GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

Private University Estd. in Karnataka State by Act No. 41 of 2013

BANGALORE

PRESIDENCY GROUP OVER 40 YEARS OF ACADEMIC WISDOM

A Project Report

On

“Patient Case Similarity”

Batch Details

Sl. No.	Roll Number	Student Name
1	20211CAI0062	Pranav Ganesh
2	20211CAI0063	Prerna Kakade
3	20211CAI0069	Bhuvana V
4	20211CAI0085	Nida Aiyman

School of Computer Science,

Presidency University, Bengaluru.

Under the guidance of,

Dr. Mohammadi Akheela Khanum

School of Computer Science and Engineering,

Presidency University, Bengaluru

CONTENTS

1. INTRODUCTION

Patient Case Similarity is used in healthcare systems, particularly in clinical decision support systems, to give similarity scores between the new and old patients. In this project, we are developing a web application designed for doctors and researchers to enhance patient care and medical research by comparing a new patient's data with a historical patient. The data is gathered from electronic health records (EHRs) and various research papers. This is done to identify similar patterns and predict the disease. Further, we can also recommend treatment options based on the patterns. The main goal is to cluster patients based on heart diseases, diabetes and other diseases. After which we will improve the diagnostic accuracy, predictive models and optimize treatment plans for the patients.

2. LITERATURE REVIEW

Research Paper	Year	Advantages	Disadvantages
i) Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases	2008	• Improve case similarity measurement • Integrate natural language processing (NLP) for feature abstraction	The study acknowledges limitations, including the focus on only four feature types and the lack of contextual information in weighing features. Future research should explore the impact of additional features and contextual factors on similarity measures to enhance their applicability.
ii) Patient Similarity: Emerging Concepts in Systems and Precision Medicine	2016	• It aims to provide a foundation for the integration of computational tools and data analytics to enhance personalized healthcare. • The primary focus is on how patient similarity algorithms can transform medical decision-making by grouping patients.	• Data Challenges • Scalability Issues
iii) Patient similarity for precision medicine: A systematic review	2018	• The primary aim of this line of research is to improve clinical outcomes for individual patients through more precise treatment targeting by leveraging on genetic, biomarker,	• Limited Database Scope • Lack of Real-World Application

		phenotypic, or psychosocial characteristics. • That distinguish a given patient from others with similar clinical presentations	
iv) A patient-similarity-based model for diagnostic prediction	2019	• To simulate the clinical reasoning of doctors, retrieve analogous patients of an index patient automatically and predict diagnoses by the similar/dissimilar patients. • The main goal is to predict patient diagnoses by comparing the similarities between the clinical features of current patients and historical patient data.	• Limited Dataset Size • High Computational Costs
v) Measurement and application of patient similarity in personalized predictive, modelling based on electronic medical records	2019	• The main goal of this research was to create a new way to measure how similar patients are, based on the data from electronic medical records (EMRs). • By measuring patient similarity, the study aimed to improve how we predict a patient's health outcomes, specifically focusing on diabetes.	• The study didn't fully use all the available data when calculating patient similarity • The models didn't include specific exclusion criteria when choosing patients for the study.
vi) Measuring Patient Similarities via a Deep Architecture with Medical Concept Embedding	2019	• Develop a framework to measure clinical similarities between patients based on EHRs. • Preserve temporal information in patient data, which is often lost in existing models.	• Loss of temporal information in existing methods: • High dimensionality and sparsity

vii) Patient-Case Similarity	2020	• Develop a system to identify patients with similar medical histories. • Improve decision-making processes in clinical settings using patient data.	• Data Quality Dependency • Complex Medical Cases
viii) Patient similarity: methods and applications	2020	• Analyze and compute similarities between patients using electronic health records (EHRs), genetic, and other data. • Improve predictive models in healthcare by integrating patient-specific data from various sources.	• Information Loss • Complexity in Implementation
ix)Patient similarity analytics for explainable clinical risk prediction	2021	• To develop an explainable and interpretable Clinical Risk Prediction Model (CRPM) by leveraging patient similarity analytics, specifically to improve explainability and interpretability. • To use real-world data from electronic medical records of patients with type-2 diabetes, hypertension, and dyslipidemia in Singapore to develop and validate the patient similarity model	• Incomplete Variable Set • Static Data Usage
x)A Novel Patient Similarity Network (PSN) Framework Based on Multi-Model Deep Learning for Precision Medicine	2022	• Utilizes multi-model deep learning to identify similarities among patients. • This patient similarity network (PSN) approach aims to enhance precision medicine by combining	• Data Heterogeneity and Dimensionality • Limited Availability of Open Datasets

		multiple data types, such as clinical records, genetic information, and imaging data.	
xi) Deep Dynamic Patient Similarity Analysis: Model Development and Validation in ICU	2022	• Develop a Novel Dynamic Patient Similarity Model • Validate Model Using Clinical Tasks	• Limited Clinical Application • Computational Complexity
xii) Patient Case Similarity	2024	• Improve healthcare analytics by leveraging data science techniques to enhance diagnostics, treatment recommendations, and patient care outcomes • The system utilizes machine learning (ML) and natural language processing (NLP) to identify similarities between patient cases, enabling more personalized, data-driven medical interventions.	• Prescription Recommendation Accuracy • Scalability Concerns

3. OBJECTIVES

- The goal is to leverage historical patient data to improve diagnosis, treatment planning, and outcomes.
- To provide a system which facilitates data driven clinical decision making. This gives insights on patients with similar symptoms and provides treatment accordingly.
- Enhancing risk predictions - Improving predictive models by analyzing patient similarities, allowing accurate diagnosis across a variety of diseases, not limited to heart diseases and diabetes.
- Contributing to reducing costs and improving the overall quality of healthcare.

EXPERIMENTAL DETAILS:

Hardwares and Softwares used:

Hardware:-
Laptop (11th Gen Intel(R) Core(TM) i5-11320H @ 3.20GHz 2.50 GHz, 16 GB RAM)

Software:-

i. Operating System:Windows
ii. Programming Language: Python 3.7 or higher - The primary language for machine learning and deep learning development.
iii. Deep Learning Libraries and Frameworks:TensorFlow with Keras API for ease of use and model building.
iv. Machine Learning Libraries:
•Scikit-learn: For classical machine learning algorithms, data preprocessing, and basic model evaluation.
•NumPy, Pandas: For data manipulation and analysis.
•Matplotlib, Seaborn: For data visualization.
v. Database Systems:SQL Database - for structured patient data storage.
vi. Jupyter Notebooks: For running interactive Python code and sharing results with the research team.
vii. Git: For version control and collaboration.

4. METHODOLOGY

i. Data collection and Preprocessing:
- Using libraries such as pandas, numpy and scikit learn.
- Loading data from electronic health records and research papers.
- Cleaning and preprocessing the data.
- Splitting the data into training and test sets.

ii. Feature Engineering:
- Using the transformer library and converting text to embeddings which can be used as features.
- Normalizing numerical features and using the one-hot encoding technique for categorical features.

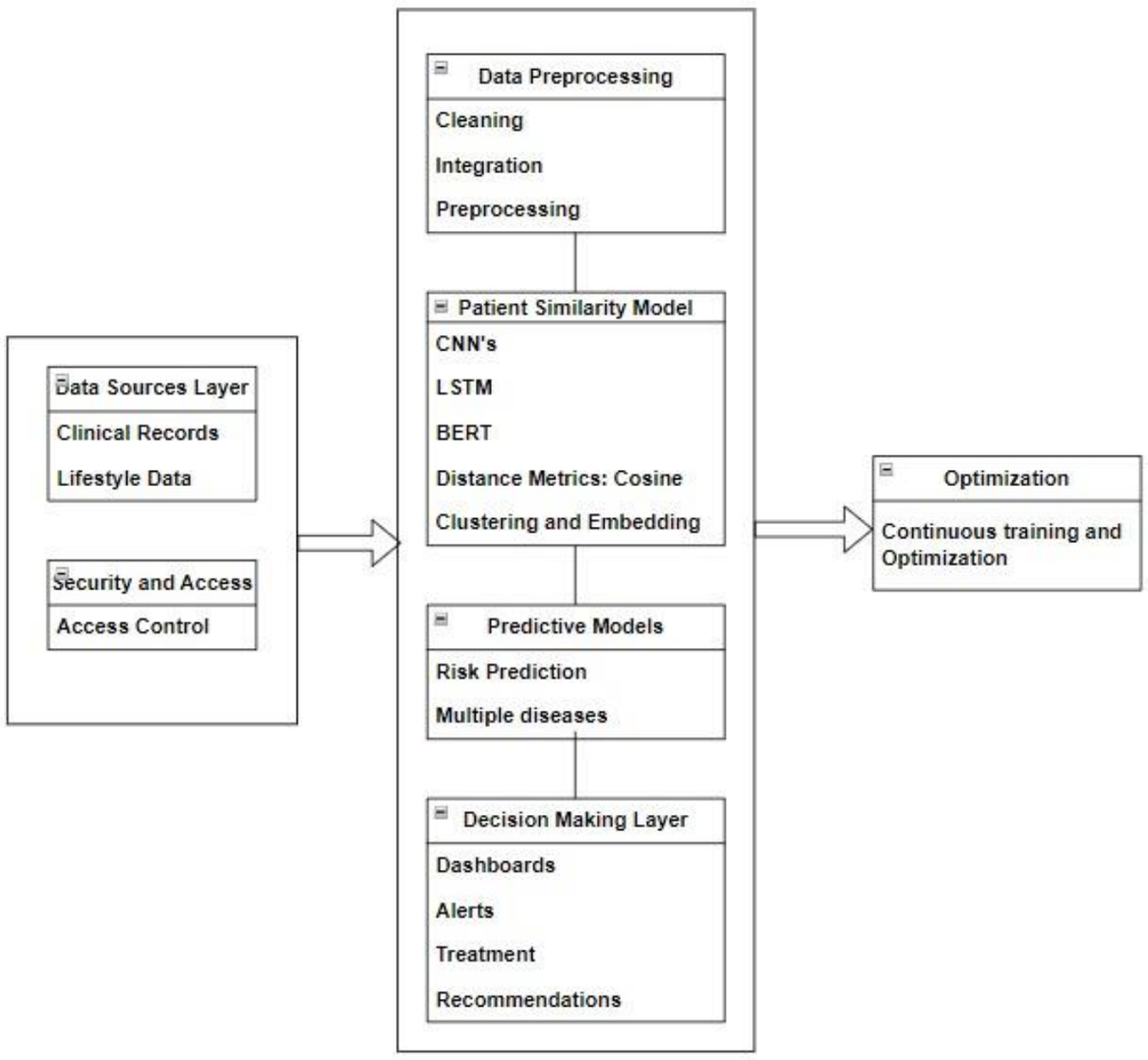
iii. Model Development:
- Exploring models such as CNN and LSTM with the help of libraries such as tensorflow or keras.
- CNN is used for extracting features from structured data.
- LSTM is used for sequential data such as time-series data of medical record.
- We plan to create a hybrid model which will integrate the outputs from CNN and LSTM and give predictions.

iv. Training and Evaluating the model:
- Training the model using test/train sets.
- For evaluation using metrics such as accuracy, precision, recall and F1 score.

v. Clusters and Similarity Score:
- Creating clusters by using clustering algorithms such as K-means and DBSCAN. K-means is a centroid based algorithm where as DBSCAN is a density based algorithm which can identify outliers and noise.
- Finding the similarity scores based on the type of diseases.

vi. Web Development:
- Frontend interfaces such as HTML and CSS will help in visualizing results and allow user interactions.
- APIs (Application Programming Interface) can be used to predict the model and cluster results. Django and flask are examples for APIs that can be used.
- The website will also display patient similarity scores and the treatment for diseases

of the patient.



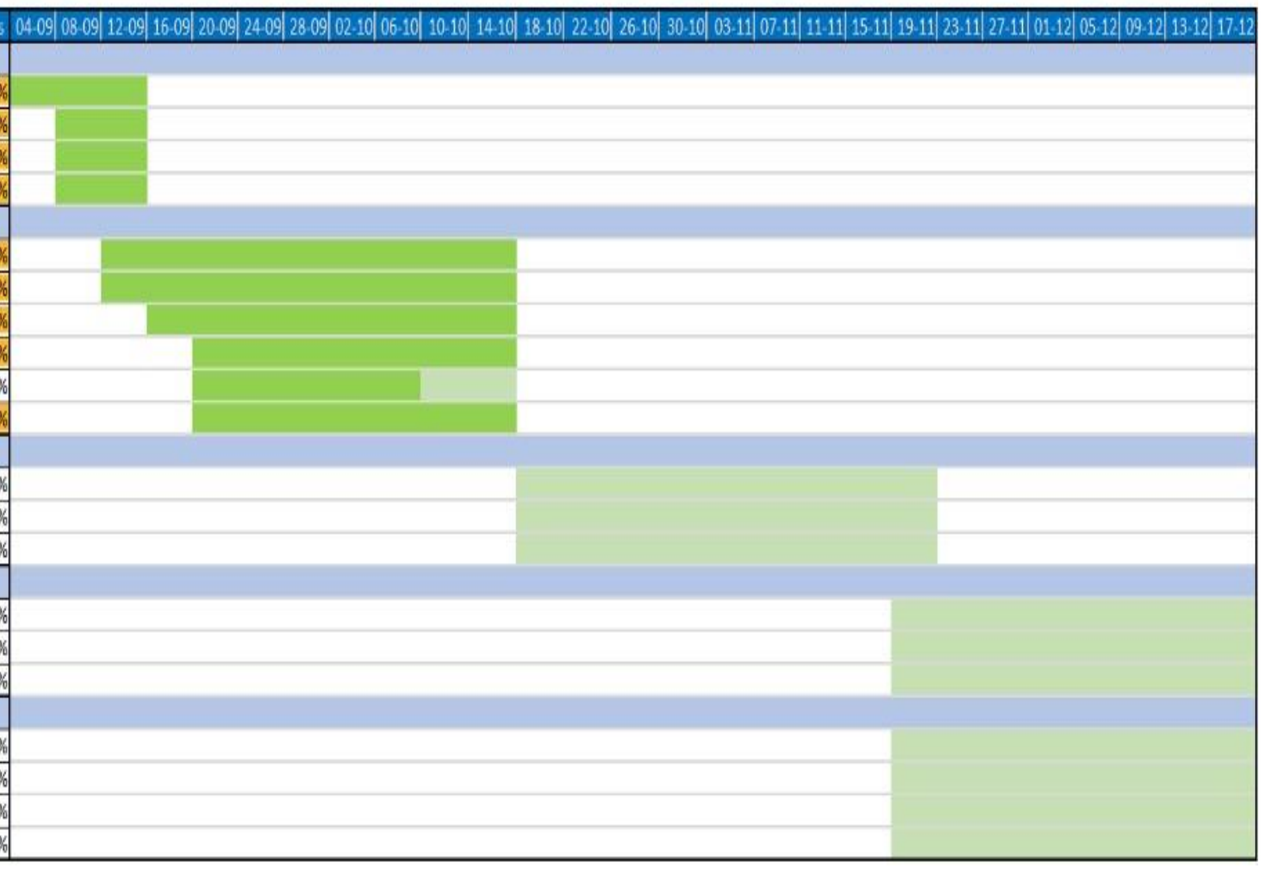
5. OUTCOMES

- To predict diseases like heart disease, diabetes, and other diseases based on patient similarity data with improved accuracy score using machine learning algorithms.
- The system will assist doctors by providing them with insights on similar patients histories, which will allow them to make data-driven diagnostic decisions.
- By implementing clustering techniques such as K-means or DBSCAN, the system will group patients based on similar symptoms, allowing for more personalized treatments.
- Our project will lead to the creation of an intelligent system that can recommend personalized treatment plans based on the analysis of similar patient cases.
- Our model will enhance the overall efficiency of healthcare processes.

6. TIMELINE OF THE PROJECT/ PROJECT EXECUTION PLAN

G24 PATIENT CASE SIMILARITY			
Project Start Date:	04-09-2024		Time allocated
Current Date:	14-10-2024		Progress
Weeks:	Wk 6		

Task	Start Date	End Date	Days	Progress
Review 0				
Title Selection & PPT	04-09-2024	13-09-2024	9	100%
GitHub Repository	05-09-2024	13-09-2024	8	100%
Finalizing Objectives	08-09-2024	13-09-2024	5	100%
Methodology	08-09-2024	13-09-2024	5	100%
Review 1				
Abstract	12-09-2024	15-10-2024	33	100%
Software Details	12-09-2024	15-10-2024	33	100%
Exploring Research papers	16-09-2024	15-10-2024	29	100%
Architecture Diagram	18-09-2024	15-10-2024	27	100%
Source Code Details	18-09-2024	15-10-2024	27	75%
Review 1 Report	18-09-2024	15-10-2024	27	100%
Review 2				
Algorithm	15-10-2024	19-11-2024	35	0%
50% Implementation of Code	15-10-2024	19-11-2024	35	0%
50% Report Completion	15-10-2024	19-11-2024	35	0%
Review 3				
100% implementation of Code	19-11-2024	17-12-2024	28	0%
100% Report Completion	19-11-2024	17-12-2024	28	0%
Live Demo	19-11-2024	17-12-2024	28	0%
Final Viva				
Live Demo	19-11-2024	17-12-2024	28	0%
Plagiarism Check	19-11-2024	17-12-2024	28	0%
Hard Copy Report Submission	19-11-2024	17-12-2024	28	0%
Publication of Research Paper	19-11-2024	17-12-2024	28	0%



7. CONCLUSION

Our project is mapped to SDG-3, that is, Sustainable Development Goals of Good Health and Well-Being. Our project combines advanced machine learning techniques and web applications to create a system for patient case similarity. Each area of healthcare can be iteratively improved based on performance metrics which helps achieve our goal of accurate similarity scores.

8. REFERENCES

i. **Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases – 2008: https://www.sciencedirect.com/science/article/pii/S153204640800044O**
ii. **Patient Similarity: Emerging Concepts in Systems and Precision Medicine – 2016: https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2016.00561/full**
iii. **Patient similarity for precision medicine: A systematic review – 2018: https://doi.org/10.1016/j.jbi.2018.06.001**
iv. **A patient-similarity-based model for diagnostic prediction – 2019: https://doi.org/10.1016/j.ijmedinf.2019.104073**
v. **Measurement and application of patient similarity in personalized predictive, modelling based on electronic medical records – 2019: https://doi.org/10.1186/s12938-019-0718-2**
vi. **Measuring Patient Similarities via a Deep Architecture with Medical Concept Embedding – 2019: https://arxiv.org/pdf/1902.03376**
vii. **Patient-Case Similarity – 2020: https://www.researchpublish.com/upload/book/Patient%20Case%20Similarity-8606.pdf**
viii. **Patient similarity: methods and applications – 2020: https://arxiv.org/pdf/2012.01976**
ix. **Patient similarity analytics for explainable clinical risk prediction – 2021: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-**

01566-y

x. A Novel Patient Similarity Network (PSN) Framework Based on Multi-Model Deep Learning for Precision Medicine – 2022: https://doi.org/10.3390/jpm12050768

xi. Deep Dynamic Patient Similarity Analysis: Model Development and Validation in ICU – 2022: https://www.sciencedirect.com/science/article/pii/S0169260722004151

xii. Patient Case Similarity - 2024: https://www.doi.org/10.56726/IRJMETS48246