PAPER NAME

DocAssist Report Group-19.docx

AUTHOR

Gladys Gince Skariah

WORD COUNT

7925 Words

CHARACTER COUNT

47003 Characters

PAGE COUNT

62 Pages

FILE SIZE

1.7MB

SUBMISSION DATE

Mar 20, 2024 3:28 PM GMT+5:30

REPORT DATE

Mar 20, 2024 3:29 PM GMT+5:30

## 7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 1% Publications database
- Crossref database
- Crossref Posted Content database
- 3% Submitted Works database

## Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 20 words)

# DocAssist: AI Doctor's Assistant

A project report submitted in fulfillment of the requirements for the degree of

**Bachelor of Engineering**

**In**

**Artificial Intelligence and Data Science**

by


**Mirza Mohammed Junaid (9459)**

**Pratham Kambli (9378)**

**Gladys Gince Skariah (9409)**


Under the guidance of

**Prof. Sarika Davare**

**(Assistant Professor)**


DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

**Fr. Conceicao Rodrigues College of Engineering,**

**Bandra (W), Mumbai - 400050**

University of Mumbai


(2023-24)

*This work is dedicated to my family.*

*I am very thankful for their motivation and support.*

# Internal Approval Sheet

## CERTIFICATE

This is to certify that the project entitled **"DocAssist: AI Doctor's Assistant"** is a bonafide work of **Mirza Mohammed Junaid (9459) , Pratham Kambli (9378), Gladys Gince Skariah (9409)** submitted to the University of Mumbai in fulfillment of the requirement for the award of the degree of Bachelor of Engineering in **Artificial Intelligence and Data Science.**

(Name and sign)
Supervisor/Guide

(Name and sign)
Head of Department

(Name and sign)
Principal

# Approval Sheet

## Project Report Approval

This project report entitled **"DocAssist: AI Doctor’s Assistant"** by **Mirza Mohammed Junaid, Pratham Kambli , Gladys Gince Skariah** is approved for the degree of Bachelor of Engineering in Artificial Intelligence and Data Science.

Examiner 1. ______________________________

Examiner 2. ______________________________

Date: 21st March, 2024

Place: Fr. Conceicao Rodrigues College of Engineering, Bandra

# Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Mirza Mohammed Junaid (9459) ________________

Pratham Kambli(9378) ________________

Gladys Gince Skariah (9409) ________________

Date: 21st March, 2024

# Abstract


In the rapidly evolving healthcare sector, efficient information management and communication are essential. This report explores the project **"DocAssist: AI Doctor’s Assistant"** , an innovative system that streamlines healthcare workflows. The project leverages cutting-edge technologies, libraries, and, notably, transformer-based deep learning models for audio and text processing.

The system integrates modules which incorporate pre-trained transformers to perform automatic speech recognition (ASR) and emotion analysis. By harnessing transformers, the project accurately transcribes audio recordings to text and assesses emotional content in spoken words. These capabilities have far-reaching implications for patient record accuracy and healthcare professionals' productivity.

A web-based interface, empowers healthcare practitioners to efficiently manage patient appointments and, potentially, record and transmit audio notes. With a user-friendly approach, the system enhances the administration of healthcare facilities.

The project's objectives include improving healthcare processes by automating transcription and sentiment analysis tasks. Transformer models, renowned for their capacity to handle sequential data, are pivotal in achieving these goals. As a result, the project accelerates processes, saves time, and fosters an accurate and comprehensive patient record system.

# Acknowledgments

We take great pleasure in presenting the report on **"DocAssist: AI Doctor's Assistant".** We wish to convey our profound gratitude to Prof. Sarika Davare, our esteemed guide at C.R.C.E, Bandra (W), Mumbai, for imparting invaluable technical expertise and offering insightful recommendations that have significantly influenced the project's course. We deeply appreciated the enriching discussions we engaged in during our departmental interactions.

Our sincere appreciation extends to Dr. Jagruti Save, the Head of the Artificial Intelligence and Data Science department, as well as the Principal and the distinguished management of C.R.C.E., Mumbai, for their steadfast encouragement and the essential infrastructure they generously provided, which played a pivotal role in ensuring the successful realization of this project.

Furthermore, we would like to express our gratitude to the dedicated non-teaching staff for their invaluable support, which has been instrumental in bringing our project to its current fruition.

Mirza Mohammed Junaid(9459)
Pratham Kambli (9378)
Gladys Gince Skariah (9409)

Date: 4th November, 2023

# Contents

# List of Figures

# Glossary

1. Audio Transcription: The process of converting spoken language in audio recordings into text format, making it accessible and searchable for healthcare professionals.

2. Emotion Analysis: The application of natural language processing techniques to determine the emotional content and sentiment expressed during healthcare appointments, helping healthcare providers understand patients' emotional states.

3. Speaker Segmentation: The identification and labeling of different speakers within audio recordings, enabling the tracking of who provided specific information or instructions during appointments.

4. User Interface (UI): The graphical interface that healthcare professionals use to interact with the system, facilitating appointment management, data retrieval, and audio recording.

5. Transformers: A type of deep learning model architecture used for a wide range of natural language processing (NLP) tasks, characterized by its attention mechanism that enables capturing contextual relationships within text data. Transformers have significantly improved the state of the art in NLP and are commonly used in tasks such as language translation, sentiment analysis, and text generation.

6. Data Analysis: The process of examining and interpreting collected data to extract valuable insights and trends, aiding healthcare professionals in decision-making.

7. Data Preprocessing: The cleaning and organization of data to ensure its quality and consistency, improving the accuracy and reliability of analysis results.

8. Natural Language Processing (NLP): The field of artificial intelligence that focuses on the interaction between computers and human language, enabling machines to understand, interpret, and generate human text.

9. Electronic Health Record (EHR): Digital records that contain a patient's medical history, diagnoses, treatments, and other health-related information, often used by healthcare providers to manage patient data.

10. Scalability: The ability of the system to handle increased data volumes and user demands as it grows, without compromising performance or responsiveness.

11. Feedback Mechanism: A feature that encourages healthcare professionals to provide input and suggestions for system improvement, fostering continuous enhancement.

12. Compliance: Adherence to healthcare regulations and standards, such as HIPAA, to ensure the secure handling of patient data and maintain data privacy.

13. Machine Learning: An artificial intelligence approach where algorithms enable systems to learn from data and improve their performance over time without being explicitly programmed.

14. Natural Language Understanding (NLU): A subset of natural language processing that focuses on the ability of a machine to understand and interpret the meaning behind human language.

15. Topic Modeling: A technique used to identify common topics or themes within a body of text, aiding in the categorization and organization of textual data.

16. Keyword Extraction: The process of identifying and extracting significant keywords or phrases from text data, helping to summarize and categorize information.

17. Data-Driven Insights: Insights and conclusions drawn from the analysis of data, allowing healthcare professionals to make informed decisions and recommendations.

18. Healthcare Data Management: The systematic collection, storage, and analysis of healthcare-related information to support patient care and administrative functions in the healthcare sector.

19. User Experience (UX): The overall experience and satisfaction of healthcare professionals when interacting with the system's user interface.

20. AER (Audio Emotion Recognition): A specialized branch of artificial intelligence that focuses on recognizing and analyzing emotions expressed in audio data. AER systems use machine learning and signal processing techniques to detect and interpret emotional cues in spoken language, which can be valuable for understanding the emotional states of individuals during conversations, such as healthcare appointments.

# Chapter 1

## Introduction

- The healthcare industry, an ever-growing and important segment of our society, faces a mammoth task: balancing patient care and operational efficiency. As the demand for healthcare services increases, so does the need for advanced tools and systems that not only reduce the workload of healthcare professionals but also improve the overall patient experience. It is in this changing context that the "DocAssist: AI Doctor’s Assistant" project emerges as a leading force, poised to revolutionize healthcare services and information systems. At the heart of this work are the simple integration of audio data, advanced information processing, and the unique capabilities of transformer-based deep learning models.
- In the following pages we begin to explore health solutions that use the power of technology to bridge the gap between spoken words and complete patient records. The most important innovation in this work is the use of pre-trained transformer models. These models, known for their expertise in natural language processing and audio analysis, are based on important aspects of the work. Speech recognition (ASR), an important but often complex task in health documentation, is performed with unprecedented accuracy due to these transformation processes. Furthermore, the task engages in, and provides, sensory analysis healthcare providers gain deeper insights from patient interactions.
- The implications of this project go far beyond digitizing health records. It is encouraging a paradigm shift in the way healthcare providers work, providing tools that not only automate manual processes but also increase the accuracy and accuracy of patient information. A web-based interface, provides an easy-to-use method for managing patient appointments, where healthcare professionals can capture and transmit audio information in a simple manner.
- Essentially it is a combination of innovation and utility. As we begin this review of "DocAssist: AI Doctor’s Assistant," we navigate the complex territory of code modules,

gain insight into the transformative impact of transformer models, and illustrate on we intend that it can be applied to healthcare beyond code and interface, this project sweeps technology integration in healthcare in line with the theme, and formulates an efficient, accurate and compassionate approach to patient care. In the following pages, we delve into the project's complexities, challenges, and applicability, all of which together contribute to the discourse of healthcare innovation. With "DocAssist: AI Doctor's Assistant," we are ushering in a new era in which technology and healthcare coexist, redefining the way healthcare services are viewed and delivered.

# Chapter 2

## Literature Review

### 2.1 Comparative study of 20 research papers

| Sr No | Title | P_Year | Name of Journal | Dataset | Algorithm | Conclusion | Research Gap |
|---|---|---|---|---|---|---|---|
| 1. | Electronic Health Records: A Systematic Review on Quality Requirements | 2015 | Journal of Medical Systems | - | - | The authors identified 203 quality requirements for electronic health records (EHRs). These requirements were categorized into the following groups: functional requirements (n=102), non-functional requirements (n=95), and security requirements (n=6). The authors also identified several research gaps, including the need for more research on the development and evaluation of EHR quality assessment tools, and the need for more research on the impact of EHR quality on patient care. | The need for more research on the development and evaluation of EHR quality assessment tools. There is a need for more research on the impact of EHR quality on patient care. The need for more research on the development of methods to improve the quality of EHR data. The need for more research on the development of methods to make EHR data more accessible to patients and |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | researchers. |
| 2. | Natural Language Processing for EHR-Based Computational Phenotyping | 2023 | Annual Review of Biomedical Engineering | - | - | The authors concluded that NLP-based computational phenotyping has the potential to revolutionize healthcare research and practice by enabling the discovery of new phenotypes, the identification of patients at risk of developing certain diseases, and the development of more personalized treatment plans. | The need for more research on the development and evaluation of NLP methods for specific computational phenotyping tasks, such as the identification of patients with rare diseases or the development of risk prediction models. The need for more research on the integration of NLP with other healthcare technologies, such as machine learning and genomics, to improve the accuracy and performance of |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | computational phenotyping models. The need for more research on the ethical and social implications of using NLP for computational phenotyping, such as the potential for bias and discrimination. |
| 3. | The Doctor-Patient Relationship and Information Seeking: The Patient's Perspective | 2017 | Patient Education and Counseling | The authors conducted a qualitative study of 36 patients with a variety of chronic diseases | - | The authors found that patients value their relationship with their doctor and trust their doctor's advice. However, patients also reported that they actively seek information about their health from a variety of sources, including the internet, other patients, and family and friends. The authors also found that patients feel empowered when they have information about their health and can make informed | The need for more research on how to improve communication between doctors and patients about online health information. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | decisions about their care. | |
| 4. | A National Study of Challenges to Electronic Health Record Adoption and Meaningful Use | 2014 | Medical Care | Regional Extension Center (REC) program dataset. | The authors used descriptive statistics to analyze the data. | The authors found that the most common challenges to EHR adoption and meaningful use were provider engagement and administrative issues. The authors also found that the most challenging meaningful use measure was the clinical summaries measure. | The need for more research on the effectiveness of different strategies for addressing the challenges to EHR adoption and meaningful use.<br>The need for more research on the impact of EHR adoption and meaningful use on patient care quality and outcomes. |
| 5. | Consent for Use of Clinical Data for Research: Can It Be Harmed? | 2020 | Journal of the American Medical Association | - | - | The authors concluded that consent for the use of clinical data for research can be harmed in a number of ways, including:<br>• Coercion: Patients may feel coerced to consent to research participation, | The need for more research on how to develop and implement consent processes that protect the rights and interests of research |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | especially if they are in a vulnerable position.<br>• Misinformation: Patients may not be given accurate or complete information about the research before they consent.<br>• Breach of confidentiality: Patient data may be breached, compromising patient privacy.<br>• Unintended consequences: Research may have unintended consequences, such as harming participants or stigmatising certain groups of people. | participants. |
| 6. | A Review of Speech-to-Text Algorithms for the Voice of the Customer Analytics | 2022 | International Journal of Speech Technology | - | - | The authors concluded that speech-to-text (STT) algorithms have the potential to revolutionize voice of the customer (VOC) analytics by making it possible to automatically transcribe and analyze customer interactions from a variety of sources, such as call center recordings, customer surveys, | The need for more research on the development and evaluation of STT algorithms that are specifically designed for VOC analytics. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | and social media posts. However, the authors also noted that there are a number of challenges that need to be addressed before STT algorithms can be widely used for VOC analytics. These challenges include:<br>Accuracy: STT algorithms can be inaccurate, especially in noisy environments or when customers are speaking with accents.<br>Scalability: STT algorithms can be computationally expensive to train and run, making it difficult to scale them to large datasets.<br>Privacy: STT algorithms may raise privacy concerns, especially if they are used to analyze customer interactions without the customers' consent. | |
| 7. | Voice Recognition System: Speec | 2015 | Journal of Applied and Fundamental Science | The dataset used in the study mentioned in | The paper mentions two algorithms used in the Voice Recognition System: Mel Frequency | The conclusion of the research paper is that the main aim of the project is to develop a Voice Recognition | A possible research gap that this paper could address is the |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | h-to-Text | | s | this paper is a sample of 5 speakers, each of whom spoke 10 digits. A database was created based on this sample, and features were extracted using MFCC. | Cepstral Coefficients (MFCC) and Vector Quantization (VQ). The MFCC algorithm is used for feature extraction from the input speech signal. It involves several steps, including framing and blocking, windowing, Fast Fourier Transform (FFT), Mel-Scale, and Discrete Cosine Transform (DCT). The extracted features are then stored in a .mat file. The VQ algorithm is used for feature matching. It involves choosing any two dimensions, inspecting the vectors, and plotting data points. The algorithm then checks whether the data regions for two different speakers are overlapping each other and in the same cluster. The Function Vqlbg is used to train the VQ | System that will allow the computer to translate voice requests and dictation into text using MFCC and VQ techniques. The extracted features will be stored in a .mat file, and models will be created using Hidden Markov Model (HMM). The desired output will be shown in the MATLAB interface. The paper compares various approaches available for developing a Voice Recognition System based on adapted feature extraction techniques and speech recognition approaches for a particular language. The study also discusses the recent progress in this field. | development of a Voice Recognition System that is specifically designed for use in hospitals. The paper could explore the challenges and requirements of developing such a system, such as the need for high accuracy and speed, the ability to recognize medical terms and jargon, and the need for secure and reliable data storage. The paper could also compare different approaches and techniques for developing such a system and evaluate |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | codebook. A distortion measure based on minimizing the Euclidean distance is used while matching the unknown speech. | | their effectiveness in a hospital setting. Additionally, the paper could discuss the potential benefits of using a Voice Recognition System in hospitals, such as improved efficiency, accuracy, and patient care. |
| 8. | Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer Using Natural Language Generation | 2020 | International Journal of Engineering and Advanced Technology | The dataset used for summarization in this model is the CNN/Dailymail corpus, which comprises news articles. | The proposed model uses two major algorithms: Deep Speech 2 and AMR graphs. However, it discusses three keyword extraction algorithms, namely TextRank, LexRank, and Latent Semantic Analysis (LSA), which were used for comparison in that study. | The main focus of the research paper is to introduce a real-time speech-to-text converter that can summarise the text and output it in audio form. The paper provides an overview of past research related to content rundown and sentiment analysis, and also references a specific version of KNN proposed by an author in a previous study. | One potential research gap that this paper does not address is the specific needs and requirements of a speech to database system for hospitals. While the techniques and algorithms used in this paper may be useful, |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | and Abstract Meaning Representation | | | | | | further research may be needed to tailor them to the specific needs of the healthcare industry. Additionally, the paper does not address the potential ethical and privacy concerns that may arise from using speech-to-text technology in a healthcare setting. |
| 9. | Speech to text and text to speech recognition systems-Areview | 2018 | IOSR Journal of Computer Engineering | - | Discrete Fourier Transform: converts each frame from time domain to frequency domain; Mel Filter Bank Algorithm: the signal is plotted against the Mel spectrum to mimic human hearing and Dynamic Time Warping | In conclusion, within the realm of Speech-to-Text (STT), Hidden Markov Models (HMM) are preferred for their computational efficiency, while in Text-to-Speech (TTS), formant synthesis methods, especially when utilizing parallel and cascade synthesis, are deemed effective in producing | A need for more research on storage of data. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | natural-sounding speech. Hybrid machine translation, by integrating rule-based and statistical approaches, offers a comprehensive solution for translation, ensuring both grammatical correctness and text fluency. These techniques represent pragmatic choices in their respective domains, addressing key challenges while optimizing performance for various applications. | |
| 10. | Speech to Text Conversion and Sentiment Analysis on Speaker Specific Data | 2021 | International Research Journal of Modernization in Engineering Technology and Science | Live generated audio inputs. | It uses many methods/algorithm such as: Artificial Neural Network Classier (ANN) based Cuckoo Search Optimization and Hidden Markov Model | In summary, this research introduces a system for automatic speech-to-text conversion and sentiment analysis. It works well with live audio inputs but has limitations in handling multiple speakers and improving accuracy. Future work will focus on enhancing system performance, addressing these limitations, and incorporating larger datasets to | It works well with live audio inputs but has limitations in handling multiple speakers and improving accuracy. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | broaden its applications. | |
| 11. | Information Extraction -a text mining approach | 2008 | IET-UK International Conference on Information and Communication Technology in Electrical Sciences | For the dataset, 600 computer-science job postings to the newsgroup austin.jobs were collected and manually annotated with correct extraction templates. | Ripper and Apriori Algorithm | In this paper, they introduce a method that combines Information Extraction (IE) and Knowledge Discovery in Databases (KDD) to extract structured data from unstructured text and mine it. The experiments show that this integration benefits both tasks: IE helps KDD with unstructured text, and KDD discovers rules to improve IE. This highlights the potential of text mining, an emerging field at the intersection of natural language processing, machine learning, data mining, and information retrieval, with computational linguistics and machine learning collaboration being essential for text-mining system development. | The research is not on real time data. |
| 12. | A Review on | 2020 | International Researc | - | Cuckoo search algorithm and Dynamic Time | HMM improves STT and TTS. For STT, combining | Using the STT and TTS by |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Methods for Speech-To-Text and Text-To-Speech Conversion | | h Journal of Engineering and Technology | | Wrapping | HMM with DNN in Python with Google's API is ideal. HMM is best for TTS using pyttsx3 or gTTS in Python. It's multilingual and adaptable for text and speech. | HMM, a web-based application can be created for sending and viewing voice-based messages. |
| 13. | Expediting Registration and Patient Identification with Face Recognition | 2020 | International Research Journal of Engineering and Technology | - | Several face recognition algorithms are also used in many different applications apart from biometrics, such as video compressions. | In conclusion, implementing a face recognition system for patient registration and identification is crucial for modern healthcare institutions. It streamlines operations, enhances user interactions, and improves efficiency. This decision benefits patients and simplifies hospital management, with many clinics already adopting and expanding these projects. | Proper dataset and algorithms aren't mentioned in the paper that have been used in the making of the project. |
| 14. | Text to SQL Query Conversion Using Deep Learning: A Comparative | 2019 | International conference on systems energy and environment | The dataset used in this research paper is called Spider. It is a large-scale,cross- | The proposed system in this research paper uses a deep neural network-based approach for text-to-SQL query conversion. Specifically, the system uses a sequence-to- | The proposed deep learning-based approach for text-to-SQL query conversion using a sequence-to-sequence model with attention mechanism is effective and outperforms | The proposed system is evaluated only on the Spider dataset, which is a specific dataset with certain |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Analysis | | | domain semantic parsing and text-to-SQL dataset that contains multi-table databases and SQL queries including many complex SQL components. | sequence model with attention mechanism, which is a type of recurrent neural network (RNN) that is commonly used in natural language processing tasks. The model takes a natural language question as input and generates a corresponding SQL query as output. | previous state-of-the-art methods on the Spider dataset. The results show that the proposed system can generalize well to new domains and achieve high accuracy on complex SQL queries with multiple tables. The authors also suggest that there is still room for improvement in this semantic parsing task and that future work could explore the use of additional features and techniques to further enhance the performance of the system. | characteristics, and it remains to be seen how well the system would perform on other datasets or in real-world applications. Therefore, future research could focus on evaluating the proposed system on other datasets and in real-world scenarios to assess its generalizability and practicality. |
| 15. | An online intelligent electronic medical record system via speech recognition | 2022 | International Journal of Distributed Sensor networks | - | Finite State Automata network as a language model to enhance recognition in the speech recognition model. Additionally, Segmentation algorithms for semantic analysis. | The brief overview of the paper's focus is on an online intelligent electronic medical record system via speech recognition, which aims to reduce the time spent by healthcare workers in filling | The proposed speech recognition system still needs to be observed in more clinical scenarios, and the authors plan to |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | out electronic medical records. The paper also mentions the use of a learning accumulation of scenarios in different medical sections and departments to achieve a formulated network in connection with the speech recognition method. | continue studying further enhancements of the speech recognition algorithm, enlargements of the medical term base, and upgrades of recognition models in the future. |
| 16. | Text and Voice Conversion for Machine Recognition using NLP | 2023 | Conference paper, August 2023, Researchgate | - | ● Text Normalization: Text normalization is the initial step in the process of converting human language into machine-level language.<br>● Bag of Words: After performing text processing, the normalized corpus is fed into the bag of words algorithm. This algorithm generates a list of unique words from | This research paper uses various NLP based algorithms such as text normalization, Bag of Words, TF-IDF, etc to extract important features and details from data sets. NLP helps overcome the difficulties of teaching computers to understand and communicate in natural languages, although there are still challenges to be addressed. | Despite advancements in NLP and its sister field, Natural Language Understanding (NLU), there are still significant challenges in fully comprehending and communicating in human language. These challenges include the complexity and inconsiste |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | the corpus along with their frequencies<br>● Term Frequency (TF): Term frequency is the frequency of a word in a document. After implementing the bag of words algorithm, the document vector table is obtained<br>● Term Frequency - Inverse Document Frequency (TF-IDF): TF-IDF is a measure used to evaluate the importance of a word in a document within a collection of documents. It is calculated by multiplying the term frequency with the inverse document frequency. | | ncy of languages, as well as the difficulty of teaching computers to comprehend and communicate in a manner similar to humans. |
| 17. | On the integration of | 2000 | Library and Information | - | The commonly used algorithms mentioned in the | The research paper proposes enhancements to the interaction | The research gap in integrating |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Text mining and Database systems | | Science Research journal | | document include:<br>• Naive Bayes: This algorithm is based on Bayes' theorem and assumes that the features (words) are conditionally independent given the class label.<br>• Rocchio: This algorithm uses a vector space model to classify documents based on their similarity to prototype documents representing each class.<br>• Widrow-Hoff: This algorithm, also known as the perceptron algorithm, is a linear classifier that adjusts its weights based on the error between predicted and actual class labels.<br>• Decision Trees: This algorithm builds a tree-like model of decisions and | between Text Mining applications and database systems by defining a primitive for the classification task in semistructured data using the XQuery language. The goal is to improve the integration between classification algorithms and the database system, as well as adapt these algorithms to handle large volumes of data. | text mining and database systems lies in the need for enhancements to the interaction between these two domains. Additionally, there is a need for implementing and testing real-world applications to validate the utility and generality of such a primitive. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | their possible consequences to classify documents.<br>● Support Vector Machines (SVM): This algorithm finds a hyperplane that separates documents into different classes with the maximum margin.<br>● Rule Induction: This algorithm generates a set of rules based on the features of the training documents to classify new documents.<br>● k-Nearest Neighbor (k-NN): This algorithm classifies documents based on the majority class of its k nearest neighbors in the feature space. | | |
| 18. | A Facial Recognition Mobile App | 2019 | Institute of Human Environment Interfac | Generic patient data | No specific algorithms were used. Android app was built and FRS was | It is possible to correctly identify both outpatients and inpatients and also reduce the unnecessary cost | It requires specific sensors and motions for patient |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | for Patient Safety and Biometric Identification: Design, Development, and Validation. | | e Biology Department of Plastic and Reconstructive Surgery | | used for data processing. | of patient verification by using the mobile facial recognition app with great accuracy. Our mobile app can provide valuable aid to patient verification, including when the patient is unconscious, as an alternative identification method. | verification. Patients in dead or unconscious state will be unable to get themselves recognized. Also, CT scan images would be most accurate but would also be expensive to use in place of mobile images. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 19. | Efficient Methods and Implementation of Automatic Speech Recognition System. | 2016 | International Journal of Computer Science and Information Technologies | Audio files to train and detect speech and voices. | Certain methods are used such as Stochastic method To deal with incomplete information, stochastic methods characterise the use of probabilistic models. Incompleteness arises from several sources in speech recognition Template Method In this method, comparison is takes place between the template set and input audio according to find most similar match. Since, last six decades, a number of methods and techniques are broadly provided for this method to sound recognition. In this approach, speech voice is handcoded from knowledge of experts as speech varies that guidance should be taken. Learning based approach This approach has overcome some disadvantages of | In this paper, the concept of ASR and various techniques are used for the system. Also focuses on the developing and creating system that should be robust with issues with variability in speaker characteristics, language characteristics, background noise, lack of vocabulary. In this paper, there are several properties and methods are define for the feature exraction such as LPC, MFCC,LDA,PCA . In the feature extraction MFCC is used in several applications. HMM is the best technique for the language modeling. And Viterbi algorithm is used to mapping for better result and reliable system. For the desirable result to make a system robust different methods have done. In our experiment accuracy is improved of system. By using Viterbi algorithm | To make a system more reliable vocabulary can be increase of various languages by that system will support more than one language at a time. The accuracy of the experimental results is never above 80% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | hidden markov model for machine learning which introduced natural language and genetic algorithm | it results to speed up a system | |
| 20. | A novel approach for learning ontology from relational database: from the | 2021 | Journal of Big Data | - | T-Box Generation and A-box Generation. | To sum up, in this paper, the autos have tried to gather the most important and contributing approaches in the subject of the mapping of the relational database to ontology. | A significant research gap in the study is the need for further exploration into scalability and efficiency challenges when mapping |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | construction to the evaluation | | | | | | relational databases to ontologies, particularly for large and complex datasets. Future research should focus on methods to address these scalability issues, given the growing volumes of data, to optimize ontology construction processes. |

## 2.2 Research Gap:

In India, existing literature lacks projects similar to DocAssist, focusing on real-time doctor-patient conversation documentation and data extraction. While Electronic Health Records (EHRs) are studied, the integration of Natural Language Processing (NLP) and speech-to-text technology is notably absent. DocAssist has the potential to address this gap with a tailored solution for the Indian healthcare context. Additionally, speech to database direct conversion is unavailable.

Currently, speech-to-text technology stands as the predominant solution for transcribing conversations, catering to a broad user base. There is no direct transformation from Speech to Database. Additionally, Existing systems do not focus on the healthcare industry and its information extraction and organization.

# Chapter 3

## Problem Statement

### 3.1 Drawbacks of the Existing System

Hospitals face a persistent challenge in maintaining accurate and organized patient records. The reliance on manual data entry and paper-based systems can result in errors, delays, and inefficiencies. Despite the progress in speech recognition and text extraction technologies, a notable gap exists in the integration of these advancements into a cohesive solution tailored for hospital environments. Existing systems often lack the ability to effectively store and manage extracted data, leaving healthcare professionals with fragmented records.

### 3.2 Solution To Above Problem

**Objectives:**

The overarching aim of this project is to develop a comprehensive system that converts doctor-patient conversations into structured, digital records stored in a database. Specific objectives include designing a user-friendly interface for healthcare professionals, implementing robust speech recognition algorithms, and ensuring the secure and compliant storage of patient data. Achieving these objectives will empower hospitals to manage patient records more effectively and improve the overall healthcare experience.

**Scope:**

The scope of this project extends to healthcare facilities of all sizes, from small clinics to large hospitals. Its significance lies in its potential to enhance patient data management, reduce administrative overhead, and facilitate data-driven decision-making in healthcare. By automating the conversion and organization of patient information, our solution has the relevance and significance to revolutionize healthcare record-keeping practices, benefiting both medical professionals and patients alike.

**Proposed Architecture:**

The proposed architecture of our system comprises three key components: speech recognition, data extraction, and database management. Speech recognition algorithms will transcribe audio conversations into text, data extraction techniques will identify and organize relevant information, and a secure database will store the structured data. This architecture ensures that the system can adapt to different healthcare environments while maintaining data integrity.

**Social Relevance:**

Our solution is highly applicable, offering usability for streamlined healthcare processes, scalability to meet varying facility sizes, and economic and environmental sustainability through efficient record-keeping practices.

**Software and Hardware Requirements:**

Software -

Frontend:

- HTML, CSS, JS, Bootstrap

Backend:

- Flask
- TensorFlow
- PyTorch
- Hugging Face

Database

- PostgreSQL

Tools:

- Visual Studio Code
- Postman
- Jupyter
- Google Colab

Hardware -

- Camera (For login using FRS)

- Microphone
- Intel Core- i5 or equivalent
- 8 GB RAM

**Timeline:**

Sem 7 -

1) Speech-to-Text conversion.
2) Removal of Noise from the audio.
3) Adding additional language for support (Hindi).
4) Designing the UI for booking appointments and displaying patient data.

Sem 8 -

1) Text to Database conversion.
2) Build the Data Extraction Model and test it. Extract patient demographics, symptoms, diseases, past-current and prescribed medications, immunization details and prescribed tests.
3) Appointment booking with different interfaces for doctor and patient for data security.
4) Training the model to improve the ‘Attention Mechanism’ in the model.
5) Improved the UI.
6) Integrate the model and frontend.
7) Deploy the final application.

# Chapter 4

# Project Description

## 4.1 Overview of the project

The flow of the "DocAssist: AI Doctor’s Assistant" project is designed to streamline healthcare workflows, particularly in the areas of audio data processing, transcription, sentiment analysis, and speaker segmentation. The project's workflow can be divided into several key steps:

1. *Data Ingestion:*
    a. The project starts with the collection of audio recordings, potentially created during patient interactions.
    b. These audio recordings serve as the primary data source for further analysis and processing.

2. *Audio Processing:*
    a. The audio data is processed using libraries such as librosa to prepare it for subsequent analysis.
    b. This may include tasks like resampling the audio to a consistent sampling rate or trimming/padding to ensure uniform duration.

3. *Speech-to-Text Conversion:*
    a. One of the central tasks in the project is converting the spoken words in the audio recordings into written text.
    b. This is achieved using Automatic Speech Recognition (ASR) powered by pre-trained transformer models, such as OpenAI's Whisper.

4. *Emotion Analysis:*
    a. The system also incorporates sentiment or emotion analysis to assess the emotional content within the transcribed text.
    b. This step provides insights into the emotional context of patient interactions,

which can be valuable for healthcare professionals.

5. *Speaker Segmentation:*
    a. Another significant aspect of the project is the identification and segmentation of speakers within the audio recordings.
    b. This is accomplished through speaker diarization, a process that distinguishes and labels different speakers in the audio.

6. *Data Storage:*
    a. The resulting transcribed text, sentiment analysis, and speaker segmentation data and extracted patient demographics, symptoms, diseases, past-current and prescribed medications, immunization details and prescribed tests are stored in a database. This data repository becomes a comprehensive record of healthcare interactions.

7. *Web Interface:*
    a. The project offers a web-based interface, where healthcare professionals can access and manage patient appointments.
    b. Additionally, the interface provides the functionality to record and transmit audio notes, simplifying the documentation process.

8. *User Interaction:*
    a. Healthcare practitioners use the web interface to view patient appointments and access related information.
    b. They can record audio notes during appointments or other interactions with patients.

9. *Reporting and Analytics:*
    a. The system incorporates reporting and analytical features to help healthcare professionals extract valuable insights from the stored data.
    b. These insights can aid in decision-making and improving patient care.

## 4.2 Module Description

### 4.2.1 Modules

The "DocAssist: AI Doctor’s Assistant" project is composed of several modules and sub-modules to efficiently manage audio data, transcribe it, analyze emotions, perform speaker segmentation, and provide a user interface. Here's a breakdown of the main modules and their sub-modules:

Main Modules:

Audio Processing Module:

- Sub-module: Audio Data Ingestion
- Sub-module: Audio Preprocessing (e.g., resampling, trimming, padding)

Speech-to-Text Module:

- Sub-module: Automatic Speech Recognition (ASR)
- Sub-module: Text Tokenization

Emotion Analysis Module:

- Sub-module: Sentiment Analysis
- Sub-module: Emotional Content Assessment

Speaker Segmentation Module:

- Sub-module: Speaker Diarization

Data Extraction Module:

- Extract patient demographics and disease, symptoms, tests, medication etc.

Data Storage and Management Module:

- Sub-module: Database Integration
- Sub-module: Data Storage
- Sub-module: Data Retrieval

User Interface Module:

- Sub-module: Web-Based User Interface Design
- Sub-module: Appointment Management
- Sub-module: Audio Recording and Transmission

Reporting and Analytics Module:

- Sub-module: Data Analysis
- Sub-module: Report Generation

- Sub-module: Insights and Analytics

Key Features and Functionality of Each Module:

1. Audio Processing Module:
    a. Responsible for collecting and preparing audio data for further analysis.
    b. Ensures audio files are standardized in terms of sampling rate and duration.
2. Speech-to-Text Module:
    a. Converts spoken words in audio recordings into written text.
    b. Sub-module, ASR, harnesses transformer models for accurate transcription.
    c. Text Tokenization enables further text analysis.
3. Emotion Analysis Module:
    a. Analyzes the emotional content of the transcribed text.
    b. Provides insights into the sentiment and emotional context of patient interactions.
4. Speaker Segmentation Module:
    a. Identifies and segments different speakers within audio recordings.
    b. Utilizes speaker diarization techniques.
5. Data Extraction Module:
    a. The resulting transcribed text, sentiment analysis, and speaker segmentation data and extracted patient demographics, symptoms, diseases, past-current and prescribed medications, immunization details and prescribed tests are stored in a database.
6. Data Storage and Management Module:
    a. Integrates with a database system for data storage.
    b. Manages the storage and retrieval of transcribed text, emotion analysis results, and speaker segmentation data.
7. User Interface Module:
    a. Designs a web-based user interface for healthcare professionals.
    b. Allows for appointment management and easy access to patient information.

c. Provides the ability to record and transmit audio notes during patient interactions.

8. Reporting and Analytics Module:

a. Analyzes stored data to extract valuable insights.

b. Generates reports for healthcare professionals to aid in decision-making and patient care improvement.

### 4.2.2 Block Diagram

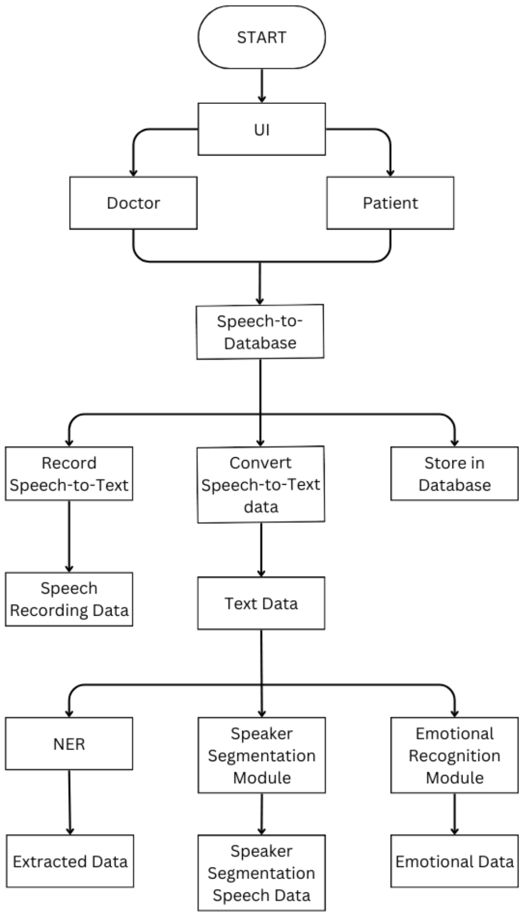


Figure 4.1: Block Diagram of DocAssist: AI Doctor’s Assistant System

### 4.2.3 UML Use Case Diagram

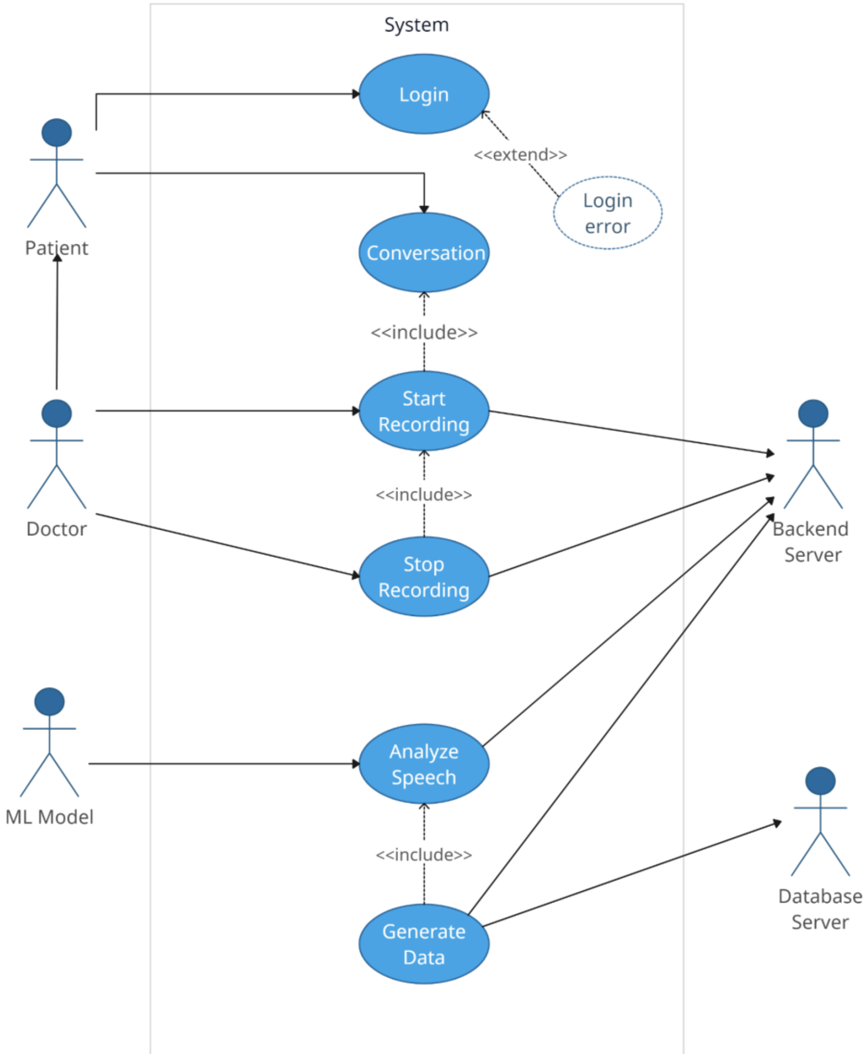


Figure 4.2: UML Use Case Diagram of DocAssist: AI Doctor's Assistant System

### 4.2.4 UML Class Diagram

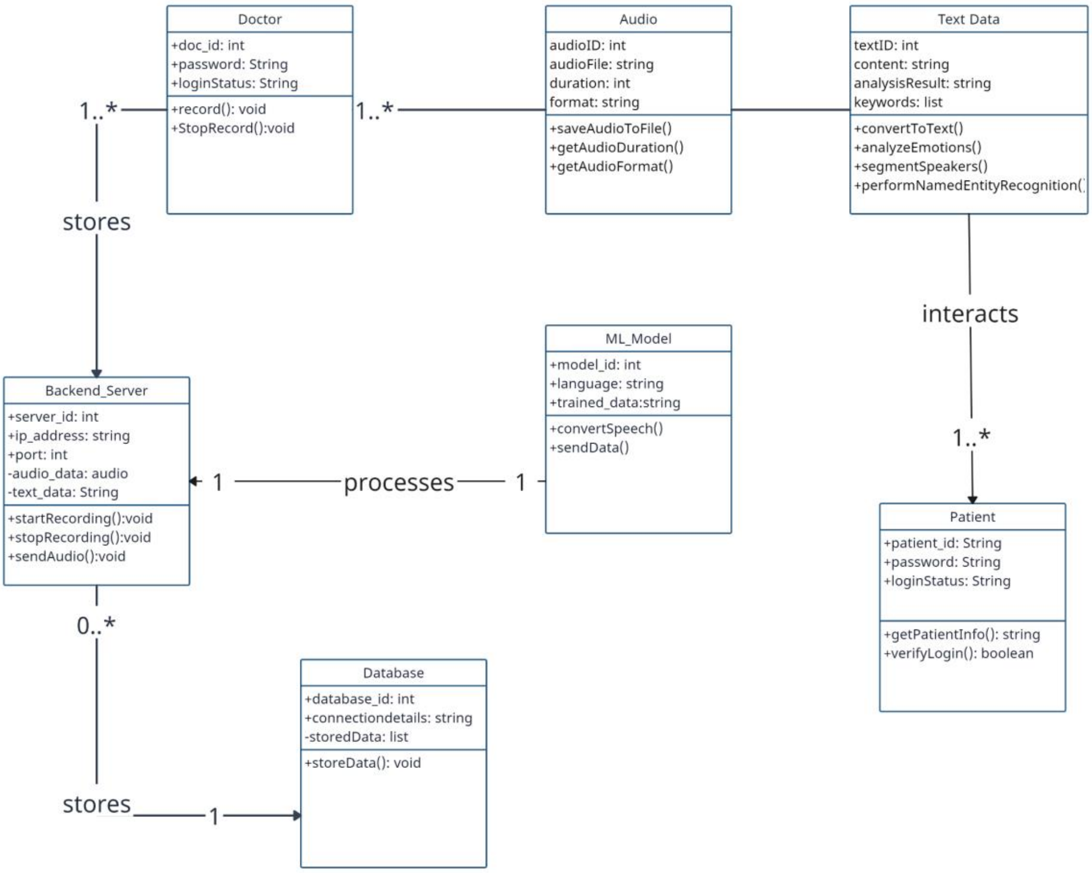


Figure 4.3: UML Class Diagram of DocAssist: AI Doctor's Assistant System

### 4.2.5 UML Sequential Diagram

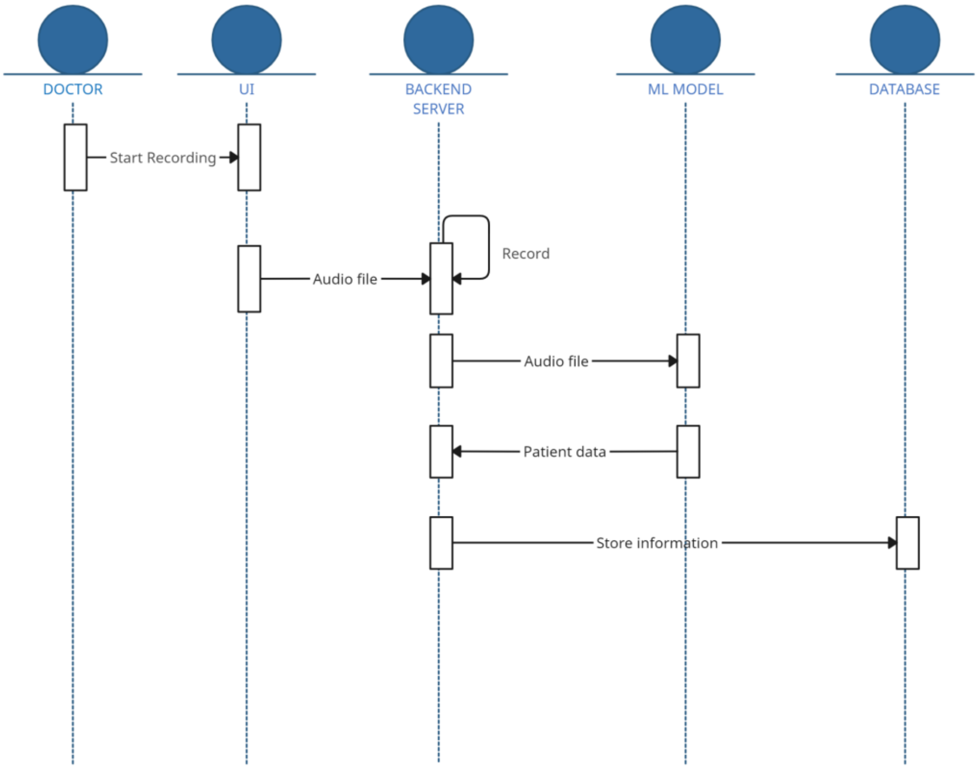


Figure 4.4: UML Sequential Diagram of DocAssist: AI Doctor's Assistant System

### 4.2.6 Data Flow Diagram

Level-0:

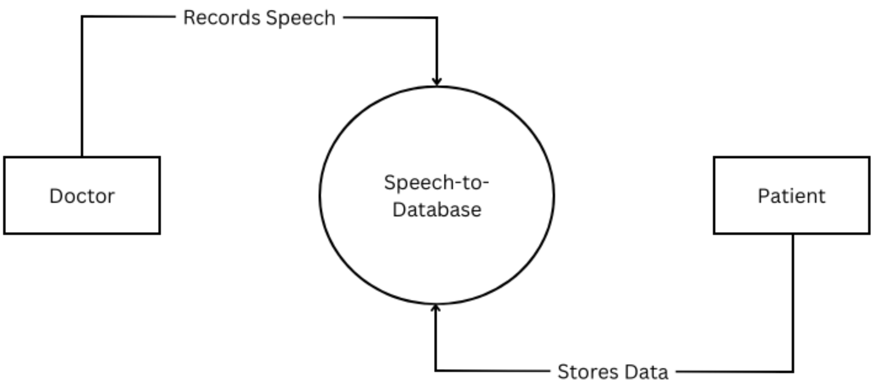


Figure 4.5: Data Flow Diagram of DocAssist: AI Doctor's Assistant System: Level-0

Level-1:

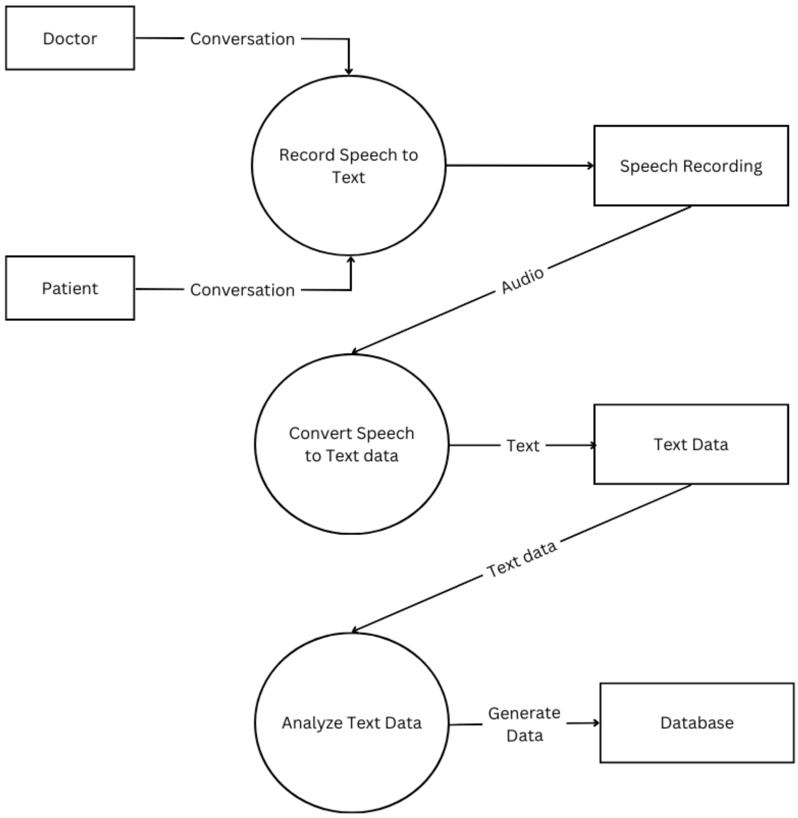


Figure 4.6: Data Flow Diagram of DocAssist: AI Doctor’s Assistant System: Level-1

Level-2:

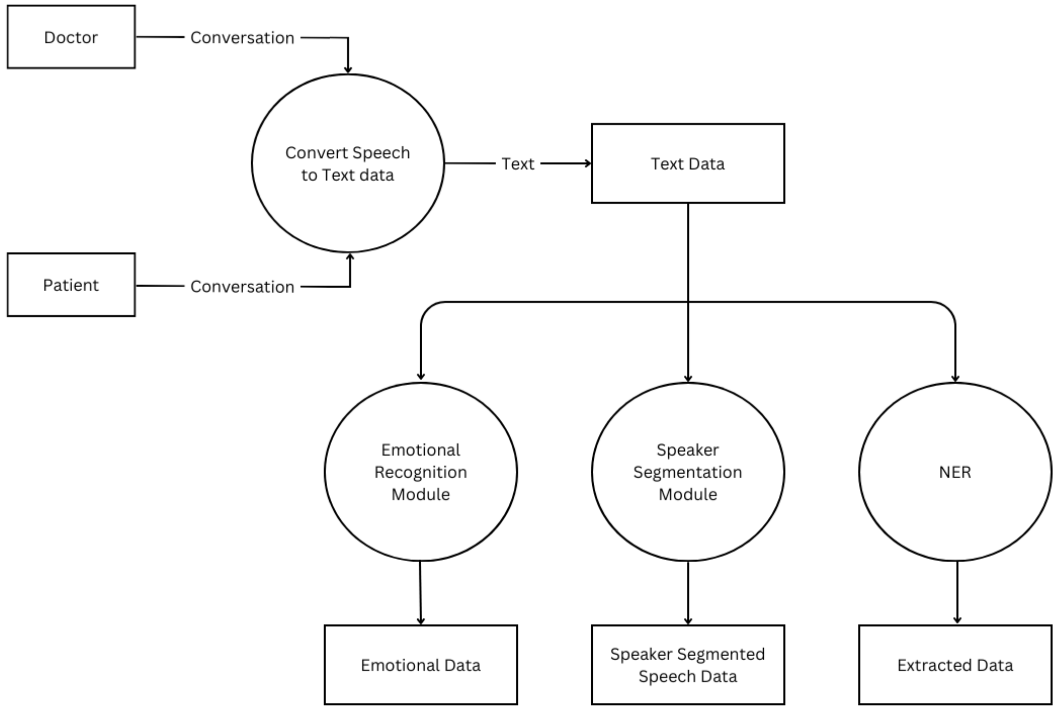


Figure 4.7: Data Flow Diagram of DocAssist: AI Doctor's Assistant System: Level-2

### 4.2.7 Transformer Model

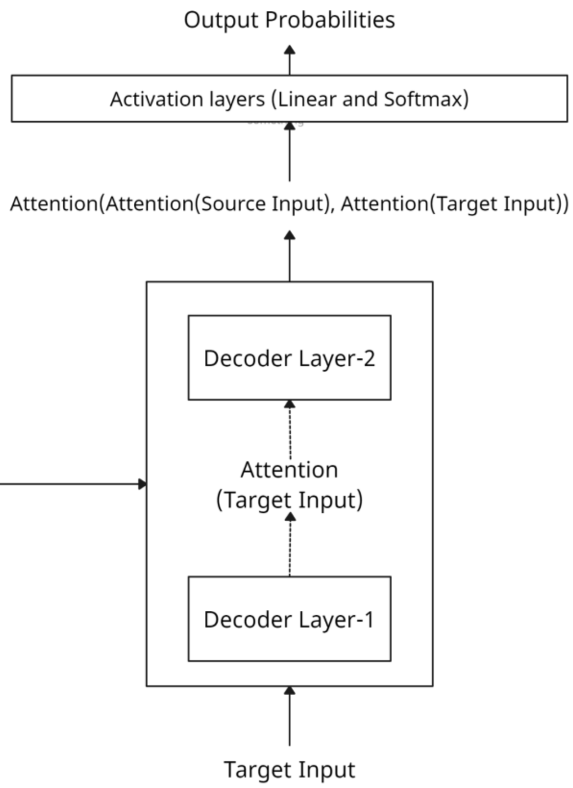


Figure 4.8: Transformer Model

### 4.2.8 Input Design

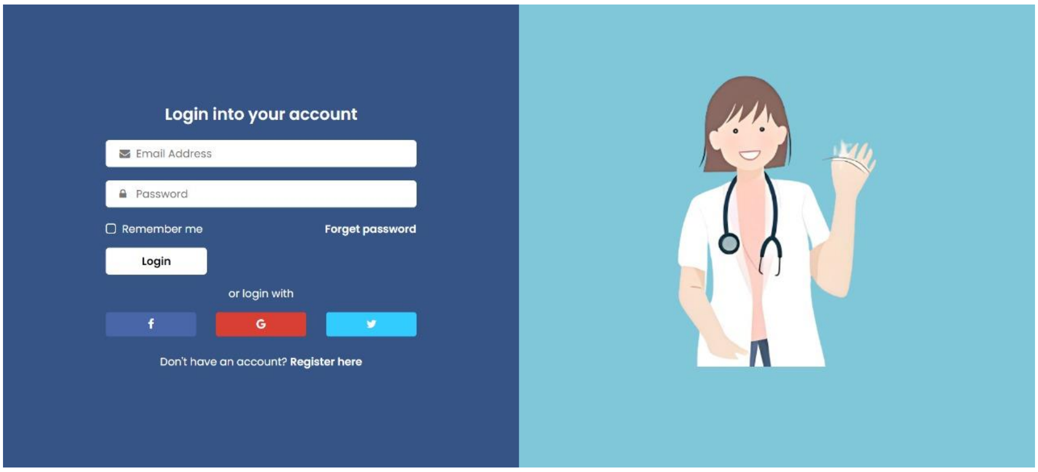


Figure 4.9: User View


Figure 4.10 : HomePage

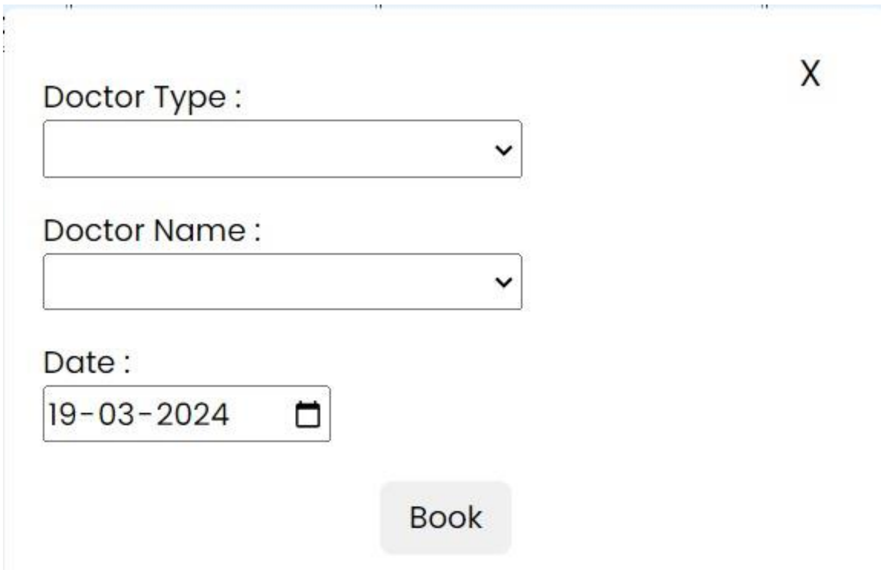


Figure 4.11 : Appointment Booking

### 4.2.9 Output Design

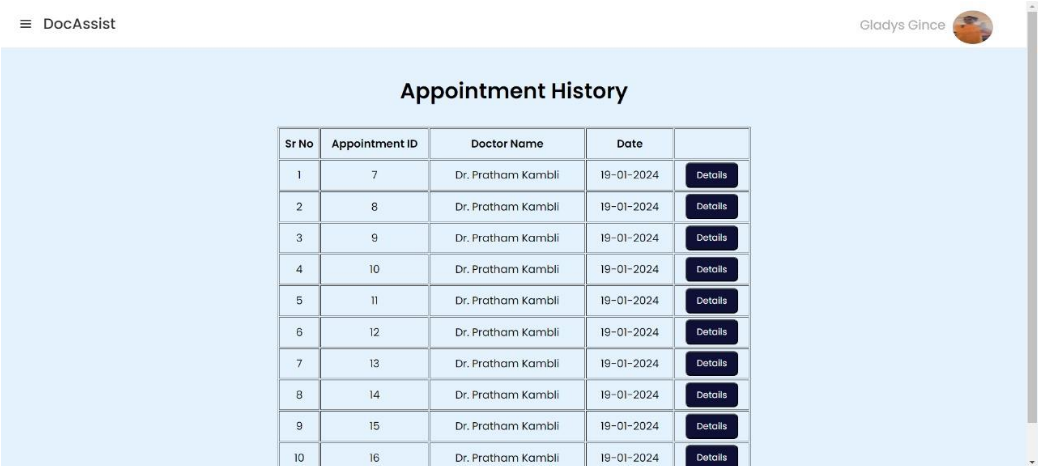


Figure 4.12: Appointment History


Figure 4.13: Profile

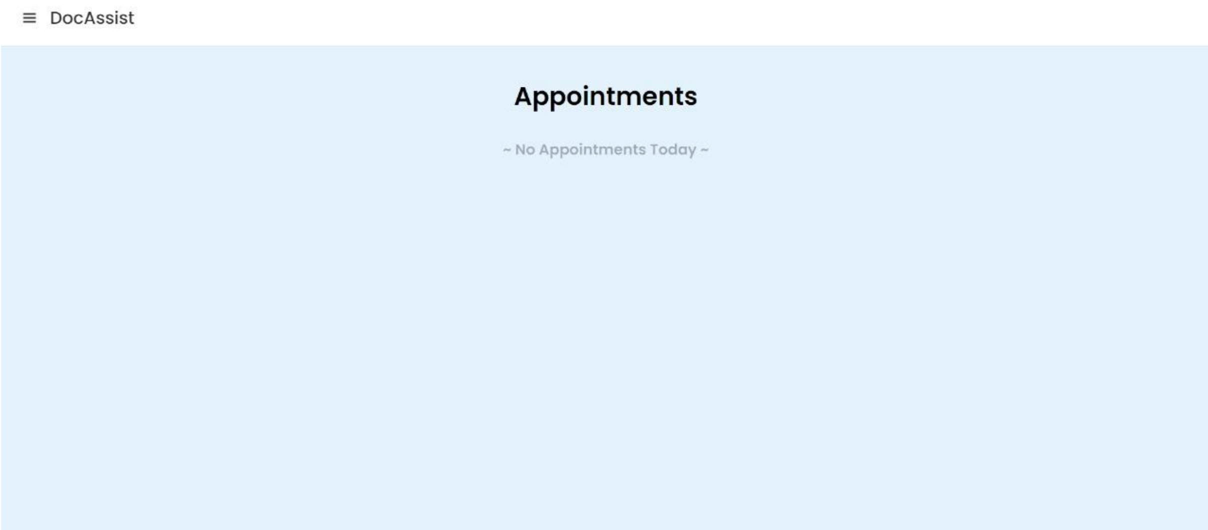


Figure 4.14: Doctor’s Home page


| Sr No | Appointment ID | Patient Name | Date | |
|---|---|---|---|---|
| 1 | 4 | 2 | 19-01-2024 | Details |
| 2 | 7 | 4 | 19-01-2024 | Details |
| 3 | 8 | 4 | 19-01-2024 | Details |
| 4 | 9 | 4 | 19-01-2024 | Details |
| 5 | 10 | 4 | 19-01-2024 | Details |
| 6 | 11 | 4 | 19-01-2024 | Details |
| 7 | 12 | 4 | 19-01-2024 | Details |
| 8 | 13 | 4 | 19-01-2024 | Details |
| 9 | 14 | 4 | 19-01-2024 | Details |
| 10 | 15 | 4 | 19-01-2024 | Details |

Figure 4.15: Doctor’s Appointment History

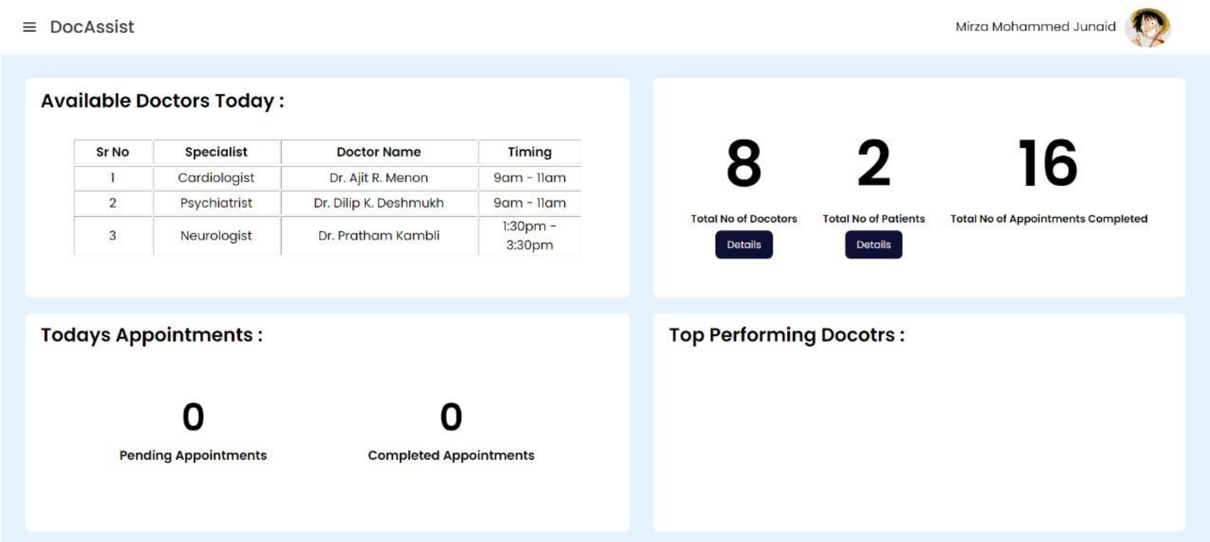


Figure 4.16: Admin Home Page

# Chapter 5

## Implementation Details

### 6.1 Sequential flow of the project:

1: Initialize the project and required libraries

2: Collect audio data from healthcare appointments

3: Preprocess audio data to ensure uniformity

4: Store the preprocessed audio data

5: Utilize Automatic Speech Recognition (ASR):

6: Choose an ASR model (e.g., OpenAI's Whisper) based on transformers

7: Transcribe audio data into text

8: Store the transcribed text data

9: Apply Emotion Analysis:

10: Apply emotion analysis to the transcribed text

11: Use a transformer-based sentiment analysis model to determine sentiment

12: Store the emotional analysis results

13: Perform Speaker Segmentation (Diarization):

14: Utilize deep learning techniques (e.g., neural networks) for speaker diarization

15: Identify and label different speakers within the audio recordings

16: Store the speaker segmentation results and data extraction of patient details.

17: Set up a database for data storage and management:

18: Integrate with an RDBMS (e.g., PostgreSQL or MySQL)

19: Create a database schema for appointment details, transcribed text, emotion analysis,

and speaker segmentation data
20:  Implement algorithms for efficient data storage and retrieval

21: Design a User Interface:
22:  Create a web-based user interface using HTML, CSS, and potentially JavaScript
23:  Develop a user-friendly interface for healthcare professionals to access and manage patient appointments
24:  Include features for recording and transmitting audio notes during appointments

25: Data Analysis and Reporting:
26:  Analyze stored data to extract insights
27:  Implement data analysis algorithms to derive trends and patterns from the transcribed text and emotional analysis results
28:  Generate comprehensive reports for healthcare professionals

29: User Interaction:
30:  Healthcare professionals access the web interface to manage appointments and review patient information
31:  They can initiate audio recordings and send them for transcription
32:  Transcribed text, emotion analysis results, and speaker segmentation data are presented through the interface

33: Data Retrieval and Reporting:
34:  Healthcare professionals can retrieve transcribed text, emotion analysis, and speaker segmentation results for specific appointments
35:  They can access reports and analytics generated from the stored data
36: End of Project

# Chapter 6

# Conclusion And Future Enhancements

## 6.1 Conclusion

In conclusion, the "DocAssist: AI Doctor's Assistant" project represents a transformational approach to health data management and analysis. By seamlessly capturing and visualizing audio from healthcare systems, using sentiment analysis and speaker classification, and providing user-friendly data, this project for healthcare professionals is able to access effective tools for real patient care and informed decision making. As it continues to evolve, the industry has the potential to advance healthcare services, enhance patient-provider relationships, and contribute to the future of healthcare through technology and data-driven insights that will be used for other purposes.

## 6.2 Future Enhancements

- Patient registration using FRS to be integrated on our website.
- Improvisation of UI.
- Adding regional languages.
- Chatbot for the website.

# References

[1] Kumar, Vijay and Singh, Hemant and Mohanty, Animesh, “Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer Using Natural Language Generation and Abstract Meaning Representation” (April 3, 2020). International Journal of Engineering and Advanced Technology (IJEAT) , Volume-9 Issue-4, April, 2020, Page no:2361-2365,

[2] Trivedi Ayushi ,Pant Navya, Shah Pinal, Sonik Simran and Agrawal Supriya. IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 2, Ver. I (Mar.- Apr. 2018), PP 36-43

[3] Das, Prerana & Acharjee, Kakali & Das, Pranab & Prasad, Vijay. (2015). “VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT.” Journal of Applied and Fundamental Sciences. 1. 2395-5562.

[4] Varathan, Kanya & Geetha, S.. (2008). “Information Extraction -a text mining approach.” 2007. 1111 - 1118. 10.1049/ic:20070576.

[5] Jain Nikhil, Goyal Manya, Gupta Agravi, Kumar Vivek, “Speech to Text Conversion and Sentiment Analysis on Speaker Specific Data”, International Research Journal of Modernization in Engineering Technology and Science, Volume:03/Issue:06/June-2021, Impact Factor- 5.354, e-ISSN: 2582-5208

[6] Jeon, Byoungjun & Jeong, Boseong & Jee, Seunghoon & Huang, Yan & Kim, Youngmin & Park, Gee Ho & Kim, Jungah & Wufuer, Maierdanjiang & Jin, Xian & Kim, Sang & Choi, Tae. (2019). “A Facial Recognition Mobile App for Patient Safety and Biometric Identification: Design, Development, and Validation”. JMIR mHealth and uHealth. 7. e11472. 10.2196/11472.

[7] Sujit, Kumar & Singh, Sujit & Kaur, Deepinder & Dhingra, Isha. (2023). “Text and Voice Conversion for Machine Recognition Using NLP.”

[8] T S, Anisha and P C, Rafeeque and Murali, Reena, “Text to SQL Query Conversion Using Deep Learning: A Comparative Analysis” (August 15, 2019). In proceedings of the International Conference on Systems, Energy & Environment (ICSEE) 2019, GCE Kannur,

Kerala, July 2019

[9] Xia, X., Ma, Y., Luo, Y., & Lu, J. (2022). “An online intelligent electronic medical record system via speech recognition”. International Journal of Distributed Sensor Networks. https://doi.org/10.1177/15501329221134479

[10] “Efficient Methods and Implementation of Automatic Speech Recognition System” Himali Junghare#1, Prof. Prashant Borkar*2

[11] Bezerra, Eduardo & Mattoso, Marta. (2000). “On the integration of Text Mining and Database Systems”.

[12] Ben Mahria, B., Chaker, I. & Zahi, A. “A novel approach for learning ontology from relational databases: from the construction to the evaluation”. J Big Data 8, 25 (2021).

[13] Nagdewani Shivangi, Jain Ashika, “A Review on Methods for Speech-To-Text and Text-To-Speech Conversion”, International Research Journal of Engineering and Technology, Volume: 07 Issue: 05, May 2020, e-ISSN: 2395-0056, p-ISSN: 2395-0072

[14] Hoerbst A, Ammenwerth E. “Electronic health records. A systematic review on quality requirements”. Methods Inf Med. 2010;49(4):320-36. doi: 10.3414/ME10-01-0038. Epub 2010 Jul 6. PMID: 20603687.

[15] Zeng Z, Deng Y, Li X, Naumann T, Luo Y. “Natural Language Processing for EHR-Based Computational Phenotyping”. IEEE/ACM Trans Comput Biol Bioinform. 2019 Jan-Feb;16(1):139-153. doi: 10.1109/TCBB.2018.2849968. Epub 2018 Jun 25. PMID: 29994486; PMCID: PMC6388621.

[16] Adamson M, Choi K, Notaro S, Cotoc C. “The Doctor-Patient Relationship and Information-Seeking Behavior: Four Orientations to Cancer Communication. J Palliat Care”. 2018 Apr;33(2):79-87. doi: 10.1177/0825859718759881. Epub 2018 Mar 7. PMID: 29514545.

[17] Heisey-Grove D, Danehy LN, Consolazio M, Lynch K, Mostashari F. “A national study of challenges to electronic health record adoption and meaningful use”. Med Care. 2014 Feb;52(2):144-8. doi: 10.1097/MLR.0000000000000038. PMID: 24309669.

[18] Gupta UC. “Informed consent in clinical research: Revisiting a few concepts and areas”. Perspect Clin Res. 2013 Jan;4(1):26-32. doi: 10.4103/2229-3485.106373. PMID: 23533976; PMCID: PMC3601699.

[19] Das, Prerana & Acharjee, Kakali & Das, Pranab & Prasad, Vijay. (2015). "VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT." Journal of Applied and Fundamental Sciences. 1. 2395-5562.

[20] Pahwa, Rohit & Tanwar, Harion & Sharma, Sachin. (2020). "Speech Recognition System: A review. International Journal of Future Generation Communication and Networking". 13. 2547-2559.

## 7% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- 1% Publications database
- Crossref database
- Crossref Posted Content database
- 3% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

| | Source | Type | Similarity |
|---|---|---|---|
| 1 | 1library.net | Internet | 2% |
| 2 | de.slideshare.net | Internet | 1% |
| 3 | ijcsit.com | Internet | <1% |
| 4 | journals.dbuniversity.ac.in | Internet | <1% |
| 5 | doaj.org | Internet | <1% |
| 6 | City University on 2023-01-04 | Submitted works | <1% |
| 7 | iosrjournals.org | Internet | <1% |
| 8 | journalofbigdata.springeropen.com | Internet | <1% |

| | | |
|---|---|---|
| 9 | University of Canberra on 2024-01-18<br>Submitted works | <1% |
| 10 | core.ac.uk<br>Internet | <1% |