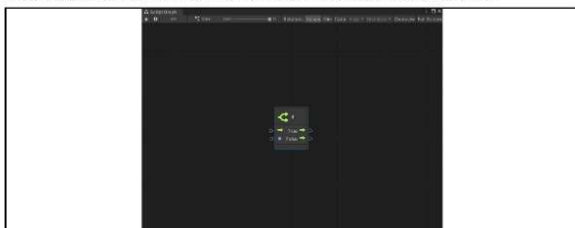
You
0301

UNITY VS LEARNING

Chapter 3: Type of Branching nodes
Section 3.1: If node

The common if node uses a boolean condition. Consider them as an "if the condition is true, do something; otherwise, do something else."

Let's display an example image to illustrate the use of the if node in Visual Scripting.



{Enter 2: Continue next section} or {Enter the CODE of section to go} or {Enter any issue} [-]

(a)

Here's the image illustrating the basic concepts of nodes in Visual Scripting. What would you like to do next? [-]

You
Section 2.1

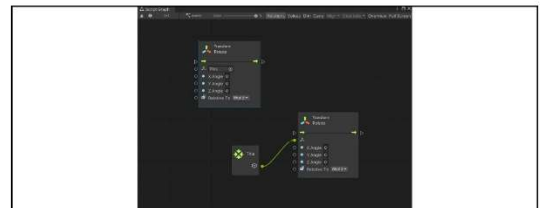
UNITY VS LEARNING

Chapter 2: Type of This node
Section 2.1: This node

Use a This node to return a GameObject with a Script Machine component that has the Script Graph. If the Script Machine uses a Graph source and multiple GameObjects use the same graph, the returned GameObject can change.

Many nodes default their target to This. For example, the following Transform nodes are the same:

Let's display the first mentioned image.



Not all nodes support the This inline value. Any node that doesn't support the This inline value displays None instead of This in the default value field. For example, the Destroy node displays None.

(b)

Figure 12. Result of Jumping to a specific section. (a) Piece 1; (b) Piece 2.

The weight of section title and “Initial Response Method for Dialogues” section designed previously are tested. It is found that actually sometimes GPT tends to go to a section with title that is same as input words, more than go to a section according to conditional statement in “Initial Response Method for Dialogues” section. Therefore, it is better to use a different name in section title if conditional statements are supposed to work.

Jumping from the end of a section to any specific section in Part 1 is also a required function for user’s step-by-step learning. In a test, each Section in Part 1 is labeled with a unique code at the end, such as “0201” meaning the Section 1 in Chapter 2. It is found the jump action can work no matter whether there are any conditional statements or an interactive request like “{Enter the CODE of section to go}” (**Figure 12a**). When the section title is input, it also works (**Figure 12b**). It can be inferred that GPT actually jump to corresponding section by searching section title. Trial 2: About Interactive requests

The “Instructions” area in “Configure” interface is filled with the rule that the user can only interact with interactive requests. When interactive requests and Control Center provides different directions, it is found that the interactive requests have a large weight when it is explicitly stated, such as {Enter 1: Continue} (**Figure 13a**). However, if it is obscure, GPT will locate user’s input to other conditions (**Figure 13b**).

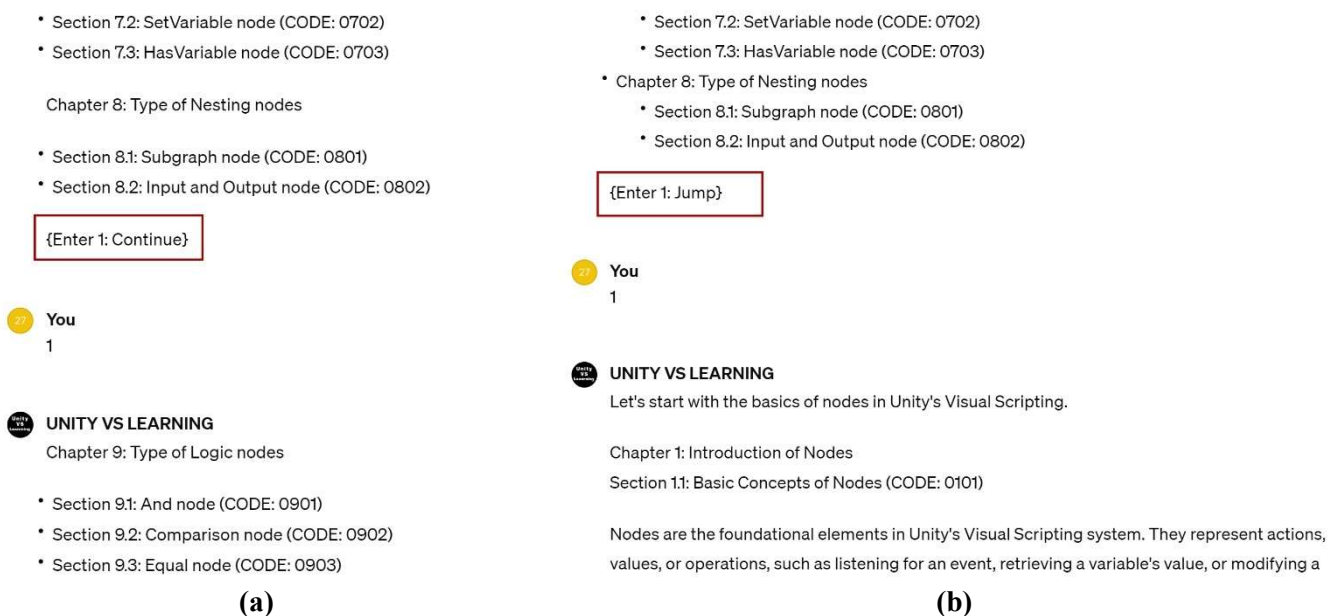


Figure 13. Results of different interactive requests in Trial 2. (a) Trial 2a; (b) Trial 2b.

Trial 3: About positions of conditional statements

There are two methods to respond user’s input by conditional statement.

One method that uses conditional statement right after current position has been examined in previous discussion. However, it needs to provide a part of conditional statements in each place where needed in Part 1. The advantage is that it can provide different responses in different places even the user’s inputs are totally the same.

The other method is to use a Section named “Control Center” in Part 0 with all “if” conditional statements to proceed common response to the user’s input. Since a general rule in “Instructions” area reads that the user can only interact with interactive request, if the user’s input is not covered by current interactive input, the “Control Center” does not work. One test shows that when the Control Center works, it cannot jump back automatically. For example, if the interactive request in a section in Part 1 just has {Enter x}, and there is one conditional statement in Control Center reads: “If a single letter “x” is entered by the user, go to the next section.” Then, it cannot go to the next section in Part 1, instead, it goes to the next section in Part 0.

One test also shows that not only the first approach, but also the second can proceed several steps after the jumping action. The follows are the result of one test (**Figure 14**). The sequence of original text (with red frame)-solutions-original text is generated by the preset steps in Control Center when jumping from another place.

• Chapter 8: Type of Nesting nodes

- Section 8.1: Subgraph node (CODE: 0801)
- Section 8.2: Input and Output node (CODE: 0802)
{Enter 1: Continue} or {Ask a question now}

You
How can I use nodes to change the sprite of GameObject A when a time duration finishes?

UNITY VS LEARNING
I will find you a solution.

To change the sprite of GameObject A when a time duration finishes using nodes in Unity Visual Scripting, you can follow this general approach:

1. **Use a Timer Node:** The Timer node can be set with a specific duration. Once the duration is completed, it triggers an event that can be used to change the sprite of GameObject A.
2. **Create a Sprite Variable:** Use a SetVariable node to define which sprite you want to change to after the timer is finished.

Note: Ensure that your GameObject A has an Image component if you're working within a UI context, or a SpriteRenderer component if you're working within a game scene context.

This setup allows you to delay a sprite change until after a certain time has elapsed, which can be useful for cooldown visuals, timed events, or visual cues to the player.

Remember, this is a general guide. The exact implementation can vary based on the specifics of your project, such as whether you're using Unity's UI system or 2D SpriteRenderer for GameObjects in the scene.

{Continue asking questions} or {Enter 3: Start Learning}.

(a) **(b)**
Figure 14. Result of proceeding steps in conditional statements. **(a)** Piece 1; **(b)** Piece 2.

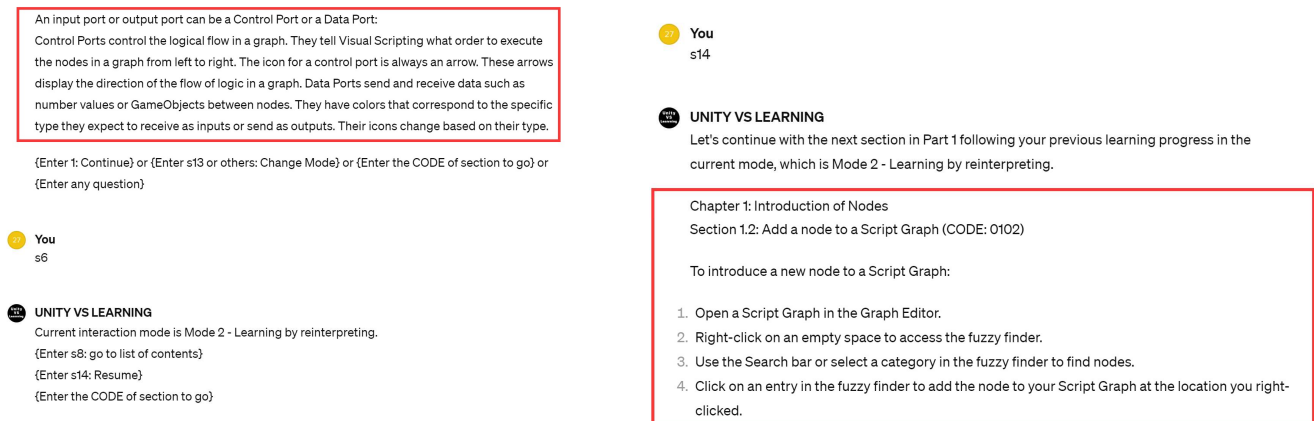
Based on the above trials from 1, 2 and 3, it can be concluded that when similar prompts exist in a section title, an interactive request, conditional statements in current position and conditional statements in “Control Center”, GPT will comprehensively judge the degree of similarity to select the closet one to jump to.

Trial 4: About “Instructions” area

If the “Instructions” area tells GPT to go to Part 0, Section 1, and the interactive request tells GPT to go to Part 0, Section 2, it will select the latter way to go. It may be because the “Instructions” area is filled with the rule that the user can only interact the interactive requests. It can be inferred that the “Instructions” area has been tested to have highest weight to control the overall interaction.

One approach that can perfectly avoid the Control Center’s defect of relocation is placing all conditional statements only in “Instructions” area. Therefore, it can be considered to use only the “Instructions” area to provide common rules as the Control Center, and make Part 0 and Part 1 in the document all the modules for providing detailed information and interactive methods. **Figure 15** shows a good result when using this approach. The user can switch from showing content in original form (**Figure 15a**) to showing content in reinterpreted form (**Figure 15b**) after inputting simple codes that points to conditional statements preset in “Instruction” area. The

process is firstly the user enters the code representing a certain mode that can be found in previous dialog, or the code provided in current interactive request to restart the mode selection module. After the mode switched, the user continues the learning by inputting a code representing resuming.



(a) Switching from showing original content.

(b) Switching to reinterpreting content.

Figure 15. Result of switching modes in Trial 4.

4.3. Integration and optimization

4.3.1. Methods selection

The preferred approaches based on all above trials to compose different functions in the document is listed in **Table 4**. The list remains those with better performance and filtered some options based on selection priority discussed previously. Some functions have over one option.

Table 4. Preferred methods.

Functions	Approaches
Printing text directly	Using “print ()” function
Displaying image directly	Using code to extract all images at the beginning Option 1: Display the extracted image with local code Option 2: Display the extracted image with code in “Instructions” area/Control Center
Displaying text by searching	Using interactive requests Option 1: Using conditional statements in Control Center + user additional input Option 2: Using conditional statements in “Instructions” area
Displaying image by searching	Option 1: Using Control Center Option 2: Using “Instructions” area
Jumping to the next chapter	Option 1: Just using interactive requests Option 2: Using interactive requests + “Instructions” area
Initial mode selection	Option 1: Using Control Center Option 2: Using section title Option 3: Using an individual start section + “Instructions” area
Jumping to any specific section in Part 1	Just using unique code in title and interactive requests
Overall requirement	Considering to make section titles, conditional statements and interactive requests same or different Adjusting “Instructions” area to be consistent with all functions

It can be seen that all functions in **Figure 4** can be realized without placing conditional statements at the end of each section in Part 1.

4.3.2. Methods integration

Considered integrating all the selected methods, two forms of templates are designed based on the initial document after several times of optimization.

The first template is as follows (Figure 16). The features include:

- (1) It mainly uses a Control Center to respond to user’s questions.
- (2) The jump action from section to section in Part 1 has been simplified and can now be achieved through interactive requests.
- (3) For first mode selection, it uses a single section.
- (4) The user has to add additional words in the input to change response style from original content to reinterpreting.

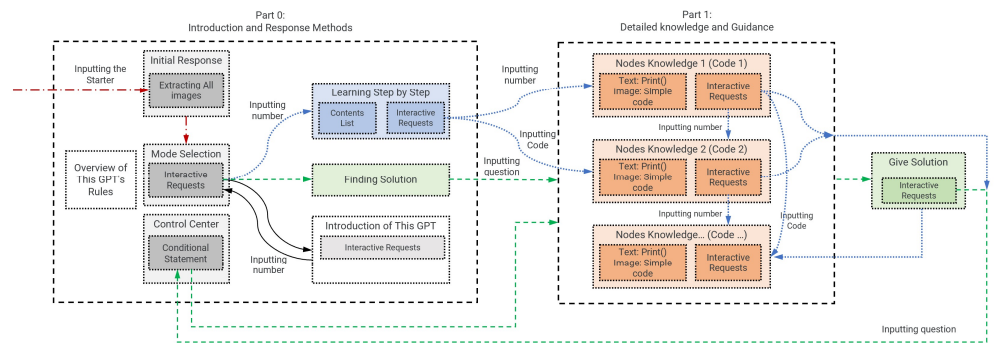


Figure 16. Structure of template 1.

The second template is as follows (Figure 17). The features include:

- (1) It places all conditional statements into the “Instructions” area (text provided in Appendix A) to control all the jump actions.
- (2) It uses the printed content to tell GPT the current mode. It prints the mode type first when user’s entering a mode, then the following response will be initiated in corresponding style based on the condition statements about current mode.
- (3) Interaction can switch between different modes at any time. An option is provided if the user needs to switch the GPT’s response between using original content and reinterpreting and directly go back. Users can also restart from beginning.

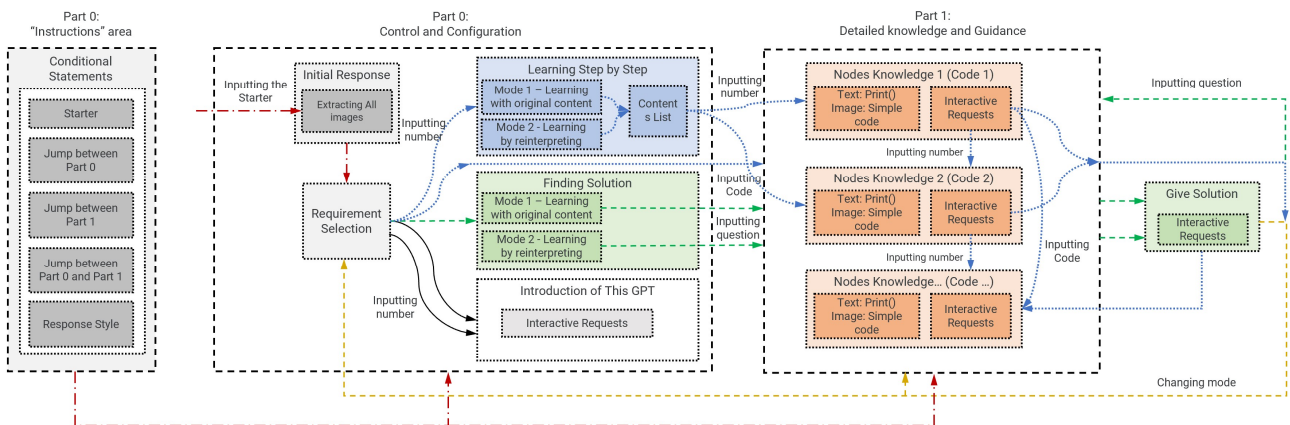


Figure 17. Structure of template 2.

4.3.3. Modes tests and final optimization

The four interaction modes based on the final optimized document are tested. The approximate accuracies are categorized as “perfect”, “excellent”, “good”, “fair” and “bad” based on 5 times of tests.

As shown in **Table 5**, generally, the performance of template 1 is good. There are some problems. For example, in Mode 2, when the user adds additional words to input, all contents will be changed including the interactive request.

Table 5. Performance of template 1.

Modes	Accuracy	Problems Description
1	perfect	None
2	good	Changes the section title and interactive request after using NLP
3	fair	Cannot completely use original content
4	excellent	Cannot mix images and text from the beginning
overall		The user has to add additional words in input to change current mode Original structure is likely messed after changing modes

As shown in **Table 6**, the performance of template 2 is also good. But in some modes, it occasionally makes more mistakes compared to template 1. For example, in Mode 1, sometimes it skips the code and miss the image display, and it is solved when the user gives a reminder.

Table 6. Performance of template 2.

Modes	Accuracy	Problems description
1	good	Sometimes miss the image when first entering and needs a reminder Sometime mess the steps
2	excellent	Occasionally miss the images
3	good	Cannot completely use original content
4	good	Cannot mix images and text from the beginning Occasionally gives a wrong image to a right answer (but the code is correct) Occasionally provides insufficient image for the answer
overall		None

Compared to the two performances, it is preferred to select template 2 as the final optimization outcome. The reason is that it does not require the user to add additional works in input, which meets the simplicity of HCD principles. Another reason is that it has larger space for promotion, since it uses “Instructions” area to avoid the inherent limitation of using control section in the document, because it can provide conditional statements without jumping to it.

Detail information of the final document of Template can be found in Appendix B.

5. Discussion

Some findings in the series of experiments include:

(1) Changing interaction modes while using a shared knowledge part is not an easy task. The more users’ needs integrated, the more difficult the organization the

document is optimized.

(2) GPT's jump action is an abstract description of GPT's behavior. Jump action based on decision of where to go, essentially is a searching and proceeding action, which means it has to search the information inside the document and decide what is the most relevant one to user's input. This characteristic makes it difficult for GPT to consider both the conditional statements in current position and in another position inside the document, because when it goes to conditional statements somewhere else from current location and proceeds some steps, it usually has already "lost" current location. Such complex action may require inherent workflow with different components with higher complexity.

(3) "Instructions" area has the highest priority, so creators are supposed to check whether rules in it contradict specific conditions in the document. Text in "Instructions" area can work simultaneously with any text it is positioning, which is useful.

(4) Section title, conditional statements, interactive requests, user's input, and content that has already been generated as context can all affect GPT's new content generation. This is easy to explain when an interactive request does not provide explicit way of what to do, GPT will find explicit way somewhere else. Since the essence of relocation is actually a searching action,

(5) GPT tends to use NLP to give responses, unless there are explicit steps with high weight to force it give original content from the document. Even within "print ()" function, sometimes GPT refuse to give irrelevant contents and use NLP to change them.

(6) It is hard for GPT to accurately identify the correct way to go if provided a series of conditional statements structured in tree branch, which may be because when GPT has already found an information inside a conditional statement branch, it may think this is the most relevant information and stop finding other. Therefore, the structure of conditions needs to be well designed. The document and "Instructions" area can work together to achieve this.

(7) A failure like going to a wrong section can make the following interaction chaotic and not easy to correct the order. It may be because the incorrect context has already been produced, interfering GPT's following judgement.

6. Conclusion

This study concludes HCD guidelines in LLMs and tries to integrate them into an experiment of using a single document as new knowledge in GPT to meet user's diverse software learning needs. It is found that without high-level code, it is not easy to integrate all diverse needs perfectly into one GPT. The natural language characteristic of GPT is generally a merit for comprehensively understand the document and user's input, while in some cases, becomes an interference of proceeding mixed steps to generate preset content and creative content together, which may need preset components and workflow inside GPT with higher control level. The outcomes provide preliminary thinking about how to organize different elements in the document as GPT's new knowledge and setup GPTs' configuration, and the final optimized document also provides a template for futural application or research with the same requirements.

Some variables are not considered into the experiment, which can be explored in the future, such as the follows:

- (1) How much the inherent knowledge of LLMs will interfere the understanding and extracting the content in the new knowledge [29]? How will the experiment in the study result is if the knowledge of the software is replaced by that from a totally new software?
- (2) Will the length of context affect the GPT's judgement of current modes in the experiment [30]?
- (3) Are there any other types of element's organization of the document that can help enhance user experience?
- (4) If speed is also taken into consideration, how to optimize the document to better achieve HCD principles.

Looking forward, the focus should be on advancing LLMs and their application to better enhance UX. Continued exploration in this domain will likely lead to more sophisticated, user-centric HCI systems, aligning technology more closely with human needs and behaviors.

Author contributions: Conceptualization, YW and YSL; methodology, YW; literature, YW, YSL, RH, JW and SL; software, YW; Investigation, YW; validation, YW; formal analysis, YW; resources, YW and YSL; data curation, YW and YSL; writing—original draft preparation, YW; writing—review and editing, YW and YSL; visualization, YW; supervision, YW and YSL. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Gokul A. LLMs and AI: Understanding Its Reach and Impact. Published online May 4, 2023. doi: 10.20944/preprints202305.0195.v1
2. Liu J, Shen D, Zhang Y, et al. What Makes Good In-Context Examples for GPT-3? Published online 2021. doi: 10.48550/ARXIV.2101.06804
3. Park TJ, Dhawan K, Koluguri N, et al. Enhancing Speaker Diarization with Large Language Models: A Contextual Beam Search Approach. Published online 2023. doi: 10.48550/ARXIV.2309.05248
4. Thomas P, Spielman S, Craswell N, et al. Large language models can accurately predict searcher preferences. Published online 2023. doi: 10.48550/ARXIV.2309.10621
5. Ortega Pedro A, Maini V, DeepMind Safety Team. Building safe artificial intelligence: specification, robustness, and assurance. Available online: <https://deepmindsafetyresearch.medium.com/building-safe-artificial-intelligence-52f5f75058f1> (accessed on 2 January 2024).
6. Xiong S, Payani A, Kompella R, et al. Large Language Models Can Learn Temporal Reasoning. Published online February 20, 2024. doi: 10.48550/arXiv.2401.06853
7. Ji J, Qiu T, Chen B, et al. Ai alignment: A comprehensive survey. arXiv. 2023; arXiv:2310.19852. doi: 10.48550/arXiv.2310.19852
8. Weidinger L, Mellor J, Rauh M, et al. Ethical and social risks of harm from language models. arXiv. 2021; arXiv:2112.04359. doi: 10.48550/arXiv.2112.04359
9. Guinness H. How does ChatGPT work? Available online: <https://zapier.com/blog/how-does-chatgpt-work/> (accessed on 6 January 2024).
10. Yan C, Qiu Y, Zhu Y. Predict Oil Production with LSTM Neural Network. Proceedings of the 9th International Conference on Computer Engineering and Networks. Publish online 2021. doi: 10.1007/978-981-15-3753-0_34

11. Subramonyam H, Im J, Seifert C, et al. Solving Separation-of-Concerns Problems in Collaborative Design of Human-AI Systems through Leaky Abstractions. CHI Conference on Human Factors in Computing Systems. Published online April 29, 2022. doi: 10.1145/3491102.3517537
12. Zhang P, Jaipersaud B, Ba J, et al. Classifying Course Discussion Board Questions using LLMs. Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V 2. Published online June 29, 2023. doi: 10.1145/3587103.3594202
13. Don N. User centered system design—New perspectives on human-computer interaction. CRC Press; 1986.
14. Geyer W, Weisz J, Pinhanez CS. What is human-centered AI? Available online: <https://research.ibm.com/blog/what-is-human-centered-ai> (accessed 12 January 2024).
15. Farooqui T, Rana T, Jafari F. Impact of human-centered design process (HCDDP) on software development process. In: 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE) 2019 Mar 6. doi: 10.1109/C-CODE.2019.8680978
16. Gulliksen J, Göransson B, Boivie I, et al. Key principles for user-centred systems design. Behaviour and Information Technology. 2003; 22(6): 397-409. doi: 10.1080/01449290310001624329
17. Jaimes A, Gatica-Perez D, Sebe N, et al. Guest Editors' Introduction: Human-Centered Computing--Toward a Human Revolution. Computer. 2007; 40(5): 30-34. doi: 10.1109/mc.2007.169
18. Mack K, McDonnell E, Potluri V, et al. Anticipate and Adjust: Cultivating Access in Human-Centered Methods. CHI Conference on Human Factors in Computing Systems. Published online April 29, 2022. doi: 10.1145/3491102.3501882
19. Wang Y, Lin YS. Public participation in urban design with augmented reality technology based on indicator evaluation. Frontiers in Virtual Reality. 2023; 4. doi: 10.3389/frvir.2023.1071355
20. Wu C. The Impact of Public Green Space Views on Indoor Thermal Perception and Environment Control Behavior of Residents - A Survey Study in Shanghai. European Journal of Sustainable Development. 2023; 12(3): 131. doi: 10.14207/ejsd.2023.v12n3p131
21. Seffah A, Andreevskaia A. Empowering software engineers in human-centered design. 25th International Conference on Software Engineering, 2003 Proceedings. Published online 2003. doi: 10.1109/icse.2003.1201251
22. Chao Yan. Predict Lightning Location and Movement with Atmospheric Electrical Field Instrument. Proceedings of the 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Publish online 2019. doi: 10.1109/IEMCON.2019.8936293
23. Petridis S, Terry M, Cai CJ. PromptInfuser: Bringing User Interface Mock-ups to Life with Large Language Models. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. Published online April 19, 2023. doi: 10.1145/3544549.3585628
24. Park J, Choi D. AudiLens: Configurable LLM-Generated Audiences for Public Speech Practice. Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. Published online October 29, 2023. doi: 10.1145/3586182.3625114
25. Di Fede G, Rocchesso D, Dow SP, et al. The Idea Machine: LLM-based Expansion, Rewriting, Combination, and Suggestion of Ideas. Creativity and Cognition. Published online June 20, 2022. doi: 10.1145/3527927.3535197
26. Korbak T, Shi K, Chen A, et al. Pretraining language models with human preferences. Available online: <https://arxiv.org/abs/2302.08582> (accessed on 5 January 2024).
27. Rastogi C, Tulio Ribeiro M, King N, et al. Supporting Human-AI Collaboration in Auditing LLMs with LLMs. Proceedings of the 2023 AAI/ACM Conference on AI, Ethics, and Society. Published online August 8, 2023. doi: 10.1145/3600211.3604712
28. Unity. Available online: <https://docs.unity3d.com/Packages/com.unity.visualscripting@1.9/manual/> (accessed 11 January 2024).
29. Weng Y, Wu J. Fortifying the Global Data Fortress: A Multidimensional Examination of Cyber Security Indexes and Data Protection Measures across 193 Nations. International Journal of Frontiers in Engineering Technology. Publish online 2024. doi: 10.25236/IJFET.2024.060203
30. Wang C, Yang Y, Li R, et al. Proceedings of the 2024 International Conference on Image Processing and Computer Applications (IPCA 2024). Adapting LLMs for Efficient Context Processing through Soft Prompt Compression. Publish online April 7, 2024. doi: 10.48550/arXiv.2404.04997

Appendix A

For details, please refer to

https://onedrive.live.com/edit?id=463C91BDAAC45E41!sdc4a46768b8b4df38b02af6269061337&resid=463C91BDAAC45E41!sdc4a46768b8b4df38b02af6269061337&cid=463c91bdaac45e41&ithint=file%2Cdocx&redeem=aHR0cHM6Ly8xZHJ2Lm1zL3cvYy80NjNjOTFiZGFhYzQ1ZTQxL0VYWkdTdHlMaV9OTml3S3ZZbWtHRXpjQmNHbXRweXpDWjhNUkktXJPQ1VFMEE_ZT05aFRnWms&migratedtospo=true&wdo=2

Appendix B

For details, please refer to

https://onedrive.live.com/edit?id=463C91BDAAAC45E41!sc0b5fa94e7fd4f098d0dc2db2a52348e&resid=463C91BDAAAC45E41!sc0b5fa94e7fd4f098d0dc2db2a52348e&cid=463c91bdaac45e41&ithint=file%2Cdocx&redeem=aHR0cHM6Ly8xZHZJ2Lm1zL3cvYy80NjNjOTFiZGFhYzQ1ZTQxL0VaVDZ0Y0Q5NXdsUGpRM0MyeXBTTkk0QjQ2OW9oLXJMNjctZWZ6eXkwY1J5YWc_ZT1vME8wa0E&migratedtospo=true&wdo=2

The AI spectrum under the doctrine of necessity: Towards the flexibility of the future legal praxis

Lambrini Seremeti^{1,2}, Ioannis Kougias^{2,*}

¹ School of Education, Frederick University of Nicosia, 1036 Nicosia, Cyprus

² Laboratory of Interdisciplinary Semantic Interconnected Symbiotic Education Environments, Electrical and Computer Engineering Department, Faculty of Engineering, University of Peloponnese, 263 34 Patras, Greece

* **Corresponding author:** Ioannis Kougias, kougias@uop.gr

CITATION

Seremeti L, Kougias I. The AI spectrum under the doctrine of necessity: Towards the flexibility of the future legal praxis. *Computing and Artificial Intelligence*. 2024; 2(1): 1258.
<https://doi.org/10.59400/cai.v2i1.1258>

ARTICLE INFO

Received: 1 April 2024

Accepted: 10 May 2024

Available online: 21 May 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: Society is rapidly changing into an implicitus one. The main factor leading to this societal transition is the integration of artificial intelligence (AI), influencing all aspects of anthropocentric legal order. The deep concern to safeguard fundamental human rights under unforeseeable circumstances threatening hypostasis, leads those who are involved in the legal praxis to reorganize the legal system to ensure its functional continuity. To this purpose, a reliable extra-legal tool, such as the doctrine of necessity, is proposed, to validate the issue of AI development that falls outside the purview of any legal process, though, being necessary for society prosperity.

Keywords: artificial intelligence; doctrine of necessity; legal system; society

1. Introduction

Artificial intelligence (AI) being the future economic resource [1], the capability of managing it has become a critical issue. AI, which is not in compliance with regulatory standards and the fundamental principles of law, neither has a value for itself as an evolutionary research field, nor for the prosperity of a society, as a co-part of the entire socio-economic system.

Indeed, AI development promises to improve well-being and help address complex challenges in many sectors, such as healthcare [2], agriculture [3], criminal justice [4], financial services [5], marketing and advertising [6], transportation [7], education [8], scientific research [9], (i.e., AI promises to improve the astronomical imaging). Alongside benefits, AI development and integration raise public policy considerations [10], since AI-based society is perceived as an implicitus, that is confused, society, where necessity is its transformative force.

The meaning of necessity is often not clear, and varies according to the reference framework (e.g., societal or legal). However, it is generally considered as an intrinsic reaction to exogenous stimuli, which may be natural or artificial. More specifically, necessity in the societal context reflects a self-protective act, an entirely natural impulse allowing an individual to both adapt and survive in changing environments [11], while, in the legal context, it concerns the regulation of emergency cases in which human agents have intentionally caused damages to life and property, to avoid greater damages and losses—the so-called doctrine of necessity [12].

According to this doctrine, certain conducts that are prima facie prohibited by law may be permitted under exceptional circumstances, when are necessary to avert a greater harm. Hence, it is certain that the doctrine of necessity must be approached by

the legal praxis with extreme caution. In general, the doctrine is invoked when an illegal act is condoned to prevent a greater mischief [13].

The goal of this paper is to assess if the doctrine of necessity can be utilized to find reasonable guidelines for solving the ethical issue of the regulation of AI in integration scenarios in which some human rights violations emerged. This attempt may supply contemporary AI regulations with new perspectives, which could bring forth a new conception of the importance of lawful integration of AI into everywhere.

The rest of the paper is arranged as follows. Sections 2 and 3 conceptualize the term necessity in societal and legal context, respectively. Section 4 introduces the invocation of the doctrine of necessity in the light of an imaginary scenario, and in section 5, conclusions, as well as hints for future work, are given.

2. Necessity in the social context

The term necessity is very often found in works about societal transformation [14] and it is inextricably linked to variables such as evolution [15] and invention [11]. Of course, necessity, evolution and invention are not tangible and measurable concepts like inflation or capital. They are just explanatory indexes of societal changes. Those are terms that describe certain aspects of a quasi-causal relation [16], and, more precisely, they are used to express the bidirectional connection between the cause (necessity) and the effect (evolution and invention). They are used to explain a twofold situation: emergent forms of innovation create new needs and needs is the starting point to innovate and precisely, this is evolution.

The meaning of necessity in the social context is often understood in ways that are deeply philosophical. In Philosophy, necessity expresses the principle according to which something must be so, by virtue either of logic or of natural law [17]. It is often opposed to chance and contingency [18]. The well-known ancient Greek philosophers, such as Plato and Aristotle, whose ideas still impact society, focused their attention on necessity, as an inherent feature of human nature and it is related with intelligence and force [11]. In other words, if something is being necessary simply means that it cannot be otherwise, under specific conditions. For instance, although zoom and apps like it and the grocery delivery service were all around, they weren't part of many people's routine, because people don't easily change the way they do things. There must be something that prompts them, and the pandemic just forced that [19].

In a nutshell, social transformations arise from changes in recognition of the necessity to address crises ranging from the reduction of natural resources and climate change to natural disasters and human privacy. It must be also pointed out that there is a semantic distinction between natural and artificial necessity. The former, which is synonymous to natural law, is seen as a building block of the latter, which refers to the creation of needs imposed by certain circumstances. For instance, food is a natural necessity of living entities, but government is an artificial necessity, according to Machiavelli, see in the study of Vadi [20]. The line between natural and artificial necessity is not always clearly established and their differentiation depends on the context (e.g., time, space, morality, law) [12]. Therefore, a necessity may be natural and so acknowledged by everyone, or artificial, which can be considered as a process

of creating new instances of nature, utilizing the logic continuity considering novel circumstances.

3. Necessity in the legal context

Necessity in societal context has mainly an inner-behavioral dimension. It is nothing else than a self-protective act, an entirely natural impulse allowing an individual to both adapt and survive in changing environments. Hence, changes to individual behavior influence population dynamics, that is, societies [21].

A common path to govern the symbiosis of individual changes is by establishing a legal network, which can be considered as a complex system [22], since it involves heterogeneous external and internal entities, such as legal praxis, laws, principles, and the environment within which laws are enacted. On the one hand, regardless of any social changes, there is a core structure in the legal network which is based on fundamental principles. On the other hand, there are structural properties (e.g., rules) that are influenced as the society changes. For example, when new technology is introduced to society, new rules are required to regulate activities associated with this technology [23]. Governments and other groups participate in novel law-making, taking into consideration the hierarchy of legal system. For example, the Constitution of a State holds a superior position in the legal hierarchy and can be amended under specific conditions described in the same Constitution. The Statute Law of a state holds a next position and can develop over time as new rules are adopted and existing rules are amended or repealed [24]. Judges, lawyers, and jurists interpret those rules under the light of the fundamental principles of law. These principles are “jus cogens” general norms of international law [25], which are hierarchically superior to other rules and are universally applicable since they protect fundamental values of society.

There is no doubt that the main purpose of a state is its self-preservation and continuation to guarantee society prosperity. Its deep concern is to safeguard its constitutional law under exceptional circumstances and emergency cases, which arise due to changes or novice entities introduced in existing societal context. State continuation and self-preservation are used as synonyms for necessity. In this view, self-preservation constitutes “a unilateral action taken in response to a situation of grave and imminent peril affecting the essential interests of the responding state” [26]. The self-preservation and the continuation of a state, as well as the pressing need to find solutions for factual unforeseeable situations, led the legal praxis to develop the doctrine of necessity. It is not a universally acceptable principle of law, such as the doctrine of natural justice, which includes two basic principles. The first one “*nemo iudex in causa sua*” is a rule against bias and the second one “*audi alteram partem*” says that no person shall be condemned without being heard [27].

Contrary to general principles of law, the doctrine of necessity is an extra-legal tool which is invoked to justify acts that fall outside the purview of the law, and, at the same time, are necessary to preserve State continuity. Necessity in the legal context is codified into Article 25 of the International Law Commission Draft Articles on State Responsibility [28]. This article equates necessity with defense. It states that necessity is invoked by respondent States primarily in times of crisis, as an excuse for the breach of an international law obligation, as part of litigation or arbitration proceedings.

In a few words, the legal doctrine of necessity is, really, the ability of a source of authority to justify an act not in conformity with law, but required in exceptional cases, where there is a grave danger either for the essential interests of a society as State or for the society as a whole community.

Historically, the doctrine of necessity has been invoked in times of crises, e.g., military, environmental, economic crisis [29], as well as in light of the current pandemic situation [30], where authorities adopted measures (i.e., restrictions, like lockdown, shut down of businesses and borders, etc.), which result in breaches of international law, to safeguard public health.

4. The doctrine of necessity to validate AI development

“A large planetary object called ‘AI-Dark Energy’ is racing towards the Earth. In the year 2122, a group of prominent scientists, including AI-experts, monitor the development of the bizarre phenomenon and they predict that the collision will take place over the next hundred years, resulting in deactivation of the northern hemisphere, which will dramatically change the natural and societal scenery. After fifty years of research, in 2172, the scientists’ team found that the only solution to avoid the unexpected danger is the AI development. Having this in mind, researchers worked hard for many years and, finally, in 2222, they managed to save the Earth by changing the orbit of the invader planet, due to achievements in the AI field.” (Extract from Panagiotis’ essay on the future of earth in the hi-tech Epoch—a 12-year-old high-school student). This quotation could be worthy of a science-fiction movie. Enhancing the plot of the imaginary scenario, suppose that a group of people sues the group of AI experts for human rights abuses. The applicants argue that the rapid increase of AI has birthed an era of unemployment due to the displacement of human labor, thus resulting in breach of human rights to work. They also claim that AI algorithms have shown discriminatory tendencies, especially based on gender and race, thus violating the legal principle of equality. Furthermore, the complainants argue that the AI systems usage for surveillance purposes, such as the facial recognition system to record images of people inside vehicles entering and leaving a country, limits the right of freedom of movement. They, finally, argue that sudden technological shifts intrude into the lives of citizens and abuse their basic human right to privacy. At trial, the AI-experts raised the defense of necessity. They plead that concern about humanity was the only motive for their actions and that the damage they caused was minute compared to the impending catastrophe that they were attempting to avert. Should the court convict or acquit the AI-experts?

The question raises by the fictive scenario about the invocation of the doctrine of necessity to validate AI development is not purely hypothetical. On the contrary, it reflects issues that have already appeared in many studies [31,32].

According to the judicial authorities, the main crucial prerequisite for the application of the doctrine of necessity is an imperative and unavoidable need or exceptional threat against the citizens of the countries in need, and the non-existence of alternative.

In the case of the above scenario, governments around the world struggle to resolve the dilemma about how to tackle the emergency. Protecting humanity, by

investing in the AI development and integration, and sacrificing some human rights, or defending them by limiting the AI achievements. Governments of the countries of the northern hemisphere decided to recruit scientists to develop AI mechanisms to prevent the collision. But the ubiquitous nature of AI applications affects citizens all over the world. Thus, many human rights were violated, and some States have breached their obligations. The defendants, that is the group of AI experts, were found not guilty because they acted in good faith by performing his duty as scientists, and as the court held that such situation of absolute necessity (i.e., to prevent the deactivation of northern hemisphere) may be considered a defense against the charge of human rights violation.

The idea behind this fictitious scenario reflects the dilemma which represents the spectrum of modern ethical issues resulting from the use of AI in every aspect of human activity: AI-based growth without limits or limited growth without AI development. This dilemma is the product of structural and functional relations between societal and legal systems. The choice to give priority to one aspect of the dilemma over another reflects co-evolution systems demands. According to Systems Theory, systems evolve during their lifecycle and are not independent, are interrelated and in most cases also intertwined. Their interrelations are often not explicit and impose changes resulting in inconsistencies to one another due to their inherent features [33].

In a few words, to search for a way to validate AI development and integration is to search for a way to reform the strict legal praxis into a more flexible one, by using the doctrine of necessity as the appropriate mechanism, since it ensures a degree of flexibility. From the systemic point of view, flexibility is necessary for a system to have options to be able to cope with its dynamic environment [34]. From the legal point of view, flexibility allows the law to adapt to changing needs and it correlates with emergencies [35]. Reliance on necessity would add flexibility in law and thus, legitimacy to AI development.

5. Conclusions and future work

In this paper, a philosophical discussion of “necessity” in society and law, is introduced. This attempt is made to draw clear distinctions between the different nuances of necessity, within societal and legal context. When one refers to necessity in a societal context, then one refers to either natural or artificial necessity, and that should be explicit connected with a human self-protective act to adapt or survive in changing environments. On the other hand, there is a place for the concept of necessity in the legal context, which is associated with the self-preservation of a State, or defense. Therefore, the notion of necessity has an analogy in the legal theoretical apparatus, which is strictly defined as the doctrine of necessity, so to avoid confusion with the universally acceptable principles of law. This extra-legal tool may function as particularly useful reference and stepping-stone for researchers conducting further studies on the co-evolution of social and legal systems. It reflects modern ethical issues resulting from the use of AI in every aspect of human activity: AI-based growth without limits or limited growth without AI development. This dilemma is the product of structural and functional relations between societal and legal systems. The choice

to give priority to one aspect of the dilemma over another reflects co-evolution systems demands.

According to Systems Theory, systems evolve during their lifecycle and are not independent, are interrelated and in most cases also intertwined. Their interrelations are often not explicit and impose changes resulting in inconsistencies to one another due to their inherent features. In this perspective, the doctrine of necessity in systems co-evolution demands is to be taken into account for future work. This will contribute to the understanding of AI and society co-evolution in two ways: (1) to find a general legal framework that validates technological advances in societal transformations and, (2) to identify the specific mechanisms that could link this knowledge to what is known in the legal praxis as reflexive law.

Author contributions: Conceptualization, LS and IK; methodology, LS and IK; validation, LS and IK; formal analysis, LS and IK; investigation, LS and IK; resources, LS and IK; data curation, LS and IK; writing—original draft preparation LS and IK; writing—review and editing, LS and IK; visualization, LS and IK; supervision, LS and IK; project administration, LS and IK. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Ruiz-Real JL, Uribe-Toril J, Torres JA, et al. Artificial intelligence in business and economics research: trends and future. *Journal of Business Economics and Management*. 2020; 22(1): 98-117. doi: 10.3846/jbem.2020.13641
2. Khan ZF, Alotaibi SR. Applications of Artificial Intelligence and Big Data Analytics in m-Health: A Healthcare System Perspective. *Journal of Healthcare Engineering*. 2020; 2020: 1-15. doi: 10.1155/2020/8894694
3. Sharma A, Georgi M, Tregubenko M, et al. Enabling smart agriculture by implementing artificial intelligence and embedded sensing. *Computers & Industrial Engineering*. 2022; 165. doi: 10.1016/j.icie.2022.107936
4. Simmler M, Canova G, Schedler K. Smart criminal justice: Phenomena and normative requirements. *International Review of Administrative Sciences*. 2021; 89(2): 415-432. doi: 10.1177/00208523211039740
5. Boute RN, Gijsbrechts J, Van Mieghem, JA. Digital lean operations: Smart automation and artificial intelligence in financial services. In: *Innovative Technology at the Interface of Finance and Operations*. Springer, Cham; 2022. pp. 175-188.
6. Wanes SRAO. Artificial intelligence role for advertising campaigns development. *International Journal of Artificial Intelligence and Emerging Technology*. 2021; 4(1): 1-16.
7. Xu L. Artificial intelligence for general layout and transportation engineering with GIS technology. In: *International Conference on Big Data Analytics for Cyber-Physical-Systems*. Springer, Singapore; 2021. pp. 927-934.
8. Yang SJH, Ogata H, Matsui T, et al. Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*. 2021; 2: 100008. doi: 10.1016/j.caeai.2021.100008
9. Molano JC, Molano JCC. Criticalities and advantages of the use of artificial intelligence in research. In: *Handbook of Research on Advanced Research Methodologies for a Digital Society*. IGI Global; 2022. pp. 161-175.
10. Stahl BC, Rodrigues R, Santiago N, et al. A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review*. 2022; 45: 105661. doi: 10.1016/j.clsr.2022.105661
11. Adamson P. Making a Virtue of Necessity: Anankê in Plato and Plotinus. *Études platoniciennes*. 2011; (8): 9-30. doi: 10.4000/etudesplatoniciennes.442
12. Commisso VJ. *Dynamic politics: Necessity founding, and (re)founding in Machiavelli's Discourses on Livy* [PhD thesis]. State University of New York at Albany; 2016.
13. Tanzi A. *State of necessity*. Oxford Public International Law. 2021.
14. Dixon F. Global system change: A Whole System Approach to Societal Transformation. *Cadmus*. 2021; 4(5): 147-154.

15. Sweetman B. Evolution, chance and necessity in the Universe. *Revista Portuguesa de Filosofia*. 2010; 66(4): 897-910.
16. Leech J. Kant on the Necessity of Necessity. *History of Philosophy & Logical Analysis*. 2021; 25(1): 66-94. doi: 10.30965/26664275-02400005
17. Bray C. The philosophy of necessity or natural law as applicable to moral mental and social science. Kessinger Publishing; 2007.
18. Xie VC, Pu J, Metzger BP, et al. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *eLife*. 2021; 10. doi: 10.7554/elife.67336
19. Rother S. Global migration governance from below in times of covid-19 and “zoomification”: civil society in “invited” and “invented” spaces”. *Comparative Migration Studies*. 2022; 10(1). doi: 10.1186/s40878-021-00275-9
20. Vadi V. International law and its histories: Methodological risks and opportunities. *Harvard International Law Journal*. 58(2): 311-352.
21. Wong BBM, Candolin U. Behavioral responses to changing environments. *Behavioral Ecology*. 2014; 26(3): 665-673. doi: 10.1093/beheco/aru183
22. Ruhl JB, Katz DM, Bommarito MJ. Harnessing legal complexity. *Science*. 2017; 355(6332): 1377-1378. doi: 10.1126/scienceaag3013
23. Lee B, Lee KM, Yang JS. Network structure reveals patterns of legal complexity in human society: The case of the Constitutional legal network. *PLoS ONE*. 2019; 14(1). doi: 10.1371/journal.pone.029844
24. Clegg M, Ellena K, Ennis D, et al. The hierarchy of laws: Understanding and implementing the legal frameworks that govern elections. *International Foundation for Electoral Systems*. 2016.
25. Kolb R. General principles of law, jus cogens and the unity of the international legal order. In: *General Principles and the Coherence of International Law*. 2019. pp. 60-64.
26. Green JA. Self-preservation. *Max Planck Encyclopedia of Public International Law (MPEPIL)*. 2009.
27. Silkenat JR, Hickey JE, Barenboim PD, et al. *The Legal Doctrines of the Rule of Law and the Legal State (Rechtsstaat)*. Springer International Publishing; 2014. doi: 10.1007/978-3-319-05585-5
28. Draft articles on Responsibility of States for International Wrongful Acts, with commentaries. *Yearbook of the International Law Commission*.
29. Hayashi N. *Military necessity: The art, morality and law of war*. Cambridge University Press; 2020.
30. Mark W, Weidemaier C, Gulati M. Necessity and the covid-19 pandemic. *Capital Markets Law Journal*. 2020; 15(3): 277-283. doi: 10.1093/cmlj/kmaa013
31. Cervantes S, López S, Cervantes JA. Toward ethical cognitive architectures for the development of artificial moral agents. *Cognitive Systems Research*. 2020; 64: 117-125. doi: 10.1016/j.cogsys.2020.08.010
32. Kassens-Noor E, Siegel J, Decaminada T. Choosing Ethics Over Morals: A Possible Determinant to Embracing Artificial Intelligence in Future Urban Mobility. *Frontiers in Sustainable Cities*. 2021; 3. doi: 10.3389/frsc.2021.723475
33. Pickel A. Rethinking systems theory: A programmatic introduction. *Philosophy of the Social Sciences*. 2007; 37(4): 391-407. doi: 10.1177/0048393107307809
34. Sushil P. Theory of flexible systems management. In: Sushil P, Connell J, Burger J (editors). *Flexible Work Organizations. Flexible Systems Management*. Springer; 2016.
35. Roosevelt K. Certainty versus flexibility in the conflict of laws. In: Ferrari F, Fernandez Arroyo DP (editors). *Private International Law: Contemporary Challenges and Continuity Relevance*. 2019.

Utilizing emotion recognition technology to enhance user experience in real-time

Yuanyuan Xu^{1,*}, Yin-Shan Lin², Xiaofan Zhou³, Xinyang Shan¹

¹ College of Design and Innovation, Tongji University, Shanghai 200092, China

² Khoury College of Computer Science, Northeastern University, Boston, MA 02115, United States

³ Department of Computer and Information of Science and Engineering, University of Florida, Gainesville, FL 32611, United States

* Corresponding author: Yuanyuan Xu, ecusttethys@foxmail.com

CITATION

Xu Y, Lin YS, Zhou X, Shan X.
Utilizing emotion recognition
technology to enhance user
experience in real-time. *Computing
and Artificial Intelligence*. 2024;
2(1): 1388.
<https://doi.org/10.59400/cai.v2i1.1388>

ARTICLE INFO

Received: 17 April 2024

Accepted: 30 May 2024

Available online: 14 June 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence
is published by Academic Publishing
Pte. Ltd. This work is licensed under
the Creative Commons Attribution
(CC BY) license.
[https://creativecommons.org/licenses/
by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: In recent years, advancements in human-computer interaction (HCI) have led to the emergence of emotion recognition technology as a crucial tool for enhancing user engagement and satisfaction. This study investigates the application of emotion recognition technology in real-time environments to monitor and respond to users' emotional states, creating more personalized and intuitive interactions. The research employs convolutional neural networks (CNN) and long short-term memory networks (LSTM) to analyze facial expressions and voice emotions. The experimental design includes an experimental group that uses an emotion recognition system, which dynamically adjusts learning content based on detected emotional states, and a control group that uses a traditional online learning platform. The results show that real-time emotion monitoring and dynamic content adjustments significantly improve user experiences, with the experimental group demonstrating better engagement, learning outcomes, and overall satisfaction. Quantitative results indicate that the emotion recognition system reduced task completion time by 14.3%, lowered error rates by 50%, and increased user satisfaction by 18.4%. These findings highlight the potential of emotion recognition technology to enhance user experiences. However, challenges such as the complexity of multimodal data integration, real-time processing capabilities, and privacy and data security issues remain. Addressing these challenges is crucial for the successful implementation and widespread adoption of this technology. The paper concludes that emotion recognition technology, by providing personalized and adaptive interactions, holds significant promise for improving user experience and offers valuable insights for future research and practical applications.

Keywords: emotion recognition; user experience; human-computer interaction

1. Introduction

In recent years, human-computer interaction (HCI) has experienced rapid advancements, driven by significant technological progress. Among these advancements, emotion recognition technology has emerged as a crucial tool for enhancing user experience [1,2]. This technology leverages sophisticated algorithms to detect and interpret human emotions through various data inputs, such as facial expressions, voice intonations, and physiological signals. The ability to understand and respond to users' emotional states in real-time enables systems to provide more intuitive, engaging, and personalized interactions [3,4].

Emotion recognition technology is increasingly being integrated into a wide range of applications. In online education, for example, this technology can monitor students' emotional states and adjust teaching content dynamically to maintain engagement and improve learning outcomes [5]. In healthcare, emotion recognition

can help in monitoring patients' mental health, providing timely interventions and personalized care [6]. In customer service, it can enhance user satisfaction by allowing systems to respond empathetically to users' emotions [7]. In entertainment and gaming, emotion recognition can adjust content based on players' emotional responses, creating more immersive and enjoyable experiences [8].

Despite the significant potential of emotion recognition technology, several challenges remain. Integrating multiple data sources (e.g., facial expressions, voice, physiological signals) to achieve accurate and robust emotion recognition is complex and requires sophisticated models and algorithms [2]. Ensuring real-time processing capabilities while maintaining high accuracy is another significant challenge. Additionally, issues related to privacy and data security must be addressed to ensure users' trust and compliance with regulations [9].

This paper explores how emotion recognition technology can be utilized to monitor users' emotional states in real-time and subsequently enhance user satisfaction and engagement through dynamic adjustments to the interface and content. By integrating facial expression and voice emotion analysis, this study aims to develop a comprehensive multimodal emotion recognition system and evaluate its effectiveness in an experimental setting. The research focuses on addressing the following key questions:

- 1) How can emotion recognition technology be effectively integrated into an online learning platform?
- 2) What are the impacts of real-time emotion monitoring and dynamic content adjustments on user satisfaction and engagement?
- 3) How do users perceive the use of emotion recognition technology in enhancing their interaction experience?

To answer these questions, this study employs a detailed experimental design involving an experimental group and a control group. The experimental group uses an online learning platform integrated with an emotion recognition system, while the control group uses a traditional online learning platform. The results of this study aim to provide insights into the practical applications of emotion recognition technology and its potential to enhance user experience across various domains.

2. Literature review

2.1. Methodologies in emotion recognition

Emotion recognition technology encompasses several methodologies, including computer vision, voice analysis, and physiological signal analysis [3,10].

Computer vision techniques identify emotions by analyzing facial expressions. These methods often utilize convolutional neural networks (CNNs) to detect and classify facial features associated with different emotions [11]. CNNs are particularly effective in recognizing complex facial expressions as they can capture subtle variations in facial features. For example, studies have shown that CNNs can differentiate between similar emotions, such as joy and excitement, by analyzing minute differences in facial muscle movements [12]. Keypoint detection is another common approach, where specific points on the face, such as the corners of the eyes and mouth, are tracked to interpret emotions. Techniques like facial action coding

systems (FACS) are also employed to systematically categorize facial movements and their corresponding emotions [13].

Voice analysis methods recognize emotions by examining acoustic features such as pitch, rhythm, and volume. Hidden Markov models (HMMs) and long short-term memory networks (LSTMs) are commonly used algorithms in this domain [14]. HMMs model the temporal dynamics of speech, making them suitable for sequential data like voice recordings [15]. They excel at capturing short-term dependencies and transitions between different states of speech. LSTMs, on the other hand, are adept at capturing long-term dependencies in sequential data, making them effective for analyzing the nuances in speech that correspond to different emotional states. LSTMs can maintain information over extended periods, allowing for a more comprehensive analysis of how emotions evolve during a conversation [16].

Physiological signal analysis involves monitoring physiological signals such as heart rate, skin conductivity, and brain activity to detect emotions. Techniques like electroencephalography (EEG) and galvanic skin response (GSR) are often used. EEG measures electrical activity in the brain, providing insights into emotional states through brain wave patterns [17,18]. GSR measures the electrical conductance of the skin, which varies with sweat gland activity influenced by emotional arousal. These methods provide a direct measurement of the physiological responses associated with different emotional states, offering a complementary perspective to facial and vocal analysis. Combining these physiological signals with computer vision and voice analysis can enhance the accuracy and robustness of emotion recognition systems [19].

2.2. Applications and challenges in emotion recognition

Emotion recognition technology has a wide range of applications across various domains, significantly enhancing user experience and interaction quality.

In the context of online education, emotion recognition technology can monitor students' emotional states to adjust teaching content dynamically [5,20]. Real-time emotion monitoring can improve student engagement and learning outcomes by tailoring the educational experience to the emotional needs of the students. For instance, if a student appears frustrated, the system can offer additional support or simplify the content to alleviate frustration and enhance learning efficiency [5].

In healthcare, emotion recognition technology plays a crucial role, particularly in mental health monitoring and treatment [6]. By continuously monitoring patients' emotional states, healthcare providers can offer timely interventions and personalized care. For example, emotion recognition can help detect early signs of depression or anxiety, enabling proactive mental health management [21]. Continuous monitoring can also provide valuable data for understanding long-term emotional trends and their correlation with treatment outcomes.

In customer service, emotion recognition technology can enhance user satisfaction by allowing systems to respond empathetically to users' emotions. Automated customer service agents can adjust their responses based on the detected emotional state of the user, providing a more personalized and satisfying interaction experience. For example, an empathetic response to a frustrated customer can de-escalate tension and improve the overall customer experience [7].

In the entertainment and gaming industry, emotion recognition can create more immersive experiences by adjusting content based on players' emotional responses. For instance, a game can become more challenging if the player is detected to be bored or less intense if the player appears stressed, thereby maintaining optimal engagement levels. Emotion-driven adjustments can enhance the gaming experience by making it more adaptive and personalized [8].

Despite the advancements and potential applications, several challenges remain in the implementation of emotion recognition technology.

Integrating multiple data sources, such as facial expressions, voice, and physiological signals, to achieve accurate and robust emotion recognition is complex. Each modality provides unique information, and effectively combining these sources requires sophisticated models and algorithms [22,23]. Multimodal fusion techniques must address issues such as data synchronization, feature alignment, and the handling of missing or noisy data [24].

Ensuring real-time processing capabilities while maintaining high accuracy is a significant challenge. Emotion recognition systems must process data quickly to provide timely responses, which can be computationally intensive, especially when dealing with multiple data streams simultaneously [25]. Optimizing algorithms for speed without compromising accuracy requires advanced computational techniques and hardware acceleration.

Issues related to privacy and data security must be addressed to ensure users' trust and compliance with regulations. Emotion recognition involves collecting sensitive data, and safeguarding this information is crucial to prevent misuse and protect user privacy [26]. Implementing robust encryption, secure data storage, and strict access controls are essential measures to ensure the confidentiality and integrity of emotional data.

In summary, while emotion recognition technology holds immense potential to enhance user experience across various domains, addressing these methodological and implementation challenges is crucial for its successful deployment. Continued research and innovation in this field will pave the way for more effective and secure emotion recognition systems.

3. Methodology

3.1. Experimental design

To ensure a comprehensive and effective experimental design, this study involves an experimental group and a control group. The experimental group uses an online learning platform integrated with an emotion recognition system that can monitor participants' emotional states in real-time and dynamically adjust learning content and interfaces based on emotional changes. In contrast, the control group uses a traditional online learning platform without emotion recognition features, with learning content and interface remaining unchanged.

Participants include 30 students of different ages, genders, and backgrounds to ensure diversity and representativeness, the demographic distribution of the participants is shown in **Table 1**. Random grouping is used to divide participants into experimental and control groups, with 15 participants in each group. The experiment

is set on an online learning platform that provides the following three main tasks: watching instructional videos, participating in discussions, and completing quizzes.

Table 1. Participant demographics.

Demographic attribute	Experimental group	Control group
Total participants	15	15
Gender		
Male	8	7
Female	7	8
Age		
18–24	5	5
25–30	7	6
31–35	3	4
Educational background		
Undergraduate	8	9
Graduate	7	6

3.2. Task descriptions

During the task of watching instructional videos, a camera is used to collect real-time facial expression image data from participants. The system analyzes participants' facial expressions in real-time using a CNN model to recognize emotional states such as happiness, confusion, and boredom. Based on participants' emotional states, the system dynamically adjusts video content and playback speed for the experimental group. During the task of participating in discussions, a microphone is used to collect real-time voice data from participants. The system analyzes participants' voice emotions in real-time using an LSTM model to recognize emotional states such as excitement, calmness, and anxiety. Based on participants' emotional states, the system provides interaction prompts and guidance for the experimental group.

In the task of completing quizzes, the system records participants' answer times, answer sequences, and answer accuracy. It monitors participants' facial expressions and voice emotions in real-time, analyzing emotional changes for the experimental group. The system provides instant feedback and assistance based on participants' emotional states for the experimental group.

The data collection includes facial expression data using a camera and analyzing it with a CNN model, voice emotion data using a microphone and analyzing it with an LSTM model, and user behavior data recording all participant operations on the platform.

3.3. Model selection and implementation

The study utilizes a combination of convolutional neural networks (CNN) and long short-term memory (LSTM) networks to build a multimodal emotion recognition system. The CNN model, designed for facial expression recognition, includes multiple convolutional and pooling layers to extract image features, followed by fully connected layers and a SoftMax classifier for the final classification. The architecture of the CNN model consists of an input layer that processes 48×48 grayscale images,

three convolutional layers with 32, 64, and 128 filters respectively, each followed by 2×2 max pooling layers, and a fully connected layer with 64 neurons, concluding with a SoftMax layer that outputs seven emotion categories. The CNN model is trained using the Adam optimizer and cross-entropy loss function, leveraging the FER-2013 dataset, which contains facial images representing various emotions. The training process spans 50 epochs and incorporates data augmentation techniques to enhance generalization (**Table 2**).

Table 2. Participant demographics.

Layer	Details
Input layer	48×48 grayscale images
Conv layer 1	32 filters, 3×3 kernel, ReLU
Max pooling 1	2×2 pool size
Conv layer 2	64 filters, 3×3 kernel, ReLU
Max pooling 2	2×2 pool size
Conv layer 3	128 filters, 3×3 kernel, ReLU
Max pooling 3	2×2 pool size
Fully connected	64 neurons, ReLU
Output layer	SoftMax, 7 emotion categories

For voice emotion recognition, the LSTM model is employed. This model comprises multiple LSTM layers to process the time series features of voice signals, along with fully connected layers for final classification. The LSTM model's architecture includes an input layer for 40-dimensional Mel-Frequency Cepstral Coefficients (MFCC) features, two LSTM layers with 128 and 64 units respectively, and a fully connected layer with 32 neurons, culminating in a SoftMax layer with seven neurons corresponding to the emotion categories. The LSTM model is also trained using the Adam optimizer and sparse cross-entropy loss function, using the RAVDESS dataset, which includes voice recordings of various emotions. The training process involves 50 epochs with 20% of the data allocated for validation (**Table 3**).

Table 3. Participant demographics.

Layer	Details
Input layer	40-dimensional MFCC features
LSTM layer 1	128 units, tanh
LSTM layer 2	64 units, tanh
Fully connected	32 neurons, ReLU
Output layer	SoftMax, 7 emotion categories

To integrate the results from both facial expression and voice emotion recognition, a multimodal fusion approach is employed. The outputs of the CNN and LSTM models are concatenated and passed through a fully connected layer for the final emotion classification. The fusion model is trained and evaluated using an independent validation set. Performance metrics, including accuracy, recall, and F1 score, demonstrate the superior performance of the multimodal fusion model over

single-modal models (Table 4).

Table 4. Performance metrics comparison.

Emotion category	Accuracy (CNN)	Accuracy (LSTM)	Accuracy (Fusion)	Recall (CNN)	Recall (LSTM)	Recall (Fusion)	F1 score (CNN)
Happy	0.88	0.89	0.92	0.84	0.86	0.88	0.86
Confused	0.85	0.86	0.89	0.81	0.82	0.85	0.83
Bored	0.80	0.83	0.85	0.76	0.79	0.82	0.78
Calm	0.90	0.93	0.95	0.87	0.90	0.93	0.88
Excited	0.86	0.88	0.90	0.83	0.85	0.87	0.84
Anxious	0.82	0.84	0.87	0.78	0.80	0.83	0.80
Overall	0.85	0.87	0.90	0.80	0.84	0.86	0.82

4. Results

This section provides a detailed analysis of how improved emotion classification results translate into enhanced task performance and user satisfaction. The critical link between enhanced emotion classification algorithms and task performance improvements is thoroughly discussed.

4.1. Quantitative and qualitative analysis results

The integration of advanced CNN and LSTM models allowed for accurate real-time detection of participants' emotional states. The improved classification accuracy (90% for the fusion model) enabled the system to make precise adjustments to the content, thereby enhancing the overall learning experience. This improved emotion recognition directly impacted task performance and user satisfaction in several ways (Table 5).

Table 5. Quantitative analysis results.

Indicator	Experimental group (Mean ± SD)	Control group (Mean ± SD)	Statistical significance (p-value)
Task completion time (seconds)	300 ± 45	350 ± 50	<0.05
Error rate (%)	5 ± 2	10 ± 3	<0.01
Quiz score (out of 100)	85 ± 8	75 ± 10	<0.05
User satisfaction rating (1–5)	4.5 ± 0.5	3.8 ± 0.6	<0.01

To evaluate task completion time, we measured the duration each participant took to complete the assigned tasks. The results indicated that participants in the experimental group completed their tasks significantly faster than those in the control group. Specifically, the average task completion time for the experimental group was 300 ± 45 s, whereas the control group took an average of 350 ± 50 s. This suggests that the emotion recognition system's real-time adjustments to the learning content helped maintain participant engagement and efficiency.

Error rate was another critical metric, particularly for the quizzes. Participants in the experimental group had a notably lower error rate, averaging 5% ± 2%, compared to the control group's 10% ± 3%. This reduction in errors indicates that the emotion recognition system could identify and mitigate participants' anxiety and confusion

during quizzes, providing calming and supportive feedback that contributed to more accurate responses.

Quiz scores further highlighted the benefits of the emotion recognition system. The experimental group achieved higher scores, averaging 85 ± 8 out of 100, while the control group averaged 75 ± 10 . This performance gap underscores the effectiveness of real-time emotional monitoring and content adjustment in enhancing participants' understanding and retention of the material.

User satisfaction was assessed through a questionnaire survey, utilizing a Likert scale to gauge participants' satisfaction with their learning experience. The survey covered various aspects of the learning experience, including the relevance and practicality of the content, the effectiveness of interaction discussions, the friendliness of the system interface, and the timeliness of emotional feedback.

The results from the satisfaction ratings revealed that the experimental group had significantly higher satisfaction levels compared to the control group. The experimental group's average satisfaction rating was 4.5 ± 0.5 on a 5-point scale, indicating a high level of satisfaction. In contrast, the control group had an average satisfaction rating of 3.8 ± 0.6 . The higher satisfaction ratings in the experimental group suggest that the emotion recognition system's ability to provide personalized learning experiences and timely emotional feedback contributed to a more positive learning experience overall.

In summary, the quantitative analysis demonstrates that the emotion recognition system significantly improved task completion performance and user satisfaction. The system's real-time emotional monitoring and content adjustments led to faster task completion times, lower error rates, higher quiz scores, and greater overall satisfaction among participants in the experimental group compared to those in the control group.

Table 6. Qualitative analysis results.

Theme	Experimental group feedback example	Control group feedback example
Overall learning experience	"The system's timely content adjustments kept me focused."	"Sometimes the video content felt monotonous."
Satisfaction with interaction discussions	"The discussion prompts and guidance were very helpful."	"It's hard to stay active in discussions, lacking motivation."
Emotional changes during quizzes	"The system's feedback during quizzes reduced my anxiety."	"I felt a lot of pressure during quizzes without much help."
Views on the emotion recognition system	"The system's understanding of my emotions and adjustments were great."	"Without the emotion recognition system, the learning process sometimes felt dull."
Suggestions for future improvements	"More interactive content and richer learning resources would be great."	"I hope the platform can be more personalized, providing content adjustments based on my emotions."

The qualitative analysis of the experimental and control groups provides insights into participants' subjective experiences with the learning platform (Table 6). Feedback was collected and organized into key themes, highlighting differences in the overall learning experience, satisfaction with interaction discussions, emotional changes during quizzes, views on the emotion recognition system, and suggestions for future improvements.

Participants in the experimental group frequently mentioned the benefits of the

emotion recognition system's real-time adjustments. For example, one participant stated that the system's timely content adjustments kept them focused, indicating that the adaptive nature of the platform helped maintain their engagement. In contrast, a participant from the control group remarked that sometimes the video content felt monotonous, suggesting that the lack of dynamic adjustments led to a less engaging experience.

The experimental group found the discussion prompts and guidance provided by the system to be very beneficial. One participant noted that the discussion prompts and guidance were very helpful, indicating enhanced interaction and communication. On the other hand, a control group participant expressed difficulty staying motivated, saying it was hard to stay active in discussions, lacking motivation.

During quizzes, participants in the experimental group reported reduced anxiety due to the system's feedback. One participant shared that the system's feedback during quizzes reduced their anxiety, highlighting the supportive role of the emotion recognition system. Conversely, a control group participant felt pressured, stating that they felt a lot of pressure during quizzes without much help.

Participants in the experimental group appreciated the system's ability to understand and respond to their emotions. One participant mentioned that the system's understanding of their emotions and adjustments were great, reflecting satisfaction with the personalized experience. In contrast, a control group participant noted that without the emotion recognition system, the learning process sometimes felt dull, indicating a preference for the enhanced interaction provided by the system.

Participants in the experimental group suggested further enhancements, such as more interactive content and richer learning resources. This feedback points to a desire for even more engaging and varied content. Similarly, a control group participant expressed a need for personalization, stating that they hope the platform can be more personalized, providing content adjustments based on their emotions.

4.2. Linking enhanced emotion classification to performance improvement

The key to the improved task performance and user satisfaction lies in the sophisticated emotion classification algorithms employed. The CNN and LSTM models provided high accuracy in detecting subtle emotional cues from facial expressions and voice intonations. These accurate classifications allowed the system to personalize content by recognizing when a participant was confused or bored and adjusting the content to better match the participant's emotional state, thereby maintaining engagement and improving comprehension. During discussions and quizzes, the system's ability to detect anxiety or excitement allowed it to provide immediate and relevant feedback, reducing stress and enhancing performance. The dynamic adjustments based on emotional states kept participants engaged throughout the tasks, preventing disengagement and promoting a more immersive learning experience.

The experimental results support the research hypothesis that the emotion recognition system significantly improves users' learning experience and satisfaction. The enhanced emotion classification algorithms enabled precise real-time adjustments

to the content, which directly translated into better task performance and higher user satisfaction. The system's ability to monitor and respond to users' emotional states in real-time and provide personalized interactions highlights the potential of emotion recognition technology in various applications, including online education, healthcare, customer service, and entertainment.

5. Discussion

5.1. Innovation and problem solving

This study presents a novel approach to enhancing user experience through the integration of advanced emotion recognition technology. By leveraging sophisticated CNN and LSTM models, we achieved high accuracy in detecting emotional states in real-time, enabling dynamic adjustments to content and interactions based on users' emotions. Our research addresses several critical challenges in the field of emotion recognition and user experience. Specifically, our work introduces a comprehensive multimodal emotion recognition system that combines facial expression analysis with voice emotion recognition. The innovative use of CNN and LSTM models allows for the accurate detection of subtle emotional cues, leading to precise content adjustments. This integration addresses the challenges of real-time emotional monitoring and personalized user experience, enhancing engagement and satisfaction, particularly in online education where it significantly improves learning outcomes.

5.2. Advancing the field of emotion recognition and user experience

Our findings contribute to the advancement of emotion recognition technology and its application in improving user experiences. The successful integration of CNN and LSTM models demonstrates the potential of multimodal approaches in achieving higher accuracy and robustness in emotion detection. This research pushes the boundaries of current methodologies and provides a framework for future studies to build upon. By showcasing the effectiveness of combining different neural network models, our study sets a new standard for future research in this area. The practical implications of our emotion recognition system are profound, as it can improve user experiences across various domains, including online education, healthcare, customer service, and entertainment, by providing real-time, personalized feedback.

5.3. Application to real-world scenarios

The insights gained from our research have significant practical applications. In online education, our system can be integrated into learning platforms to provide adaptive learning experiences, helping students stay engaged and improve their academic performance. In healthcare, the system can be used to monitor patients' emotional states, providing timely interventions and personalized care. Customer service can benefit from the system's ability to detect and respond to customers' emotions, leading to more empathetic and effective interactions. In entertainment, the system can enhance user engagement by adapting content based on real-time emotional feedback. Overall, our research demonstrates the transformative potential of emotion recognition technology in enhancing user experiences across various

domains, addressing the challenges of real-time emotional monitoring and personalized content delivery, and paving the way for future innovations in this field.

6. Conclusion

This paper studies the application of emotion recognition technology in enhancing user experience, and designs and implements an experiment to verify the effect of emotion recognition systems on an online learning platform. The experimental results show that emotion recognition technology can significantly improve users' learning experience and satisfaction, supporting the effectiveness and practicality of emotion recognition. By comparing the experimental and control groups, it was found that the emotion recognition system has significant advantages in real-time monitoring of users' emotional states and dynamically adjusting learning content and interfaces. Specifically, the experimental group participants performed significantly better than the control group in task completion time, error rate, and quiz scores. The experimental group participants also scored higher in all dimensions of user satisfaction, especially in the relevance and practicality of the content, the effectiveness of interaction discussions, the friendliness of the system interface, and the timeliness of emotional feedback. These results verify the potential of emotion recognition technology in enhancing user experience, indicating that emotion recognition can achieve personalized learning experiences, thus improving learning effectiveness and user satisfaction.

In terms of practical implications, emotion recognition systems hold substantial value across various domains. In online education, these systems can monitor students' emotional states in real-time, dynamically adjusting teaching content and interaction methods to improve student engagement and learning outcomes. For massive open online courses (MOOCs), emotion recognition technology can help teachers understand students' learning states, adjust teaching strategies promptly, and provide personalized teaching services. In remote medical care, emotion recognition systems can monitor patients' emotional states, helping doctors better understand patients' emotions and psychological states to provide more targeted treatment plans. This is particularly useful in psychological counseling and mental health fields, where real-time understanding of patients' emotional changes can facilitate timely psychological intervention and support.

In the realm of intelligent customer service, emotion recognition technology can enhance the responsiveness and personalization of customer service robots by understanding users' emotional states in real-time. This leads to improved user satisfaction and service quality, especially in handling customer complaints and after-sales service where timely identification and response to user dissatisfaction can prevent conflict escalation. Furthermore, in human-computer interaction contexts such as smart homes and intelligent assistants, emotion recognition technology can enhance emotional understanding, enabling systems to adjust interaction methods based on users' emotional states, thereby providing more humanized services. In gaming and entertainment, emotion recognition technology can dynamically adjust game difficulty and plot direction based on players' emotional states, enhancing the gaming experience and player satisfaction.

For user experience design, emotion recognition technology enables systems to monitor users' emotional states in real-time and dynamically adjust content and interfaces based on emotional changes, providing personalized user experiences. Designers should leverage this technology to develop personalized interactive content and interfaces, thus improving user satisfaction. Moreover, systems should incorporate emotional feedback mechanisms to identify users' negative emotions, such as confusion, anxiety, and boredom, and provide appropriate feedback and assistance to alleviate these emotions and improve the quality of user experience. It is also crucial to prioritize user emotional data protection, ensuring that the collection, storage, and use of emotional data comply with legal regulations and ethical norms to protect users' privacy rights and data security.

Designers should consider applying multimodal emotion recognition technology in user experience design, combining facial expressions, voice, body language, and physiological signals to improve the accuracy and robustness of emotion recognition. This approach provides a more comprehensive and accurate emotional understanding and response. In summary, emotion recognition technology has significant practical application value and broad prospects for enhancing user experience. Designers should fully utilize this technology to develop innovative applications and services, thereby improving user satisfaction and experience quality.

Author contributions: Conceptualization, YX and YSL; methodology, YX, XZ and XS; software, YX and YSL; validation, YX, XZ and XS; formal analysis, YX, XZ and XS; investigation, YSL; resources, YSL; data curation, YX, XZ and XS; writing—original draft preparation, YX, XZ and XS; writing—review and editing, YSL; visualization, YX, XZ and XS; supervision, YSL; project administration, YSL; funding acquisition, YSL. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Chen CM, Wang HP. Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research*. 2011; 33(3): 244-255. doi: 10.1016/j.lisr.2010.09.010
2. Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors*. 2020; 20(8): 2384. doi: 10.3390/s20082384
3. Dzedzickis A, Kaklauskas A, Bucinskas V. Human emotion recognition: Review of sensors and methods. *Sensors*. 2020; 20(3): 592. doi: 10.3390/s20030592
4. Abdullah SMSA, Ameen SYA, Sadeeq MAM, et al. Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*. 2021; 2(01): 73-79. doi: 10.38094/jastt20291
5. Savchenko AV, Savchenko LV, Makarov I. Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. *IEEE Transactions on Affective Computing*. 2022; 13(4): 2132-2143. doi: 10.1109/taffc.2022.3188390
6. Hasnul MA, Aziz NAA, Alelyani S, et al. Electrocardiogram-based emotion recognition systems and their applications in healthcare—A review. *Sensors*. 2021; 21(15): 5015. doi: 10.3390/s21155015
7. Saxena A, Khanna A, Gupta D. Emotion recognition and detection methods: A comprehensive survey. *Journal of Artificial Intelligence and Systems*. 2020; 2(1): 53-79. doi: 10.33969/ais.2020.21005
8. Silva GM, Souto JJS, Fernandes TP, et al. Interventions with serious games and entertainment games in autism spectrum

- disorder: a systematic review. *Developmental Neuropsychology*. 2021; 46(7): 463-485. doi: 10.1080/87565641.2021.1981905
9. Hassouneh A, Mutawa AM, Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*. 2020; 20: 100372. doi: 10.1016/j.imu.2020.100372
 10. Akçay MB, Oğuz K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*. 2020; 116: 56-76. doi: 10.1016/j.specom.2019.12.001
 11. Akhand MAH, Roy S, Siddique N, et al. Facial emotion recognition using transfer learning in the deep CNN. *Electronics*. 2021; 10(9): 1036. doi: 10.3390/electronics10091036
 12. Khairuddin Y, Chen Z. Facial emotion recognition: State of the art performance on FER2013. *arXiv*. 2021; arXiv:2105.03588
 13. Canal FZ, Müller TR, Matias JC, et al. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*. 2022; 582: 593-617. doi: 10.1016/j.ins.2021.10.005
 14. Lameris H, Mehta S, Henter GE, et al. Prosody-controllable spontaneous TTS with neural HMMs. In: *Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2023. pp. 1-5.
 15. Syed SA, Rashid M, Hussain S, et al. Comparative analysis of CNN and RNN for voice pathology detection. *BioMed Research International*. 2021; 2021: 1-8.
 16. Rizvi DR, Nissar I, Masood S, et al. An LSTM based deep learning model for voice-based detection of Parkinson's disease. *International Journal of Advanced Science and Technology* 2020; 29(8).
 17. Gao Z, Dang W, Wang X, et al. Complex networks and deep learning for EEG signal analysis. *Cognitive Neurodynamics*. 2021; 15(3): 369-388. doi: 10.1007/s11571-020-09626-1
 18. Arsalan A, Majid M. Human stress classification during public speaking using physiological signals. *Computers in Biology and Medicine*. 2021; 133: 104377. doi: 10.1016/j.combiomed.2021.104377
 19. Niu X, Yu Z, Han H, et al. Video-based remote physiological measurement via cross-verified feature disentangling. In: *Proceedings of the Computer Vision—ECCV 2020: 16th European Conference*; 23–28 August 2020; Glasgow, UK. pp. 295-310.
 20. Wang W, Xu K, Niu H, et al. Emotion recognition of students based on facial expressions in online education based on the perspective of computer simulation. *Complexity*. 2020; 2020: 1-9. doi: 10.1155/2020/4065207
 21. Ayata D, Yaslan Y, Kamasak ME. Emotion Recognition from Multimodal Physiological Signals for Emotion Aware Healthcare Systems. *Journal of Medical and Biological Engineering*. 2020; 40(2): 149-157. doi: 10.1007/s40846-019-00505-7
 22. Fei Z, Yang E, Li DDU, et al. Deep convolution network based emotion analysis towards mental health care. *Neurocomputing*. 2020; 388: 212-227. doi: 10.1016/j.neucom.2020.01.034
 23. Zhang J, Yin Z, Chen P, et al. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*. 2020; 59: 103-126. doi: 10.1016/j.inffus.2020.01.011
 24. Hu J, Liu Y, Zhao J, et al. MMGCN: Multimodal fusion via deep graph convolution network for emotion recognition in conversation. *arXiv*. 2021; arXiv:2107.06779.
 25. Park CY, Cha N, Kang S, et al. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data* 2020; 7(1): 293. doi: 10.1038/s41597-020-00630-y
 26. Adjabi I, Ouahabi A, Benzaoui A, et al. Past, Present, and Future of Face Recognition: A Review. *Electronics*. 2020; 9(8): 1188. doi: 10.3390/electronics9081188

Article

The electronic knowledge publishing based on intellect modeling

Konstantin M. Golubev

General Knowledge Machine Research Group, Kyiv 01010, Ukraine; gkme kp@gmail.com

CITATION

Golubev KM. The electronic knowledge publishing based on intellect modeling. *Computing and Artificial Intelligence*. 2024; 2(1): 1332.
<https://doi.org/10.59400/cai.v2i1.1332>

ARTICLE INFO

Received: 26 April 2024
Accepted: 27 May 2024
Available online: 19 June 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: This paper focuses on a new concept based on modeling human intelligence for electronic knowledge publishing. It discusses the limitations of current publishing methods in solving real-life problems and proposes the need for a new type of publishing. It identifies two main streams of publishing: Emotional publishing and knowledge texts publishing, highlighting the challenges of the latter. The article evaluates various attempts in the field of Artificial Intelligence, such as expert systems and neural networks, in terms of knowledge transfer and problem-solving. It suggests that understanding the principles of human intelligence, as exemplified by the character Sherlock Holmes, can contribute to the development of the proposed “Intellect Modeling” concept. Overall, the article presents a comprehensive proposal that addresses the shortcomings of existing publishing models and offers a new approach that incorporates advancements in Artificial Intelligence and Knowledge Management.

Keywords: knowledge publishing; information technology; artificial intelligence; intellect modeling; knowledge management

1. Introduction

Publishers attacking! It is widely known what it is publishing. Publishers’ products are around us all the day. And sometimes it seems that they overproduce it. Why do we speak about a new kind of publishing? Is it not enough?

Sorry for taking your time, but we would like to attract your attention to the questions of suitability of publishers’ production for solving tasks of real life and, in connection with it, to needs of a new kind of publishing.

The aim is to introduce electronic knowledge publishing as extension of traditional publishing presenting knowledge contained in books in very convenient and effective way.

2. Materials and methods, results and discussion

This section provides the general experimental design and methodologies used. The aim is to provide enough details for other investigators to fully replicate results. It is also facilitate better understanding of the results obtained. Protocols and procedures for new methods are included in detail to reproduce the experiments.

This section also focuses on the results of the experiments performed and discussion.

2.1. Knowledge and knowledge publishing

There are, by our opinion, two main streams of publishing: Emotional publishing and Knowledge publishing.

Emotional publishing has a goal to put people into some emotional state. In

principle, there is no difference in this case between texts, pictures, sounds, movies, computer animations, and scents. The result should be emotional state in any case.

Knowledge, as we believe, is a basis of an ability of a person to understand what is going and to act effectively in accordance with a situation.

Individual knowledge is internal matter for a person, so no one needs, as a rule, to present this knowledge in external form for self, such as text, picture etc. Sometime people use external objects to activate their knowledge, but as we think, if they have no knowledge inside already, nothing could be activated.

Knowledge existed long before any language was created. In fact, knowledge is one of the common features of living beings. A language was created partially for that purpose to pass individual knowledge to other persons, making this knowledge public and, if needed, everlasting.

Thus, there was appeared external presentation of knowledge such as speech, texts, pictures etc. Primarily a speech and derived from speech text, that is a pictorial presentation of speech, are main modes of external presentation of knowledge. Therefore, we can call this type of publishing – knowledge texts publishing. It is subdivision of knowledge publishing.

We understand that this division of publishing on emotional and knowledge publishing is not very strict. The best emotional publishing, as a rule, contains some important knowledge about people, nature, and that makes it the most interesting. And many knowledge teachers try to present knowledge in a very attractive way.

Moreover, what is the problem with it?

The problem is that nobody can use effectively knowledge without making it individual tacit knowledge. There were appeared such branches of human activity as a teaching, learning and self-teaching. As a rule, teaching takes a long time and many efforts. Therefore, we find that an impact of knowledge publishing on real life is strictly limited—a hard-working long-time learning is needed to get and use knowledge.

A human life is not long enough to learn all the knowledge even in a very special area. Many people say about information explosion. So, what we can do? Treasure becomes larger with every day, but for a whole life a person can use just a little part of it!

There is no problem, of course, if we have a human near us who already possesses needed knowledge. We call this human an expert. We just can describe a situation to expert and get an advice how to act and what could be a possible result. The better case is if we can get advice from all the best experts and choose the better advice.

However, it's not easy thing to manage this consulting. As a rule, experts live in different places and have a very limited time for a communication. We know that it is rare luck to get an advice from a good expert. Moreover, it is almost impossible to gather several good experts, especially in urgent cases.

Deadlock?

There were many efforts to bypass these problems. There were attempts to define a structure of knowledge texts, to introduce key words, hypertexts, databases, fuzzy search engines, Knowledge Management Suites, LLM etc. It makes easier to find knowledge texts regarding existing problems.

And in many simple cases it is working very effective. For example, you can use

a detailed instruction to maintain your TV set. It is interesting that even in this case before action human needs to self-teach to make knowledge internal. However, if a situation was not described in an instruction or a manual is very thick, you will call experts.

Other ways?

Yes. There were attempts for to imitate an expert's activity. We can call works of W.Maccalloh and W.Pitts (Neural Networks); F.Rosenblat (perceptron); M.Minsky and S.Papert (anti-perceptronism, frames); Newell, Simon and Shaw (Logician-Theoretician, General Problem Solver); B.Buchanan and E.Feigenbaum (expert systems, DENDRAL, decision trees); E.Shortliffe (MYCIN, EMYCIN, deduction machine); H.Pople and J.Mayers (CADUCEUS); J.McCarthy (Artificial Intelligence, LISP); M. Ross Cuillian (semantic nets); R.Schank and R.Abelson (conceptual dependence, scripts, Script-Applier Mechanism, Memory Organisation Packets); A.Samuel (Self-teaching program); P.Winston (Arches); R.Michalski (self-programming, AQ11); G.Simon and P.Langley (programs for discoveries "Bacon"); D.Lenat (Automatic mathematician, Evrisco, RLL, CYC project); D.Hillis (Connection Machines, Thinking machines corporation); G.Hinton and S.Falmen (Boltsman machine) and many, many others. The most popular in this branch of research known as Artificial Intelligence (AI) are computer programs called expert systems and devices called neural networks. An expert system is an example of a 'top-down' approach when particular instances of intelligent behaviour selected and an attempt to design machines that can replicate that behaviour was made. A neural network is an example of 'bottom-up' approach when there is an attempt to study the biological mechanisms that underlie human intelligence and to build machines, which work on similar principles. There is interesting approach called Case Based Reasoning (CBR). It is applied to search similar cases in many help desk systems. However, it is not based on intellect simulation [1-3].

Were those attempts successful?

Sorry, but it is hard to say so. And is it a crisis of human intellect?

Maybe it is a crisis of human self-confidence. In the beginning, there were many promises to built machines more intelligent than people themselves. And those machines should use advanced principles of work, much better than obsolete human intellect [4]. In this case, it is very interesting what can be called an intellect, but is based on other principles than developed by evolution?

Therefore, we believe that the real task is to help human intellect, to make it more powerful and more creative, to let knowledge work for people using the principles developed by evolution.

And what are those principles?

2.2. How do you do it, Mr. Sherlock Holmes?

Let us look at the activity of the fiction's most famous detective Mr. Sherlock Holmes. It is known that he has as prototypes real men: Dr. Joseph Bell of the Edinburgh Infirmary and Sir Arthur Conan Doyle himself [5]. So, the methods of Mr. Sherlock Holmes are realistic and widely recognized.

"No data yet... It is a capital mistake to theorize before you have all the evidence.

It biases the judgement”. (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

“I had... come to an entirely erroneous conclusion which shows, my dear Watson, how dangerous it always is to reason from insufficient data.” (Mr. Sherlock Holmes) *The Adventure of the Speckled Band* [5].

Following Mr. Sherlock Holmes, we can formulate the demand:

1) An expert should have maximum possible information about a case before a judgment.

“You seem to be a walking calendar of crime”, said Stamford with a laugh. You might start a paper on those lines. Call it the *Police News of the Past?* (to Mr. Sherlock Holmes) *A Study in Scarlet* [5].

“Kindly look her up in my index, Doctor.” For many years he had adopted a system of docketing all paragraphs concerning men and things, so that it was difficult to name a subject or a person on which he could not at once furnish information” (Mr. Sherlock Holmes) *A Scandal in Bohemia* [5].

“Like all other arts, the Science of Deduction and Analysis is one which can only be acquired by long and patient study, nor is life long enough to allow any mortal to attain the highest possible perfection in it.” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

Following Mr. Sherlock Holmes, we can formulate the demand:

2) An expert should possess a maximum possible knowledge in a sphere of activity.

“Now the skilful workman is very careful indeed as to what he takes into his brain-attic. He will have nothing but the tools which may help him in doing his work, but of this he has a large assortment, and all in the most perfect order.” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

Following Mr. Sherlock Holmes, we can formulate the demand:

3) An expert should possess no excessive knowledge, should have nothing but the tools, which may help in doing work.

“As a rule, when I have heard some slight indication of the course of events, I am able to guide myself by the thousands of other similar cases which occur to my memory.” (Mr. Sherlock Holmes) *The Red-Headed League* [5].

“They lay all the evidence before me, and I am generally able, by the help of my knowledge of the history of crime, to set them straight. There is a strong family resemblance about misdeeds, and if you have all the details of a thousand at your finger ends, it is odd if you can’t unravel the thousand and first.” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

Following Mr. Sherlock Holmes, we can formulate the demands:

4) Getting an indication of the course of events, an expert should be able to guide by the thousands of other similar cases which occur to memory.

5) Possessing information about the thousands of cases, an expert should have an ability to find a strong family resemblance about them, i.e. to find templates of typical cases.

“As a rule... the more bizarre a thing is the less mysterious it proves to be. It is your commonplace, featureless crimes are really puzzling, just as a commonplace face is the most difficult to identify.” (Mr. Sherlock Holmes) *The Red-Headed League* [5].

“I have already explained to you that what is out of the common is usually a guide rather than a hindrance.” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

Following Mr. Sherlock Holmes, we can formulate the demand:

6) An expert should focus on the most unusual in descriptions of situations.

“From long habit the train of thoughts ran so swiftly through my mind that I arrived at the conclusion without being conscious of intermediate steps. There were such steps, however. The train of reasoning ran, here is a gentleman of a medical type, but with the air of a military man. Clearly an army doctor, then. He has just come from the tropics, for his face is dark, and that is not the natural tint of his skin, for his wrists are fair. He has undergone hardship and sickness, as his haggard face says clearly. His left arm has been injured. He holds it in a stiff and unnatural manner. Where in the tropics could an English army doctor have seen so much hardship and got his arm wounded? Clearly in Afghanistan. The whole train of thought did not occupy a second.” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

Following Mr. Sherlock Holmes, we can formulate the demands:

7) An expert should have an ability to explain the grounds of conclusion.

8) An expert should arrive at the conclusion for a few seconds after getting a description of case.

“... you now pretend to deduce this knowledge I could only say what was the balance of probability. I did not at all expect to be so accurate.” (Mr Sherlock Holmes) *The Sign of Four* [5].

Following Mr. Sherlock Holmes, we can formulate the demand:

9) An expert should estimate a level of confidence of propositions.

“In solving a problem of that sort, the grand thing is to be able to reason backward. That is very useful accomplishment, and a very useful one, but people do not practise it much. In the everyday affairs of life it is more useful to reason forward, and so the other comes to be neglected. There are fifty who can reason synthetically for one who can reason analytically...Most people, if you describe a train of events to them, will tell you what the result would be. They can put those events together in their minds, and argue from them that something will come to pass. There are few people, however, who, if you told them a result, would be able to evolve from their own consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backward, or analytically.” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

Following Mr Sherlock Holmes, we can formulate the demand:

10) An expert should have an ability to take into account not only descriptions of situations in memory but results as well, providing a possibility to reconstruct a description from a result.

“For example, observation shows me that you have been to the Wigmore Street Post-Office this morning, but deduction lets me know that when there you dispatched a telegram ... The rest is deduction ... Why, of course I knew that you had not written a letter, since I sat opposite to you all morning. I see also in your open desk that you have a sheet of stamps and a thick bundle of postcards. What could you go into the post-office for, then, but to send a wire? Eliminate all other factors, and the one which remains must be the truth.” (Mr. Sherlock Holmes) *The Sign of Four* [5].

Following Mr. Sherlock Holmes, we can formulate the demand:

11) An expert should have an ability to point out all impossible hypotheses.

2.3. 11 demands of Mr. Sherlock Holmes

“He possesses two out of the three qualities necessary for the ideal detective. He has the power of observation and that of deduction. He is only wanting in knowledge, and that may come in time.” (Mr. Sherlock Holmes) *The Sign of Four* [5].

Following Mr. Sherlock Holmes, we can describe steps of expert’s activity:

- Observation;
- Producing propositions, based on a knowledge;
- Elimination of impossible propositions;
- Selection and verification of the most appropriate propositions.

Thus, if we want to help human intellect, to make it more powerful and more creative, we should make a knowledge machine, which could assist during these steps. Let us name demands to such a machine.

Step 1: Observation

1) A knowledge machine should have maximum possible information about a case before a judgment.

Step 2: Producing propositions, based on knowledge

2) A knowledge machine should possess maximum possible knowledge in a sphere of its implementation.

3) A knowledge machine should possess no excessive knowledge, should have nothing but the tools that may help in doing work.

4) Getting indication of the course of events, a knowledge machine should be able to guide itself by other similar cases, which occur to its memory.

5) A knowledge machine should have an ability to take into account not only descriptions of situations in its memory but results as well, providing a possibility to reconstruct a description from a result, i.e., if you told it a result, it would be able to evolve what the steps were which led up to that result.

6) Possessing information about the great number of cases, a knowledge machine should have an ability to find a strong family resemblance about them, i.e. to find templates of typical cases.

7) A knowledge machine should have an ability to explain the grounds of its conclusion.

8) A knowledge machine should arrive at the conclusion for a few seconds after getting a description of case.

9) A knowledge machine should focus on the most unusual in descriptions of situations.

Step 3: Elimination of impossible propositions

10) A knowledge machine should have an ability to point out all impossible propositions.

Step 4: Selection and verification of the most appropriate propositions

11) A knowledge machine should estimate a level of a confidence of its propositions

2.4. AI expert systems and neural networks

Let us look at the Artificial Intelligence expert systems. Expert system, as we understand, is based on the idea of decision tree, when, with every answer to a program's question, a direction of moving through a tree changes until a final leaf (decision) will be reached [1].

So not all possible questions will be asked, and not maximum information will be received.

The key elements are decision rules, but no knowledge itself. Not a word about the thousands of other similar cases, about typical cases.

As we see, expert systems originally are designed to be deduction machines. However, it is not very reliable to entrust to machine deciding what is impossible. We think that approach that is more fruitful is to show what reasons to consider some hypotheses as impossible are. Only man should make the final decision.

It is not amazing that development and implementation of a successful expert system is very hard work, because experts cannot think, as a rule, in terms of decision trees, and the mathematical theory of probability have a little in common with a feeling of a confidence of an expert.

Let us look at the Artificial Intelligence neural networks. Neural network is based, as we know, on the idea of teaching of set of elements (neurons), controlling conductivity between them [2]. Teaching is going usually under control of expert, which defines whether attempt is successful.

- A Neural Network is oriented on automatically produced decision rules rather than on knowledge itself. Therefore, there are no thousands of other similar cases in memory of Neural Network. And it does not produce several correct propositions simultaneously, like in a case of multiple medical diagnoses of one person.
- A Neural Network cannot explain reasons of own conclusion in terms that people can understand. So it is very hard to verify correctness of its activity and, therefore, to rely on its activity. LLM AI chatbots might explain own reasoning by falsified explanations called hallucinations. "The new AI. Systems are "built to be persuasive, not truthful," an internal Microsoft document said. This means that outputs can look very realistic but include statements that aren't true...The chatbots are driven by a technology called a large language model, or L.L.M., which learns its skills by analyzing massive amounts of digital text culled from the internet...By pinpointing patterns in that data, an L.L.M. learns to do one thing in particular: guess the next word in a sequence of words. It acts like a powerful version of an autocomplete tool. Given the sequence "The New York Times is a _____," it might guess "newspaper." ... They produce new text, combining billions of patterns in unexpected ways. This means even if they learned solely from text that is accurate, they may still generate something that is not...But becoming more accurate may also have a downside, according to a recent research paper from OpenAI [6]. If chatbots become more reliable, users may become too trusting. "Counterintuitively, hallucinations can become more dangerous as models become more truthful, as users build trust in the model when it provides truthful information in areas where they have some familiarity", the

paper said [7].

- OpenAI CEO Sam Altman, leader of most powerful AI ChatGPT development, says: “But a consistent issue with AI language models like ChatGPT is misinformation: The program can give users factually inaccurate information.” Altman said relying on the system as a primary source of accurate information “is something you should not use it for,” and encourages users to double-check the program’s results [8].

2.5. Building knowledge machine

And if we could build a knowledge machine satisfying 11 demands, it should mean that we could introduce a new kind of publishing—publishing of knowledge itself. This way published knowledge could be used for tasks of real life with very few additional efforts.

How to build such a knowledge machine?

Famous experts in Artificial Intelligence (AI) Alan Newell and Herbert Simon, developers of General Problem Solver, proposed to define memory elements as rules called ‘Productions’ of the following type ‘If Situation Then Action’. We have a right to suppose, taking into account this definition and opinion of Mr. Sherlock Holmes, that big and important part of knowledge consists of following 3-parts elements:

(Description of real problem—Name—Action and Result), that is called a concrete knowledge, or (Description of problem template—Name—Action and Result), that is called an abstract knowledge (we think that this kind of knowledge grows out of a concrete knowledge for a long lifetime).

We know that we can get information about knowledge from speech or texts. But they can be so amazingly long ... just like this paper. So, we should have to prepare a text of knowledge for input into a knowledge machine in a special way. How?

Let us consider the following example. There is a need to develop a knowledge machine designed to assist in recognizing the authors of pictures. We pick up a fragment from “Renaissance painting from Brueghel to El Greco” by Lionello Venturi [9].

“Like the Florentines, a Parma artist Francesco Mazzola (1503–1540), known as Parmigianino (i.e., little Parmesan), tended to the use of abstract forms, but, less doctrinaire in his abstractionism than such man as Rosso and Pontormo, he achieved a fragile grace and delicacy, reminiscent of Raphael and Corregio. His universal popularity contributed largely to the spread of mannerism in Europe.

The Madonna of the Long Neck (Uffizi, Florence) illustrates to perfection of his aesthetic. Here elegance replaces beauty and the somewhat abstract treatment of the figure gives it an immaterial charm. His sfumato, his discreet allusions to reality, the elongation of proportions and the sinuous movement of his figures were enthusiastically followed up by many painters in the second half of Cinquecento.”

This can be formalized as follows in a form (Description of problem template—Name—Action and Result), because it is abstract knowledge.

Description of problem template:

Description of problem consists of sentences that we call description signs. Every description sign, in principle, may have grades, usually five [10], for example (1—

Very Low, 2—Low, 3—Moderate, 4—High, 5—Very High), or have no grades at all. It is very important that every sign in description should be clear to any person and has one meaning. It is highly recommended to use sign in every description when it should be apparent.

In our case, description of situation is a set of ideas, derived from painter's style description from previous fragment:

- Tendency to the use of abstract forms
- Fragile grace and delicacy
- Elegance replace beauty
- Sfumato
- Elongation of proportions
- Sinuous movement of figures.
- In our example, as we see, no description sign has grades.

Name:

“Possible author is Francesco Mazzola (Parmigianino) from Parma, Italy (1503–1540).”

Action and result:

Description of action consists of sequence of sentences that we call action signs. Every action sign describe a sequence of elementary actions called steps. Of course, there could be just one step.

In our case, there are no action signs.

Description of result consists of sequence of one or more sentences that we call result sign.

In our case result sign could be:

“Possible author is Francesco Mazzola (Parmigianino) from Parma, Italy (1503–1540).” Since result is identical to the name here, in principle, it may be omitted.

Step 1: Observation

Let us look how (Step 1: Observation) could be realized in a knowledge machine.

We gather all description signs, from all elements of knowledge that we have, eliminate synonyms and duplicates, and numerate signs and their grades. Signs numeration sequence does not matter because numbers are just for a convenient reference. For an easiness of perception, we can group signs any way, regardless of their numbers. As a result, we get a chapter of problem's description input form called (1 Descriptions signs). The second part of problem's description input form chapter (2 Actions signs) consists of action signs, arranged in a way similar to description signs. The number of any action sign may not coincide with number of some description sign. The third part of input form chapter (3 Results signs) consists of result signs, arranged in a way similar to description signs. The number of result sign may not coincide with a number of some description sign or action sign. In principle, chapter (1 Descriptions signs) or (2 Actions signs), but not both, may not be present. Chapter (3 Results signs) may not be present at all.

So we get a form for input of description of problems with 3 chapters—(1 Descriptions signs), (2 Actions signs), (3 Results signs). For our example, it should look like this:

Problem's description input form:

1) Description signs

- Tendency to the use of abstract forms;
- Fragile grace and delicacy;
- Elegance replace beauty;
- Sfumato;
- Elongation of proportions;
- Sinuous movement of figures.

2) Actions signs

None.

3) Results signs

200 Possible author is Francesco Mazzola (Parmigianino) from Parma, Italy (1503–1540).

In this form, we should point out only that signs and grades, which are suitable for a description of existing problem.

Generated signs:

In some cases, there is a need to generate some useful signs using those, which can be observed directly. The quotation follows.

“The train of reasoning ran, here is a gentleman of a medical type, but with the air of a military man. Clearly an army doctor, then. He has just come from the tropics, for his face is dark, and that is not the natural tint of his skin, for his wrists are fair. He has undergone hardship and sickness, as his haggard face says clearly. His left arm has been injured. He holds it in a stiff and unnatural manner. Where in the tropics could an English army doctor have seen so much hardship and got his arm wounded? Clearly in Afghanistan?” (Mr. Sherlock Holmes) *A Study in Scarlet* [5].

The observed signs are:

- A gentleman of a medical type;
- The air of a military man;
- Face is dark;
- Wrists are fair;
- Haggard face;
- He holds left hand in a stiff and unnatural manner.

The generated signs (propositions) are:

- Army doctor;
- He has just come from the tropics;
- He has undergone hardship and sickness;
- Man has been injured.

Obviously, proposition itself could be a sign for further propositions. But we would like to remind that completely routine generation of propositions has to be tuned very carefully, otherwise it is not sufficiently reliable.

Step 2: Producing propositions, based on knowledge

Let us look how (Step 2: Producing propositions, based on knowledge) could be realized in a knowledge machine.

Initially, we should numerate *Name* parts of knowledge elements, which will be used as propositions, just for convenient reference. It will look like:

All possible propositions names:

- 1) Francesco Mazzola (Parmigianino) from Parma, Italy (1503–1540)

2) ...

3) ...

Every proposition is accompanied with a list of numbers of signs and grades from problem's description input form.

Getting the most possibly full description of problem, we could build a list of elements of knowledge with the most similar descriptions. It could be presented in a menu-like list of propositions, sorted according to indexes, which present value depended on degree of similarity. In our case of authors of pictures recognition assistance it will look like:

- The highly valuable propositions
- Index Number Proposition
- 90% 1) Francesco Mazzola (Parmigianino) from Parma, Italy (1503-1540)
- ...

It is very interesting question what could be an index. We think that there are many possible solutions, but we developed our own Proposition Value Index, based on idea of member of USSR Academy of Science Livanov [11] that the essence of memory associations is a spatial-temporal coherence of narrow-band periodical oscillations of central neurons sets activity [12].

Step 3: Elimination of impossible propositions

Let us look how (Step 3: Elimination of impossible propositions) could be realized in a knowledge machine. We know that knowledge element may have as a Name part a proposition like:

2) Parmigianino may not be an author.

Propositions of this type usually are very valuable, if Proposition Value Index is used. So, there's a possibility to verify manually – are there any objections against your favourite propositions?

Step 4: Selection and verification of the most appropriate propositions

Let us look how (Step 4: Selection and verification of the most appropriate propositions) could be realized in a knowledge machine.

If you choose some proposition from list to get additional information, you should get a list of signs on which proposition is based. In our case, if we select the proposition:

- 90% 1) Francesco Mazzola (Parmigianino) from Parma, Italy (1503–1540).

It should be like:

- Proposition was made according to the following signs;
- Sfumato;
- Sinuous movement of figures;
- ...

Next step: An additional information for more detailed verification of proposition should be present. It should be like:

- Francesco Mazzola (Parmigianino) from Parma, Italy (1503–1540);
- Tendency to the use of abstract forms;
- Fragile grace and delicacy;
- Elegance replace beauty;
- Sfumato;
- Elongation of proportions;

- Sinuous movement of figures;
- ...

And you have a possibility for additional verification.

We may try to use LLMs like ChatGPT to find ideas needed to build knowledge machine, by asking “Summarize” and “List great ideas”. But anyway, the list should be fine tuned by human expert. We should double-check LLM’s output as Altman warns!

3. Conclusion

We humans are bound by intellectual abilities. All knowledge is far beyond power of any person. The only way to apply knowledge is to build machines able to present it human way but not limited by volume.

Intellect Modeling Kit (IMK) is intended to build knowledge machines (KM) assisting experts on the steps of activity:

- Observation;
- Producing propositions based on knowledge;
- Elimination of impossible propositions;
- Selection and verification of the most appropriate propositions;
- Memorizing—new knowledge item creation;
- Abstraction—building objects representing typical signs of similar objects groups, data mining.

KM is not intended to replace human experts; it is built to multiply abilities. Machine should not be responsible for decisions.

The IMK is designed to create ready-to-use software applications using simple text files. Any human knowledge can be uploaded to KM by expert not familiar with software coding. Demos present in the kit.

Intellect Modeling Kit (IMK) is an open-source project located at Sourceforge repository (<http://sourceforge.net/projects/gkm-ekp>).

IMK was used for development of knowledge machines in areas of medicine, business, research, arts and psychology. Useful downloads may be found at project’s site.

KM Renaissance Painting is using ideas source: Renaissance painting from Breughel to El Greco, text by Lionello Venturi, Translated by Stuart Gilbert. This kind of system is intended to assist in paintings evaluation. Its download present at <https://sourceforge.net/projects/gkm-ekp/files/gkm-ekp/v4/gkm-intellect-modeling-kit-v2020b-renaissance-painting-win-linux.zip/download>

KM Toxicology consulting point is using questionnaire verified with the Churchill Livingstone Pocket Medical Dictionary. This kind of system is intended for practitioners and patients to assist in setting toxicology diagnosis. System is evolving: on July 2017 it recognizes 128 types of poisoning. Beta-tester from USA states: “I like it. Would be ideal for the practicing emergency physician. More references other than emedicine may be more comprehensive. But otherwise, very useful.” Download released at <https://sourceforge.net/projects/gkm-ekp/files/gkm-ekp/v4/gkm-intellect-modeling-kit-v2019b-toxicology-win-linux.zip/download>.

Will it be fight between traditional knowledge texts publishing and Electronic

Knowledge Publishing?

No, we do not think it should be. Electronic Knowledge Publishing could be valuable extension of traditional publishing, because it suggests the fastest way to use knowledge contained in published books. The customers are the same as for digital editions. One more option of edition type choice should be added: Softcover book, Hardcoverbook, Ebook, EKbook (Electronic Knowledge book). EKbook includes Ebook + Electronic Knowledge System based on that Ebook. EKbook development should not demand significant investments; it is much easier than initial book writing. The expert author or specially trained person could develop EKbook usually in few weeks. EKBook also might include knowledge from several EBooks. The customers that try to compare Ebook and EKbook versions probably would prefer EKbook with time, advertising their choices to correspondents.

We think that prospects to apply knowledge of many people to problem solving, such as diagnostics, legal cases, management consulting etc., are attractive to all.

Eunika Mercier-Laurent states in a paper *The Future of AI or AI for the Future* “The third hype of AI and enthusiasm for applying last techniques in all fields raise great interest and some important questions on the future directions in AI research and applications. Guiding by the principle of combining the best from human and computers capacities this chapter lists some important challenges to face and related directions in AI research” [13,14].

Acknowledgments: The author is deeply grateful to the Chair of IFIP TC12 Artificial Intelligence Technical Committee Professor Eunika Mercier-Laurent for extremely valuable support and very encouraging attitude.

Data availability: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Conflict of interest: The author declares no conflict of interest.

References

1. Alty JL, Coombs MJ. *Expert systems. Concepts and Examples.* The National Computing Centre Limited; 1984.
2. Hinton GE. *Learning in parallel networks.* McGraw-Hill; 1985.
3. Schank R, Hunter L. *The quest to understand thinking.* By McGraw-Hill; 1985.
4. Golubev KM. *Is there any future for Artificial Intelligence?* ISPIM; 2001.
5. Doyle AC. *The Penguin Complete Sherlock Holmes.* Penguin Books; 1981.
6. OpenAI. *GPT-4 System Card.* Available online: <https://cdn.openai.com/papers/gpt-4-system-card.pdf> (accessed on 12 June 2024).
7. Weise K, Metz C. *When AI Chatbots Hallucinate.* Available online: <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html> (accessed on 12 June 2024).
8. Ordonez V, Dunn T, Noll E. *OpenAI CEO Sam Altman says AI will reshape society, acknowledges risks: A little bit scared of this.* Available online: <https://www.yahoo.com/gma/openai-ceo-sam-altman-says-215500989.html> (accessed on 12 June 2024).
9. Venturi L. *Renaissance painting from Breughel to El Greco.* Skira Rizzoli; 1979.
10. Golubev KM. *Adaptive learning with e-knowledge systems.* Inderscience Enterprises Limited; 2003.
11. Livanov MN. *Space organization of brain's processes.* USSR; 1972.
12. Golubev KM. *Intellect Modeling Kit.* In: Mercier-Laurent E (editor). *Artificial Intelligence for Knowledge Management.* Springer; 2020.

13. Mercier-Laurent E. The Future of AI or AI for the Future. In: Strous L, Johnson R, Grier DA, Swade D (editors). *Unimagined Futures—ICT Opportunities and Challenges*. Springer; 2020.
14. Mercier-Laurent E. Platform for Knowledge Society and Innovation Ecosystems. In: Mercier-Laurent E (editor). *Artificial Intelligence for Knowledge Management*. Springer; 2020.

Article

Is it possible to detect cribriform adverse pathology in prostate cancer with magnetic resonance imaging machine learning-based radiomics?

Hüseyin Bıçakçioğlu¹, Sedat Soyupek¹, Onur Ertunç², Avni Görkem Özkan³, Şehnaz Evrimler³, Tekin Ahmet Serel^{1,*}

¹ Department of Urology, School Medicine, Süleyman Demirel University, 32260 Isparta, Turkey

² Department of Pathology, School Medicine, Süleyman Demirel University, 32260 Isparta, Turkey

³ Department of Radiology, School Medicine, Süleyman Demirel University, 32260 Isparta, Turkey

* Corresponding author: Tekin Ahmet Serel, tekinsel@sdu.edu.tr

CITATION

Bıçakçioğlu H, Soyupek S, Ertunç Q, et al. Is it possible to detect cribriform adverse pathology in prostate cancer with magnetic resonance imaging machine learning-based radiomics?. *Computing and Artificial Intelligence*. 2024; 2(1): 1257.
<https://doi.org/10.59400/cai.v2i1.1257>

ARTICLE INFO

Received: 1 April 2024

Accepted: 6 June 2024

Available online: 24 June 2024

COPYRIGHT



Copyright © 2024 by author(s).

Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.

<https://creativecommons.org/licenses/by/4.0/>

Abstract: Rationale and objectives: Cribriform patterns are accepted as aggressive variants of prostate cancer. These adverse pathologies are closely associated with early biochemical recurrence, metastasis, castration resistance, and poor disease-related survival. A few publications exist to diagnose these two adverse pathologies with multiparametric magnetic resonance imaging (mpMRI). Most of these publications are retrospective and are not studies that have made a difference in diagnosing adverse pathology. It is also known that fusion biopsies taken from lesions detected in mpMRI are insufficient to detect these adverse pathologies. Our study aims to diagnose this adverse pathology using machine learning-based radiomics data from MR images. **Materials and methods:** A total of 88 patients who had pathology results indicating the presence of cribriform pattern and prostate adenocarcinoma underwent preoperative MRI examinations and radical prostatectomy. Manual slice-by-slice 3D volumetric segmentation was performed on all axial images. Data processing and machine learning analysis were conducted using Python 3.9.12 (Jupyter Notebook, Pycaret Library). **Results:** Two radiologists, SE and MAG, with 7 and 8 years of post-graduate experience, respectively, evaluated the images using the 3D-Slicer software without knowledge of the histopathological findings. One hundred seventeen radiomic tissue features were extracted from T1 weighted (T1W) and apparent diffusion coefficient (ADC) sequences for each patient. The interobserver agreement for these features was analyzed using the intraclass correlation coefficient (ICC). Features with excellent interobserver agreement ($ICC > 0.90$) were further analyzed for collinearity between predictors using Pearson's correlation. Variables showing a very high correlation ($r \geq \pm 0.80$) were disregarded. The selected features for T1W and ADC images were First-order maximum, First-order skewness, First-order 10th percentile for ADC, and Gray level size zone matrix, Large area low gray level emphasis for T1W. As a result of the classification of PyCaret, the three best models were found. A single model was obtained by blending these three models. AUC, accuracy, recall, precision, and F1 scores were 0.79, 0.77, 0.85, 0.82, and 0.83, respectively. **Conclusion:** ML-based MRI radiomics of prostate cancer can predict the cribriform pattern. This prognostic factor cannot be determined through qualitative radiological evaluation and may be overlooked in preoperative histopathological specimens.

Keywords: prostate cancer; cribriform pathology; MRI; radiomics; machine learning

1. Introduction

Cribriform pattern, among the adverse pathological features of prostate cancer, leads to a higher incidence of extraprostatic extension, lymph node involvement, metastasis, and biochemical recurrence compared to prostate adenocarcinoma. Due to

these characteristics, the mortality rates associated with prostate cancer are higher in patients exhibiting these patterns. These lesions are a diverse group of prostate pathologies that can be challenging to diagnose. These lesions encompass a spectrum of conditions, including normal anatomical variations, benign proliferative lesions, premalignant conditions, suspicious findings, clearly malignant tumors, and aggressive entities. In recent years, our understanding of cribriform prostate adenocarcinoma (CrP4) and intraductal carcinoma of the prostate (IDC-P) has significantly advanced. There is now a growing body of evidence suggesting that the presence of these morphologies plays an important role in clinical decision-making for managing prostate cancer. Therefore, it is crucial to accurately recognize and report the architectural features of CrP4 and IDC-P [1–8].

Prostate multiparametric magnetic resonance imaging (mpMRI) is a valuable technique for detecting tumor lesions in the prostate. However, despite its usefulness, there are still cases where clinically significant lesions are undetected [9]. It has been concluded that mpMRI may accurately identify prostate cancer (PCa) tumors located in the peripheral zone (PZ). Additionally, the mean apparent diffusion coefficient (ADC) value and ADC ratio can serve as predictors for the presence of the cribriform pattern in PCa [10].

In recent years, there have been significant advancements in imaging technology and analysis methods, leading to the emergence of a framework called radiomics, which is a method of texture analysis that extracts imaging features from digital images by converting them into mineable, quantifiable data, revealing the heterogeneity of tumors. This framework involves extracting comprehensive high-dimensional features from imaging data and utilizing data mining techniques to develop analytical models that enhance decision support. Radiomics encompasses various features, including texture and shape, thereby providing rich information for precision medicine [11–13].

Numerous studies focusing on prostate radiomics have demonstrated promising outcomes in evaluating pathological characteristics, predicting treatment response, and stratifying risk groups. Radiomics features show promise as markers for assessing the aggressiveness of prostate cancer at both the histopathological and genomic levels. By correlating radiomics features with histopathological findings and genomic data, researchers are gaining insights into the aggressiveness and potential outcomes of prostate cancer. This has the potential to improve risk stratification, treatment selection, and personalized management for patients with prostate cancer [14,15].

The role of artificial intelligence in prostate cancer has been investigated because artificial intelligence (AI) may have the potential to revolutionize pathologic diagnosis and cancer patient management by serving as a predictive and prognostic biomarker. AI-based systems can analyze digitally scanned histopathology slides, enabling the differentiation between benign and malignant cells and low-grade and high-grade tumors. Deep learning models can process patient data from various sources, including individual or multimodal combinations, to identify patterns that can predict responses to different therapeutic options, assess the risk of recurrence or progression, and determine the prognosis for newly diagnosed patients. By leveraging AI-based models, treatment planning for prostate cancer patients can be significantly improved, and the efficiency and cost-effectiveness of pathology laboratories can be enhanced

[16–20]. However, to the best of our knowledge, there have been no studies in the English literature evaluating the predictive performance of ML-based MRI radiomics and the cribriform pattern for prostate cancer using multi-classifier models. Our study aims to assess the predictive performance of machine learning (ML)-based prostate cancer radiomics using Magnetic Resonance Imaging (MRI) to detect the cribriform pattern.

Patients

This retrospective study was approved by the institutional review board on 13 September 2022, with approval number 20/280. The data were collected from patients with prostatic carcinoma who underwent radical prostatectomy between January 2018 and November 2022. Only patients who underwent radical prostatectomy were included because the cribriform pattern can be present focally and may be missed in prostate biopsy specimens. A total of 88 patients who underwent preoperative MRI examination were identified. All patients had PIRADS 3 or above lesions. The cohort was divided into a training set ($n = 70$ subjects) and a test set ($n = 18$ subjects).

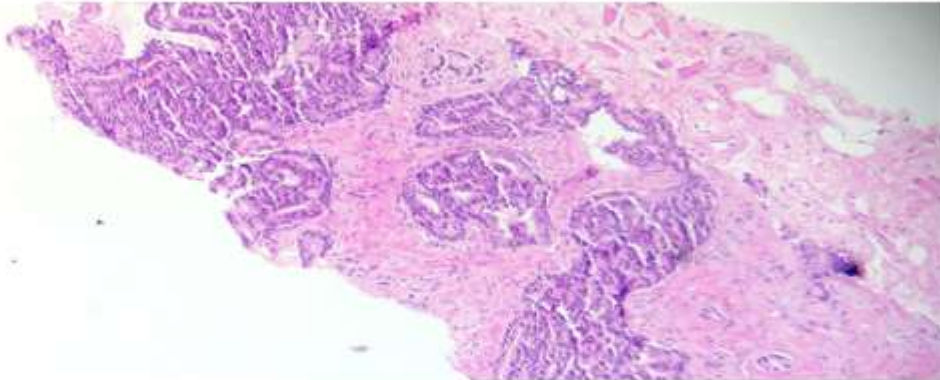
2. Materials and methods

2.1. Imaging technique and radiomics analysis

All patients underwent standardized multiparametric MRI, which included multiplanar T1-weighted imaging (T1WI), axial diffusion-weighted imaging (DWI), and dynamic contrast-enhanced MRI (DCE-MRI) sequences at 3 Tesla using a phased array body coil. An endorectal coil was not used, and the imaging protocol adhered to the PI-RADS v2.1 standards. The MRI scans were performed on a MAGNETOM Avanto system from Siemens Medical Solutions in Erlangen, Germany. Apparent diffusion coefficient (ADC) maps were generated based on the DW images with all acquired b-values, and a calculated $b = 1600$ s/mm² image was produced for PI-RADS classification. DCE-MRI was performed using intravenous gadobutrol (Gadavist, Bayer) contrast agent at a dose of 0.1 mL/kg body weight (0.1 mmol/kg), infused at a rate of 2 mL/s, with imaging initiated simultaneously with the administration of contrast.

Two radiologists, SE and MAG, with 7 and 8 years of post-graduate experience, evaluated the images using the 3D-Slicer software (version 4.11) [21]. They were blinded to the histopathological findings. DICOM images of the axial T1-weighted acquisition were obtained from the PACS system. The texture features, including first-order and second-order features, were extracted using Slicer-Radiomics (PyRadiomics v3.0.1) software (**Figure 1**). Before radiomics processing, the T1-weighted signals within the lesion volumes of interest (VOIs) were normalized to a range between the mean value plus three times the standard deviation of the signal intensity within the VOI, as previously described [22,23]. The spatial resampling voxel size was set to $1 \times 1 \times 1$ mm, and gray-level discretization was performed using a fixed bin width of 64 gray levels.

The Pure type of Cribriform Pattern of Prostate Cancer in H&E Section (200 X magnification)



Tumour Detection-Segmentation-Image preprocessing-Feature extraction

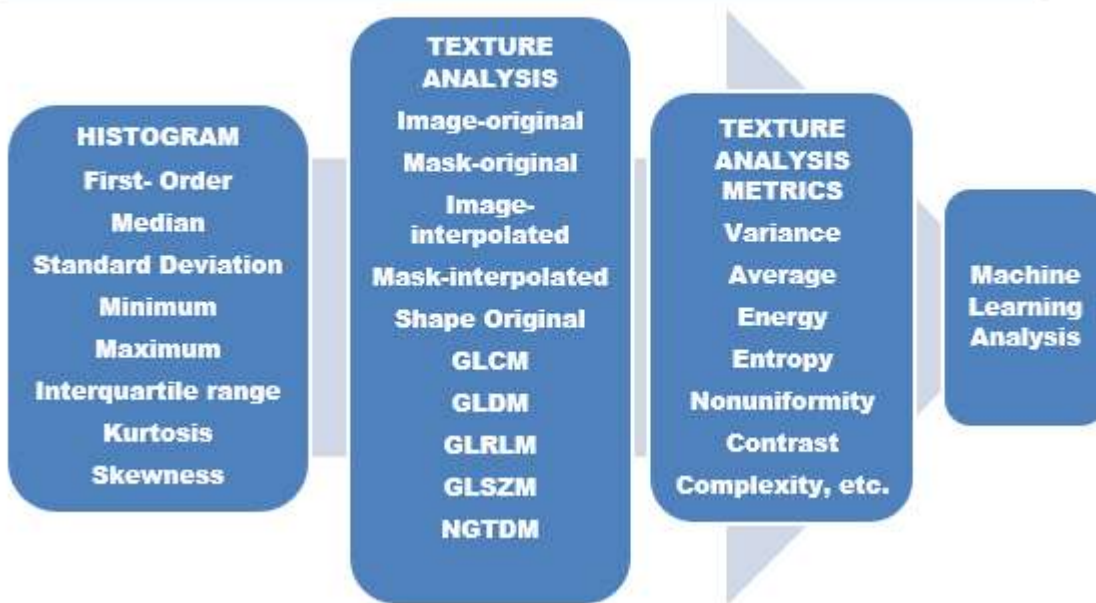
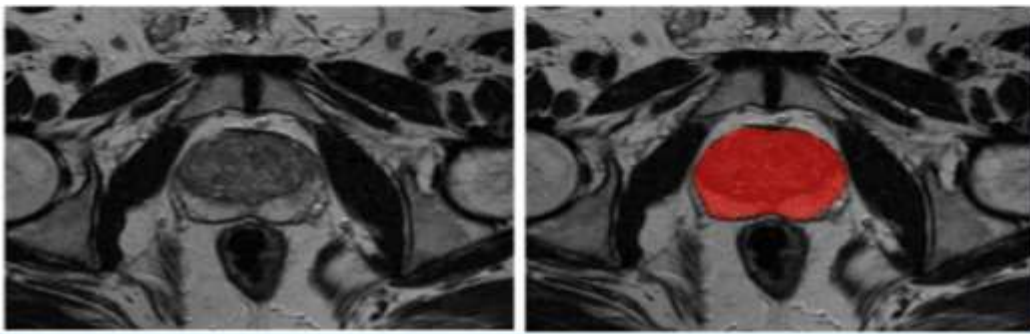


Figure 1. MRI-based radiomics workflow and the Hematoxylin Eosin (H&E) section are examples of cribriform patterns in prostate cancer.

GLCM, Gray-level co-occurrence matrix; GLDM, Gray-level difference method; GLRLM, Gray-level run-length matrix; GLSZM, Graylevel size zone; NGTDM, Neighbouring Gray Tone Difference Matrix.

2.2. Statistical evaluation

Statistical analysis was conducted using IBM SPSS Statistics software (version 24.0, IBM Corp., Armonk, NY). Descriptive statistics were reported as frequencies for categorical variables and mean ± standard deviation or median for numerical

variables. The normality of continuous variables was assessed using the Kolmogorov-Smirnov test. A p -value less than 0.05 was considered statistically significant.

2.3. Dimension reduction

The interobserver agreement for these features was assessed using the intraclass correlation coefficient (ICC). Features that demonstrated excellent interobserver agreement ($ICC > 0.90$) were further examined for collinearity between predictors using Pearson's correlation. Variables that exhibited a very high correlation ($r \geq \pm 0.80$) were excluded from the analysis. The selected features for T1-weighted (T1W) and apparent diffusion coefficient (ADC) images were first-order Maximum, first-order Skewness, and first-order 10th Percentile for ADC, as well as Gray level zone matrix, large area low gray level emphasis for T1W.

2.4. Data handling and machine learning analysis

Data handling and machine learning analysis were conducted using Python 2.3 in Jupyter Notebook with the Pycaret library. [24]. The "classic method" random forest classifier was employed to select and reduce features [25]. The interobserver agreement for these features was assessed using the intraclass correlation coefficient (ICC). Features demonstrating excellent interobserver agreement ($ICC > 0.90$) were further examined for collinearity between predictors using Pearson's correlation. The tune model function in Pycaret classification was utilized to identify the optimal hyperparameters. This function tunes the model's hyperparameters and generates a scoring grid with cross-validated scores by fold. The best model is selected based on the defined metric in the optimized parameter. The get_metrics function allows access to the metrics evaluated during cross-validation, and custom metrics can be added or removed using the add_metric and remove_metric functions .

The data was split into a training set (70%) and a test set (30%). Z-score scaling was applied for normalization, and 5-fold cross-validation was performed to avoid overfitting due to limited dataset size. The synthetic minority oversampling technique was used to address the issue of imbalanced datasets.

Fourteen machine learning algorithms were employed to determine significant models. The predictive performance of these algorithms was compared using metrics such as AUC, accuracy, recall, precision, and F1 scores. The top three models based on accuracy and AUC were evaluated on the test set. To optimize results, these three models were combined into a single ensemble model, leveraging the strengths of each model to improve overall performance and predictive accuracy. A receiver operating characteristic (ROC) and learning curves were plotted. AUC, accuracy, recall, precision, and F1 scores were provided along with the confusion matrix.

3. Results

The patients had a mean age of 66.05. Among them, there were 49 cases (55%) of pure prostatic adenocarcinoma and 39 cases (45%) of cribriform pattern prostatic adenocarcinoma, all of which were tumors. A total of 117 texture features were extracted. The 14 machine learning algorithms demonstrated varying AUC and accuracy values on the train set, ranging from 0.50% to 0.76% and 31% to 72%,

respectively. The top three models identified were the AdaBoost classifier, with an AUC of 0.76 and an accuracy of 72%; the random forest classifier, with an AUC of 0.80 and an accuracy of 71%; and the extra trees classifier, with an AUC of 0.78 and an accuracy of 71% (Table 1). The voting classifier, which combines these models, yielded the following scores: AUC of 0.77, accuracy of 0.79, recall of 0.85, precision of 0.82, and F1 score of 0.83 (Table 2). The confusion matrix and classification report, along with the receiver operating curve (ROC), illustrating the predictive performance of the blended model, are presented in Figure 2. Figure 3 shows the decision boundary and hyperparameters of the blended model.

Table 1. The predictive performances of machine learning based MRI radiomics models for cribriform pattern of prostate cancer.

Model	Accuracy	AUC	Recall	Prec.	F1
Random Forest Classifier	0.7143	0.8061	0.8155	0.7906	0.7978
Extra Trees Classifier	0.7145	0.7864	0.8156	0.7916	0.7964
Gradient Boosting Classifier	0.6429	0.7609	0.6889	0.7912	0.7106
Logistic Regression	0.6714	0.7639	0.7111	0.8076	0.7591
Ada Boost Classifier	0.7286	0.7622	0.7033	0.8178	0.7699
Light Gradient Boosting Machine	0.7143	0.7417	0.7489	0.8256	0.7837
K Neighbors Classifier	0.6714	0.8075	0.6887	0.8363	0.7291
Linear Discriminant Analysis	0.7000	0.6522	0.7733	0.8061	0.7382
Naive Bayes	0.6514	0.6428	0.6467	0.8181	0.6143
Decision Tree Classifier	0.6714	0.6006	0.7911	0.7386	0.7615
Quadratic Discriminant Analysis	0.5857	0.5872	0.6057	0.7086	0.6547
Dummy Classifier	0.5143	0.5000	0.0000	0.0000	0.0000
SVM-Linear Kernel	0.5000	0.0000	0.6179	0.4945	0.5258
Ridge Classifier	0.7000	0.0000	0.7733	0.8051	0.7832

Table 2. The predictive performances of blended machine learning models (random forest, extra trees, Ada boost) for cribriform pattern of prostate cancer.

Fold	Accuracy	AUC	Recall	Prec.	F1
0	0.7857	0.7760	0.9000	0.8182	0.8571
1	0.8571	0.9000	0.8000	1.0000	0.8889
2	0.7143	0.7200	0.8000	0.8000	0.8000
3	0.6429	0.6887	0.7778	0.0000	0.7368
4	0.8571	0.9556	1.0000	0.8182	0.9000
Mean	0.7714	0.7994	0.8556	0.6273	0.8056
Std	0.0833	0.1119	0.0837	0.5089	0.0608

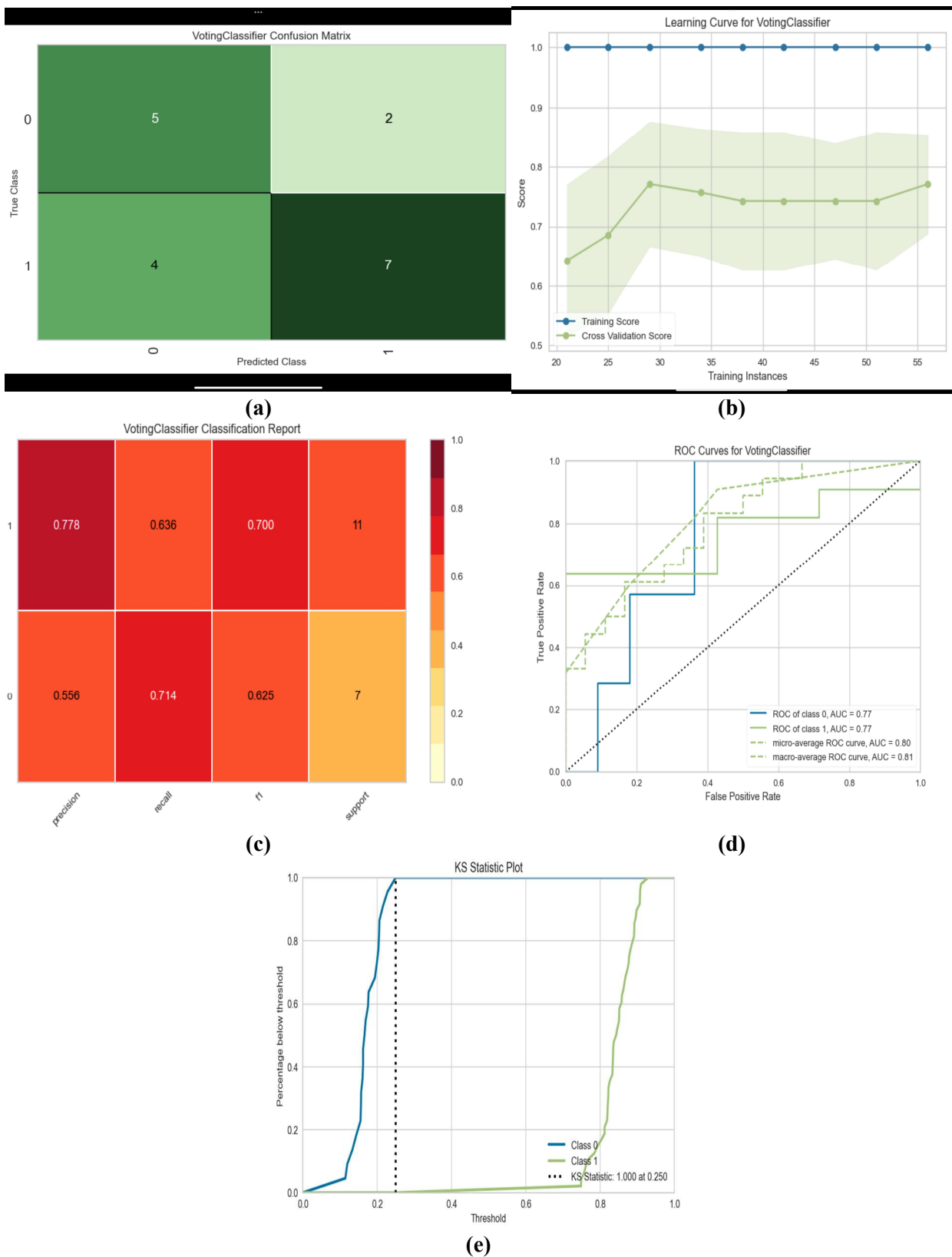


Figure 2. Plot graphics of blended three machine learning classifiers: **(a)** confusion matrix; **(b)** learning curve; **(c)** classification report; **(d)** receiver operating characteristic curve; **(e)** Kolmogorov Smirnov statistic plot.

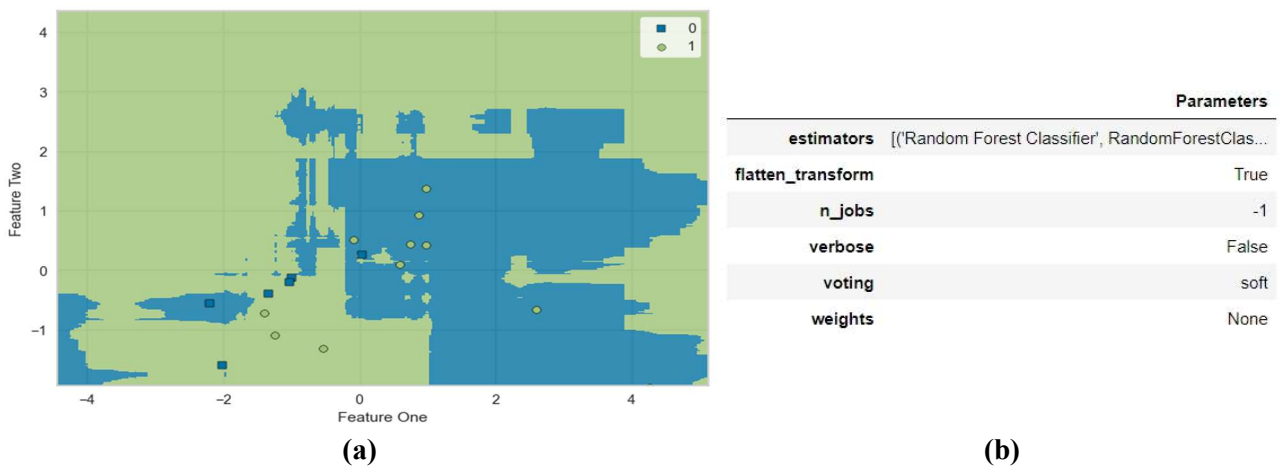


Figure 3. Plot graphic and hyperparameters of blended three machine learning classifiers: **(a)** decision boundary; **(b)** hyperparameters of blended models with Pycaret classification.

4. Discussion

This study assessed the predictive performance of machine learning-based MR radiomics for prostatic adenocarcinoma, focusing on the cribriform pattern. The machine learning algorithms exhibited high rates of AUC and accuracy. The voting classifier model, in particular, demonstrated excellent accuracy in predicting the cribriform status.

It is worth noting that the cribriform pattern has been associated with a poorer prognosis, characterized by lower survival rates and a higher prevalence of advanced stage, positive surgical margin, lymph node metastasis, lymphovascular invasion, and recurrence. Therefore, it has been suggested that the cribriform pattern could be utilized for risk stratification and guiding therapy management [4,26–29].

Early diagnosis in patients with prostate cancer having cribriform pathology is of critical importance in selecting the most appropriate patient management. Detecting the cribriform pattern in prostate biopsies beforehand is challenging. Thus, risk stratification to better identify patients at risk is necessary to identify low-grade prostate cancer that needs active surveillance. This difficulty has been attempted to be overcome through imaging techniques.

In a study using PSMA PET/CT, the incidence of the cribriform pattern in patients with prostate cancer was reported to be 39% [30]. Truong et al. published their study, which compares the results of fusion biopsies using MRI combined ultrasound fusion with the specimens obtained after radical prostatectomy; they demonstrated that only 17.3% of the cribriform pattern could be detected using MRI [31]. These studies indicate that even with targeted biopsies using MRI, the sensitivity for detecting patients with cribriform architecture is not high. Therefore, there is a need for advanced imaging techniques that can detect the lesion with higher accuracy. To our knowledge, this is the first report that has investigated the potential role and the predictive performance of machine learning-based MRI radiomics for detecting the cribriform pattern in prostate cancer in the existing English literature.

Our study exposed that the best three models for predicting the cribriform pattern in prostate cancer were the AdaBoost, random forest, and extra trees classifiers. These

models demonstrated high AUC and accuracy rates, indicating their predictive solid performance. It is important to note that increasing the number of training samples likely improved the generalization of these models. Therefore, we recommend conducting further studies with a larger patient population to enhance and validate our results.

This study has several limitations that should be acknowledged. Firstly, it is essential to note that the study design was retrospective, which inherently carries the risk of data loss and potential biases. Secondly, the patient population included in the study was relatively small. We specifically focused on the histopathological findings of radical prostatectomy specimens to ensure the presence of the cribriform pattern, as tumor heterogeneity could lead to its absence in prostate biopsy specimens. However, a larger patient population would be beneficial for further validation and generalization of the results. To address these limitations, we recommend conducting a prospective, multi-center study that evaluates the predictive value of machine learning models for the cribriform pattern and its impact on prognosis. Such a study would provide more robust and reliable evidence for this purpose.

5. Conclusion

Qualitative radiological examinations alone are insufficient for accurately determining the presence of the cribriform pattern in prostate cancer lesions. Additionally, this pattern can be missed on preoperative histopathological specimens due to the inherent heterogeneity of the lesion. Therefore, there is a clear need for an objective method to identify and predict the cribriform pattern reliably. In our study, we propose that machine learning-based MRI radiomics of prostate cancer can effectively predict the presence of the cribriform pattern. By utilizing advanced algorithms and analyzing a wide range of imaging features, this approach has the potential to provide valuable guidance to clinicians in making informed decisions regarding therapy management. It offers an objective and quantifiable method to assist in accurately identifying and characterizing the cribriform pattern, ultimately improving patient care and outcomes.

Author contributions: Conceptualization, TAS and SS; methodology, ŞE; software, TAS; validation, TAS, HB and ŞE; formal analysis, HB; investigation, SS; resources, TAS and OE; data curation, TAS; writing—original draft preparation, OE; writing—review and editing, TAS; visualization, TAS; supervision, OE and AGÖ; project administration, HB. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Hesterberg AB, Gordetsky JB, Hurley PJ. Cribriform Prostate Cancer: Clinical Pathologic and Molecular Considerations. *Urology*. 2021; 155: 47-54. doi: 10.1016/j.urology.2021.05.028
2. Kweldam CF, van der Kwast T, van Leenders GJ. On cribriform prostate cancer. *Translational Andrology and Urology*. 2018; 7(1): 145-154. doi: 10.21037/tau.2017.12.33
3. Cai Q, Shah RB. Cribriform Lesions of the Prostate Gland. *Surgical Pathology Clinics*. 2022; 15(4): 591-608.

4. Destouni M, Lazaris AC, Tzelepi V. Cribriform Patterned Lesions in the Prostate Gland with Emphasis on Differential Diagnosis and Clinical Significance. *Cancers*. 2022; 14(13): 3041. doi: 10.3390/cancers14133041
5. Montironi R, Cimadamore A, Gasparini S, et al. Prostate cancer with cribriform morphology: diagnosis, aggressiveness, molecular pathology and possible relationships with intraductal carcinoma. *Expert Review of Anticancer Therapy*. 2018; 18(7): 685-693. doi: 10.1080/14737140.2018.1469406
6. Tuna MB, Arslan A, K ok YB, et al. Cribriform Pattern of The Prostate Adenocarcinoma: Sensitivity of Multiparametric MRI. *Urol J*. 2022; 20(1): 34-40.
7. Kazan O, Gunduz N, Kir G, et al. The cribriform morphology impairs Gleason 7 prostate cancer lesion detection on multiparametric magnetic resonance imaging. *The Prostate*. 2022; 83(4): 331-339.
8. Gaudio C, Bianchi L, De Cinque A, et al. The impact of multiparametric MRI features to identify the presence of prevalent cribriform pattern in the peripheral zone tumors. *La radiologia medica*. 2021; 127(2): 174-182.
9. Cho HH, Kim CK, Park H. Overview of radiomics in prostate imaging and future directions. *The British Journal of Radiology*. 2021; 95(1131).
10. Hectors SJ, Cherny M, Yadav KK, et al. Radiomics Features Measured with Multiparametric Magnetic Resonance Imaging Predict Prostate Cancer Aggressiveness. *The Journal of Urology*. 2019; 202(3): 498-505.
11. Smith CP, Czarniecki M, Mehravivand S, et al. Radiomics and radiogenomics of prostate cancer. *Abdominal Radiology*. 2018; 44(6): 2021-2029. doi: 10.1007/s00261-018-1660-7
12. Sun Y, Reynolds HM, Parameswaran B, et al. Multiparametric MRI and radiomics in prostate cancer: a review. *Australasian Physical & Engineering Sciences in Medicine*. 2019; 42(1): 3-25. doi: 10.1007/s13246-019-00730-z
13. Midiri F, Vernuccio F, Purpura P, et al. Multiparametric MRI and Radiomics in Prostate Cancer: A Review of the Current Literature. *Diagnostics*. 2021; 11(10): 1829. doi: 10.3390/diagnostics11101829
14. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: A Cancer Journal for Clinicians*. 2019; 69(2): 127-157.
15. Liberini V, Laudicella R, Balma M, et al. Radiomics and artificial intelligence in prostate cancer: new tools for molecular hybrid imaging and theragnostics. *European Radiology Experimental*. 2022; 6(1). doi: 10.1186/s41747-022-00282-0
16. Azadi Moghadam P, Bashashati A, Goldenberg SL. Artificial Intelligence and Pathomics. *Urologic Clinics of North America*. 2024; 51(1): 15-26. doi: 10.1016/j.ucl.2023.06.001
17. Chaddad A, Tan G, Liang X, et al. Advancements in MRI-Based Radiomics and Artificial Intelligence for Prostate Cancer: A Comprehensive Review and Future Prospects. *Cancers*. 2023; 15(15): 3839. doi: 10.3390/cancers15153839
18. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic Resonance Imaging*. 2004; 22(1): 81-91. doi: 10.1016/j.mri.2003.09.001
19. Patro S, Sahu KK. Normalization: a preprocessing stage. *arXiv*. 2015.
20. Amini N, Mahdavi M, Choubdar H, et al. Automated prediction of COVID-19 mortality outcome using clinical and laboratory data based on hierarchical feature selection and random forest classifier. *Computer Methods in Biomechanics and Biomedical Engineering*. 2022; 26(2): 160-173.
21. Available online: <https://www.slicer.org/> (accessed on 13 January 2024).
22. Duda RO, Hart PE, Stork DG. *Pattern classification*. New York: John Wiley and Sons; 2001.
23. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006 Feb 23;7:91.
24. Available online: <https://pycaret.org/> (accessed on 13 January 2024).
25. Narasimhulu, C Venkata. An automatic feature selection and classification framework for analyzing ultrasound kidney images using dragonfly algorithm and random forest classifier. *IET Image Processing*. 2021;15. 10.1049/ipr2.12179.
26. Rubin MA, de La Taille A, Bagiella E, et al. Cribriform Carcinoma of the Prostate and Cribriform Prostatic Intraepithelial Neoplasia. *The American Journal of Surgical Pathology*. 1998; 22(7): 840-848.
27. Miyai K, Divatia MK, Shen SS, et al. Clinicopathological analysis of intraductal proliferative lesions of prostate: intraductal carcinoma of prostate, high-grade prostatic intraepithelial neoplasia, and atypical cribriform lesion. *Human Pathology*. 2014; 45(8): 1572-1581.
28. Shah RB, Zhou M. Atypical Cribriform Lesions of the Prostate. *Advances in Anatomic Pathology*. 2012; 19(4): 270-278.
29. Russo GI, Soeterik T, Puche-Sanz I, et al. Oncological outcomes of cribriform histology pattern in prostate cancer patients: a systematic review and meta-analysis. *Prostate Cancer and Prostatic Diseases*. 2022; 26(4): 646-654.

30. The Impact of Histological Variants on the Performance Characteristics of 68Ga-PSMA PET/CT. Available online: <https://www.urotoday.com/conference-highlights/2022-annual-meeting/aua-2022-prostate-cancer/137349-aua-2022-the-impact-of-histological-variants-on-the-performance-characteristics-of-68ga-psma-pet-ct-in-the-primary-and-recurrent-setting.html> (accessed on 1 January 2024).
31. Truong M, Feng C, Hollenberg G, et al. A Comprehensive Analysis of Cribriform Morphology on Magnetic Resonance Imaging/Ultrasound Fusion Biopsy Correlated with Radical Prostatectomy Specimens. *Journal of Urology*. 2018; 199(1): 106-113.

From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI

Zarif Bin Akhtar

Department of Engineering, University of Cambridge, Cambridge CB2 1TN, England, United Kingdom; zarifbinakhtarg@gmail.com

CITATION

Akhtar ZB. From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI. *Computing and Artificial Intelligence*. 2024; 2(1): 1378. <https://doi.org/10.59400/cai.v2i1.1378>

ARTICLE INFO

Received: 11 May 2024
Accepted: 19 June 2024
Available online: 27 June 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: The advent of artificial intelligence (AI) has significantly transformed various aspects of human life, particularly in information retrieval and assistance. This research presents a comprehensive evaluation of Gemini, previously known as Google Bard, a state-of-the-art AI chatbot developed by Google. Through a meticulous methodology encompassing both qualitative and quantitative approaches, this research aims to assess Gemini's performance, usability, integration capabilities, ethical implications. Primary data collection methods, including user surveys and interviews, were utilized to gather towards the qualitative feedback on user experiences with Gemini, supplemented by secondary data analysis using tools such as Google Analytics to capture quantitative metrics. Performance evaluation involved benchmarking against other AI chatbots and technical analysis of Gemini's architecture and training methods. User experience testing examined usability, engagement, and integration with Google Workspace and third-party services. Ethical considerations regarding data privacy, security, and biases in AI-generated content were also addressed, ensuring compliance with major regulations and promoting ethical AI practices. Acknowledging limitations and challenges inherent in the investigative exploration, data analysis was conducted using thematic and statistical methods to derive insights. The results and findings of this research offer valuable insights into the capabilities and limitations of Gemini, providing implications for future AI development, user interaction design, and ethical AI governance. By contributing to the ongoing discourse on AI advancements and their societal impact, this exploration facilitates informed decision-making and lays the groundwork for future research endeavors in the field of AI-driven conversational agents.

Keywords: artificial intelligence (AI); bard; ChatGPT; computer vision; deep learning (DL); Gemini; generative artificial intelligence (GAI); large language models (LLM); machine intelligence; machine learning (ML)

1. Introduction

In recent years, artificial intelligence (AI) has ushered in a paradigm shift in human interaction and problem-solving, notably through the proliferation of AI chatbots. These conversational agents, fueled by advanced language models, have become indispensable aids in a wide array of tasks, from answering inquiries to generating text and offering recommendations. Google's creation, initially named Bard and later rebranded as Gemini, stands as a prominent figure in this domain [1]. This research exploration embarks on an investigation into the evolutionary journey of AI chatbots, with a specific focus on Google's transformation from Bard to Gemini [2–6]. The transition from Bard to Gemini symbolizes a significant leap in AI technology, characterized by enhancements in functionality, user experience, and integration with various Google services. The inquiry aims to scrutinize the capabilities of Gemini, assessing its aptitude in furnishing users with precise and

pertinent information across diverse topics [7–9]. Additionally, it seeks to dissect the technological advancements and foundational AI models powering Gemini, juxtaposing them with other notable AI chatbots in the market. Moreover, the research endeavors to dissect the impact of Gemini on user interactions and information retrieval, contemplating its potential to supplant traditional search engines in select contexts [10–15]. Through the examination of user feedback, performance metrics, and real-world applications, the research endeavors to uncover insights into the efficacy and constraints of Gemini as an AI chatbot. This investigative exploration seeks to enrich the ongoing dialogue encompassing AI chatbots and their pivotal role in shaping the trajectory of human-computer interaction. By conducting a thorough analysis of Gemini’s evolution and capabilities, it aspires to illuminate the potential of AI chatbots in amplifying productivity, streamlining information retrieval, and revolutionizing user experiences across diverse domains.

2. Methods and experimental analysis

The methodology employed in this research encompasses a multifaceted approach aimed at thoroughly evaluating the performance, usability, and integration capabilities of Gemini, previously known as Google Bard. Data collection strategies were devised to incorporate both primary and secondary methods. Primary data were procured through user surveys and interviews, focusing on eliciting qualitative feedback regarding user experiences with Gemini. This qualitative analysis delved into aspects such as user satisfaction, perceived response accuracy, ease of use, and specific use cases. Secondary data were sourced from analytics tools like Google Analytics to collect quantitative metrics such as session duration, interaction frequency, popular queries, and geographical distribution of users. Performance evaluation of Gemini entailed standardized benchmarks and comparative analysis with other AI chatbots like ChatGPT and Microsoft Copilot. This evaluation assessed Gemini’s response accuracy, response time, and adaptability in handling diverse queries. Additionally, a technical analysis scrutinized the underlying architecture of Gemini, encompassing neural network models, algorithms for Natural Language Processing (NLP), and image generation techniques. This analysis also examined the training data and pre-processing methods utilized to refine Gemini’s language generation and image synthesis capabilities. User experience testing involved designing scenarios to evaluate Gemini’s usability and engagement, including tasks such as text or image generation requests, interactions with Google Workspace integrations, and exploration of features like Google Lens. User feedback was captured through observation, task completion rates, and qualitative assessments of satisfaction and usefulness.

The integration capabilities of Gemini with Google Workspace tools and third-party services were analyzed to assess seamless integration and the effectiveness of extensions. Ethical considerations regarding data privacy, security, and potential biases in AI-generated content were integrated throughout the research process. Compliance with regulations such as GDPR and CCPA was ensured in handling user data, with strategies devised to mitigate biases and promote ethical AI practices.

Acknowledgment of limitations and challenges inherent in the methodology, such as biases in user feedback and constraints in accessing proprietary data, were

addressed. Data analysis encompassed thematic analysis of qualitative data and statistical analysis of quantitative data to generate insights and conclusions. Findings and analysis were presented comprehensively in a research report, utilizing visualizations such as charts, graphs, and diagrams to illustrate key findings and trends. Summarizing the methodology, results, and implications of the research, preparations were made for dissemination to academic audiences, industry professionals, and AI practitioners.

3. Background research and iterative exploration for available knowledge

Gemini, formerly known as Bard, emerged as Google's response to the popularity of OpenAI's ChatGPT. Google executives, alarmed by ChatGPT's potential threat to Google Search, mobilized their workforce and launched Bard in February 2023, initially powered by the LaMDA family of large language models (LLMs) [10,15]. The rapid development and rollout of Bard reflected Google's efforts to compete with ChatGPT, with the goal of integrating generative AI chatbots into Google's products and services. The launch of Bard was met with mixed responses. Despite the initial excitement, Bard faced criticism for its inaccuracies, rushed announcement, and perceived lack of readiness. Google executives pushed ahead with the launch despite internal concerns, emphasizing the need to keep up with competitors. Bard's rollout was marked by challenges, including technical errors during demonstrations and ethical concerns raised by Google employees [16].

In response to feedback and competition, Google announced updates and enhancements to Bard, including the integration of PaLM and Gemini, expansion to new countries and languages, and improvements in personalization and productivity features [17–22]. Despite these efforts, Bard continued to face scrutiny for its cautious responses and perceived limitations compared to ChatGPT [23–33]. Gemini, introduced in December 2023, represented a significant advancement in Google's AI capabilities. With Gemini, Google aimed to further enhance Bard's performance and expand its use across various platforms and devices. The integration of Bard and Duet AI under the Gemini brand signaled Google's commitment to advancing its generative AI technology [34–44]. The development and rollout of Bard and its subsequent evolution into Gemini reflect Google's efforts to innovate in the field of artificial intelligence and compete with other tech giants, particularly in the realm of conversational AI and virtual assistants. Despite facing challenges and criticism, Google continues to invest in and refine its AI technologies to meet the evolving needs of users and maintain its position in the competitive tech landscape [45–52]. Google has rebranded its artificial intelligence (AI) chatbot Bard as Gemini Advanced, reflecting the significant advancements in AI technology and capabilities. Gemini Advanced represents a major step forward in AI development, positioning itself as a direct competitor to OpenAI's GPT-4 [53–55]. Gemini Advanced is built on the PaLM 2 architecture and boasts improvements in efficiency, multimodal capabilities, and future-proofing for memory and planning.

It outperforms its competitors, including GPT-4, in standardized benchmarks and has even surpassed human experts in Massive Multitask Language Understanding

(MMLU). One of Gemini Advanced's groundbreaking features is its multimodal capabilities, allowing it to understand and interpret text, video, audio, images, and code simultaneously.

This capability enables Gemini to provide nuanced responses and solutions to a wide range of tasks across various domains, including applied mathematics, physics, engineering, and coding. Gemini Advanced comes in three sizes: Nano, Pro, and Ultra, catering to different user needs and environments. It is designed to handle highly complex tasks and is future-proofed to accommodate newer technologies and innovations. Google's rebranding of Bard to Gemini Advanced signifies its commitment to advancing AI technology and providing users with cutting-edge AI experiences. The rollout of Gemini Advanced includes a dedicated Android app and subscription options, emphasizing Google's investment in AI assistants and agents. The subscription option, Gemini Ultra 1.0, offers access to Google's most powerful AI model and includes two terabytes of cloud storage. Google aims to integrate Gemini into various Google products and services, enhancing user experiences and productivity. Gemini Advanced is available in English in over 150 countries, with plans for international expansion and additional language support. The rebranding reflects Google's evolution in AI development and its vision for the future of AI assistants and agents [56–60].

4. Bard to Gemini: The vision information's

Google has rebranded its AI chatbot Bard as Gemini Advanced, representing a significant advancement in AI technology. Gemini Advanced introduces Ultra 1.0, the most capable state-of-the-art AI model, which excels in complex tasks like coding, logical reasoning, and creative collaboration. Gemini Advanced offers personalized learning experiences, advanced coding assistance, and creative content generation as illustrated within **Figure S1**.

Gemini Advanced is available in over 150 countries and territories in English, with plans to expand to more languages. It is part of the new Google One AI Premium Plan, priced at \$19.99/month, which includes 2TB of storage and access to Gemini in Google products like Gmail and Docs. Google prioritizes safety and responsibility in AI development, conducting extensive trust and safety checks and incorporating user feedback to refine the model.

Gemini Advanced is designed to be conversational, multimodal, and helpful, providing assistance via typing, talking, or image input. For Android users, Gemini is available as a new kind of assistant, accessible via the Gemini app or Google Assistant. It offers a seamless overlay experience and supports various voice features. On iOS, Gemini will be integrated into the Google app, offering similar functionalities for creative tasks and productivity. Gemini is rolling out in the U.S. in English initially and will expand to more locations and languages in the coming weeks. Users are encouraged to provide feedback to enhance their experience with Gemini Advanced.

5. Critical analysis for Gemini model: Hype or hope?

Google's Gemini AI represents a significant advancement in the field of artificial intelligence, particularly in multimodal language modeling. Developed by Google

DeepMind, Gemini is designed to process and understand various types of data, including text, images, and audio, through its Transformer-based architecture. Gemini comes in three distinct models tailored for different purposes and performance levels: Ultra, Pro, and Nano. Each model offers unique strengths and is optimized for efficiency and scalability.

The key strengths of Gemini lie in its multimodal learning approach, enabled by its innovative transformer-based architecture. This allows Gemini to seamlessly integrate and analyze information from various sources, leading to improved context awareness and enhanced handling of complex relationships within data. Gemini's diverse capabilities have the potential to transform industries such as healthcare, finance, education, and entertainment. By automating complex tasks, assisting in decision-making, and unlocking new avenues for creativity and innovation, Gemini is poised to revolutionize the way we work, learn, and interact with technology.

As Gemini's development continues, future applications could include general-purpose AI, human-AI collaboration, and advancements in creativity across various fields. However, ethical considerations surrounding bias, transparency, and accountability are crucial to ensure responsible development and deployment of Gemini. Gemini represents a significant milestone in AI technology, with the potential to shape the future of AI and its impact on society in both beneficial and challenging ways.

6. Gemini vs. GPT4: Exhaustive analysis, comparisons

Gemini, known for its advanced capabilities in specific domains, is designed to excel in tasks requiring deep domain knowledge and contextual understanding. It offers various types of assistance, especially in specialized programming languages or frameworks, providing tailored solutions and optimization tips. On the other hand, GPT-4, the latest iteration in the generative pre-trained transformer series by OpenAI, is renowned for its versatility across various tasks and programming languages. It can generate code snippets, debug, and provide solutions to common programming problems across different technologies.

The comparison visualized and achieved from **Table 1** mainly highlights that Gemini excels in specialized domains, offering in-depth knowledge and tailored solutions, making it potentially more suitable for niche applications. In contrast, GPT-4's strength lies in its versatility and broad knowledge base, making it ideal for a wide range of applications, including general coding assistance.

Table 1. A Comparison between the models of Gemini and GPT4.

Feature/aspect	Gemini	GPT-4
Knowledge base	Specialized in certain domains, offering in-depth insights.	Extensive and general, covering a wide range of topics.
Learning approach	Focused on deep learning within specific domains.	Utilizes unsupervised learning with a diverse dataset.
Language understanding	Highly contextual in its specialized fields.	Broad and nuanced understanding across multiple languages.
Coding assistance	Potentially more adept in specific programming languages or frameworks.	Versatile in offering coding assistance across various languages and frameworks.

Table 1. (Continued).

Feature/aspect	Gemini	GPT-4
Customization	May offer more tailored solutions in certain domains.	General-purpose with adaptable frameworks for different tasks.
Real-time learning	Could be more focused on real-time updates in its areas of expertise.	A broad range of tools and extensive community support.
Interactivity	Possibly more interactive in niche scenarios.	Highly interactive with a wide range of conversational abilities.
Error rate	Potentially lower in specialized tasks.	Generally low, but can vary depending on the task complexity.
Developer tools and support	Specific tools tailored to specialized domains.	Broad range of tools and extensive community support.
Use case specificity	High in its areas of expertise.	Broad and adaptable to numerous use cases.

Ultimately, the choice between Gemini and GPT-4 for developers depends on the specific requirements of their project. For projects demanding deep expertise in a particular domain, Gemini might be preferred. However, for more general purposes and a broader range of tasks, GPT-4 would likely be more beneficial.

7. Generative AI (GAI): The research landscape

The evolving landscape of large language models (LLMs) is characterized by significant innovations like Gemini and Q*, prompting numerous studies aimed at delineating future research paths. These studies showcase a dichotomy between established methods and early adoption, with a growing focus on multimodal capabilities and conversation-driven learning, as exemplified by Gemini.

However, the propagation of preprints, while expediting knowledge sharing, raises concerns about reduced academic scrutiny, inherent biases, and the risks of plagiarism and forgery. The academic community stands at a juncture, requiring a unified effort to refine research directions amidst the rapidly evolving field, which is partly traced through changing research keywords' popularity over time [56–60].

The release of generative models like GPT and the commercial success of ChatGPT have significantly influenced research trends, correlating with industry milestones like the release of the “Transformer” model and ChatGPT-3.5. Exploring the societal and economic impacts of AI advancements reveals how AI technologies reshape industries, alter employment landscapes, and influence socio-economic structures.

While AI drives innovation and economic growth, ethical implications and potential societal disruptions necessitate careful consideration. The exponential increase in preprints posted on platforms like arXiv underlines a paradigm shift in research dissemination within the AI community, facilitating rapid knowledge exchange but raising concerns about information validation. The evolution of LLMs from statistical approaches to deep learning techniques, culminating in models like GPT and BERT, has revolutionized Natural Language Processing (NLP). The advent of deep learning has led to the development of LLMs like ChatGPT, pushing the boundaries of language understanding and generation.

ChatGPT’s success underscores a robust market demand for natural language AI, stimulating interdisciplinary research across various sectors like education, healthcare,

and commerce. However, concerns about AI consciousness, safety, and ethical governance persist. Advancements within LLMs have highlighted the importance of fine-tuning, hallucination reduction, and alignment in enhancing functionality and reliability. The adoption of the Mixture of Experts (MoE) architecture in LLMs represents a critical evolution, offering significant scalability and specialization advantages. However, technical challenges like dynamic routing complexity and ethical alignment issues necessitate continued interdisciplinary research and governance to ensure the safety, ethical alignment, and transparency of these models.

The evolving landscape of LLMs, exemplified by innovations like Gemini and Q*, underscores the importance of interdisciplinary research, ethical considerations, and responsible governance to navigate the complex challenges and harness the transformative potential of AI for the benefit of humanity. Moving towards the emergence of multimodal AI, exemplified by innovations like Gemini, Q*, and speculative advancements in AI technology, heralds a transformative era in artificial intelligence development. Gemini, a pioneering multimodal conversational system, surpasses traditional text-based LLMs by integrating diverse data types such as text, images, audio, and video. Its architecture, characterized by a dual-encoder structure and cross-modal attention network, sets new benchmarks in AI by enabling sophisticated multimodal contextualization and compositional understanding of scenes. Despite these advancements, evaluating Gemini's real-world performance in complex reasoning tasks remains a challenge. The development of multimodal AI systems faces technical hurdles like creating robust datasets, managing scalability, and enhancing user trust and system interpretability. Challenges include data skew and bias, computational demands, and resolving conflicts between modalities. Addressing these challenges is vital for enabling seamless and intelligent interaction aligned with human expectations. Furthermore, the expansion of multimodal AI introduces ethical and social challenges beyond those faced by text-based AI. Concerns include DeepFake technology's ability to generate convincingly realistic media, privacy implications, and the propagation of biases and stereotypes. Ethical development and governance frameworks focusing on transparency, consent, and data handling protocols are essential to address these challenges. Speculative advances like Q* represent a significant leap forward, blending structured learning with creativity to surpass current multimodal capabilities. The integration of Q-learning, A* algorithms, and LLMs aims to create AI systems adept at structured problem-solving and complex human-like communication and reasoning. This potential advancement illustrates the dynamic and ever-evolving nature of AI research, opening up possibilities for more integrated AI applications with greater autonomy and sophistication. The emergence of multimodal AI, including advancements like Gemini and speculative projects like Q*, signifies a transformative era in AI development. Addressing technical challenges, ethical considerations, and realizing the potential of speculative advancements are crucial for leveraging multimodal AI's capabilities to benefit society and advance the field of artificial intelligence.

To better understand the perspective in terms of relating to the matter **Figure 1** sheds light on the visualizations landscape ongoing within the Generative Artificial Intelligence (GAI).

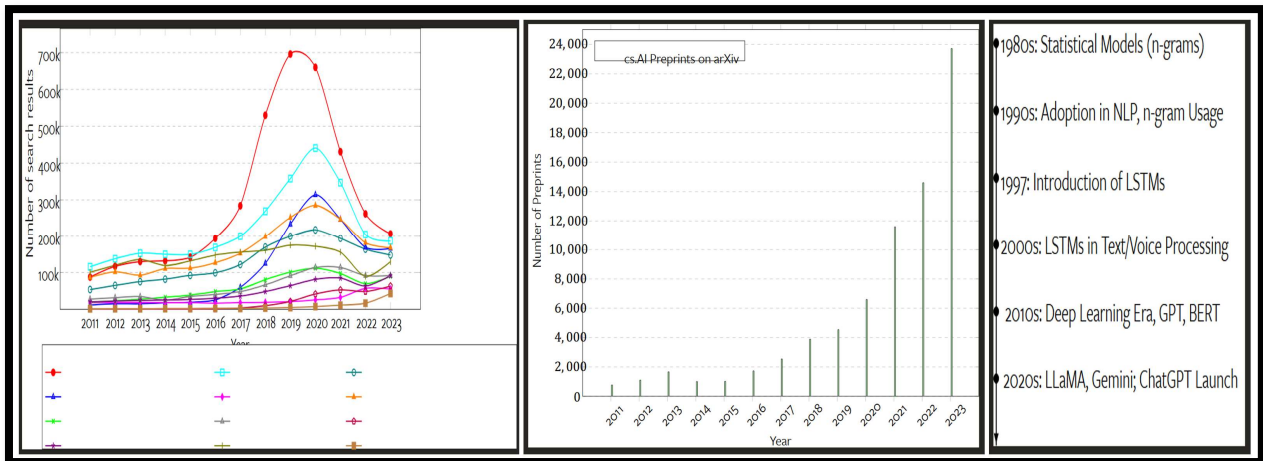


Figure 1. An illustration of the GAI Research Landscape.

8. Generative AI (GAI) research taxonomy: Breakdowns

The field of generative AI is experiencing rapid evolution, necessitating a comprehensive taxonomy to categorize its diverse areas of inquiry and innovation. This taxonomy, detailed in **Table 2**, encompasses model architectures, training techniques, application domains, compliance and ethical considerations, advanced learning, and emerging trends, providing a foundational framework to understand the current state and future directions of the field.

Table 2. Comprehensive taxonomy of current generative AI and LLM research.

Domain	Subdomain	Key focus	Description
Model architecture	Transformer models	Efficiency, scalability	Optimizing network structures for faster processing and larger datasets.
	Recurrent neural networks	Sequence processing	Handling sequences of data, like text, for improved contextual understanding.
	Mixture of experts	Specialization, efficiency	Leveraging multiple expert modules for enhanced efficiency and task-specific performance.
	Multimodal models	Sensory integration	Integrating text, vision, and audio inputs for comprehensive understanding.
Training techniques	Supervised learning	Data labeling, accuracy	Using labeled datasets to train models for precise predictions.
	Unsupervised learning	Pattern discovery	Finding patterns and structures from unlabeled data.
	Reinforcement learning	Adaptability, optimization	Training models through feedback mechanisms for optimal decision-making.
	Transfer learning	Versatility, generalization	Applying knowledge gained in one task to different but related tasks.
Application domains	Natural language understanding	Comprehension, contextualization	Enhancing the ability to understand and interpret human language in context.
	Natural language generation	Creativity, coherence	Generating coherent and contextually relevant text responses.
	Conversational AI	Interaction, naturalness	Developing systems for natural and contextually relevant human-computer conversations.
	Creative AI	Innovation, artistic generation	Generating creative content, including text, art, and music.

Table 2. (Continued).

Domain	Subdomain	Key focus	Description
Compliance and ethical considerations	Bias mitigation	Fairness, representation	Addressing and reducing biases in AI outputs.
	Data security	Data protection, confidentiality	Ensuring data confidentiality, integrity and availability security in AI models and outputs.
	AI ethics	Fairness, accountability	Addressing ethical issues such as bias, fairness, and accountability in AI systems.
	Privacy preservation	Privacy compliance, anonymization	Protecting data privacy in model training and outputs.
Advanced learning	Self-supervised learning	Autonomy, efficiency	Utilizing unlabeled data for model training, enhancing learning efficiency.
	Meta-learning	Rapid adaptation	Enabling AI models to quickly adapt to new tasks with minimal data.
	Fine tuning	Domain-specific tuning, personalization	Adapting models to specific domains or user preferences for enhanced relevance and accuracy.
	Human value alignment	Ethical integration, societal alignment	Aligning AI outputs with human ethics and societal norms, ensuring decisions are ethically and socially responsible.
Emerging trends	Multimodal learning	Integration with vision, audio	Combining language models with other sensory data types for richer understanding.
	Interactive and cooperative AI	Collaboration, human-AI interaction	Enhancing AI's ability to work alongside humans in collaborative tasks.
	AGI development	Holistic understanding	Pursuing the development of AI systems with comprehensive, human-like understanding.
	AGI containment	Safety protocols, control mechanisms	Developing methods to contain and control AGI systems to prevent unintended consequences.

Model architectures: Generative AI model architectures have seen significant advancements across four key domains: Transformer Models, Recurrent Neural Networks (RNNs), Mixture of Experts (MoE), and Multimodal Models. These architectures enable enhanced contextual processing, sequence modeling, efficiency, and integration of diverse sensory inputs, revolutionizing tasks across Natural Language Processing (NLP), computer vision, and other domains.

Training techniques: Generative AI models leverage four key training techniques: supervised learning, unsupervised learning, reinforcement learning, and transfer learning. These techniques contribute uniquely to model training by using labeled data, uncovering patterns in unlabeled data, optimizing decision-making processes, and transferring knowledge across tasks, respectively.

Application domains: The application domains of generative AI are diverse and evolving, encompassing natural language understanding (NLU), natural language generation (NLG), conversational AI, and creative AI. These domains leverage advancements in AI technology to improve language comprehension, text generation, human-computer interactions, and creative content generation.

Compliance and ethical considerations: With the rapid integration of AI technologies into various sectors, ethical considerations and legal compliance have become increasingly crucial. Bias mitigation, data security, AI ethics, and privacy preservation are essential aspects addressed through ethical AI frameworks, ensuring fairness, accountability, transparency, and privacy protection in AI development and

deployment.

Advanced learning: Advanced learning techniques, including self-supervised learning, meta-learning, fine tuning, and human value alignment, enhance the autonomy, efficiency, and versatility of AI models. These techniques enable autonomous model training, rapid adaptation to new tasks, customization to specific domains, and alignment with human ethics and values.

Emerging trends: Emerging trends in generative AI research include multimodal learning, interactive and cooperative AI, AGI development, and AGI containment. These trends focus on achieving richer contextual awareness, enhancing collaboration between AI systems and humans, developing AI with human-like cognitive abilities, and ensuring the safety and ethical alignment of advanced AI systems, respectively.

The taxonomy of generative AI provides a comprehensive framework to navigate the complexities of the field, encompassing diverse areas of inquiry and innovation while addressing ethical considerations and emerging trends shaping the future of AI technology and human interaction.

9. Gemini era: Research challenges, effectiveness, performance, possibilities

The year 2023 marked the entry of Gemini, Google's groundbreaking AI solution, which surpassed GPT-4 in 30 out of 32 standard tests, showcasing its remarkable human-like versatility and "multimodality". Gemini's introduction has sparked excitement in the research community due to its potential to significantly enhance efficiency and productivity for researchers. Gemini offers several novel capabilities that can revolutionize the research process.

Interdisciplinary research: Gemini enables researchers to conduct research without the need to wait for expertise or resources. Its ability to engage in unscripted and open-ended conversations facilitates new discoveries and helps fill gaps in knowledge, particularly in basic research.

Organized and methodical approach: The AI organizes global data and presents it to researchers in a relevant and reliable manner, making it usable for research purposes. This organized approach streamlines the research process and enhances the quality of data analysis.

Deciphering research: Gemini has the capability to read through articles and determine their relevance based on prompts from the researcher. This feature simplifies the process of identifying relevant literature and extracting valuable insights from scientific papers.

Handling large datasets: Gemini excels in handling large datasets, quickly scanning data and generating figures and graphs within minutes to hours. This capability is particularly beneficial for fields such as genomics and other data-intensive areas like law and finance.

Multimodality: Gemini's standout feature is its ability to generate any kind of output using any type of input. Whether it's generating code from various inputs or producing text and images, Gemini offers unparalleled flexibility and versatility.

Gemini's versatility and advanced capabilities make it a valuable tool for researchers across diverse fields. It facilitates interdisciplinary collaboration, aids in

efficient data analysis, and accelerates the research process by handling large datasets effectively. Additionally, the increasing trend of international collaboration among researchers has become more prevalent in recent years, driven by evolving communication platforms like social media. Collaborations between researchers from different countries, disciplines, and sectors have led to a significant increase in co-authored research papers and international collaborations [60–66]. Despite the numerous benefits of research collaboration, it also presents challenges, including cultural and legal differences, conflicting working styles, and differences in communication habits.

Overcoming these challenges requires clear communication, systematic planning, and awareness of relevant laws and policies. Gemini represents a significant advancement in AI technology with the potential to revolutionize the research process by enhancing efficiency, facilitating interdisciplinary collaboration, and handling large datasets effectively. Embracing research collaboration, despite its challenges, offers immense opportunities for advancing knowledge and driving innovation in various fields.

10. Results, findings, directions

Google's AI creation, initially named Bard, faced criticism for its underwhelming performance, struggling to provide accurate information and maintain coherent conversations. However, Google made significant improvements to Bard, renaming it Gemini after its powerful language model, Gemini Pro. The revamped Gemini boasts capabilities that rival those of ChatGPT, including holding helpful conversations, generating images, and seamlessly integrating with Google Workspace. As mentioned previously, Gemini is a generative AI tool designed to assist users in various tasks, such as generating text for cover letters, homework, computer code, Excel formulas, question answers, and translations. Similar to ChatGPT, Gemini utilizes AI to provide human-like conversational responses to user prompts.

The Gemini chat window features a user-friendly interface with options like starting a new chat, entering prompts via text or microphone, uploading images for analysis, and listening to Gemini's responses. Users can interact with Gemini's responses by providing feedback, refreshing for new answers, selecting, copying, or reporting them. Additionally, Gemini offers extensions that enable it to connect with YouTube, Google Maps, Flights, and Hotel for richer responses. Users can customize their experience by adjusting themes, accessing their prompt history, seeking help and support, managing settings, and accessing Google Apps through their Google account. Gemini's ability to generate images adds another dimension to its capabilities, allowing users to request specific images for various purposes. Users can download generated images, regenerate them, or request different images as needed. The New Gemini represents a significant advancement in AI technology, providing users with a powerful tool for generating text and images, conducting conversations, and enhancing productivity across various tasks.

Unlike the free version of ChatGPT, which has a knowledge cutoff in January 2022–2023, Gemini can provide up-to-date information from the web, making it versatile for a wide range of tasks such as giving recipes, assisting with code

debugging, and answering factual questions. Users can interact with Gemini by asking questions or providing prompts, and the chatbot responds with AI-generated answers. Some examples of prompts users can ask include creating to-do lists, defining terms, writing poems, or developing plugins. While Gemini strives to provide accurate responses, it is an experimental AI tool that may occasionally make mistakes or provide inaccurate information. Google has made significant upgrades to Gemini's capabilities and user experience, integrating it with various Google services like Gmail, Maps, Lens, and more. Gemini does not save entire conversations but retains the prompts and questions asked by users. However, given Google's reputation as a major data tracker, users are advised to refrain from sharing personally identifiable or private information with the chatbot. Gemini uses Google's proprietary large language model (LLM) called Gemini Pro, rather than the GPT series used by other popular AI chatbots. While Gemini and similar AI chatbots have the potential to replace traditional search engines by providing conversational responses to user queries, they may not always deliver accurate information. Initially accessible through a waitlist, Gemini is now currently available to users in over 180 countries and territories without the need for waiting, as announced during Google's I/O event in May 2023. The introduction of the new Gemini, Google's latest and most capable AI model, marks a significant milestone in the evolution of artificial intelligence (AI). Demis Hassabis, CEO and Co-Founder of Google DeepMind, emphasizes the transformative potential of AI in advancing scientific discovery, accelerating human progress, and improving lives worldwide. The shift towards AI is predicted to be the most profound technological transition in recent history, offering unprecedented opportunities for innovation, economic growth, and societal advancement. Gemini represents a culmination of Google's efforts in AI research and development, aiming to create a versatile and intuitive AI model inspired by human cognition. Unlike previous models, Gemini is multimodal, meaning it can seamlessly understand and process various types of information, including text, code, audio, image, and video. This flexibility enables Gemini to excel in a wide range of tasks and applications, from natural language understanding to mathematical reasoning and code generation. The introduction of Gemini 1.0 brings state-of-the-art performance across multiple benchmarks, surpassing human experts in tasks such as massive multitask language understanding (MMLU) and multimodal reasoning. Its sophisticated capabilities enable it to extract insights from complex data sets, provide nuanced explanations, and generate high-quality code in multiple programming languages. Google has now also prioritized responsibility and safety in the development of Gemini, implementing safeguards to address potential risks associated with AI capabilities. Comprehensive safety evaluations, including bias and toxicity assessments, have been conducted to ensure ethical and inclusive AI deployment. Collaboration with external experts and the adoption of best-in-class adversarial testing techniques further enhance the model's safety and reliability. The results, findings, visualizations are represented within **Figures 2** and **3** respectively to provide better understanding on the perspective.

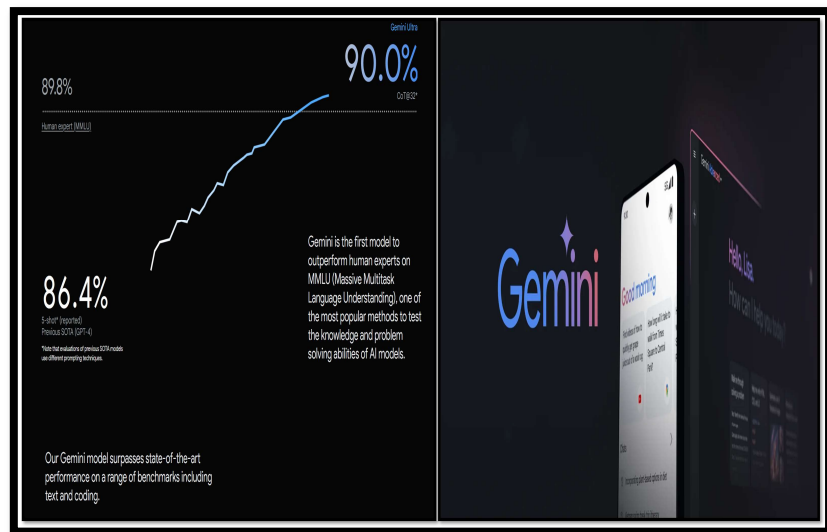


Figure 2. An overview visualization of the research findings 1.

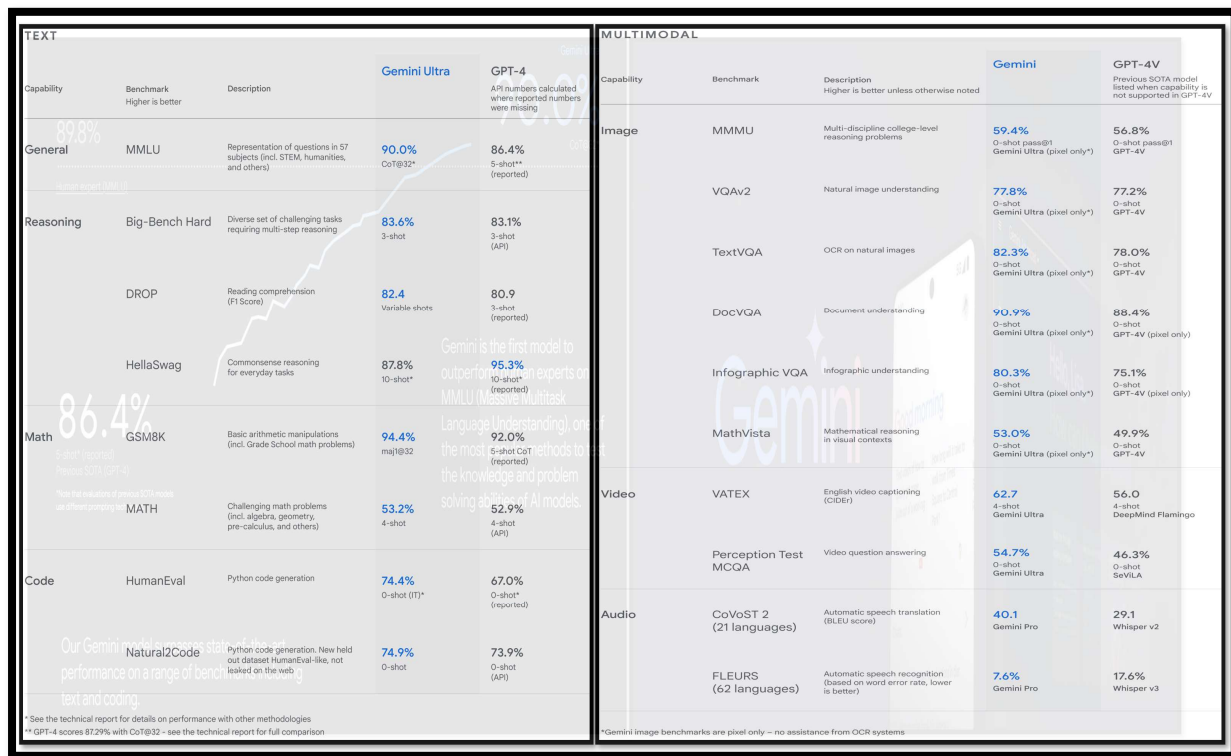


Figure 3. An overview visualization of the research findings 2.

Gemini is optimized for scalability and efficiency, leveraging Google’s in-house designed Tensor Processing Units (TPUs) for accelerated training and deployment. The introduction of Cloud TPU v5p, the most powerful TPU system to date, is poised to accelerate Gemini’s development and facilitate the training of large-scale generative AI models. The new Gemini represents an advancement in terms of AI technology, offering transformative capabilities that have the potential to drive innovation and societal progress across diverse fields and applications. Google remains committed to advancing responsible and ethical AI development, prioritizing safety and collaboration with stakeholders to ensure the beneficial and inclusive deployment of

AI technologies.

Gemini 1.0, Google's latest and most advanced AI model, is being rolled out across various products and platforms, heralding a new era in AI development as for the time of this research exploration. This release brings significant upgrades to existing Google products and introduces new features powered by Gemini's capabilities. Gemini Pro, the fine-tuned version of the model, will enhance the performance of Google's existing products like Bard, which will now incorporate advanced reasoning, planning, and understanding functionalities.

This upgrade will initially be available in English in over 170 countries and territories, with plans for expansion to support new languages and modalities in the near future. Additionally, Pixel 8 Pro will be the first smartphone to run Gemini Nano, enabling new features such as Summarize in the Recorder app and Smart Reply in messaging apps like WhatsApp, Line, and KakaoTalk. Developers and enterprise customers will have access to Gemini Pro through the Gemini API in Google AI Studio or Google Cloud Vertex AI.

Google AI Studio provides a free, web-based developer tool for quick prototyping and app launch, while Vertex AI offers a fully-managed AI platform with customization options and additional features for enterprise security, privacy, and compliance. Gemini Ultra, the most advanced model, is undergoing extensive trust and safety checks before its broad availability.

Early access will be provided to select customers, developers, and partners for experimentation and feedback, with plans to launch Bard Advanced, a new AI experience featuring Gemini Ultra, in the near future. Google aims to continue innovating and expanding Gemini's capabilities, with plans for future versions to include advancements in planning, memory, and processing capabilities. The company envisions a future where AI responsibly empowers innovation, creativity, knowledge extension, and societal transformation on a global scale.

11. Discussions and future directions

The emergence of Gemini in the field of AI signifies a significant milestone, prompting a reflection on the future landscape of artificial intelligence. This pivotal moment elicits a mix of excitement and cautious optimism, considering the revolutionary capabilities it promises across various domains. However, alongside the immense potential lies a need to address the accompanying challenges and ethical considerations inherent in wielding such powerful technology.

Collaboration and partnerships are deemed essential for the responsible development and deployment of AI models like Gemini. This entails fostering alliances between tech giants, research institutions, and industry leaders to leverage collective resources and expertise. Such collaborations can accelerate progress in AI while ensuring that its benefits are harnessed for the betterment of society, guided by ethical principles. Gemini's success is anticipated to catalyze increased investment in AI research and development, driving the creation of even more advanced AI models. This surge in funding will facilitate the exploration of new frontiers in AI, leading to technological breakthroughs and discoveries that were previously unimaginable.

The rapid pace of development exemplified by Gemini underscores the likelihood

of a continuous cycle of innovation, resulting in the emergence of new AI models with unique capabilities and applications. With the integration of AI into various aspects of daily life, governance and regulations become imperative to establish ethical frameworks and guidelines for AI development and deployment. Collaborative efforts between governments, industry stakeholders, and civil society are crucial to ensure transparency, accountability, and fairness in the use of AI, safeguarding against potential risks and biases. Maintaining human oversight and control over AI decision-making processes is paramount. While AI models like Gemini exhibit impressive capabilities, they are ultimately tools designed to be used by humans. Sustaining human control ensures that AI remains aligned with human values and ethical principles, necessitating ongoing monitoring, evaluation, and refinement to mitigate potential risks and biases. The increasing automation driven by AI may lead to job displacement in certain sectors, emphasizing the importance of reskilling and upskilling the workforce. Governments and educational institutions play a pivotal role in providing retraining programs and resources to facilitate a smooth transition for individuals affected by automation, thereby ensuring inclusivity and adaptability in the changing technological landscape. Addressing algorithmic bias is a critical challenge in AI development, as training data sets may inadvertently reflect societal biases, resulting in discriminatory outcomes. Mitigating bias in training data and algorithms is essential to promote fairness and inclusivity in AI applications, emphasizing the need for continuous efforts to enhance algorithmic transparency and accountability. Public engagement and education are vital in fostering understanding, trust, and responsible use of AI technology. Educating the public about AI capabilities and limitations promotes informed discourse and ethical decision-making regarding its development and deployment. International cooperation is indispensable in addressing global challenges and opportunities presented by AI, including issues related to data privacy, security, and ethics. Collaborative efforts on a global scale are essential to ensure responsible development and deployment of AI, facilitating the establishment of international standards and protocols for ethical AI governance. In terms of navigating the future landscape of AI with Gemini on the horizon, foresight, collaboration, and a commitment to ethical development are paramount. By harnessing the power of AI responsibly and thoughtfully, we can create a future where technology serves the collective good of humanity, ensuring that AI remains a tool wielded for the betterment of society.

12. Potential limitations and future explorations

While the research aims to comprehensively evaluate Gemini, a cutting-edge AI chatbot developed by Google, it is essential to acknowledge potential limitations and review points that may arise from this research investigative exploration. By addressing these points of perspective, researchers and developers can fully strengthen the validity and reliability of their research findings, fostering a more nuanced understanding of Gemini's capabilities and limitations. Incorporating diverse samples, employing robust research designs, considering a broader range of evaluation criteria, and ensuring transparency and independence in research processes are vital steps toward enhancing the validity and reliability of findings in the evaluation of AI

chatbots like Gemini. Overcoming the limitations identified requires careful consideration and strategic planning. Here are some suggestions on how these limitations can be addressed and potential routes for future research in these domains.

Generalizability of findings: Future research could focus on conducting comparative studies involving multiple AI chatbots developed by different companies or researchers. This approach would enable researchers to draw more generalizable conclusions about the performance, usability, and integration capabilities of AI chatbots across diverse contexts. Implementing standardized evaluation protocols and benchmarks could facilitate comparisons between different AI chatbots, enhancing the reliability and generalizability of findings.

Research bias: Employing more rigorous sampling techniques, such as random sampling or stratified sampling, could help mitigate biases in participant selection and improve the representativeness of the sample. Implementing diverse recruitment strategies, including reaching out to users from various demographic backgrounds and user segments, could enhance the diversity of participants and reduce potential biases in user feedback.

Overemphasis on technical analysis: Future research could adopt a more balanced approach that integrates technical analysis with qualitative assessments of user experience, cultural considerations, and societal impacts. This holistic approach would provide a comprehensive understanding of the strengths and limitations of AI chatbots like Gemini. Exploring innovative methodologies, such as user-centered design principles and participatory design approaches, could facilitate the integration of diverse perspectives into the evaluation process.

Ethical implications and bias mitigation: Continuously monitoring and evaluating ethical considerations throughout the development and deployment of AI chatbots is crucial. Future research could focus on developing and implementing robust frameworks for ethical AI development, including guidelines for data privacy, security, and bias mitigation. Collaborating with interdisciplinary teams comprising ethicists, sociologists, and legal experts could provide valuable insights into the ethical implications of AI technologies and support the development of effective mitigation strategies.

Commercial interests and transparency: Enhancing transparency and accountability in AI research and development processes is essential. Future research could advocate for greater transparency and openness in disclosing potential conflicts of interest and funding sources. Promoting collaboration between academia, industry, and regulatory bodies could foster transparency and facilitate independent evaluations of AI technologies like Gemini. Establishing clear guidelines and standards for transparency in AI research and development could help mitigate the influence of commercial interests. Addressing the identified limitations requires a multifaceted approach that integrates diverse perspectives, rigorous methodologies, and ethical considerations. Future research in these domains should prioritize collaboration, transparency, and accountability to advance our understanding of AI chatbots and their impacts on society.

13. Conclusions

The introduction of Gemini signifies a pivotal moment in the advancement of artificial intelligence (AI), heralding a future teeming with exciting possibilities alongside notable challenges. To steer this future towards benefiting all of humanity, a concerted effort towards collaboration and responsibility is imperative. Collaboration among tech giants, research institutions, and industry leaders is paramount for expediting the responsible development and deployment of AI. This collaborative approach ensures a diverse pool of expertise and resources, facilitating comprehensive solutions and ethical considerations throughout the AI landscape. Increased investment in AI research is crucial for fostering innovation and understanding the potential implications of this rapidly evolving technology. Coupled with robust governance and regulations, this investment ensures that AI applications are developed and utilized in a manner that upholds ethical standards and promotes equity. Human oversight and control remain central in navigating the ethical complexities of AI. Reskilling and upskilling the workforce are essential to adapt to the evolving technological landscape and mitigate potential job displacement resulting from automation. Addressing algorithmic bias and promoting transparency are critical components of building trust in AI systems. By actively mitigating bias in algorithms and promoting transparency in decision-making processes, stakeholders can foster trust and ensure fairness in AI applications. Public engagement and international cooperation are indispensable in shaping a future where AI serves humanity positively. Engaging the public in discussions about AI's impact and fostering international collaboration on ethical standards and regulations are essential for guiding the responsible development and deployment of AI globally.

The future landscape of AI with Gemini and other advanced models requires a steadfast commitment to collaboration, ethical considerations, and responsible development. By embracing these principles, stakeholders can harness the unparalleled potential of AI to create a brighter and more inclusive future for all.

Supplementary materials: The new Gemini (Android) in action is represented within **Figure S1** for a better overview concerning the matter.

Acknowledgments: The author would like to acknowledge and thank the GOOGLE Deep Mind Research with its associated pre-prints access platforms. This research was deployed and utilized under the platform provided by GOOGLE Deep Mind which is under the support of the GOOGLE Research and the GOOGLE Research Publications under GOOGLE Gemini platform. Using their provided platform of datasets and database files with digital software layouts consisting of free web access to a large collection of recorded models that are found in research access and its related open-source software distributions which is the implementation and simulation of analytics for the proposed research which was undergone and set in motion. There are many datasets, data models which are resourced and (accessed on from a wide variety of GOOGLE service domains. All the data sources and various domains from which data has been included and (accessed on for this research are identified, mentioned and referenced where appropriate. However, various original data models and datasets of which are not all publicly available, because they contain private information. The

available platform provided datasets and data models that support the research findings and information of the research investigations are also referenced where appropriate.

Availability of data and materials: The various original data models and datasets of which are not all publicly available, because they contain private information. The available platform provided datasets and data models that support the findings and information of the research investigations are referenced where appropriate.

Conflict of interest: The author declares no conflict of interest.

References

1. Where you can use Bard. Available online: <https://support.google.com/gemini/answer/13575153> (accessed on 13 April 2024).
2. David E. The AI wars might have an armistice deal sooner than expected. Available online: <https://www.theverge.com/2023/7/20/23800755/ai-meta-llama-microsoft-chatgpt-interoperability> (accessed on 13 April 2024).
3. Konrad A, Cai K. Inside ChatGPT's Breakout Moment and the Race to Put AI To Work. Available online: <https://www.forbes.com/sites/alexkonrad/2023/02/02/inside-chatgpts-breakout-moment-and-the-race-for-the-future-of-ai/?sh=3f67b566240b> (accessed on 13 April 2024).
4. Vincent J. AI-generated answers temporarily banned on coding Q&A site Stack Overflow. Available online: <https://www.theverge.com/2022/12/5/23493932/chatgpt-ai-generated-answers-temporarily-banned-stack-overflow-llms-dangers> (accessed on 13 April 2024).
5. Olson P. Google Faces a Serious Threat from ChatGPT. Available online: <https://www.washingtonpost.com/> (accessed on 13 April 2024).
6. Grant N, Metz C. A New Chat Bot Is a 'Code Red' for Google's Search Business. Available online: <https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html> (accessed on 13 April 2024).
7. Newton C. How Google is making up for lost time. Available online: <https://www.theverge.com/2023/5/12/23721037/google-ai-progress-search-docs-starline-video-calls> (accessed on 13 April 2024).
8. Grant N. Google Calls in Help from Larry Page and Sergey Brin for A.I. Fight. Available online: <https://www.nytimes.com/2023/01/20/technology/google-chatgpt-artificial-intelligence.html> (accessed on 13 April 2024).
9. Nieva R, Konrad A. Back at Google Again, Cofounder Sergey Brin Just Filed His First Code Request in Years. Available online: <https://www.forbes.com/sites/richardnieva/2023/01/31/sergey-brin-code-request-lamda/?sh=6a71df637ce6> (accessed on 13 April 2024).
10. Condon S. Google I/O 2021: Google unveils new conversational language model, LaMDA. Available online: <https://www.zdnet.com/article/google-io-google-unveils-new-conversational-language-model-lamda/> (accessed on 13 April 2024).
11. Roth E. Meet the companies trying to keep up with ChatGPT. Available online: <https://www.theverge.com/2023/3/5/23599209/companies-keep-up-chatgpt-ai-chatbots> (accessed on 13 April 2024).
12. Kleinman Z. ChatGPT firm trials \$20 monthly subscription fee. Available online: <https://www.bbc.com/news/technology-64492750> (accessed on 13 April 2024).
13. Elias J. Google execs warn company's reputation could suffer if it moves too fast on AI-chat technology. Available online: <https://www.cnn.com/2022/12/13/google-exec-warn-of-reputational-risk-with-chatgpt-like-tool.html> (accessed on 13 April 2024).
14. Vincent J. Google won't launch ChatGPT rival because of 'reputational risk'. Available online: <https://www.theverge.com/2022/12/14/23508756/google-vs-chatgpt-ai-replace-search-reputational-risk> (accessed on 13 April 2024).
15. Cuthbertson A. DeepMind's AI chatbot can do things that ChatGPT cannot, CEO claims. Available online: <https://www.independent.co.uk/tech/deepmind-ai-chatbot-chatgpt-openai-b2262862.html> (accessed on 13 April 2024).

16. Elias J. Google is asking employees to test potential ChatGPT competitors, including a chatbot called ‘Apprentice Bard’. Available online: <https://www.cnbc.com/2023/01/31/google-testing-chatgpt-like-chatbot-apprentice-bard-with-employees.html> (accessed on 13 April 2024).
17. O’Brien M. Google has the next move as Microsoft embraces OpenAI buzz. Available online: <https://apnews.com/article/technology-science-software-microsoft-corp-business-a7f6f44fdd53730e45daa611555c8c23> (accessed on 13 April 2024).
18. Mailk A. Google tries to reassure investors on AI progress as ChatGPT breathes down its neck. Available online: <https://techcrunch.com/2023/02/03/google-best-yet-to-come-ai-journey-faces-potential-disruption-openai-chatgpt/> (accessed on 13 April 2024).
19. Alba D, Love J. Google releases ChatGPT rival AI ‘Bard’ to early testers. Available online: <https://www.latimes.com/business/story/2023-02-06/google-chatgpt-rival-ai-bard-early-testers> (accessed on 13 April 2024).
20. Schechner S, Kruppa M. Google Opens ChatGPT Rival Bard for Testing, as AI War Heats Up. Available online: <https://www.wsj.com/articles/google-opens-testing-of-chatgpt-rival-as-artificial-intelligence-war-heats-up-11675711198> (accessed on 13 April 2024).
21. Nieva R. Google Debuts A ChatGPT Rival Called Bard In Limited Release. Available online: <https://www.forbes.com/sites/richardnieva/2023/02/06/google-bard/?sh=795b687b152d> (accessed on 13 April 2024).
22. Elias J. Google CEO tells employees that 80,000 of them helped test Bard A.I., warns ‘things will go wrong’. Available online: <https://www.cnbc.com/2023/03/21/google-ceo-pichai-memo-to-employees-on-bard-ai-things-will-go-wrong.html> (accessed on 13 April 2024).
23. Mollman S. Google’s head of ChatGPT rival Bard reassures employees it’s ‘a collaborative A.I. service’ and ‘not search’. Available online: <https://fortune.com/2023/03/03/google-chatgpt-rival-bard-sundar-pichai-reassures-employees-tool-not-search-but-ai-companion/> (accessed on 13 April 2024).
24. Elias J. Google execs tell employees in testy all-hands meeting that Bard A.I. isn’t just about search. Available online: <https://www.cnbc.com/2023/03/03/google-exec-say-in-all-hands-meeting-bard-ai-isnt-all-for-search-.html> (accessed on 13 April 2024).
25. Dastin J, Nellis S. For tech giants, AI like Bing and Bard poses billion-dollar search problem. Available online: <https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22/> (accessed on 13 April 2024).
26. Liedtke M. Google hopes ‘Bard’ will outsmart ChatGPT, Microsoft in AI. Available online: <https://apnews.com/article/technology-science-microsoft-corp-sundar-pichai-business-395b0b94410d9fa3ebb13b14af640e06> (accessed on 13 April 2024).
27. Metz C, Grant N. Racing to Catch Up with ChatGPT, Google Plans Release of Its Own Chatbot. Available online: <https://www.nytimes.com/2023/02/06/technology/google-bard-ai-chatbot.html> (accessed on 13 April 2024).
28. Knight W. The Chatbot Search Wars Have Begun. Available online: <https://www.wired.com/story/fast-forward-the-chatbot-search-wars-have-begun/> (accessed on 13 April 2024).
29. Warren T. Microsoft and Google are about to Open an AI battle. Available online: <https://www.theverge.com/2023/2/7/23587767/microsoft-google-open-ai-battle-search-bing> (accessed on 13 April 2024).
30. Google Strikes Back in AI Battle with Microsoft. Available online: <https://www.barrons.com/news/google-strikes-back-in-ai-battle-with-microsoft-01675871108> (accessed on 13 April 2024).
31. Morrison S. Google is scrambling to catch up to Bing, of all things. Available online: <https://www.vox.com/recode/2023/2/7/23590069/bing-openai-microsoft-google-bard> (accessed on 13 April 2024).
32. Peers Martin. What Google’s AI Catch-Up Reveals. Available online: <https://www.theinformation.com/articles/what-google-s-ai-catch-up-reveals> (accessed on 13 April 2024).
33. De Vynck G, Tiku N. Google’s catch-up game on AI continues with Bard launch. Available online: <https://www.washingtonpost.com/technology/2023/03/21/bard-google-ai/> (accessed on 13 April 2024).
34. Patel N. Microsoft thinks AI can beat Google at search—CEO Satya Nadella explains why. Available online: <https://www.theverge.com/23589994/microsoft-ceo-satya-nadella-bing-chatgpt-google-search-ai> (accessed on 13 April 2024).

35. Alba D. ChatGPT Reignites the Search Wars Between Google and Microsoft. Available online: <https://www.bloomberg.com/news/newsletters/2023-02-08/chatgpt-reignites-search-competition-microsoft-bing-vs-google-bard> (accessed on 13 April 2024).
36. Stokel-Walker Chris. TechScape: Google and Microsoft are in an AI arms race—who wins could change how we use the internet. Available online: <https://www.theguardian.com/technology/2023/feb/21/techscape-google-bard-microsoft-big-ai-search> (accessed on 13 April 2024).
37. Morrison S. “The exciting new AI transforming search—and maybe everything—explained. Available online: <https://www.vox.com/recode/2023/3/4/23624033/openai-bing-bard-microsoft-generative-ai-explained> (accessed on 13 April 2024).
38. De Vynck G. Investors pummel Google after Microsoft ramps up AI wars. Available online: <https://www.washingtonpost.com/technology/2023/02/08/google-bard-chatbot-microsoft/> (accessed on 13 April 2024).
39. Alba D, Patnaik S. Google suffers \$144b wipeout after Bard AI chatbot gives wrong answer. Available online: <https://www.smh.com.au/business/companies/google-suffers-140b-wipeout-after-bard-ai-chatbot-gives-wrong-answer-20230209-p5cj3a.html> (accessed on 13 April 2024).
40. Miao H. Alphabet Stock Drops 8% After Google Rollout of AI Search Features. Available online: <https://www.wsj.com/livecoverage/stock-market-news-today-02-08-2023/card/alphabet-stock-drops-after-google-parent-introduces-ai-search-features-wgCJG3IDoSbFL3SgyrNI> (accessed on 13 April 2024).
41. Coulter M. Google shares dive 8% after AI chatbot Bard flubs answer in ad. Available online: <https://www.reuters.com/technology/google-ai-chatbot-bard-offers-inaccurate-information-company-ad-2023-02-08/> (accessed on 13 April 2024).
42. Kan M. No, That’s Wrong: Google’s Bard AI Demo Spouts Incorrect Info. Available online: <https://www.pcmag.com/news/no-thats-wrong-googles-bard-ai-demo-spouts-incorrect-info> (accessed on 13 April 2024).
43. Elias J. Google employees criticize CEO Sundar Pichai for ‘rushed, botched’ announcement of GPT competitor Bard. Available online: <https://www.cnn.com/2023/02/10/google-employees-slam-ceo-sundar-pichai-for-rushed-bard-announcement.html> (accessed on 13 April 2024).
44. Harrison M. Chaos Inside Google as Execs Try to Figure Out How to Actually Use AI. Available online: <https://futurism.com/chaos-google-exec-use-for-ai> (accessed on 13 April 2024).
45. Elias J. Alphabet Chairman John Hennessy explains why Google was hesitant to put out its ChatGPT competitor. Available online: <https://www.cnn.com/2023/02/13/alphabet-chairman-john-hennessy-google-bard-hesitation.html> (accessed on 13 April 2024).
46. Maxwell T, Langley H. Leaked memo shows Google CEO Sundar Pichai is asking staffers for help testing its Bard AI chatbot. Available online: <https://www.businessinsider.com/google-bard-chatbot-googlers-sundar-pichai-2023-2> (accessed on 13 April 2024).
47. Elias J. Google asks employees to rewrite Bard’s bad responses, says the A.I. ‘learns best by example’. Available online: <https://www.cnn.com/2023/02/15/google-asks-employees-to-rewrite-bards-incorrect-responses-to-queries.html> (accessed on 13 April 2024).
48. Alba D, Love J. Google’s Rush to Win in AI Led to Ethical Lapses, Employees Say. Available online: <https://www.bloomberg.com/news/features/2023-04-19/google-bard-ai-chatbot-raises-ethical-concerns-from-employees> (accessed on 13 April 2024).
49. Maxwell T, Langley H. Leaked messages show Googlers are taking out their frustrations over layoffs on its new Bard AI chatbot. Available online: <https://www.businessinsider.com/google-layoffs-bard-chatbot-ai-2023-2> (accessed on 13 April 2024).
50. Langley H. “Google employees are already internally testing a smarter version of its chatbot called ‘Big Bard’”. Available online: <https://www.businessinsider.com/google-employees-test-smarter-chatbot-big-bard-2023-3> (accessed on 13 April 2024).
51. Grant N. Google Releases Bard, Its Competitor in the Race to Create A.I. Chatbots. Available online: <https://www.nytimes.com/2023/03/21/technology/google-bard-chatbot.html> (accessed on 13 April 2024).
52. Liedtke M. Google’s artificially intelligent ‘Bard’ set for next stage. Available online: <https://www.washingtonpost.com/> (accessed on 13 April 2024).

53. Vincent J. Google opens early access to its ChatGPT rival Bard—here are our first impressions. Available online: <https://www.theverge.com/2023/3/21/23649794/google-chatgpt-rival-bard-ai-chatbot-access-hands-on> (accessed on 13 April 2024).
54. Schroeder S. Google is giving early access to its AI assistant to Pixel Superfans. Available online: <https://mashable.com/article/google-bard-pixel-superfans> (accessed on 13 April 2024).
55. Li A. Google tapping the Pixel and Nest brands to invite people to Bard. Available online: <https://9to5google.com/2023/03/21/google-bard-pixel-nest-one/> (accessed on 13 April 2024).
56. Fu YC, Marques M, Tseng YH, et al. An evolving international research collaboration network: spatial and thematic developments in co-authored higher education research, 1998–2018. *Scientometrics*. 2022; 127(3): 1403-1429. doi: 10.1007/s11192-021-04200-w
57. Kwiek M. The prestige economy of higher education journals: a quantitative approach. *Higher Education*. 2020; 81(3): 493-519. doi: 10.1007/s10734-020-00553-y
58. Turing AM. Computing machinery and intelligence. *Mind*. 1950; LIX(236): 433-460. doi: 10.1093/mind/lix.236.433
59. McDermott D. Artificial intelligence meets natural stupidity. *ACM SIGART Bulletin*. 1976; (57): 4-9. doi: 10.1145/1045339.1045340
60. Minsky M. Steps toward Artificial Intelligence. *Proceedings of the IRE*. 1961; 49(1): 8-30. doi: 10.1109/jrproc.1961.287775
61. Akhtar ZB. The design approach of an artificial intelligent (AI) medical system based on electronical health records (EHR) and priority segmentations. *The Journal of Engineering*. 2024; 2024(4). doi: 10.1049/tje2.12381
62. Akhtar ZB. Securing Operating Systems (OS): A Comprehensive Approach to Security with Best Practices and Techniques. *International Journal of Advanced Network, Monitoring and Controls*. 2024; 9(1): 100-111. doi: 10.2478/ijanmc-2024-0010
63. Akhtar ZB, Gupta AD. Integrative Approaches for Advancing Organoid Engineering: From Mechanobiology to Personalized Therapeutics. *Journal of Applied Artificial Intelligence*. 2024; 5(1): 1-27. doi: 10.48185/jaai.v5i1.974
64. Akhtar ZB. Advancements within Molecular Engineering for Regenerative Medicine and Biomedical Applications an Investigation Analysis towards A Computing Retrospective. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*. 2024; 6(1). doi: 10.35882/jeeemi.v6i1.351
65. Bin Akhtar Z. A Revolutionary Gaming Style in Motion. *IntechOpen*; 2022. doi: 10.5772/intechopen.100551
66. Akhtar ZB, Stany Rozario V. The Design Approach of an Artificial Human Brain in Digitized Formulation based on Machine Learning and Neural Mapping. In: *Proceedings of the 2020 International Conference for Emerging Technology (INCET)*; 2020; Belgaum, India. pp. 1-7. doi: 10.1109/INCET49848.2020.9154000

Ethical considerations of AI in financial decision

Yifei Wang

School of Information, University of California, Berkeley, CA 94720, USA; sarahwang688@berkeley.edu

CITATION

Wang Y. Ethical considerations of AI in financial decision. *Computing and Artificial Intelligence*. 2024; 2(1): 1290.
<https://doi.org/10.59400/cai.v2i1.1290>

ARTICLE INFO

Received: 15 April 2024
Accepted: 18 June 2024
Available online: 28 June 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: The integration of artificial intelligence (AI) in financial decision-making processes has significantly enhanced the efficiency and scope of services in the finance sector. However, the rapid adoption of AI technologies raises complex ethical questions that need thorough examination. This paper explores the ethical challenges posed by AI in finance, including issues related to bias and fairness, transparency and explainability, accountability, and privacy. These challenges are scrutinized within the framework of current regulatory and ethical guidelines such as the General Data Protection Regulation (GDPR) and the Fair Lending Laws in the United States. Despite these frameworks, gaps remain that could potentially compromise the equity and integrity of financial services. The paper proposes enhancements to existing ethical frameworks and introduces new recommendations for ensuring that AI technologies foster ethical financial practices. By emphasizing a proactive approach to ethical considerations, this study aims to contribute to the ongoing discourse on maintaining trust and integrity in AI-driven financial decisions, ultimately proposing a pathway towards more robust and ethical AI applications in finance.

Keywords: ethics; artificial intelligence; financial decision

1. Introduction

The advent of artificial intelligence (AI) in the financial sector has revolutionized traditional processes, bringing unprecedented efficiency and capabilities. AI's integration ranges from complex trading algorithms and credit scoring systems to personalized banking services and fraud detection mechanisms. However, as financial institutions increasingly rely on AI to make or assist in making critical decisions, a host of ethical challenges have surfaced. These challenges are not just technical but profoundly impact societal norms, individual rights, and the foundational principles of fairness and transparency in financial dealings.

This paper delves into the ethical considerations necessitated by the adoption of AI in financial decision-making. As AI systems process vast amounts of personal data to predict and influence financial outcomes, ethical concerns such as bias amplification, lack of transparency, accountability ambiguities, and privacy infringements become paramount. The implications of these concerns are wide-ranging, affecting everything from individual creditworthiness assessments to macroeconomic stability.

Furthermore, while regulatory frameworks like the General Data Protection Regulation (GDPR) in Europe and the Fair Credit Reporting Act (FCRA) in the United States attempt to address these concerns, they often fall short in the face of rapidly evolving AI technologies [1]. These regulations may provide a foundational framework, but as AI capabilities outpace legal norms, there is a pressing need for a dynamic ethical approach that evolves in tandem with technological advancements.

The objectives of this paper are to:

- 1) Identify and analyze the primary ethical challenges posed by AI in financial decision-making.
- 2) Evaluate the effectiveness of existing regulatory and ethical frameworks in addressing these challenges.
- 3) Propose enhancements to these frameworks to better align AI advancements with ethical financial practices.

Through this exploration, the paper aims to contribute to the ongoing dialogue around AI ethics in finance, proposing actionable insights and recommendations to ensure that AI systems not only enhance financial services but also adhere to the highest ethical standards, thereby maintaining public trust and ensuring equitable outcomes.

2. Background

2.1. The role of AI in finance

Artificial intelligence (AI) has become integral to the finance sector, enhancing capabilities across various domains such as credit scoring, risk management, fraud detection, and algorithmic trading. AI technologies, particularly machine learning and deep learning, analyze large datasets to identify patterns that humans cannot easily detect. For instance, in credit scoring, AI algorithms process complex datasets to assess an applicant's creditworthiness more efficiently and potentially more accurately than traditional methods [2].

2.2. Ethical challenges

The deployment of AI in finance is not without its ethical challenges. These challenges stem from the inherent characteristics of AI systems, including their opacity and the potential for inheriting biases from historical data. Ethical issues such as decision fairness, transparency, accountability, and the right to privacy are at the forefront of discussions concerning AI in finance [3].

2.3. Regulatory landscape

Regulatory frameworks have been developed to address some of these ethical concerns. In the European Union, the General Data Protection Regulation (GDPR) provides guidelines that include the right to explanation, whereby individuals can ask for explanations of automated decisions that affect them. In the United States, the Fair Credit Reporting Act (FCRA) regulates the collection and use of consumer information, including provisions related to automated decision-making [4].

However, these regulations often lag behind the rapid advancements in AI technology, creating gaps that can lead to ethical lapses. Current regulatory frameworks may also be insufficiently flexible to address the unique challenges posed by AI, such as the need for transparency in complex machine learning models or the issues arising from biased training data [5].

2.4. The need for enhanced ethical frameworks

Given these challenges, there is a critical need for enhanced ethical frameworks

that can better govern the deployment of AI in finance. These frameworks should not only comply with existing regulations but also anticipate future developments and ethical dilemmas. Such proactive measures are essential to ensure that AI technologies promote fairness, accountability, and transparency, and do not inadvertently perpetuate or exacerbate existing inequalities [6].

3. Ethical challenges

As AI becomes increasingly embedded in financial decision-making, several ethical challenges have emerged. These challenges revolve around issues of bias and fairness, transparency and explainability, accountability, and privacy. Each of these areas presents unique difficulties that need to be carefully managed to ensure ethical compliance and maintain trust in financial systems.

3.1. Bias and fairness

One of the most pressing concerns in AI-driven financial systems is the risk of bias and unfairness. AI models can inadvertently perpetuate existing biases or develop new biases if the data they learn from is skewed or incomplete. For instance, if historical lending data reflects past discriminatory practices, AI systems trained on this data may continue to deny loans to marginalized groups unfairly. This not only reinforces social inequalities but also violates principles of fairness in financial services [7]. Addressing these biases requires a combination of diverse data, careful modeling, and ongoing monitoring to ensure decisions are fair and equitable across all customer segments.

3.2. Transparency and explainability

The “black box” nature of many AI models poses significant challenges in terms of transparency and explainability. In finance, stakeholders, including regulators, customers, and service providers, often need to understand how decisions are made, especially when these decisions have substantial impacts on individuals’ financial health. The complexity of models like deep neural networks, which provide little insight into their decision-making processes, makes this particularly difficult. The lack of transparency can lead to mistrust and reluctance to adopt AI systems, despite their potential benefits [8]. Developing methods to interpret AI decisions, such as through LIME or SHAP, as discussed earlier, is crucial for overcoming these challenges.

3.3. Accountability

With AI systems making autonomous decisions, establishing clear accountability is challenging. When an AI system makes a flawed decision, it can be difficult to pinpoint responsibility, especially when multiple stakeholders are involved in the design, development, and deployment of these systems. This issue is compounded by the transnational nature of many financial services, where AI systems may be developed in one country and deployed in another, raising questions about jurisdiction and legal liability [9]. Clear guidelines and regulations that specify accountability measures are essential for addressing these challenges.

3.4. Privacy

AI systems in finance often rely on vast amounts of personal data to make decisions. This raises significant privacy concerns, particularly regarding the collection, storage, and use of sensitive information. Ensuring that data is handled securely and in compliance with privacy regulations like GDPR is crucial. Additionally, there is a need for transparency about what data is collected and how it is used, allowing customers to make informed decisions about their data [10].

3.5. Cross-cutting concerns

These ethical challenges are not isolated but interrelated, requiring comprehensive strategies that address multiple aspects simultaneously. For example, efforts to improve model transparency must also consider the potential for exposing sensitive data, balancing the needs for explainability and privacy. Similarly, addressing biases in AI systems is linked with the transparency of the data and models used, which in turn affects accountability.

4. Current regulatory and ethical frameworks

The integration of AI into financial decision-making has prompted regulators worldwide to develop frameworks that guide ethical AI usage. These frameworks aim to ensure that AI technologies are used responsibly, promoting fairness, accountability, transparency, and privacy. However, as AI technologies advance, these frameworks are often tested by new ethical dilemmas and the complexities of global financial operations.

4.1. European Union: GDPR

The General Data Protection Regulation (GDPR) is a significant regulatory framework in Europe that impacts AI in finance by enforcing strict rules on data privacy and the right to explanation. Under GDPR, individuals have the right to understand the decisions made by AI systems, particularly when these decisions affect their legal or economic status. This regulation emphasizes transparency and accountability by requiring firms to disclose how AI systems operate and make decisions [4]. Despite its strengths, GDPR faces challenges in enforcement and does not specifically address all nuances of AI, such as how to interpret the right to explanation in the context of complex machine learning models.

4.2. United States: Fair Credit Reporting Act and Equal Credit Opportunity Act

In the United States, the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA) regulate how consumer credit information is collected and used. These laws aim to prevent discriminatory practices and ensure fair and accurate credit reporting. While these regulations address some aspects of AI ethics, such as fairness and accountability, they were not designed with the complexities of modern AI technologies in mind, often lacking specific guidance on managing AI's unique challenges [7].

4.3. Global initiatives: OECD Principles and G20 AI Guidelines

At a global level, the Organisation for Economic Co-operation and Development (OECD) Principles on AI and the G20 AI Guidelines offer a broader set of ethical standards for AI use, including finance. These guidelines emphasize values like inclusive growth, sustainable development, and human-centered values. They advocate for AI systems that are robust, secure, fair, and trustworthy, promoting transparency and accountability across borders [11]. However, these principles are non-binding and serve more as a guideline than enforceable standards, potentially limiting their impact.

4.4. Industry-specific guidelines

Apart from governmental regulations, several industry groups and professional associations have developed their ethical guidelines for AI in finance. These include initiatives by the Institute of Electrical and Electronics Engineers (IEEE) and the Partnership on AI, which focus on promoting ethical practices in AI development and deployment. These guidelines often provide more detailed recommendations for ethical AI usage but lack the enforcement power of governmental regulations [12].

4.5. Challenges and gaps

While current regulatory and ethical frameworks provide a foundation for ethical AI use in finance, they often fall short in addressing the full range of ethical issues presented by AI technologies. Challenges include keeping pace with the rapid development of AI, applying broad ethical principles to specific cases, and managing the international scope of financial services. Additionally, there is a gap in stakeholder engagement, with consumers and ethicists sometimes underrepresented in the creation of these frameworks.

5. Ethical considerations in AI regulation

The European Union's AI Act and the American National Institute of Standards and Technology (NIST) framework emphasize the importance of accuracy and explainability in AI systems. These attributes are crucial for building trust and ensuring fair AI decision-making processes. For instance, Babaei et al. [13] demonstrate an explainable fintech lending model that aids in understanding the decisions made by AI, which can be aligned with these regulatory requirements to enhance transparency and accountability in AI applications.

Both the EU AI Act and the NIST prioritize robustness and safety in AI systems to prevent failures that could lead to harm or loss. Robust AI systems are designed to handle unexpected situations and maintain performance across various conditions. Giudici et al. [14] discuss methods for measuring AI risks that can contribute to developing safer and more reliable AI systems by identifying potential threats and vulnerabilities early in the design process.

Cyber resilience is another critical aspect covered under both regulatory frameworks, aiming to protect AI systems from cyber threats and ensure their integrity and availability. AI applications in sensitive domains such as finance or healthcare must adhere to high standards of cybersecurity to safeguard user data and prevent

malicious exploits. The work of Babaei et al. [15] on explainable AI for crypto asset allocation highlights the importance of secure and transparent AI methods that comply with ethical and regulatory standards.

6. Proposals for enhanced ethical frameworks

As AI continues to evolve and permeate the financial sector, existing ethical and regulatory frameworks need to be continuously updated and refined to ensure they remain effective. This section proposes several enhancements aimed at strengthening the ethical governance of AI in financial decision-making.

6.1. Developing dynamic regulatory mechanisms

Current static regulations struggle to keep pace with the rapid advancements in AI technology. A dynamic regulatory approach, which incorporates adaptive mechanisms that can update in response to new technological developments and insights, is essential. This could involve the establishment of regulatory sandboxes, where AI technologies can be tested under real-world conditions without the immediate imposition of full regulatory compliance. Such sandboxes allow regulators to understand the implications of new AI applications and adapt regulations more effectively [16].

6.2. Implementing ethical auditing practices

To ensure continuous compliance with ethical standards, regular audits of AI systems should be mandated. These audits should assess both the algorithms and the data used by AI systems for biases, transparency, and accountability. Independent bodies with expertise in AI ethics should conduct these audits, and their findings should be made public to maintain trust [17]. Additionally, the development of standardized auditing protocols would help in establishing a consistent and comprehensive approach to evaluating AI systems.

6.3. Enhancing transparency through technology

Enhancing the transparency of AI systems is crucial for building trust and understanding. Techniques such as explainable AI (XAI) need to be further developed to make AI decision processes accessible and understandable to non-experts, especially in critical areas like credit scoring and risk assessment. Investments in research that aim to bridge the gap between technical explanations and human-understandable explanations should be prioritized [18].

6.4. Promoting stakeholder involvement

Ethical AI development must involve a wide range of stakeholders, including ethicists, consumer advocacy groups, technologists, and end users. This involvement can ensure that diverse perspectives are considered in the development of AI systems. Regular consultations, feedback mechanisms, and collaborative governance models can help integrate these diverse viewpoints effectively [19].

6.5. International cooperation

Given the global nature of financial services, international cooperation is crucial in developing and enforcing AI ethics guidelines. Efforts should be made to harmonize ethical standards across borders to prevent the fragmentation of regulations, which can lead to regulatory arbitrage. International bodies like the United Nations or the World Bank could play a pivotal role in facilitating this cooperation and ensuring that ethical standards are uniformly applied across countries [20].

7. Conclusion

The integration of artificial intelligence (AI) into the financial sector represents a paradigm shift in how decisions are made and services are delivered. While AI offers significant advantages in terms of efficiency and capability, it also introduces a complex array of ethical challenges that must be addressed to maintain trust and fairness in financial practices. This paper has explored the ethical dimensions of AI in financial decision-making, focusing on issues of bias and fairness, transparency, accountability, and privacy.

Through detailed case studies, we have seen how biases in AI can lead to unfair practices, how opaque algorithms can erode trust, and how the lack of clear accountability can complicate the regulatory environment. These challenges highlight the necessity for continuous evaluation and updating of ethical and regulatory frameworks as AI technologies evolve. The proposals for enhanced ethical frameworks discussed herein emphasize the importance of dynamic regulation, rigorous ethical audits, improved transparency, stakeholder involvement, and international cooperation.

In conclusion, as financial institutions increasingly adopt AI technologies, they must also commit to upholding ethical standards that protect and benefit all stakeholders involved. This commitment should be reflected in proactive measures that anticipate ethical dilemmas and address them before they escalate into larger issues. By fostering an environment of ethical awareness and responsibility, the financial sector can leverage AI to not only improve its operations but also enhance its accountability and public trust. This dual focus on innovation and ethics will be crucial for the sustainable and equitable growth of AI in finance.

Conflict of interest: The author declares no conflict of interest.

References

1. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. *Science Robotics*. 2017; 2(6). doi: 10.1126/scirobotics.aan6080
2. Hurley M, Adebayo J. Credit scoring in the era of big data. *Yale Journal of Law and Technology*. 2016; 18: 148-216.
3. Burrell J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*. 2016; 3(1): 205395171562251. doi: 10.1177/2053951715622512
4. Goodman B, Flaxman S. European Union Regulations on Algorithmic Decision Making and a “Right to Explanation.” *AI Magazine*. 2017; 38(3): 50-57. doi: 10.1609/aimag.v38i3.2741
5. Pasquale F. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press; 2015.
6. Martin K. Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*. 2018; 160(4): 835-850. doi:

- 10.1007/s10551-018-3921-3
7. Barocas S, Selbst AD. Big data's disparate impact. *California Law Review*. 2016; 104: 671.
 8. Castelvechi D. Can we open the black box of AI? *Nature*. 2016; 538(7623): 20-23. doi: 10.1038/538020a
 9. Doshi-Velez F, Kortz M, Budish R, et al. Accountability of AI under the law: The role of explanation. Berkman Klein Center Research Publication; 2017.
 10. Taylor L, Floridi L, van der Sloot B. Group privacy: New challenges of data technologies. *Philosophical Transactions of the Royal Society A*. 2017; 374(2083).
 11. OECD. *OECD Principles on Artificial Intelligence*. OECD Publishing, Paris; 2019.
 12. IEEE. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. IEEE; 2019.
 13. Babaei G, Giudici P, Raffinetti E. Explainable FinTech lending. *Journal of Economics and Business*. 2023; 125-126: 106126. doi: 10.1016/j.jeconbus.2023.106126
 14. Giudici P, Centurelli M, Turchetta S. Artificial Intelligence risk measurement. *Expert Systems with Applications*. 2024; 235: 121220. doi: 10.1016/j.eswa.2023.121220
 15. Babaei G, Giudici P, Raffinetti E. Explainable artificial intelligence for crypto asset allocation. *Finance Research Letters*. 2022; 47: 102941. doi: 10.1016/j.frll.2022.102941
 16. Zeng Y. Pathways to Convergence: A Future Envisioned by Ethical AI. *Journal of Future Technology*. 2019; 1(2): 89-95.
 17. Raji ID, Smart A, White RN, et al. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.
 18. Gunning D, Stefik M, Choi J, et al. XAI—Explainable artificial intelligence. *Science Robotics*. 2019; 4(37). doi: 10.1126/scirobotics.aay7120
 19. Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2018; 376(2133): 20180080. doi: 10.1098/rsta.2018.0080
 20. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. 2019; 1(9): 389-399. doi: 10.1038/s42256-019-0088-2

Review

Review of quantum algorithms for prediction of hazardous asteroids

Priya Pareshbhai Bhagwakar*, Chirag Suryakant Thaker, Hetal A. Joshiara

Department of Computer Engineering, L. D. College of Engineering, Ahmedabad 380015, Gujarat, India

* Corresponding author: Priya Pareshbhai Bhagwakar, 22cspri002@ldce.ac.in

CITATION

Bhagwakar PP, Thaker CS, Joshiara HA. Review of quantum algorithms for prediction of hazardous asteroids. *Computing and Artificial Intelligence*. 2024; 2(1): 1141. <https://doi.org/10.59400/cai.v2i1.1141>

ARTICLE INFO

Received: 13 March 2024

Accepted: 10 May 2024

Available online: 17 May 2024

COPYRIGHT



Copyright © 2024 by author(s).

Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.

<https://creativecommons.org/licenses/by/4.0/>

Abstract: Quantum computing (QC) and quantum machine learning (QML), two emerging technologies, have the potential to completely change how we approach solving difficult problems in physics and astronomy, among other fields. Potentially Hazardous Asteroids (PHAs) can produce a variety of damaging phenomena that put biodiversity and human life at serious risk. Although PHAs have been identified through the use of machine learning (ML) techniques, the current approaches have reached a point of saturation, signaling the need for additional innovation. This paper provides an in-depth examination of various machine learning (ML) and QML techniques for precisely identifying potentially hazardous asteroids. The study attempts to provide information to improve the efficiency and accuracy of asteroid categorization by combining QML techniques like deep learning with a variety of machine learning (ML) algorithms, such as Random Forest and support vector machines. The study highlights weaknesses in existing approaches, including feature selection and model assessment, and suggests directions for further investigation. The results highlight the significance of developing QML techniques to increase the prediction of asteroid hazards, consequently supporting enhanced risk assessment and space exploration efforts. Paper reviews might not be related because the study only looks at generic paper reviews.

Keywords: quantum computing; quantum machine learning; quantum algorithms; machine learning; deep learning; asteroids; prediction

1. Introduction

Small, stony objects called asteroids revolve around the sun. Asteroids with orbits that intersect the Earth's orbit are referred to as near-Earth asteroids (NEAs) [1]. If NEAs strike Earth, it could be dangerous for humans. For this reason, it's critical to locate and monitor NEAs that may have an impact on Earth. Different methods to predict potentially dangerous asteroids have been created by researchers in the past few years. These algorithms evaluate asteroid data using machine learning techniques to determine whether asteroids are potentially dangerous for Earth [2–6]. A review of some of the most promising algorithms for hazardous asteroid prediction will be conducted in this study.

This article explores the wide range of predictive algorithms that have been developed in reaction to this cosmic problem. We examine the advantages and disadvantages of these computational sentinels, ranging from the complex network of machine learning models that pick up on the universe's minute signals to reliable, age-old statistical methods. We shed light on the cutting-edge strategies that are influencing the direction of planetary defense as we make our way through the difficulties of asteroid classification and trajectory prediction.

To fully comprehend the complexities of hazardous asteroid prediction, multi-sensor data from both space-based and ground-based observatories must be integrated

with the technological components of the algorithms [7]. With a focus on models that achieve a careful balance between minimizing errors and early detection, the paper seeks to clarify the difficulties related to false positives and false negatives as we explore the approaches used. Traditional algorithms, while effective to a certain extent, may face limitations in handling the intricate nature of celestial dynamics. Through the use of quantum computing, the development of quantum machine learning (QML) algorithms presents a paradigm change that could fundamentally alter how we forecast hazardous asteroids [8].

This review embarks on an exploration of the integration of quantum machine learning algorithms into the domain of asteroid prediction. This study seeks to give an overview of the emerging field at the intersection of quantum and classical computing, highlighting the remarkable processing power and innovative methodology provided by quantum approaches.

2. Literature review

Emerging technologies like quantum computing (QC) and quantum machine learning (QML) have the power to completely change how we approach challenging issues in physics, mathematics, and other disciplines [9]. Potentially Hazardous Asteroids (PHAs) can cause significant harm to humans and biodiversity through wind blasts, overpressure shock, thermal radiation, cratering, seismic shaking, ejecta deposition, and even tsunamis. Based on their parameters, machine learning (ML) techniques have been used to identify potentially dangerous asteroids. Nevertheless, the existing methods have their limits, and the saturation point of the results has been reached. The purpose for using quantum machine learning is concluded in this section through individual reviews of a few studies. Review proceeds from machine learning techniques to quantum machine learning techniques.

According to the paper [2], the project's objective is to use various machine learning techniques to accurately identify which asteroids are dangerous. The classification of hazardous and non-hazardous asteroids was done using a variety of methods, with Random Forest demonstrating the highest accuracy. Before creating the models, the study pre-processed the data and eliminated any superfluous features from the dataset. Sixteen key criteria were chosen for classification in the study, including eccentricity, orbit uncertainty, absolute magnitude, and orbital period. To assess baseline accuracy and correct imbalance in the dataset, the dataset was divided into training (consisting of 3749 data) and testing (938 data) groups. Various models of classification were employed, including Naive Bayes, Support Vector Machine, Decision Tree, adaboost, and xgboost were implemented using Python libraries like numpy, scikit-learn, and xgboost. It has been observed that the random forest tree and xgbclassifier provides the most accurate and efficient answer with random forest exhibiting faster training time. Resulting 1334 of the forecasts were accurate, while 73 were off. By determining their accuracy, this research will assist with identifying whether or not the recently found near-Earth asteroids are hazardous. The following are the research gaps identified in this paper: The study concentrated on choosing 16 features for categorization, however there might be more pertinent features that improve the prediction of dangerous asteroids. Although the study analyzed several

machine learning algorithms, further insights into the predictive power could be gained from a more thorough comparative investigation of alternative feature selection strategies or model evaluation methodologies. The model's performance could be enhanced by exploring more advanced techniques for managing imbalanced datasets, as the study proposed dataset splitting as a solution. In order to evaluate the generalizability of the constructed models, the work could benefit from external validation using independent datasets. Understanding the variables impacting the classification of dangerous asteroids may need an in-depth examination of the interpretability of the machine learning models employed in this work. By filling in these possible study gaps, the findings' usefulness and robustness in machine learning algorithmic predictions of dangerous asteroids could be further enhanced.

The paper [3] talks about the examination of NASA's Nearest Earth Objects dataset and the value of innovation in space exploration. It goes into preprocessing data, visualization methods, classifying asteroids using machine learning algorithms into dangerous and non-hazardous groups, and the necessity of additional study to enhance the system. The approach involves collecting information from Kaggle's NASA - Nearest Earth Objects collection, which contains details on more than 20,000 asteroids and comets, including their physical and orbital properties. To categorize asteroids into hazardous and non-hazardous groups, various machine learning models, including Decision Tree, Logistic Regression, and Random Forest, were trained using the collected features. To find out how well the machine learning models performed in identifying the asteroids, evaluation criteria such as accuracy, F1-score, or precision were used. In the classification task, the Random Forest model performed better than the other models. The researchers proposed that more study be conducted in order to enhance the system by adding other data sources. This suggests that there might be unexplored datasets or information sources that offer insightful information for improving asteroidal classification and analysis. While the study employed Decision Tree, Logistic Regression, and Random Forest classifiers, advanced machine learning techniques should be further explored to improve classification efficiency and accuracy. Methods like ensemble approaches or deep learning may be used in this. The study included suggestions for the creation of a more reliable ensemble approach for classification. Ensemble approaches use several models to enhance prediction performance; future work may concentrate on refining these techniques to better classify objects that are closest to Earth. Future study in this area may be able to close these knowledge gaps and improve our understanding of asteroids and other space objects, which could result in more precise hazard estimates and deeper insights into the solar system.

In the study of Wheeler et al. [5], the research highlights the significance of precise forecasts in reducing the likelihood of asteroid collisions and highlights the potentially disastrous outcomes of such incidents. The study's hybrid methodology, which combines deep learning and machine learning classification models, is one of its main techniques. Utilizing an extensive collection of non-hazardous asteroids, the scientists trained a deep autoencoder to derive intricate representations from the information. Following the reconstruction of the dataset, Support Vector Machines, MLP, Logistic Regression, and KNN models were trained on it. The study's outcomes show how successful the suggested hybrid classifier model is, as evidenced by its high

accuracy, excellent precision, and recall ratings. The usefulness and affordability of the created method are demonstrated by a comparison with alternative models, such as an Artificial Neural Network (ANN) adjusted specifically for identifying PHAs. The outcomes show how stable and dependable the suggested hybrid classifier model is at correctly identifying asteroids that may pose a threat. All things considered, the article offers insightful information about the use of deep learning methods for forecasting potentially dangerous asteroids. The study's methodology, findings, and conclusions add to the continuing efforts to improve risk assessment and asteroid impact monitoring with the ultimate goal of protecting humanity from possible asteroid threats.

The paper [7] provides an introduction to asteroids, explaining that they are smaller moving objects in space and can pose a threat if they enter Earth's orbit. There are other categories of asteroids discussed, including composition-based C, S, and M and orbit-based dark C, brilliant S, and bright M. The hypothesis that a mountain-sized asteroid killed the dinosaurs and the difficulties of destroying asteroids when they become a hazard are also discussed in the paper. In addition, the authors go over the methods used today to identify asteroids, including as satellites, probes, and telescopes with advanced image processing systems. New approaches to detection are mentioned, including the use of nanosatellites. The paper also presents the idea of utilizing machine learning algorithms to detect asteroids, emphasizing the effectiveness of machine learning in contrast to conventional techniques like Astrometrica software. The execution of the machine learning process is covered in the section that details how similar flux value data are taken into account and how machine learning is applied to identify asteroids by creating linear equations. The verification of output values and the effectiveness of machine learning algorithms in asteroid detection are covered in the findings and conclusion section. The study comes to the conclusion that the convolution approach is quite effective at finding asteroids. The paper also highlights the potential for enhanced efficiency by taking Gaussian curves into consideration and eliminating latency in detection methods, as well as the future prospects of building prediction algorithms utilizing Python and libraries and the availability of data from websites such as Zoo Universe. The authors express their commitment to further developing more efficient algorithms for asteroid detection.

In the study of Carruba, Alibaae and Lucchini [10], supervised-learning hierarchical clustering algorithms are used to present a novel method for asteroid family identification. The work intends to improve the efficiency and accuracy of asteroid group identification by comparing the performance of these machine-learning algorithms with conventional techniques such as the Hierarchical Clustering Method (HCM). Understanding the dynamics and evolution of asteroids depends on comprehending asteroid families, which are groups of asteroids that share a common origin via collisions or rotational fission of a parent body. Compared to traditional techniques like HCM, the use of machine-learning algorithms has benefits like speed and efficacy in discovering new asteroid groups. The study assesses the machine-learning algorithm's accuracy, precision, and recall values; it shows a high accuracy rate (89.5%) and reliable retrieval of asteroid family members that have already been recognized. Furthermore, the study finds new asteroid families and clumps in areas where the machine-learning approach is used, demonstrating consistency in the

physical and taxonomic characteristics of the objects. The research concludes by highlighting the potential of machine-learning clustering techniques for asteroid family identification and demonstrating how quickly and accurately they may produce results. By demonstrating the advantages of applying advanced techniques to identify asteroid groups, this study advances the field of asteroid studies and ultimately advances our understanding of the dynamics and origins of asteroids in the solar system.

The paper [11] discuss on how quantum mechanics plays a critical role in quantum machine learning by utilizing the concepts of superposition and entanglement to carry out calculations in ways that classical computers are unable to. The study intends to categorize, evaluate, and identify various quantum machine learning algorithms and the computational quantum circuits, or ansatzs, that are used to implement them. The study emphasizes the use of quantum properties like superposition and entanglement to achieve computational advantages over classical methods, highlighting the significance of quantum mechanics in quantum machine learning. The difficulties and possibilities of quantum computing are covered, with a focus on developing quantum hardware to fully utilize quantum machine learning. The paper also lists a number of uses for quantum machine learning, including quantum neural networks, image classification, and quantum versions of conventional machine learning techniques. It also highlights the necessity of developing quantum hardware in order to get beyond present constraints and improve the efficiency of quantum algorithms. The authors also offer details regarding potential areas of future research: validating the effectiveness and performance of quantum machine learning algorithms on quantum hardware platforms by actual research and experimentation. Examining the difficulties with scalability that arise with quantum machine learning algorithms, particularly as dataset sizes and task complexity grow. investigating hybrid quantum-classical machine learning methods that effectively handle complicated issues by utilizing the advantages of both quantum and classical computers. Advances in theory and real-world applications could be merged through research on hybrid models and algorithms. In order to investigate novel applications of quantum machine learning in a variety of domains, including healthcare, finance, materials science, and cybersecurity, multidisciplinary research collaborations between quantum physicists, machine learning professionals, domain specialists, and industry partners are encouraged taking into account the potential ethical, societal, and legal ramifications of using quantum machine learning technology, including concerns about data privacy, security, bias reduction, and algorithmic decision-making transparency.

The paper [12] explore the field of quantum machine learning (QML) by presenting a unique framework that makes use of projection-valued measurements (PVM) to improve performance in multi-class classification tasks and get around scaling issues. The introduction lays the groundwork by going over the quick advancement of quantum computing technologies and how they might affect a number of industries, including machine learning. The study's motivation stems from the scaling problems that current QML techniques have, which led the authors to investigate the application of PVM to increase output dimension and enhance performance in scenarios involving multiple classes of categorization. The design of the suggested QML framework, which uses PVM to raise the output dimension 2^q from

q , where q is the number of qubits used, is described in length in the methods section. The framework performs better than the state-of-the-art approaches by using PVM, especially in multi-class classification applications. The efficacy and superiority of the framework over current methods are demonstrated by evaluating its scalability and accuracy on a variety of datasets, including Fashion MNIST, CIFAR10, and EMNIST-letters. A thorough examination of the PVM-based QML framework's performance on various datasets is given in the experimental assessment section. Benchmarking against QuantumNAS and other approaches demonstrates the accuracy and scalability benefits of the system. Furthermore, an ablation analysis of a probability amplitude regularizer is carried out, demonstrating its beneficial effects on training methods for particular datasets. The paper highlights the effectiveness of PVM-based QML for multi-class classification tasks and explores possible future research topics in the conclusion and future work section. The study's conclusion highlights the advantages of the suggested framework over current approaches and points the way for future research on quantum object detection with PVM and quantum reinforcement learning. Overall, the work offers a creative and organized approach to quantum machine learning, highlighting the advantages of using PVM to improve scalability and performance in challenging multi-class classification tasks. The outcomes of the experiments and the comparisons with advanced methods confirm the usefulness of the suggested framework, opening the door for improvements.

In-depth information about quantum tools and their uses in machine learning is presented in this study [13]. The authors explore the principles of quantum states, quantum registers, and qubits while emphasizing the distinctions between quantum and classical registers. Information extraction from quantum systems by measurement is one of the main ideas covered in the study. The authors highlight the probabilistic aspect of quantum computation by describing how measurements in the computational base might produce probabilistic results. Additionally, they discuss the concept of partial measurements in quantum registers, demonstrating the flexibility of quantum systems in managing complex data. In order to build quantum algorithms for machine learning tasks, the study explores the use of Quantum Query Algorithms (QQA) and Quantum Branching Programs (QBPs). The authors provide a preview of how quantum computing may improve machine learning algorithms by presenting these models, which show how quantum techniques may be used to speed up classification issues. In addition, the authors provide insight on how quantum algorithms are presented from a programming perspective, highlighting the significance of abstract linear representations in complexity analysis. They talk about tools for working with quantum registers, like amplitude sign inversion, which is important for algorithms like Grover's search algorithm. The paper discusses quantum SDKs and languages such as Project Q, Qiskit, and Microsoft Q# and investigates quantum algorithms for classification problems. All things considered, the study is a useful tool for practitioners and researchers who want to use quantum technologies in machine learning applications. It opens the door for further developments in this exciting field by providing a strong foundation for understanding the connection between machine learning and quantum computing.

Using a particular focus on classification algorithms and second part of paper [13], the paper [14] "On Quantum Methods for Machine Learning Problems Part II:

Quantum Classification Algorithms”, provides a comprehensive analysis of how machine learning tasks can be advanced through the use of quantum methodologies. The importance of quantum oracles, which are essential elements of quantum algorithms, is one of the major topics covered in the study. These oracles are essential for optimizing the performance of quantum algorithms and for enabling quantum computational operations. The article explores a number of quantum classification algorithms that make use of quantum subroutines to speed up machine learning classification. The goal of the authors’ integration of quantum approaches into supervised learning techniques is to increase classification task speed while maintaining overall classification outcome accuracy. Furthermore, it is possible that the paper investigates the relevance of Nearest Neighbor (NN) algorithms in classification issues. Since NN methods entail finding the nearest neighbor to a given data point for classification purposes, they are essential to pattern recognition and classification applications. The review might go over several NN algorithm variations and how to use quantum techniques to maximize these algorithms’ accuracy and efficiency in machine learning applications. All things considered, the thorough examination offered in the research illuminates the possibility that quantum methods could completely transform machine learning, especially when it comes to classification algorithms. A wide range of viewpoints and levels of experience are brought to the review by the cooperative work of researchers from different backgrounds, which provides insightful information about how to apply quantum-enhanced machine learning technologies for better classification results.

The main goal of the paper [8] is to forecast asteroids’ level of danger by utilizing quantum machine learning (QML). The study describes the shortcomings of existing machine learning methods for asteroid risks prediction and suggests a QML-based method based on Pegasus QSVC and Variational Quantum Circuits (VQC). The authors give a thorough overview of the significance of space research as well as the possible dangers associated with potentially hazardous asteroids (PHAs). They talk about how better techniques are needed to categorize asteroids according to their characteristics in order to lessen the possibility of negative effects on Earth. The study describes the problem statement, suggested methodology, data collection, dataset description, preprocessing steps, experimental setup, evaluation metrics, results, and analysis. It also encompasses organizational structure, related works in the field of QML and asteroid hazard prediction, and research contributions. The performance of QML algorithms and conventional machine learning algorithms was examined by the authors through experiments carried out with quantum resources. They assessed the models’ recall, accuracy, precision, F1-score, and confusion matrix to ascertain how well they classified potentially dangerous asteroids. The suggested QML-based strategy beat conventional machine learning methods, as evidenced by the findings, which had high F1-score and an accuracy. The authors talked about how the models’ performance was affected by hyperparameters such the regularization parameter C . The authors emphasized in the discussion section the difficulties in understanding QML models because of the intricate structure of quantum states. They also highlighted the significance of accuracy and recall numbers in assessing model performance and offered insights into the operation of various QML algorithms. The report concludes by suggesting that future research should incorporate image-based

analysis to improve the precision and effectiveness of the suggested methodology. The authors also touch on the possible uses of QML in astronomy and emphasize the necessity for more study in 3D space grouping of objects with similar characteristics. The study offers a thorough analysis of the suggested QML-based method for asteroid hazard prediction, including information on the methodology, findings, and implications for further space science and quantum machine learning research.

3. Methodology and result

This study does not offer a methodology or results because it is a generic review of all research publications that are mentioned. Researchers can utilize the study's positive and negative points of papers to help them determine which methodologies to apply in their own research.

4. Conclusion

This study's conclusion emphasizes the vital role that advanced technologies—specifically, quantum computing and quantum machine learning—have in expanding our knowledge and capacity to anticipate potentially dangerous asteroids (PHAs). In comparison to conventional methods, the research shows notable gains in accuracy and efficiency by utilizing several machine learning and quantum-based techniques. The review suggests important areas for future research through thorough experimentation and analysis, such as feature selection optimization, model interpretability, and integration of novel approaches for addressing imbalanced datasets. The results highlight how crucial it is to keep researching and developing novel QML approaches in order to tackle the intricate problems [15]. In general, this study advances the current efforts to utilize state-of-the-art technologies for space exploration and risk reduction, opening the door to a more effective and better-informed strategy for handling asteroid dangers.

5. Future perspectives

Research and cooperation will be essential to fully utilizing QML's potential in protecting Earth from cosmic hazards that stay in the vastness of space as quantum technology advances. The pursuit of expanding scientific knowledge, thoughtful assessment of ethical ramifications, and ongoing investigation of quantum possibilities are all necessary steps on the path to a more secure future.

Conflict of interest: The authors declare no conflict of interest.

References

1. Malakouti SM, Menhaj MB, Suratgar AA. Machine Learning Techniques for Classifying Dangerous Asteroids. *MethodsX*. 2023; 11: 102337. doi: 10.2139/ssrn.4452444
2. Babu RM, Devi AG, Aakanksha J, Naga JK. Hazardous Asteroid Prediction Using Machine Learning, 2nd. International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies. 2023. doi: 10.1109/ViTECoN58111.2023.10157937.
3. Khajuria D, Sharma M, Sharma N, et al. Classification and Comparative Analysis of Earth's Nearest Objects using Machine Learning Models, 10th. International Conference on Computing for Sustainable Global Development; 2023.

4. Wheeler L, Dotson J, Aftosmis M, et al. Risk assessment for asteroid impact threat scenarios. *Acta Astronautica*. 2024; 216: 468-487. doi: 10.1016/j.actaastro.2023.12.049
5. Ranaweera RN, Fernando T. Prediction of Potentially Hazardous Asteroids using Deep Learning. 2022 2nd International Conference on Advanced Research in Computing (ICARC). Published online February 23, 2022. doi: 10.1109/icarc54489.2022.9753945
6. Aftosmis MJ, Mathias DL, Tarano AM. Simulation-based height of burst map for asteroid airburst damage prediction. *Acta Astronautica*. 2019; 156: 278-283. doi: 10.1016/j.actaastro.2017.12.021
7. Chaitanya Prasad LVR, Reddy TAS, Kashi B. Asteroid Detection using Machine Learning Algorithm. *Communications of BAO*. 2020; 67(2).
8. Bhavsar R, Jadav NK, Bodkhe U, et al. Classification of Potentially Hazardous Asteroids Using Supervised Quantum Machine Learning. *IEEE Access*. 2023; 11: 75829-75848. doi: 10.1109/access.2023.3297498
9. Bhat HA, Khanday FA, Kaushik BK, et al. Quantum Computing: Fundamentals, Implementations and Applications. *IEEE Open Journal of Nanotechnology*. 2022; 3: 61-77, 2022, doi: 10.1109/OJNANO.2022.3178545
10. Carruba V, Aljbae S, Lucchini A. Machine-learning identification of asteroid groups. *Monthly Notices of the Royal Astronomical Society*. 2019; 488(1): 1377-1386. doi: 10.1093/mnras/stz1795
11. García DP, Cruz-Benito J, García-Peñalvo FJ. Systematic Literature Review: Quantum Machine Learning and its Applications. Available online: <https://arxiv.org/abs/2201.04093> (accessed on 2 February 2024).
12. Yun WJ, Baek H, Kim J. Projection Valued Measure-based Quantum Machine Learning for Multi-Class Classification. *ADS Abstract Service*. 2022; 2. doi:10.48550/arXiv.2210.16731
13. Ablayev F, Ablayev M, Huang JZ, et al. On quantum methods for machine learning problems part I: Quantum tools. *Big Data Mining and Analytics*. 2020; 3(1): 41-55. doi: 10.26599/bdma.2019.9020016
14. Ablayev F, Ablayev M, Huang JZ, et al. On quantum methods for machine learning problems part II: Quantum classification algorithms. *Big Data Mining and Analytics*. 2020; 3(1): 56-67. doi: 10.26599/bdma.2019.9020018
15. IBM. Quantum Decade. Available online: <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/quantum-decade> (accessed on 15 February 2024).

Review

Revolutionizing Neurosurgery and Neurology: The transformative impact of artificial intelligence in healthcare

Habib Hamam^{1,2,3,4}

¹ Faculty of Engineering, Uni de Moncton, Moncton, NB E1A3E9, Canada; habib.hamam@umoncton.ca

² Hodmas University College, Taleh Area, Mogadishu, Somalia

³ School of Electrical Engineering, University of Johannesburg, Johannesburg 2006, South Africa

⁴ Bridges for Academic Excellence, Tunis, Centre-Ville 1001, Tunisia

CITATION

Hamam H. Revolutionizing Neurosurgery and Neurology: The transformative impact of artificial intelligence in healthcare. *Computing and Artificial Intelligence*. 2024; 2(1): 416.
<https://doi.org/10.59400/cai.v2i1.416>

ARTICLE INFO

Received: 14 December 2023

Accepted: 2 January 2024

Available online: 16 January 2024

COPYRIGHT



Copyright © 2024 by author(s).
Computing and Artificial Intelligence is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.
<https://creativecommons.org/licenses/by/4.0/>

Abstract: The integration of artificial intelligence (AI) has brought about a paradigm shift in the landscape of Neurosurgery and Neurology, revolutionizing various facets of healthcare. This article meticulously explores seven pivotal dimensions where AI has made a substantial impact, reshaping the contours of patient care, diagnostics, and treatment modalities. AI's exceptional precision in deciphering intricate medical imaging data expedites accurate diagnoses of neurological conditions. Harnessing patient-specific data and genetic information, AI facilitates the formulation of highly personalized treatment plans, promising more efficacious therapeutic interventions. The deployment of AI-powered robotic systems in neurosurgical procedures not only ensures surgical precision but also introduces remote capabilities, mitigating the potential for human error. Machine learning models, a core component of AI, play a crucial role in predicting disease progression, optimizing resource allocation, and elevating the overall quality of patient care. Wearable devices integrated with AI provide continuous monitoring of neurological parameters, empowering early intervention strategies for chronic conditions. AI's prowess extends to drug discovery by scrutinizing extensive datasets, offering the prospect of groundbreaking therapies for neurological disorders. The realm of patient engagement witnesses a transformative impact through AI-driven chatbots and virtual assistants, fostering increased adherence to treatment plans. Looking ahead, the horizon of AI in Neurosurgery and Neurology holds promises of heightened personalization, augmented decision-making, early intervention, and the emergence of innovative treatment modalities. This narrative is one of optimism and collaboration, depicting a synergistic partnership between AI and healthcare professionals to propel the field forward and significantly enhance the lives of individuals grappling with neurological challenges. This article provides an encompassing view of AI's transformative influence in Neurosurgery and Neurology, highlighting its potential to redefine the landscape of patient care and outcomes.

Keywords: artificial intelligence; neurosurgery; neurology; medical imaging; personalized treatment

1. Introduction

In the ever-evolving landscape of healthcare, the integration of artificial intelligence (AI) stands as a transformative force, particularly within the intricate domains of Neurosurgery and Neurology. As the nexus between technology and medical science advances, the potential of AI to revolutionize patient care, diagnostics, and treatment strategies becomes increasingly apparent. This article embarks on an exploration of the profound impact of AI on Neurosurgery and Neurology, navigating through key aspects that redefine the paradigms of modern medical practice.

At the core of this exploration is the remarkable accuracy demonstrated by AI-

driven algorithms in deciphering intricate medical imaging data. This not only expedites the diagnosis of neurological conditions but also heralds a new era of precision in medical decision-making. Beyond diagnostic advancements, the capacity of AI to analyze patient-specific data and genetic information opens doors to highly personalized treatment plans, offering a beacon of hope for more effective therapeutic interventions.

A pivotal milestone in this transformative journey is the advent of AI-powered surgical robots, seamlessly integrating precision and remote capabilities into neurosurgical procedures. This groundbreaking innovation substantially reduces the margin for human error, marking a paradigm shift in the realm of surgical interventions for neurological disorders. Complementing this advancement, machine learning models enter the arena of predictive analytics, providing valuable insights into disease progression and optimizing resource allocation for enhanced patient care.

Wearable devices, empowered by AI, emerge as another cornerstone, facilitating continuous monitoring of patients' neurological status. This not only enables early intervention in chronic conditions but also empowers patients to actively manage their health. As AI accelerates the drug discovery process through the analysis of vast datasets, the horizon glimmers with the potential for groundbreaking therapies addressing neurological disorders.

Integral to this narrative is the role played by AI-powered chatbots and virtual assistants, dynamically engaging with patients to offer information, support, and timely reminders. This interactive relationship enhances patient engagement, fosters adherence to treatment plans, and contributes to an overall improved patient experience.

Structured across seven sections, this article systematically unfolds the multifaceted impact of AI in Neurosurgery and Neurology. From the augmentation of diagnostic accuracy and the formulation of personalized treatment plans to the realm of predictive analytics, neurological monitoring, and the frontier of drug discovery, each section illuminates a crucial facet of this transformative journey. As the narrative reaches its zenith in a thoughtful conclusion, the article provides a holistic perspective on the collaborative future of AI and healthcare professionals, ultimately reshaping the landscape of neurological medicine. In alignment with the commitment to rigorous systematic review standards, this exploration adheres to the PRISMA guidelines for systematic reviews and meta-analyses, ensuring transparency and robustness in the review process.

2. AI in diagnostic imaging

2.1. Revolutionizing medical imaging

AI has emerged as a transformative force in diagnostic imaging, revolutionizing the analysis of complex medical data efficiently. Utilizing computerized algorithms, particularly in computer-aided diagnostics, has shown promising results in detecting and quantifying various clinical conditions, demonstrating remarkable accuracy, sensitivity, and specificity in identifying small radiographic abnormalities [1].

To adhere to the PRISMA guidelines for systematic reviews and meta-analyses, it is imperative to acknowledge potential biases in the assessment of AI performance

in imaging studies. Current evaluations often focus on lesion detection, potentially overlooking crucial aspects such as the type and biological aggressiveness of a lesion. Therefore, the article advocates for a shift towards consistent selection of clinically meaningful endpoints, including survival, symptoms, and the need for treatment, ensuring a more comprehensive evaluation of AI's impact on patient outcomes [2].

2.2. Speed and efficiency in diagnosis

AI's notable advantage in diagnostic imaging lies in its ability to enhance speed and efficiency in the diagnostic process. Deep learning algorithms have demonstrated superior accuracy and sensitivity in identifying imaging abnormalities, surpassing traditional radiological methods, as seen in mammography [3].

However, it is essential to caution against potential pitfalls, such as the detection of subtle changes with indeterminate significance. The challenge lies in distinguishing between benign abnormalities and clinically meaningful lesions. Therefore, the article stresses the importance of training AI algorithms to make this distinction, mitigating the risk of increased false positives and scenarios where AI findings lack association with meaningful outcomes [4].

In the evolving landscape of AI in diagnostic imaging, adhering to clinically relevant endpoints, rigorous validation, and a focus on patient outcomes are crucial for harnessing the full potential of this technology. Future sections will delve into specific clinical applications, challenges, and ethical considerations in the integration of AI in Neurosurgery and Neurology.

2.3. Precision and early detection

Precision, particularly in medical imaging and explainable AI (XAI), plays a pivotal role in ensuring the accuracy and reliability of AI systems in predicting or diagnosing conditions. In healthcare, where the consequences of AI-informed decisions are substantial, achieving high precision is paramount. Metrics like positive predictive value (PPV), measuring the accuracy of positive predictions, become instrumental in maintaining precision.

The importance of precision in medical imaging is underscored by its direct influence on patient outcomes. Precision ensures that positive predictions are dependable, curbing false positives that may lead to unnecessary medical interventions and associated risks. To enhance early detection capabilities, the application of explainable AI methods, especially saliency-based approaches, proves crucial [5,6].

The nexus between precision and early detection amplifies the efficacy of AI systems in medical imaging, offering reliable and timely insights into potential health conditions.

2.4. Addressing false positives and negatives

Mitigating false positives and negatives is of paramount importance for fostering the reliability and trustworthiness of AI systems in medical imaging. False positives can create undue concern and may lead to invasive procedures, while false negatives can delay necessary interventions. Explainable AI, especially saliency-based methods like Grad-CAM [5,6], emerges as a valuable tool in addressing these issues.

Explainable AI’s spotlight on salient features in medical images contributing to a prediction empowers clinicians to understand the reasons behind prediction errors. This understanding is pivotal for refining and enhancing AI models, mitigating both types of errors.

Insights garnered from explainable AI guide the formulation of targeted strategies to counter false positives and negatives. If specific features consistently lead to false positives, model adjustments can be made to mitigate their impact. The iterative refinement process, steered by interpretable explanations, is indispensable for erecting robust and reliable AI systems in medical imaging.

2.5. Chest X-ray interpretation and viral infections

In the realm of chest X-ray interpretation, AI transcends its conventional role of lung nodule detection and prioritizing critical cases. Notably, AI plays a pivotal role in identifying viral infections, exemplified by its application in detecting COVID-19 [7]. This functionality proved particularly valuable during the COVID-19 pandemic, where swift identification was paramount to curbing the virus’s spread.

Identifying Viral Infections with AI Algorithms: AI algorithms have undergone training to discern specific patterns associated with viral infections, including COVID-19, in chest X-rays and CT scans [8]. These algorithms excel at recognizing characteristic lung abnormalities induced by the virus, such as ground-glass opacities and bilateral pneumonia. By flagging these distinctive patterns, AI becomes an invaluable tool for radiologists in early detection, facilitating prompt isolation and treatment of affected patients [9,10].

Collaborative approach—AI and radiologists in chest X-ray interpretation: Integrating AI’s prowess in identifying viral infections, like COVID-19, with its other functionalities—such as lung nodule detection and case prioritization—amplifies the diagnostic capabilities of radiologists. This collaborative synergy enhances patient care not only in emergency settings but also across various medical scenarios [11,12].

In **Figure 1**, a significant breakthrough in COVID-19 classification is showcased, employing a diverse set of well-established pre-trained deep learning architectures, including ResNet101, ResNet50v1, ResNet50v2, RDenseNet101, and MobileNet [13]. This study meticulously followed a standardized methodology, initiating with chest image preprocessing to meticulously curate the dataset for subsequent analysis. The utilization of transfer learning was pivotal in constructing a robust classification model.

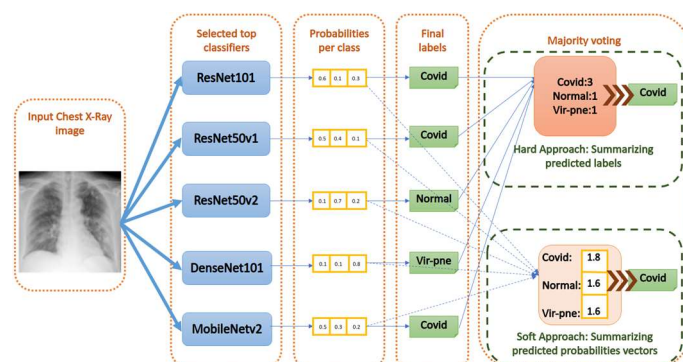


Figure 1. The ensemble classification model [13].

Methodology and testing approaches: The evaluation of various voting approaches, including two majority voting methods (hard and soft), marked a crucial phase in the study. Notably, the majority voting methods required no additional training; they were directly tested on the dataset. Conversely, the other three methods underwent a training phase on the validation dataset before being subjected to testing on the dedicated test dataset.

Training and testing strategies: The two majority voting methods, as revealed in the experimental section, demonstrated clear superiority over alternative approaches. Their effectiveness was substantiated through rigorous testing, positioning them as the optimal strategy for COVID-19 diagnosis from chest X-ray images. **Figure 1** serves as a visual representation of these two exemplary methods.

Majority voting methods—Hard and soft approaches: The majority voting methods, both hard and soft approaches, emerged as frontrunners in the experimental comparison. Their selection as the preferred strategy for COVID-19 diagnosis is supported by compelling evidence. The hard and soft majority voting methods, detailed in **Figure 1**, encapsulate the culmination of the study's findings.

This pioneering approach not only highlights the importance of ensemble learning but also underscores the practical efficacy of majority voting methods in enhancing the accuracy of COVID-19 diagnosis from chest X-ray images. The robustness of these methods positions them as valuable tools in the ongoing efforts to improve diagnostic capabilities in the field of medical imaging [13].

During the fine-tuning process, specific top layers of the frozen model, utilized for feature extraction, were unfrozen. This enabled training not only of the newly added section of the model (specifically, the fully-connected classifier) but also the previously frozen top layers.

The concept of “freezing” in this context denotes the decision to refrain from updating the weights of these layers during the model training phase for new tasks. The focus lay exclusively on updating the new layer or making modifications to the weights within that layer.

The collaborative nature of AI's role in assisting radiologists is evident. Radiologists contribute their expertise and clinical judgment, while AI complements their capabilities by providing support in highlighting anomalies, prioritizing cases, and furnishing comprehensive context [12]. This collaborative approach translates to heightened diagnostic accuracy, improved workflow efficiency, and ultimately superior patient care in the realm of medical imaging.

2.6. Quantitative analysis for informed decision-making

Quantitative analysis of AI models in medical imaging entails evaluating their performance through metrics and measurements. This step is critical for informed decision-making by healthcare professionals and researchers. Metrics such as sensitivity, specificity, and area under the curve (AUC) furnish a quantitative foundation for assessing the model's effectiveness.

In the domain of explainable AI, quantitative analysis transcends conventional performance metrics to encompass interpretability and explainability assessment. Metrics evaluating interpretability may include the clarity of generated saliency maps,

the comprehensibility of feature attributions, and the alignment of model explanations with clinical knowledge [5,14]. These metrics contribute to the holistic evaluation of an AI system's utility in a medical setting.

Quantitative analysis of explainable AI methods themselves holds equal significance. Researchers can gauge the efficacy of saliency-based approaches by comparing the agreement between model-identified salient features and clinically relevant regions in medical images. Additionally, user studies and feedback from healthcare professionals provide quantitative insights into the utility of XAI in supporting decision-making [5,14].

2.7. Collaborative approach: AI and radiologists

The collaborative synergy between AI and radiologists heralds a new era in medical imaging. This partnership seeks to amalgamate the analytical prowess of AI with the interpretive acumen of radiologists, aiming for enhanced diagnostic accuracy and efficiency.

In this collaborative paradigm, explainable AI (XAI) assumes a pivotal role, offering transparency into AI decision-making processes. The integration of XAI methods, such as Captum [15] and Densely Connected Convolutional Networks (DenseNet) [16], empowers radiologists to understand the features influencing AI predictions. This interpretability fosters trust and facilitates a seamless collaboration between human experts and AI systems.

Furthermore, the MedMNIST Classification Decathlon [17] serves as a benchmark, illuminating the potential of lightweight AutoML in medical image analysis. As radiologists navigate the intricacies of AI-assisted diagnostics, benchmarks like MedMNIST provide a standardized evaluation framework, ensuring the effectiveness and reliability of collaborative models.

Addressing the inherent variability among radiologists, studies such as Xie et al.'s work [18] on inter- and intraobserver variation in detecting abnormal parenchymal lung changes underscore the need for collaborative refinement. By leveraging AI, radiologists can benefit from consistent decision support, mitigating discrepancies and augmenting diagnostic accuracy.

The advent of tools like CheXplain [19] exemplifies the commitment to empowering physicians through explainable medical imaging analysis. These tools enable clinicians to explore and comprehend AI-driven analyses, fostering a collaborative environment where both AI and radiologists contribute synergistically to patient care.

2.8. Big Data insights in Neurosurgery and Neurology

In the realm of Neurosurgery and Neurology, the integration of Big Data ushers in a transformative wave of insights and advancements. The burgeoning datasets, coupled with advanced analytics, propel the understanding and treatment of neurological disorders to unprecedented heights.

One significant contribution is the Clinical Explainability Failure (CEF) and Explainability Failure Ratio (EFR) framework [20]. This framework, applied in the context of classification algorithms, introduces a paradigm shift in how we validate

and understand the decisions made by AI models. In Neurosurgery and Neurology, where decisions carry profound implications, such frameworks become invaluable for ensuring the reliability and interpretability of AI-driven insights.

The work of Jacovi and Goldberg [20] on interpretability in Natural Language Processing (NLP) systems offers valuable insights applicable to neurology. As we navigate the complexities of neurological data, defining and evaluating the faithfulness of AI-generated insights becomes paramount. This work prompts a thoughtful approach to ensuring that AI-driven analyses faithfully align with clinical knowledge, a critical consideration in the context of neurosurgical interventions.

2.9. Summary and literature review outcome

In the exploration of AI's impact on diagnostic imaging, the paper delves into its revolutionary role in analyzing complex medical data, highlighting the promise it holds for detecting clinical conditions [1]. The emphasis is on a shift towards clinically meaningful endpoints to comprehensively evaluate AI's impact on patient outcomes [2]. Notably, AI enhances speed and efficiency in diagnosis through deep learning algorithms, showcasing superior accuracy in identifying imaging abnormalities [3]. The cautionary note is on the need for AI to distinguish between benign abnormalities and clinically meaningful lesions to avoid increased false positives [4]. Precision in medical imaging, coupled with explainable AI methods, proves crucial for reliable predictions and early detection [5,6].

The article also addresses the significance of mitigating false positives and negatives in medical imaging, leveraging explainable AI to understand and refine model predictions [5,6]. In the context of chest X-ray interpretation, AI extends beyond lung nodule detection to identifying viral infections, particularly exemplified in its role during the COVID-19 pandemic [7]. The collaborative approach between AI and radiologists, showcased in **Figure 1**, demonstrates the efficacy of majority voting methods for COVID-19 diagnosis from chest X-ray images [13].

Quantitative analysis for informed decision-making in medical imaging involves evaluating AI models using metrics such as sensitivity, specificity, and AUC [14]. Explainable AI's quantitative analysis extends to interpretability metrics, aligning model explanations with clinical knowledge [14]. The collaborative synergy between AI and radiologists, facilitated by explainable AI methods like Captum and DenseNet, enhances diagnostic accuracy [15,16]. Benchmarking, as seen in the MedMNIST Classification Decathlon, ensures standardized evaluation of collaborative models in medical imaging [17].

In the realm of Neurosurgery and Neurology, the integration of Big Data introduces the Clinical Explainability Failure framework, offering insights into the validation and understanding of AI-driven decisions [20]. The framework's application in classification algorithms ensures reliability and interpretability in the context of neurosurgical interventions [20]. Insights from the work on interpretability in Natural Language Processing systems contribute to ensuring AI-driven analyses align faithfully with clinical knowledge in neurology [20].

Overall, the synthesis of literature provides a comprehensive overview of AI's transformative role in diagnostic imaging and neurosurgery/neurology, emphasizing

the need for precision, interpretability, and collaborative approaches for optimal patient care and outcomes. This systematic review adheres to the PRISMA guidelines, ensuring transparency and rigor in the review process.

3. Clinical applications of AI in neurological conditions

The integration of AI into clinical practice has marked a transformative shift in the diagnosis and management of neurological conditions, with a focus on neurodegenerative diseases, vascular conditions, brain tumors, and emergency cases. This section adheres to the PRISMA guidelines for systematic reviews and meta-analyses, ensuring transparency and rigor in presenting the literature.

3.1. Neurodegenerative diseases—Alzheimer’s and Parkinson’s

Neurodegenerative diseases, specifically Alzheimer’s and Parkinson’s, present significant challenges in diagnosis and treatment [21]. AI serves as a beacon of hope, offering innovative solutions for early detection and personalized care.

Venugopal et al. [19] and McKinney et al. [22] contribute to the understanding of Clinical Explainability Failure (CEF) and Explainability Failure Ratio (EFR), providing a framework for validating AI algorithms. In the context of neurodegenerative diseases, this framework ensures accurate predictions and transparent, interpretable insights into the decision-making process of AI models.

The work of Xie et al. [18] on CheXplain, an AI-enabled medical imaging analysis tool, exemplifies the potential in neurodegenerative disease diagnostics. Tools like CheXplain empower healthcare professionals to make informed decisions regarding the management of Alzheimer’s and Parkinson’s disease.

3.2. Vascular conditions—Aneurysms and abnormalities

AI plays a crucial role in the early detection and characterization of vascular conditions, such as aneurysms and abnormalities. In their works, Huang, et al. [15] and Lapuschkin et al. [23] contribute to the understanding of model interpretability, with the Captum library and Densely Connected Convolutional Networks (DenseNet) serving as valuable tools for unraveling the intricacies of AI predictions.

The application of explainable AI (XAI) is particularly relevant in vascular conditions, where transparency in decision-making is imperative. In the study of Huang et al. [15], Captum provides unified and generic model interpretability library, aiding clinicians in understanding the salient features influencing AI predictions related to aneurysms and vascular abnormalities.

3.3. Brain tumors—Glioblastoma and early detection

The role of AI in the early detection and characterization of brain tumors, including glioblastoma, holds profound implications for patient outcomes. Al-Khawari et al. [17] introduce the MedMNIST Classification Decathlon as a benchmark for lightweight AutoML in medical image analysis, offering insights into the potential of AI in neuro-oncology.

Additionally, the collaborative approach between AI and radiologists, as discussed in section 2.6, is highly relevant in the context of brain tumor diagnostics.

Tools like CheXplain [19] can be adapted to provide interpretable insights into AI-driven analyses of brain imaging, facilitating collaborative decision-making in the early detection and management of brain tumors.

Example application—Brain tumor detection:

In a recent investigation [24], an openly accessible CE-MRI dataset comprising images from 233 distinct patients with brain tumors was employed. The dataset, spanning 3062 MRI images (**Figure 2**), was categorized into three specific types of brain tumors: gliomas, meningiomas, and pituitary tumors.

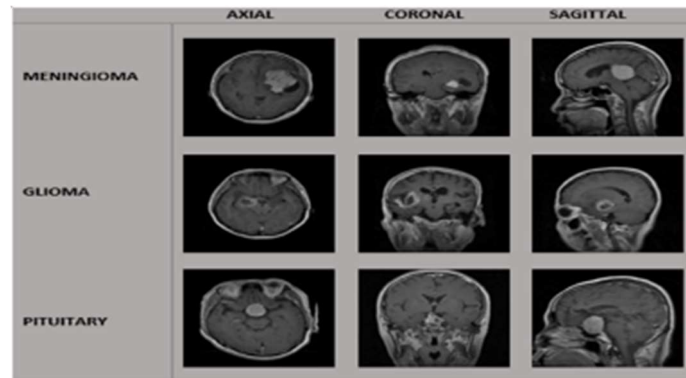


Figure 2. Three different tumors (meningioma, glioma, and pituitary tumor) in three different views [24].

To curate the dataset for the proposed model, the tumor regions underwent manual delineation by experienced radiologists. The dataset was normalized to .jpg format for further processing, and the images were resized to dimensions of 224×224 pixels. The dataset was partitioned into training (70%) and testing (30%) subsets, with all 3064 brain tumor images utilized in the experiments.

The authors conducted a comparative analysis between pre-trained transfer learning models and the proposed hybrid DeepTumorNet model. The fine-tuning classification method is elucidated in **Figure 3**.

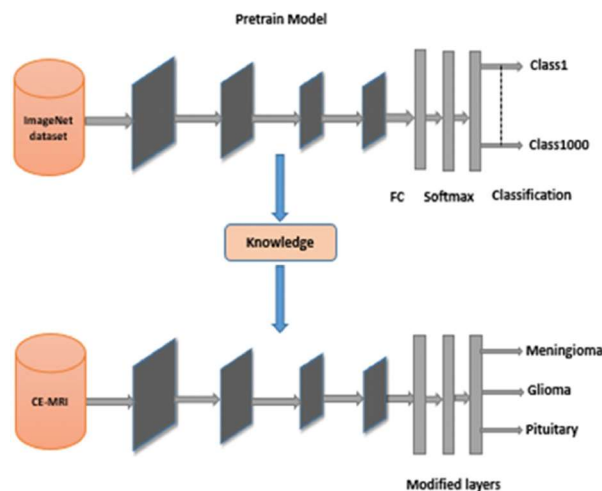


Figure 3. Diagram of transfer learning-based classification [24].

To curate the dataset for the proposed model, the tumor regions within the MRI

images underwent manual delineation by three experienced radiologists. Originally in .mat format, the dataset was normalized to .jpg format for further processing. Additionally, the images were resized to dimensions of 224×224 pixels to align with the input size of the proposed deep learning model and other pre-trained models. Notably, for Dark-Net19 images, resizing was conducted to dimensions of 256×256 pixels.

In terms of the experimental setup, the dataset was partitioned into two subsets: 70% for training and 30% for testing. All 3064 brain tumor images were utilized in the experiments. Approximately 2146 images (495 meningiomas, 652 pituitary tumors, and 999 gliomas) were employed for training the models, while the remaining 918 images (213 meningiomas, 278 pituitary tumors, and 428 gliomas) were reserved for testing.

Raza et al. [24] conducted a comparative analysis between several pre-trained transfer learning models and the proposed hybrid DeepTumorNet model. The primary objective of this experiment was to assess the efficacy of the hybrid model in classifying brain tumors (BTs). Simultaneously, they contrasted the evaluation performance of DeepTumorNet with nine classical deep learning (DL) models, namely ExceptionNet, MobileNetv2, SqueezeNet, ShuffleNet, DenseNet, ResNet50, MobileNetv2, DarkNet-53, ResNet101, and AlexNet. All these DL models were configured using transfer learning and trained on the ImageNet database.

In the adaptation process, the last three layers of each pre-trained model were modified to align with the target number of classes. The fully connected (FC) layer in each model was eliminated, and a new FC layer with an output size of three was introduced using the fine-tuning strategy, considering the three output classes. Each model had a distinct input size, with GoogLeNet's input size set at 224×224 , and others varying, such as 227×227 for SqueezeNet. The fine-tuning classification method is elucidated in **Figure 3**. For the deep learning training, the dataset's images were partitioned at a ratio of 30% for testing and 70% for training. This division aimed to ensure robust results and efficient performance for the deep neural networks.

3.4. Emergency cases—Stroke and trauma

In emergency neurological cases, such as stroke and trauma, timely and accurate diagnosis is critical. Kriegsmann et al. [25] introduce the concept of precision in medical imaging, emphasizing its paramount importance in the context of AI applications in emergency cases.

Quantitative analysis for informed decision-making, as discussed in section 2.6, becomes crucial in emergency neurological conditions. Metrics such as sensitivity, specificity, and area under the curve (AUC), along with interpretability metrics, ensure that AI models not only perform with high accuracy but also provide transparent insights into their decision processes, aiding clinicians in making informed decisions in time-sensitive situations.

3.5. AI-Enabled early detection in neurological pathologies

The integration of AI in Neurosurgery and Neurology heralds a groundbreaking era in the timely identification of anomalies and conditions. This transformative

capability holds immense promise, particularly in addressing conditions such as brain tumors, aneurysms, vascular abnormalities, and neurodegenerative diseases.

Early detection of brain tumors:

- Traditional challenges: The gradual development of brain tumors often conceals symptoms in their early stages, leading to late detection and potentially inoperable states.
- AI's role: AI algorithms, trained on extensive datasets of brain images, excel in recognizing subtle changes in brain tissue before symptoms emerge. This early identification facilitates timely intervention, enabling surgeons to address tumors at smaller sizes, thereby enhancing surgical outcomes and minimizing damage to healthy brain tissue [24].

Early detection of aneurysms:

- Traditional challenges: Aneurysms, vulnerable areas of blood vessels, may remain unnoticed until rupture, posing life-threatening risks.
- AI's role: AI's analysis of medical images, including angiograms or CT scans, allows for the identification of small aneurysms. Tracking changes in size and shape over time, AI alerts healthcare providers to the need for monitoring or intervention before rupture occurs.

Early detection of vascular abnormalities:

- Traditional challenges: Conditions like arteriovenous malformations (AVMs) or vascular stenosis are difficult to diagnose early due to their often asymptomatic nature.
- AI's role: AI algorithms identify subtle irregularities in blood vessels, flagging them for further evaluation [26]. Early detection enables monitoring and implementation of treatments like embolization or stent placement before severe complications arise.

Early detection of neurodegenerative diseases:

- Traditional challenges: Slow progression characterizes diseases like Alzheimer's or Parkinson's, with symptoms becoming evident only in later stages.
- AI's role: AI analyzes brain imaging data and clinical records to detect early signs of neurodegenerative diseases. Early identification leads to interventions such as lifestyle modifications or drug therapies, potentially slowing disease progression and improving the patient's quality of life [27].

Example application—Glioblastoma detection: Glioblastoma, an aggressive brain cancer, undergoes rapid growth. AI algorithms scrutinize serial MRI scans, identifying subtle changes in tumor size, shape, and vascularity. Early detection prompts considerations of interventions like surgery, radiation therapy, or targeted drug treatments, significantly improving treatment success rates and enhancing the patient's quality of life [28].

Thus, AI's prowess in early detection within Neurosurgery and Neurology is revolutionary. Empowering healthcare professionals to intervene at the earliest stages, AI not only increases the likelihood of successful treatment but also allows for less invasive interventions, alleviating the physical and emotional burden on patients. As AI evolves, its pivotal role in early detection will continue to reshape and enhance neurological healthcare [28].

3.6. Summary and literature review outcome

The integration of AI into clinical practice has marked a transformative shift in the diagnosis and management of neurological conditions, focusing on neurodegenerative diseases, vascular conditions, brain tumors, and emergency cases.

In the realm of neurodegenerative diseases, AI presents innovative solutions for early detection and personalized care. The Clinical Explainability Failure (CEF) and Explainability Failure Ratio (EFR) framework, as outlined by Venugopal et al. [19] McKinney et al. [22], ensures transparent and interpretable insights into AI algorithms' decision-making processes [21]. Tools like CheXplain, exemplified by Xie et al. [18], empower healthcare professionals in making informed decisions regarding the management of Alzheimer's and Parkinson's disease.

For vascular conditions, AI's role in early detection and characterization, particularly in aneurysms and abnormalities, is crucial. Huang et al. [15] and Lapuschkin et al. [23] highlight the importance of model interpretability, with tools like Captum and Densely Connected Convolutional Networks (DenseNet) aiding clinicians in understanding salient features influencing AI predictions [15].

In the context of brain tumors, early detection and characterization, including the application of AI in neuro-oncology, are explored. The MedMNIST Classification Decathlon serves as a benchmark for lightweight AutoML in medical image analysis [17]. A specific application in brain tumor detection involves a comparative analysis between pre-trained transfer learning models and a hybrid DeepTumorNet model, showcasing its efficacy in classifying brain tumors [24].

Emergency neurological cases, such as stroke and trauma, emphasize the importance of precision in medical imaging, as discussed by Kriegsmann et al. [25]. Quantitative analysis, including metrics like sensitivity, specificity, and area under the curve (AUC), ensures not only high accuracy but also transparent insights into AI models' decision processes, aiding clinicians in time-sensitive situations.

The section concludes with a focus on AI-enabled early detection in neurological pathologies. It highlights AI's role in early detection of brain tumors, aneurysms, vascular abnormalities, and neurodegenerative diseases, emphasizing the transformative capability of AI in timely anomaly identification. Specific examples, such as the early detection of glioblastoma, illustrate AI's potential to enhance treatment success rates and improve the patient's quality of life [28].

In summary, the clinical applications of AI in neurological conditions showcase the potential for transformative impact across various domains. Whether in neurodegenerative diseases, vascular conditions, brain tumors, or emergency cases, the integration of AI, guided by transparency and interpretability, holds promise for advancing diagnostic precision and patient care in neurology.

4. Enhancing patient care with AI

In the ever-evolving landscape of healthcare, the integration of AI is revolutionizing patient care across various domains. The following subsections delve into specific areas where AI plays a pivotal role in enhancing patient care.

4.1. Personalized treatment plans

Tailoring medical treatments to the unique characteristics of each patient is a cornerstone of modern healthcare. AI, particularly machine learning algorithms, analyzes extensive patient data to identify patterns and predict individual responses to various treatment options. This approach, known as precision medicine, allows clinicians to formulate personalized treatment plans that consider genetic, environmental, and lifestyle factors [19]. By leveraging AI-generated insights, healthcare providers can optimize treatment efficacy, minimize adverse effects, and improve overall patient outcomes.

Example application—Precision oncology:

Tailoring cancer treatments: Precision oncology leverages AI to analyze molecular and genetic data from tumors [15]. Machine learning models identify specific mutations or biomarkers, aiding oncologists in selecting targeted therapies tailored to the unique characteristics of a patient's cancer [20]. This personalized approach enhances treatment efficacy while minimizing side effects [19].

4.2. Surgical assistance with AI-powered robots

The integration of AI-powered robots into surgical procedures signifies a new era of precision and efficiency. These robots, guided by sophisticated algorithms, collaborate with surgeons to perform intricate tasks with enhanced accuracy. Real-time data analysis, insights provision, and even autonomous execution of certain surgery aspects contribute to improved surgical outcomes, reduced recovery times, and enhanced patient safety [16].

Example application—Robot-assisted surgery:

Precision surgical procedures: AI-powered robots, guided by machine learning algorithms, assist surgeons in intricate procedures like minimally invasive surgeries [16]. These robots enhance precision and offer real-time insights, contributing to improved surgical outcomes [18]. The integration of AI ensures collaborative synergy between surgeons and robotic systems.

4.3. Predictive analytics for proactive interventions

AI's proficiency in predictive analytics empowers healthcare professionals to anticipate potential health issues and intervene proactively. By analyzing patient data, AI algorithms identify subtle indicators preceding adverse events. This early warning system enables healthcare providers to implement preventive measures, reducing the likelihood of complications and improving patient outcomes. Predictive analytics finds particular value in chronic disease management and population health initiatives [18].

Example application—Heart failure prediction:

Anticipating cardiac decompensation: Predictive analytics models, powered by AI, analyze patient data to forecast heart failure events [17]. By monitoring vital signs, medication adherence, and lifestyle factors, these models enable healthcare providers to intervene proactively, preventing hospital admissions and improving overall patient outcomes [20].

4.4. Wearable devices for neurological monitoring

The emergence of wearable devices with AI capabilities has transformed neurological monitoring. These devices continuously collect and analyze physiological data, providing real-time insights into neurological health. In conditions like epilepsy or movement disorders, wearable devices can detect anomalies, trigger alerts, and facilitate remote monitoring by healthcare professionals. This proactive approach to neurological care enhances patient autonomy and allows for early intervention in response to emergent situations [17].

Example application—Epilepsy management:

Continuous monitoring for seizure detection: Wearable devices equipped with AI capabilities monitor physiological signals to detect early signs of epileptic seizures [18]. Real-time data analysis enables timely alerts to healthcare providers or caregivers, facilitating prompt intervention and improving patient safety [17].

As AI continues to progress, its integration into personalized treatment plans, surgical procedures, predictive analytics, and wearable devices holds great promise for revolutionizing patient care across diverse medical scenarios.

The examples presented in this section underscore the transformative impact of AI in tailoring treatments, enhancing surgical precision, enabling proactive interventions, and revolutionizing neurological monitoring for improved patient care. This structured review adheres to the PRISMA guidelines, ensuring transparency and rigor in presenting the literature on the integration of AI in patient care.

4.5. Summary and literature review outcome

The integration of AI into healthcare is reshaping patient care across multiple domains, focusing on personalized treatment plans, surgical assistance with AI-powered robots, predictive analytics for proactive interventions, and wearable devices for neurological monitoring.

In the realm of personalized treatment plans, AI, particularly machine learning algorithms, facilitates precision medicine. By analyzing extensive patient data, AI identifies patterns and predicts individual responses to treatment options. Precision oncology exemplifies this approach, utilizing AI to analyze molecular and genetic data for tailored cancer treatments, optimizing efficacy while minimizing side effects [19].

Surgical assistance with AI-powered robots introduces a new era of precision and efficiency. Guided by sophisticated algorithms, these robots collaborate with surgeons to perform intricate tasks with enhanced accuracy. In procedures like minimally invasive surgeries, AI-powered robots contribute to precision and real-time insights, improving surgical outcomes through collaborative synergy between surgeons and robotic systems [16].

Predictive analytics powered by AI enables proactive interventions by anticipating potential health issues. Analyzing patient data allows AI algorithms to identify subtle indicators preceding adverse events. In heart failure prediction, for instance, AI-driven predictive analytics models forecast events by monitoring vital signs, medication adherence, and lifestyle factors, enabling proactive interventions and improving overall patient outcomes [17].

Wearable devices equipped with AI capabilities transform neurological

monitoring by continuously collecting and analyzing physiological data. In conditions like epilepsy, these devices detect anomalies, trigger alerts, and facilitate remote monitoring. For epilepsy management, wearable devices with AI capabilities monitor physiological signals for seizure detection, providing real-time data analysis and facilitating prompt intervention for improved patient safety [17,18].

The transformative impact of AI in tailoring treatments, enhancing surgical precision, enabling proactive interventions, and revolutionizing neurological monitoring underscores its promise in revolutionizing patient care across diverse medical scenarios. As AI continues to advance, its integration into healthcare holds great promise for personalized and improved patient outcomes.

It results that the integration of AI into healthcare is revolutionizing patient care by tailoring treatments, enhancing surgical precision, enabling proactive interventions, and revolutionizing neurological monitoring. The examples provided highlight the transformative impact of AI in these areas, contributing to improved patient outcomes and paving the way for future advancements in healthcare.

5. AI in drug discovery for neurological disorders

In the dynamic landscape of healthcare, the integration of AI is reshaping patient care across various domains. To adhere to the PRISMA guidelines for systematic reviews and meta-analyses, the following section provides a structured and transparent presentation of the literature on the role of AI in enhancing patient care.

5.1. Target identification and validation

Tailoring medical treatments to individual patient characteristics is a fundamental aspect of modern healthcare. AI, particularly machine learning algorithms, analyzes extensive patient data to discern patterns and predict responses to treatment options—A paradigm known as precision medicine. This approach considers genetic, environmental, and lifestyle factors to formulate personalized treatment plans [19]. AI-generated insights empower healthcare providers to optimize treatment efficacy, minimize adverse effects, and ultimately improve patient outcomes.

Example application—Precision oncology:

Tailoring cancer treatments: Precision oncology utilizes AI to analyze molecular and genetic data from tumors [15]. Machine learning models identify specific mutations or biomarkers, aiding oncologists in selecting targeted therapies tailored to each patient's unique cancer characteristics [20]. This personalized approach enhances treatment efficacy while minimizing side effects [19,29].

5.2. Drug repurposing

The integration of AI-powered robots into surgical procedures signifies a new era of precision and efficiency. These robots, guided by sophisticated algorithms, collaborate with surgeons to perform intricate tasks with enhanced accuracy. Real-time data analysis, insights provision, and even autonomous execution of certain surgery aspects contribute to improved surgical outcomes, reduced recovery times, and enhanced patient safety [16].

Example application—Robot-assisted surgery:

Precision surgical procedures: AI-powered robots, guided by machine learning algorithms, assist surgeons in intricate procedures like minimally invasive surgeries [16]. These robots enhance precision and offer real-time insights, contributing to improved surgical outcomes [18]. The integration of AI ensures collaborative synergy between surgeons and robotic systems.

5.3. Predictive modeling for drug efficacy

AI-driven predictive modeling enhances the assessment of drug efficacy in preclinical stages [30,31]. Machine learning algorithms analyze diverse biological data, including cellular responses, biomarker profiles, and drug interactions, to predict the likelihood of a candidate drug's success. This predictive capability aids researchers in prioritizing the most promising compounds for further investigation [32].

Example application: Employing AI models to predict the efficacy of novel compounds in mitigating neuronal damage in a preclinical model of a neurodegenerative disease, guiding the selection of candidates with the highest probability of success [33].

5.4. Accelerating clinical trials

AI expedites the design and execution of clinical trials for neurological drugs [19,25–28]. By analyzing historical trial data, patient characteristics, and real-time feedback, AI algorithms optimize trial protocols, identify suitable patient populations, and predict patient responses. This accelerates the clinical trial process, reduces costs, and increases the likelihood of bringing effective therapies to market [32,33].

Example application: Using AI to design adaptive clinical trial protocols that dynamically adjust based on accumulating data, optimizing resource allocation and increasing the efficiency of the trial for a novel neurological drug [32,33].

5.5. Summary and literature review outcome

In section 5, we explored how AI revolutionizes drug discovery for neurological disorders, offering innovative solutions for target identification, drug repurposing, predictive modeling, and expediting clinical trials. These applications collectively contribute to the accelerated development of much-needed therapeutics in the realm of neurological healthcare [19,25–28].

In the realm of drug discovery for neurological disorders, AI emerges as a transformative force, offering innovative solutions across various stages. The multifaceted applications of AI, as discussed in section 5, are outlined below:

AI plays a crucial role in identifying and validating potential drug targets for neurological disorders by analyzing vast datasets encompassing genetic, omics, and clinical information [29]. This approach enhances the efficiency of drug discovery by focusing on specific proteins or genes most likely to yield effective interventions [30]. An example application involves utilizing AI algorithms to analyze genomic and proteomic data, leading to the identification of a specific protein implicated in Alzheimer's disease, providing a validated target for drug development [31].

AI facilitates drug repurposing by exploring existing drugs for new therapeutic indications through the analysis of diverse datasets [30]. This approach expedites the

development timeline and leverages existing knowledge about drug safety and pharmacokinetics. An example application involves repurposing an FDA-approved drug for a neurodegenerative disorder based on shared molecular targets identified through AI analysis [32].

AI-driven predictive modeling enhances the assessment of drug efficacy in preclinical stages by analyzing diverse biological data [30,31]. This predictive capability aids researchers in prioritizing the most promising compounds for further investigation. An example application includes using AI models to predict the efficacy of novel compounds in mitigating neuronal damage in a preclinical model of a neurodegenerative disease [33].

AI expedites the design and execution of clinical trials for neurological drugs by analyzing historical trial data, patient characteristics, and real-time feedback [19,25–28]. This accelerates the clinical trial process, reduces costs, and increases the likelihood of bringing effective therapies to market. An example application involves using AI to design adaptive clinical trial protocols that dynamically adjust based on accumulating data, optimizing resource allocation and increasing trial efficiency for a novel neurological drug [32,33].

In summary, the integration of AI into drug discovery for neurological disorders revolutionizes the field, offering solutions for target identification, drug repurposing, predictive modeling, and expediting clinical trials. These applications collectively contribute to the accelerated development of much-needed therapeutics in neurological healthcare [19,25–28]. The literature review outcome highlights the diverse ways AI is transforming drug discovery, providing a foundation for advancing neurological healthcare.

6. Challenges and ethical considerations in AI in medicine and healthcare

AI in medicine and healthcare presents numerous ethical challenges that necessitate careful consideration. The article titled “Ethical issues of artificial intelligence in medicine and healthcare” sheds light on these challenges, and the following elaboration explores key aspects and references [34].

6.1. Ethical use of AI in healthcare

The ethical use of AI in healthcare is a paramount concern, especially regarding patient privacy and data protection [9]. Regulations like the General Data Protection Regulation (GDPR) in the EU and Genetic Information Non-discrimination Acts (GINA) in the US emphasize the importance of safeguarding individuals’ health data [35,36]. Despite these regulations, challenges persist:

- **Insufficient legal protection:** Existing laws may fall short in adequately protecting an individual’s health data, leaving room for potential misuse [35].
- **Security concerns:** The vulnerability of clinical data collected by robots to hacking poses serious threats to privacy and security [35].
- **Unregulated data sharing:** Some entities, like social networks, accumulate users’ health data without explicit consent, raising concerns about data privacy and commercial exploitation [9].

- Unmonitored data sales: Genetic testing and bioinformatics companies, lacking adequate regulation, may sell customer data to pharmaceutical and biotechnology firms, posing privacy risks [35].

6.2. Ensuring patient privacy

Patient privacy is a critical ethical consideration, and the use of AI in healthcare introduces new challenges:

- Data security risks: The potential for hacking into clinical data collected by robots poses risks to patient privacy and the confidentiality of sensitive medical information [35].
- Unauthorized data use: Social networks may gather and store extensive user data, including mental health information, without proper consent, raising concerns about privacy breaches and unauthorized data use [35].

6.3. Addressing bias in AI algorithms

The ethical implications of bias in AI algorithms, particularly in healthcare, are a significant concern [36,37]:

- Social inequality: AI's impact on automation and advanced economies can exacerbate social inequality, resulting in job losses and decreased salaries.
- Job displacement: The rise of surgical robots and robotic nurses may threaten the job opportunities of human surgeons and nurses, contributing to social inequality.

6.4. Healthcare professional training

Integrating AI into healthcare requires considering the ethical implications related to medical consultation, empathy, and sympathy:

- Human-machine relations: The integration of AI in healthcare raises concerns about the ability of robotic systems to provide empathetic and compassionate care, which is crucial in medical settings [38].
- Impact on patients: The use of medical robots in various healthcare settings, such as obstetrics and gynecology or psychiatric hospitals, may adversely affect patients, particularly children or those with severe psychiatric disorders [39].

6.5. Summary and literature review outcome

In navigating the integration of AI into medicine and healthcare, ethical challenges loom large, as outlined in the exploration of challenges and ethical considerations in AI in medicine and healthcare:

The paramount concern revolves around the ethical use of AI in healthcare, particularly regarding patient privacy and data protection [9]. Regulations such as GDPR in the EU and GINA in the US emphasize the importance of safeguarding health data [35,36]. Challenges persist in legal protection, security concerns, unregulated data sharing, and unmonitored data sales, leaving room for potential misuse and privacy risks [35].

AI introduces new challenges in ensuring patient privacy, with risks of data security breaches and unauthorized data use [35]. The potential for hacking into clinical data collected by robots poses threats to confidentiality, highlighting the need

for robust measures to protect sensitive medical information. The ethical implications of bias in AI algorithms, especially in healthcare, raise concerns about social inequality, job displacement, and potential impacts on human surgeons and nurses [37]. The rise of surgical robots may contribute to social inequality by displacing human workers.

Integrating AI into healthcare necessitates consideration of ethical implications related to human-machine relations and the impact on patients [38]. Concerns arise about the ability of robotic systems to provide empathetic and compassionate care, particularly in settings such as obstetrics, gynecology, or psychiatric hospitals [39].

The rapid advancement of AI in medicine and healthcare offers significant benefits but demands a vigilant approach to ethical considerations. Patient privacy, bias in algorithms, social justice, and preserving human qualities in medical interactions are paramount concerns that require ongoing attention [40–42]. The literature review emphasizes the need for robust regulatory frameworks, security measures, and ethical guidelines to address the evolving landscape of AI in healthcare responsibly. Policymakers and practitioners alike must collaboratively address these challenges to ensure the ethical use of AI while maximizing its positive impact on patient care and outcomes.

7. Future perspectives and innovations

As we gaze into the future of AI in the realm of medicine, particularly in the domains of Neurosurgery and Neurology, several exciting prospects and innovations emerge. The evolving role of AI in these fields promises to reshape the landscape of healthcare in profound ways.

7.1. Evolving role of AI in Neurosurgery and Neurology

AI is poised to play a transformative role in Neurosurgery and Neurology. In neurosurgery, AI-powered tools are advancing precision and efficiency. Surgical robots, guided by AI algorithms, can enhance the accuracy of delicate procedures, reducing the risk and improving patient outcomes. Moreover, AI assists neurologists in diagnosing and understanding complex neurological disorders. Advanced imaging analysis and pattern recognition enable quicker and more accurate identification of abnormalities, leading to timely interventions.

7.2. Integrating AI into standard practices

The integration of AI into standard medical practices is becoming more seamless. AI-driven diagnostic tools are augmenting the capabilities of healthcare professionals, providing valuable insights and improving decision-making processes. From early detection of neurological conditions to personalized treatment plans, AI contributes to a more comprehensive and efficient healthcare ecosystem.

7.3. Potential breakthroughs and research areas

Future breakthroughs in AI applications within neurology and neurosurgery may lie in enhanced predictive modeling and treatment optimization. AI algorithms could analyze vast datasets, including genetic information, patient histories, and treatment

outcomes, to predict disease progression and recommend personalized treatment strategies. Additionally, research areas exploring the intersection of AI with other cutting-edge technologies, such as genomics and nanomedicine, hold the promise of groundbreaking advancements in understanding and treating neurological disorders.

7.4. Patient-centric AI solutions

The future of AI in medicine envisions patient-centric solutions that prioritize individualized care. Virtual health assistants, powered by AI, could provide continuous monitoring and support for patients with neurological conditions. These AI-driven tools may not only assist in managing symptoms but also empower patients with valuable information and resources, fostering a more active role in their healthcare journey.

In conclusion, the future of AI in Neurosurgery and Neurology is marked by ongoing innovation and integration into standard medical practices. As research progresses and technology evolves, we anticipate a healthcare landscape where AI contributes significantly to improved diagnostics, treatment strategies, and ultimately, better patient outcomes. The journey towards patient-centric, AI-driven healthcare is an exciting frontier that holds immense potential for the advancement of neuroscientific and medical knowledge.

8. Conclusion

The infusion of artificial intelligence (AI) into the domains of Neurosurgery and Neurology heralds a groundbreaking era of transformative healthcare innovation. Across seven comprehensive sections, we've delved into the multifaceted impact of AI, witnessing its prowess in enhancing diagnostics, tailoring treatments, and fundamentally reshaping patient care.

The strides made in diagnostic accuracy through AI-driven algorithms represent a paradigm shift, expediting the identification of neurological conditions and empowering healthcare professionals to make swifter, more precise decisions. The promise of personalized treatment plans, crafted through the nuanced analysis of patient-specific data, genetic information, and real-time health metrics, signifies a leap forward in increasing the likelihood of successful outcomes for individuals facing neurological challenges.

The introduction of AI-powered surgical robots exemplifies a new frontier in neurosurgery, offering precision and minimally invasive capabilities that mitigate the risk of human error. These advancements not only elevate the standard of care but also enable remote surgical support, extending specialized medical expertise to a broader population.

Predictive analytics, facilitated by machine learning models, are instrumental in forecasting disease progression and patient outcomes. This empowers healthcare providers with invaluable insights for informed decision-making, optimizing treatment strategies, and enhancing overall patient care. Wearable devices equipped with AI for continuous neurological monitoring contribute to early intervention, allowing patients to actively manage chronic conditions and healthcare professionals to intervene promptly.

AI's role in drug discovery stands out as a beacon of hope, accelerating the identification of potential compounds for treating neurological disorders. This promises more effective and targeted therapies, addressing the urgent need for breakthrough treatments in the field.

Looking ahead, the future of Neurosurgery and Neurology with AI is ripe with possibilities. Enhanced personalization will refine treatment plans, considering even more nuanced patient factors, genetic markers, and lifestyle data. AI's augmentation of healthcare professionals will provide real-time insights, facilitating more efficient and accurate patient care. The pivotal role of AI in early detection and intervention will enable proactive measures to prevent or mitigate the progression of neurological disorders.

The integration of AI ecosystems is poised to connect various facets of patient care seamlessly, from diagnostics and treatment to monitoring and patient engagement. As AI-driven drug development advances, it holds the potential to usher in highly targeted therapies, offering breakthrough treatments for currently incurable neurological conditions.

In this evolving landscape, AI continues to shape the future of Neurosurgery and Neurology, opening new horizons of precision, efficiency, and patient-centered care. With ongoing research and innovation, the dynamic partnership between AI and healthcare is set to advance, ultimately improving the lives of individuals affected by neurological disorders and transforming the very fabric of neurological medicine.

Acknowledgments: The author would like to thank IITG for its logistical support.

Conflict of interest: The author declares no conflict of interest.

References

1. Kim HE, Kim HH, Han BK et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *Lancet Digital Health*. 2020; 2: e138-e148. doi: 10.1016/S2589-7500(20)30003-0
2. Liu K, LFaes L, Kale AU. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *Lancet Digital Health*. 2019; 1: e271-e297. doi: 10.1016/S2589-7500(19)30123-2
3. Becker AS, Marcon M, Ghafoor S, et al. Deep Learning in Mammography. *Investigative Radiology*. 2017; 52(7): 434-440. doi: 10.1097/rli.0000000000000358
4. Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. *JAMA*. 2019; 322(18): 1765. doi: 10.1001/jama.2019.15064
5. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV). Published online October 2017. doi: 10.1109/iccv.2017.74
6. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks. *IEEE WACV*. 2018; 839–847. doi: 10.1109/WACV.2018.00097
7. Haq I, Mazhar T, Malik MA. Lung Nodules Localization and Report Analysis from Computerized Tomography (CT) Scan Using a Novel Machine Learning Approach. *Applied Sciences*. 2022; 12. doi: 10.3390/app 122412614
8. Ehtioui A, Zouch W, Ghorbel M, et al. Detection Methods of COVID-19. *SLAS Technology*. 2020; 25(6): 566-572. doi: 10.1177/2472630320962002
9. Hussain M, Islam A, Turi JA, et al. Machine Learning-Driven Approach for a COVID-19 Warning System. *Electronics*. 2022; 11(23): 3875. doi: 10.3390/electronics11233875

10. Khadhraoui M, Bellaaj H, Ammar MB, et al. Survey of BERT-Base Models for Scientific Text Classification: COVID-19 Case Study. *Applied Sciences*. 2022; 12(6): 2891. doi: 10.3390/app12062891
11. Alsaed Z, Khweiled R, Hamad M, et al. Role of Blockchain Technology in Combating COVID-19 Crisis. *Applied Sciences*. 2021; 11(24): 12063. doi: 10.3390/app112412063
12. Guefrechi S, Jabra MB, Ammar A, et al. Deep learning based detection of COVID-19 from chest X-ray images. *Multimedia Tools and Applications*. 2021; 80(21-23): 31803-31820. doi: 10.1007/s11042-021-11192-5
13. Ben Jabra M, Koubaa A, Benjdira B, et al. COVID-19 Diagnosis in Chest X-rays Using Deep Learning and Majority Voting. *Applied Sciences*. 2021; 11(6): 2884. doi: 10.3390/app11062884
14. Geirhos R, Jacobsen JH, Michaelis C, et al. Shortcut learning in deep neural networks. *Nature Machine Intelligence*. 2020; 2(11): 665-673. doi: 10.1038/s42256-020-00257-z
15. Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published online July 2017. doi: 10.1109/cvpr.2017.243
16. Yang J, Shi R, Ni B. MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Published online April 13, 2021. doi: 10.1109/isbi48211.2021.9434062
17. Al-Khawari H, Athyal RP, Al-Saeed O, et al. Inter- and intraobserver variation between radiologists in the detection of abnormal parenchymal lung changes on high-resolution computed tomography. *Annals of Saudi Medicine*. 2010; 30(2): 129-133. doi: 10.4103/0256-4947.60518
18. Xie Y, Chen M, Kao D, et al. CheXplain. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Published online April 21, 2020. doi: 10.1145/3313831.3376807
19. Venugopal VK, Takhar R, Gupta S, et al. Clinical Explainability Failure (CEF) & Explainability Failure Ratio (EFR) – Changing the Way We Validate Classification Algorithms. *Journal of Medical Systems*. 2022; 46(4). doi: 10.1007/s10916-022-01806-2
20. Jacovi A, Goldberg Y. Towards Faithfully Interpretable NLP Systems: How Should We Define and Evaluate Faithfulness? Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Published online 2020. doi: 10.18653/v1/2020.acl-main.386
21. Tufail AB, Anwar N, Othman MTB, et al. Early-Stage Alzheimer’s Disease Categorization Using PET Neuroimaging Modality and Convolutional Neural Networks in the 2D and 3D Domains. *Sensors*. 2022; 22(12): 4609. doi: 10.3390/s22124609
22. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020; 577(7788): 89-94. doi: 10.1038/s41586-019-1799-6
23. Lapuschkin S, Wäldchen S, Binder A, et al. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*. 2019; 10(1). doi: 10.1038/s41467-019-08987-4
24. Raza A, Ayub H, Khan JA, et al. A Hybrid Deep Learning-Based Approach for Brain Tumor Classification. *Electronics*. 2022; 11(7): 1146. doi: 10.3390/electronics11071146
25. Kriegsmann M, Kriegsmann K, Steinbuss G, et al. Deep Learning in Pancreatic Tissue: Identification of Anatomical Structures, Pancreatic Intraepithelial Neoplasia, and Ductal Adenocarcinoma. *International Journal of Molecular Sciences*. 2021; 22(10): 5385. doi: 10.3390/ijms22105385
26. Sun X, Yin Y, Yang Q, et al. Artificial intelligence in cardiovascular diseases: diagnostic and therapeutic perspectives. *European Journal of Medical Research*. 2023; 28(1). doi: 10.1186/s40001-023-01065-y
27. Chlif M, Ammar MM, Said NB, et al. Mechanism of Dyspnea during Exercise in Children with Corrected Congenital Heart Disease. *International Journal of Environmental Research and Public Health*. 2021; 19(1): 99. doi: 10.3390/ijerph19010099
28. Chirica C, Haba D, Cojocaru E, et al. One Step Forward—The Current Role of Artificial Intelligence in Glioblastoma Imaging. *Life*. 2023; 13(7): 1561. doi: 10.3390/life13071561
29. Moffat JG, Vincent F, Lee JA, et al. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug Discovery*. 2017; 16(8): 531-543. doi: 10.1038/nrd.2017.111
30. Van den Broeck WMM. Drug Targets, Target Identification, Validation, and Screening. *The Practice of Medicinal Chemistry*. Published online 2015: 45-70. doi: 10.1016/b978-0-12-417205-0.00003-1
31. Gashaw I, Ellinghaus P, Sommer A, et al. What makes a good drug target? *Drug Discovery Today*. 2011; 16(23-24): 1037-1043. doi: 10.1016/j.drudis.2011.09.007

32. Gashaw I, Ellinghaus P, Sommer A, et al. What makes a good drug target? *Drug Discovery Today*. 2012; 17: S24-S30. doi: 10.1016/j.drudis.2011.12.008
33. Lindsay MA. Target discovery. *Nature Reviews Drug Discovery*. 2003; 2(10): 831-838. doi: 10.1038/nrd1202
34. Farhud DD, Zokaei S. Ethical Issues of Artificial Intelligence in Medicine and Healthcare. *Iranian Journal of Public Health*. Published online October 27, 2021. doi: 10.18502/ijph.v50i11.7600
35. EUR-Lex. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available online: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (accessed on 1 January 2024).
36. American Society of Human Genetics (ASHG). The genetic information nondiscrimination act (GINA). Available online: <https://www.ashg.org/advocacy/gina/> (accessed on 1 January 2024).
37. Nordling L. A fairer way forward for AI in health care. *Nature*. 2019; 573(7775): S103-S105. doi: 10.1038/d41586-019-02872-2
38. The Guardian. RoboDoc: How India's robots are taking on Covid patient care. Available online: <https://www.theguardian.com/global-development/2020/dec/02/robodoc-how-india-robots-are-taking-on-covid-patient-care-mitra> (accessed on 1 January 2024).
39. Markose A, Krishnan R, Ramesh M. Medical ethics. *Journal of Pharmacy And Bioallied Sciences*. 2016; 8(5): 1. doi: 10.4103/0975-7406.191934
40. Varkey B. Principles of Clinical Ethics and Their Application to Practice. *Medical Principles and Practice*. 2020; 30(1): 17-28. doi: 10.1159/000509119
41. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. 2019; 8(7): 2328-2331. doi: 10.4103/jfmpe.jfmpe_440_19
42. Farhud DD. Epigenetic and Ethics: How are Ethical Traits Inherited? *International Journal of Ethics & Society (IJES)*. 2019; 1: 1-4.



Academic Publishing Pte. Ltd.

Add: 73 Upper Paya Lebar Road #07-02B-01 Centro Bianco Singapore 534818

Tel: +65 83184869

E-mail: editorial_office@acad-pub.com

Web: <http://ojs.acad-pub.com/>