PATIENT CASE SIMILARITY

A PROJECT REPORT

Submitted by,

Mr. Pranav Ganesh- 20211CAI0062
Ms. Prerna Kakade – 20211CAI0063
Ms. Bhuvana V – 20211CAI0069
Ms. Nida Aiyman – 20211CAI0085

Under the guidance of,

Dr. Mohammadi Akheela Khanum

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report **“PATIENT CASE SIMILARITY”** being submitted by “PRANAV GANESH, PRERNA KAKADE, BHUVANA V, NIDA AIYMAN” bearing roll numbers “20211CAI0062, 20211CAI0063, 20211CAI0069, 20211CAI0085” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

Dr. Mohammadi Akheela Khanum
Professor
School of CSE&IS
Presidency University

Dr. Zafar Ali Khan
Professor & HoD
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University

Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-Vc School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **PATIENT CASE SIMILARITY** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering – Artificial Intelligence and Machine Learning**, is a record of our own investigations carried under the guidance of **MOHAMMADI AKHEELA KHANUM, PROFESSOR, School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Name	**Roll Number**	**Signature**
Pranav Ganesh	20211CAI0062	
Prerna Kakade	20211CAI0063	
Bhuvana V	20211CAI0069	
Nida Aiyman	20211CAI0085	

ABSTRACT

The Patient Case Similarity web app helps doctors and researchers by comparing new patient data with past cases.

It uses electronic health records (EHRs) and medical research to find patterns, and predict diseases.

The web app groups patients with similar conditions, like heart diseases, to spot trends and improve diagnosis accuracy. It helps build better prediction tools to improve patient care.

First, we load the data into a program. Then, we look at the data to see patterns or connections. The data is initially loaded into the program. The patterns and connections are found out by looking at the data. We fix the missing values by removing them or filling them. We then make use of charts so that the data is more understandable. The data is split into two parts: training and testing. The model is trained with the first part and then checks how well it works with the second part. In the end, patterns are found out by grouping similar data.

Our innovative approach enhances the modern day medical decision which leads to better patient outcomes.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair,** School of Computer Science Engineering & Information Science, Presidency University, and Dr. Zafar Ali Khan, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr. Mohammadi Akheela Khanum, Professor,** and Reviewer **Dr. Afroz Pasha, Associate Professor,** School of Computer Science Engineering & Information Science, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman,** department Project Coordinator **Dr. Afroz Pasha** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Pranav Ganesh – 20211CAI0062

Prerna Kakade – 20211CAI0063

Bhuvana V – 20211CAI0069

Nida Aiyman – 20211CAI0085

LIST OF TABLES

Sl. No.	Table Name	Table Caption	Page No.
1	Table 2.1	Literature Survey	4-9

LIST OF FIGURES

Sl. No.	Figure Name	Caption	Page No.
1	Figure 3.1	Bar Plot of Age VS. Heart Rate	13
2	Figure 3.2	Histogram of Heart Rate	14
3	Figure 3.3	Pair plot of patient data	14
4	Figure 3.4	Scatter plot for K-Means Clustering	17
5	Figure 6.1	System Design	20
6	Figure 7.1	Timeline using Gantt Chart	22

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	Abstract	**iv**
	Acknowledgement	**v**
	List of Tables	**vi**
	List of Figures	**vii**
1.	**Introduction**	**1-3**
	1.1 General Introduction	1
	1.2 Project Specifics	1-2
	1.3 Services Provided	2-3
2.	**Literature Review**	**4-9**
3.	**Research Gaps of existing methods**	**10-11**
4.	**Proposed Methodology**	**12-17**
	4.1 Loading Data	12
	4.2 Exploring Data	12
	4.3 Missing Values	12-13
	4.4 Preprocessing	13
	4.5 Data Visualization	13-14
	4.6 Training and Testing	15
	4.7 Models	15-16
	4.8 Clustering	16
	4.9 Silhouette Score	17
5.	**Objectives**	**18-19**
6.	**System Design and Implementation**	**20-21**
7.	**Timeline for execution of project**	**22-23**
8.	**Outcomes**	**24-25**
	8.1 Outcome1	24
	8.2 Outcome2	24-25
9.	**Results and Discussions**	**26-27**
10.	**Conclusion**	**28-29**
11.	**References**	**30**

12	**Appendix-a**	**31-39**
	Front end	31-36
	Back end	37-39
13.	**Appendix-b**	**40-44**
14.	**Appendix-c**	**45-57**
	14.1 Journal Publication	45-50
	14.2 Plagiarism Check	51-55
.	14.3 SDG mapping	56-57

CHAPTER-1

INTRODUCTION

1.1 General Introduction

Artificial Intelligence and Machine Learning has had a great influence on humanity. It has effected and brought change to almost every single industry. It has also brought massive change and developments into the healthcare field. One of its most important applications in the healthcare field has been its ability to learn and predict from given data.

The Patient Case Similarity project helps the doctors and researchers in enhancing diagnostic accuracy and predicting diseases by leveraging these technological advancements.

Healthcare officials usually depend on Electronic Health Records (EHRs) and vast datasets to make reasoned judgements about the patient treatment plans. Regardless of how, analyzing this data manually can consume a lot of time and be more prone to errors. This challenge has been addressed by the Patient Similarity project by developing a web-based application which is extremely user-friendly, analyzes the patient data, identifies patterns, and predicts the outcomes based on how similar current case is compared to past cases.

The Patient Case Similarity project can be represented as a significant step ahead promoting Artificial Intelligence for healthcare analytics. By making use of several Machine Learning techniques and several data visualization tools, the system provides a trustable, user-friendly solution for improving diagnosis, the predictions and the research. In the future, many enhancements could be incorporated in the project, such as Deep Learning models, more comprehensive datasets for improving scalability and accuracy.

This system has many benefits not only limited to supporting medical professionals in making informed decisions but also by encouraging preventive care strategies which will ultimately lead to a better and healthier life of the patient.

1.2 Project Specifics

The system makes use of a publicly available datasets on Kaggle related to medical histories. The key attributes of this dataset include demographics, such gender and age. Medical history includes heart rate, glucose levels, cholesterol and blood pressure. Lifestyle factors like smoking habits and medication usage. The model provides accurate and reliable predictions using this dataset.

The methodology consists of various steps starting with data loading and data cleaning. Data is imported from CSV files and then structured into Data Frames. The missing values of the dataset are handled through imputation techniques such as median filling. Next, the Exploratory Data Analysis (EDA) is performed and visualized by making us of Histograms, Scatter Plots and Pair Plots in order to detect patterns and correlations. Finally, the data is split into training (70%) and testing (30%) to make sure of the reliability and accuracy in the Machine Learning models training and testing stage.

Machine Learning models like Logistic Regression, Decision Trees and Random Forest have been used for predicting the outcomes. Logistic Regression predicts binary outcomes. In this project, it can be used to predict the presence or absence of diseases. Decision Trees basically incorporate flowchart-based techniques which lead to classification. In this project, it can use factors like Cholesterol, Heart rate, Diabetes, to make a tree like structure which leads to an ultimate prediction. Random Forest is an algorithm that combines multiple Decision Trees in order to achieve greater stability and accuracy. Furthermore, K-means clustering has been employed to identify similar patterns and group similar patients together. The validity of clustering has been evaluated using Silhouette scores in order to ensure well-defined formations of groups.

The outcome of the prediction models provide likelihood scores of diseases and similarity assessments for new patients. Eventually, aiding doctors in diagnosing and treating patients more beneficially. This approach not only improves the state of diagnostic accuracy but also reduces the effort and time which manual analysis asks of.

1.3 Services Provided

- Assistance in Diagnosis: By clustering patients who have similar profiles, making it easier to identify the patients who are at risk of specific diseases like heart disease and diabetes, the system improves diagnostic accuracy.
- Predictive Modeling: The system makes use of historical data to train the Machine Learning models in order to predict patient outcomes. This helps the patient to find

out about the disease in early stages reducing the need for extensive diagnostic tests which leads to lowering of costs. It also helps them to come up with a plan.

- Decision Support System: This application acts as a Clinical Decision Support System (CDSS) by assisting healthcare officials such as doctors and nurses by providing diagnosis of diseases based on evidence.
- Facilitation of Research: The system can be used by researchers in order to identify trends and patterns in patient data enabling the development of new insights and solutions in the field of medicine. Hypothesis testing and validation is supported by the clustering feature.
- Accessibility: The system can be accessed from different devices and locations through the web-based application. The web application has a user friendly interface designed for medical professionals, researchers and also students.

CHAPTER-2

LITERATURE SURVEY

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 admits limitations and focuses on only four feature types and the lack of contextual information in weighing features. • Future research should examine the impact of additional features and contextual factors on similarity measures to upgrade their suitability.
ii) "Patient Similarity: Emerging Concepts in Systems and Precision Medicine"	**2016**	• The goal is to set the foundation for the integration of computational tools and data analytics to enhance personalized healthcare. • The main focus is how patient similarity algorithms can improve medical decisions by	• Data Difficulties • Growth Challenges

		grouping similar patients.	
iii) "Patient similarity for precision medicine: A systematic review"	**2018**	• The primary aim is to improve clinical results for individual patients through more accurate treatments by focusing on genetic, biomarker, phenotypic, or psychosocial characteristics. • This aim allows us to distinguish a given patient from others with similar clinical presentations.	• Limited Database Scope • Lack of Real-World Application
iv) "A patient-similarity-based model for diagnostic prediction"	2019	• To generate the clinical reasoning of doctors, retrieve data of an index patient automatically and predict diagnoses by similar or dissimilar patients. • The main goal is to predict patient diagnoses by comparing the similarities between the clinical features	• Limited Dataset Size • High Computational Costs

		of current patients and historical patient data.	
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 way to measure the similarity of patients using data from electronic medical records (EMRs). • By measuring similarity, the study aimed to improve predictions of health outcomes, particularly for 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 using a deep learning model with medical concept embedding”	2019	• Created a framework to measure clinical similarities between patients using EHRs. • Kept track of time-related information in patient data, which is often	• Loss of temporal information in existing methods: • High dimensionality and sparsity

		missed in other model.	
vii) "Patient-Case Similarity"	2020	• Build a system to identify patients with similar medical histories. • Improve decision-making processes in clinical environments using patient data.	• Data Quality Dependency • Complex Medical Cases
viii) "Patient similarity: methods and applications"	2020	• Study 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	2021	• To develop an understandable and interpretable Clinical Risk Prediction Model	• Incomplete Variable Set • Static Data Usage

prediction”		(CRPM) by leveraging patient similarity analytics, specifically to improve explain ability and interpretability. • To use real-world data from EMRs of patients with type-2 diabetes, hypertension, and dyslipidaemia in Singapore to develop and validate the patient similarity model	
x)”A Novel Patient Similarity Network (PSN) Framework Based on Multi-Model Deep Learning for Precision Medicine”	2022	• Uses multi-model deep learning to identify similarities among patients. • The patient Similarity Network (PSN) approach aims to improve precision medicine by using different types of data, like clinical records, genetic data, and imaging	• Different types and sizes of data • Limited access to public datasets
xi) “Deep Dynamic	2022	• Developed a Novel	• Limited Clinical

Patient Similarity Analysis: Model Development and Validation in ICU”		Dynamic Patient Similarity Model. • Validate Model Using Clinical Tasks.	Application • Computational Complexity
xii) “Patient Case Similarity”	2024	• Improve healthcare analytics by utilizing data science techniques to enhance diagnostics, treatment recommendations, and patient care outcomes • The system uses machine learning (ML) and natural language processing (NLP) to identify similarities between patient cases, enabling more personalized, data-driven medical action.	• Prescription Recommendation Accuracy • Scalability Concerns

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

i. “Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases” – 2008:-

The study only looks at four types of features and doesn't consider the context when weighing them. Future research should look at how adding more features and context can improve similarity measures

ii. “Patient Similarity: Emerging Concepts in Systems and Precision Medicine – 2016”:-

Complexity of algorithms – many of the proposed algorithms are not yet optimized for real-world clinical use due to their complexity and reliance on high-end computing infrastructure.

iii. “Patient similarity for precision medicine: A systematic review” – 2018:-

Lack of Deep Learning Exploration - The paper talks very little about deep learning, which is now important for analyzing complex medical data and finding patient similarities. This might be a missed chance to use better methods.

iv. “A patient-similarity-based model for diagnostic prediction” – 2019:-

Low Success Percentage - The model's success percentage (the percentage of patients for whom diagnoses were correctly predicted) is low (19%).

v. “Measurement and application of patient similarity in personalized predictive, modeling based on electronic medical records” – 2019 :-

The models didn’t include specific exclusion criteria when choosing patients for the study. This could affect the accuracy of the predictions because not all patients may be equally relevant for the predictive task

vi. “Measuring Patient Similarities via a Deep Architecture with Medical Concept Embedding” – 2019:

Electronic Health Records are complex, and patient records contain sparse and high-dimensional data.

vii. “Patient-Case Similarity” – 2020:

The use of machine learning models leads to over fitting, where the model performs well on training data but fails to generalize new or unseen data.

viii. “Patient similarity: methods and applications” – 2020

During data transformation and integration, particularly in early integration strategies, there is a risk of losing valuable patient information.

ix. “Patient Similarity Analytics for Explainable Clinical Risk Prediction” – 2021

The model doesn't include important factors like gender, race, diet, and lifestyle, which are linked to complications of diabetes, hypertension, and high cholesterol. Missing this data makes the model less complete.

x. “A Novel Patient Similarity Network (PSN) Framework Based on Multi-Model Deep Learning for Precision Medicine” – 2022

The mix of different types of clinical data, both structured and unstructured, makes it hard to create accurate models. Handling this complex data that can cause information to be lost during processes like auto encoders, which reduce the data's size. Using auto encoders for this can lead to a loss in accuracy.

xi. “Deep Dynamic Patient Similarity Analysis: Model Development and Validation in ICU” – 2022

Need for Clinical Protocol - The model requires a clinical protocol for practical implementation, which hasn't been discussed in this research

xii. “Patient Case Similarity” – 2024: Prescription Recommendation Accuracy: The system shows lower accuracy in prescription recommendations (61.27%), indicating room for improvement in this area.

CHAPTER-4
PROPOSED METHODOLOGY

4.1 Loading data:

- We have first imported the pandas library in Python which is used for data manipulation and analysis.
- The pd.read_csv() method has been used in order to load data from CSV files into pandas DataFrames.
- The DataFrame is a fundamental data structure in pandas so it becomes simple to read CSV data directly into the DataFrame.
- We have initially loaded two datasets : test data and train data. These two DataFrames are concatenated into a single DataFrame using the concat() function.

4.2 Exploring:

- Data.info() is used in the exploration step of the algorithm. This step gives us information about all the columns present in the dataframe.
- It confirms the data structure by displaying the class of the object.
- It lists down all the columns name is order for easy reference.
- It also displays the count of non-null values in each column which helps us to easily identify missing data.
- It shows the memory that the data frame has consumed.
- Data.describe() provides summary of statistics of all the numerical data in the dataframe such as count, mean, std, min, max and the percentiles.

4.3 Missing values:

- This method lets us know how many missing values are present in each column of the dataframe.
- The isnull() function generates missing values in the data frame with True indicating missing values and False indicating no missing values.
- The .sum() method gives the total number of missing values in each column by considering the number of True values.

- Columns that show 0 count are the columns that have no missing values.

4.4 Preprocessing:

- This is a crucial step in data analysis which prepares raw data for further analysis and modeling.
- An effective way to handle these missing values is by filling them by finding out the median value of each column.
- Certain columns like TenYearCHD, is_smoking and education are dropped because they provide unnecessary information.

4.5 Data visualization:

- In our project, we have used libraries such as matplotlib and seaborn.
- Matplotlib and seaborn are popular python libraries.
- Matplotlib is used for generating charts and plots.
- It allows customization of the charts and plots by adding titles, changing color of graph and giving labels on X axis and Y axis.
- Seaborn is built on top of matplotlib and is a library that is used to make complicated statistical visuals easier to create and understand.
- A bar plot has been created in order to compare the Age vs. Heart Rate of the patient and give us a detailed visual view of the same.

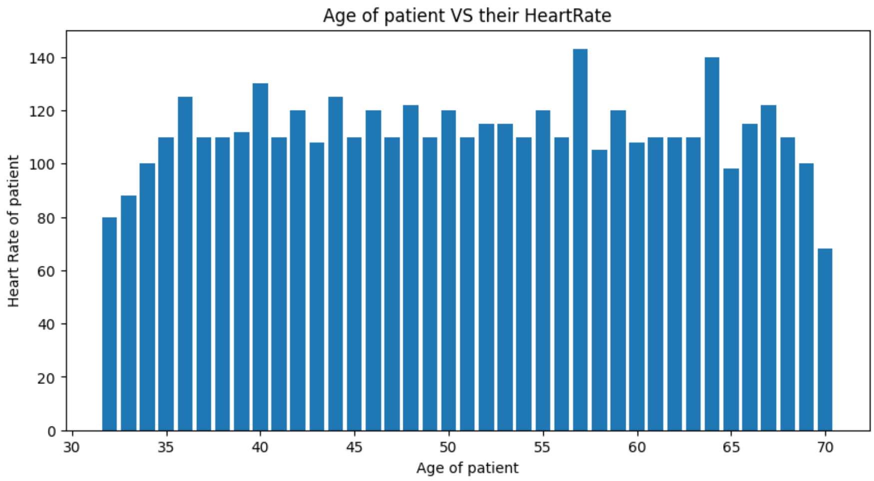


Fig 3.1: Bar Plot of Age VS. Heart Rate

- A Histogram has been created for checking the distribution of Heart Rates across various Frequencies. By looking at the Histogram it can be checked whether the heart rate is normally distributed, skewed or bimodal. If the histogram shows peak, then the data might be normally distributed.

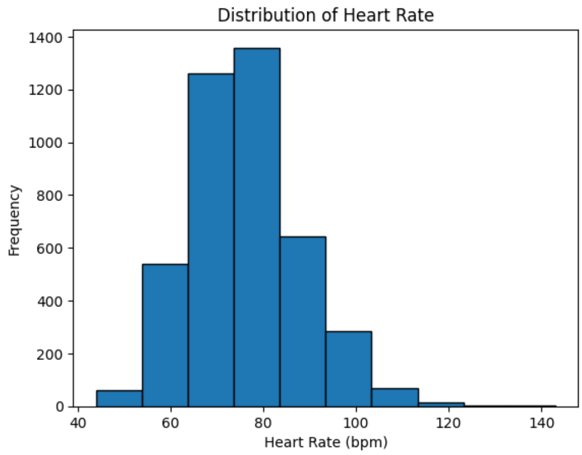


Fig 3.2: Histogram of Heart Rate

- A pairplot has also been created to show the relationships between variables in the dataset.

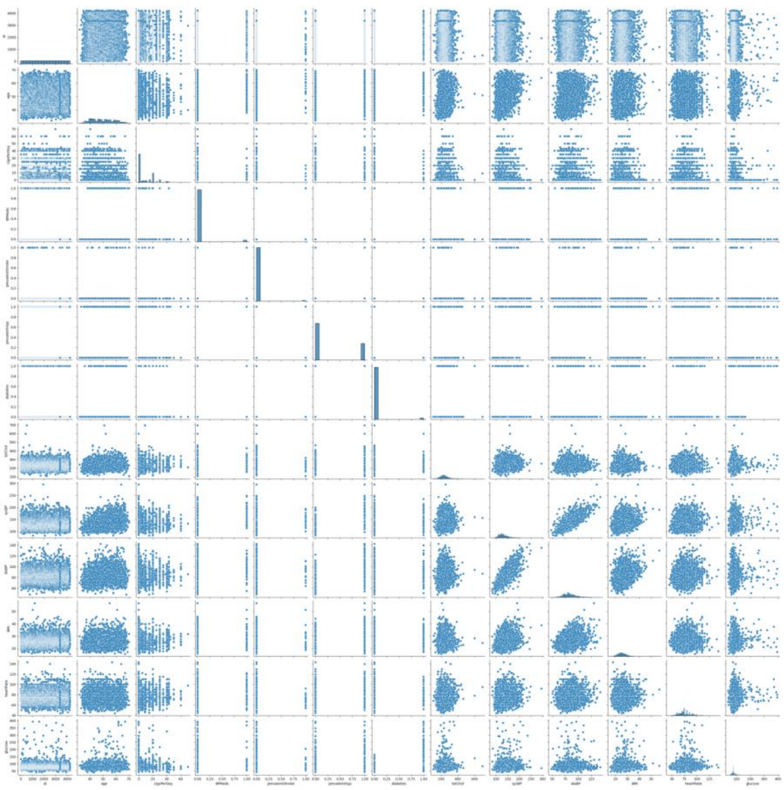


Fig 3.3: Pairplot of patient data

4.6 Training and Testing:

- Before training the model, the data is split into two subsets: training data and testing data.
- In most cases, about 70-80% of the data is used for training and the remaining 20-30% is used for testing. In our project, we have utilized 70% of the data for training and 30% for testing in order to prevent over fitting.
- During training, the model is fed data from which it learns patterns. The input data contains features such as :
    1. Age
    2. cigsPerDay
    3. BPMeds
    4. prevalentStroke
    5. prevalentHyp
    6. Diabetes
    7. totChol
    8. sysBP
    9. diaBP
    10. BMI
    11. heartrate
    12. Glucose.
- If the model performs poorly on training data then it may show issues such as under fitting which did not happen in our case.
- Once the training is done, the model is tested on new, unseen data. Predictions are made by the model and the results are compared to the actual outcomes to evaluate its performance.
- The testing phase is very crucial as the model is not just memorizing the training data but it is also making accurate predictions on new unseen data.

4.7 Models:

- In our project, the models that we have used in our backend are "Logistic Regression, Decision Tree and Random Forest".
- Logistic Regression: This is used to predict if the patient has heart diseases or not. '1' is used to indicate heart diseases of the patient and '0' is used to

indicate if the patient does not have heart diseases based on features such as Age and cholesterol.

- Decision Tree: This is a flowchart which is used to make decisions based on different features. In our patient case similarity model, this model uses patient features to split the data into branches which leads to an accurate prediction.
- Random Forest: It is a collection of multiple decision trees. An ensemble method that combines multiple decision tress instead of just one to make an accurate and stable prediction.

```
The accuracy score of the Logistic Model is  0.8584070796460177
/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:469: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Decision Tree Accuracy is  0.7541789577187807
```

```
Random Forest Accuracy is  0.8505408062930186
```

4.8 Clustering:

- Clustering is a method in data analysis that groups similar objects or data points together.
- In our project, we have implemented K-Means clustering. K-Means Clustering is a well-known machine learning method used to divide data into a number of groups.
- It is an unsupervised machine learning technique. Our code groups patient data into three clusters using K-Means Clustering and shows the results with a chart.
- Next, the average values of each feature for each cluster are calculated and shown.
- A scatter plot is made with patient age on the x-axis and cholesterol levels on the y-axis, and the points are colored based on their cluster.

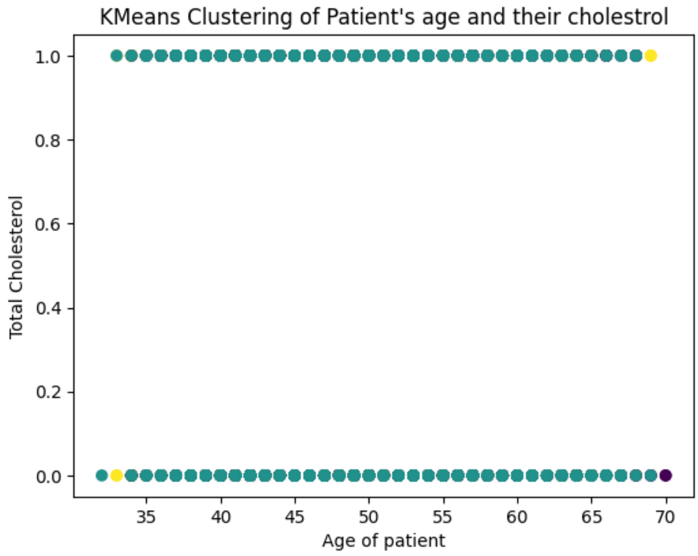


Fig 3.4: Scatter plot for K-Means Clustering

4.9 Silhouette score:

- The Silhouette Score checks how well data points are grouped together by calculating the score for different numbers of clusters using K-Means.
- It compares the similarity of each point to its own group compared to other groups.
- It prints the score for each number of clusters to help us choose the best one.
- Higher score means the groups are clear and better formed.

CHAPTER-5

OBJECTIVES

- The goal of this project is to make use of historical patient data in order to make accurate and useful predictions about heart diseases. The system analyzes trends, patterns and outcomes from the previous data and then aims to recognize similarities between patients. This gives a deeper understanding of all the factors which come together to contribute to heart diseases. It also provides a foundation which is data driven for the predictive analysis. Advanced Machine Learning algorithms have been used to provide well aimed, valuable insights for clinical applications.
- The main focus of the project is data driven decision making in the healthcare sector. By making use of various tools and Machine Learning models, the project calls attention to the value of factual information in medical practices. Biases can be minimized and reliability can be enhanced by using this particular approach. It works towards bridging the gap between the importance of adapting modern technology into traditional medical practices in order to improve overall outcomes.
- To develop predictive models which aim to analyze patient similarities early and accurately predict heart diseases. By early detection, not only can the patients survival rate be increased, but it also acts as a path for planning a better preventive care and treatment strategy. Accuracy in predictions has been given high importance so that the system is reliable. Also because of how crucial accuracy is in medical applications since errors could have consequences.
- By improving the diagnostic accuracy, the model aims to reduce healthcare costs. It can reduce the cost of unnecessary test costs and hospital fees. It also leads to saving of resources. Therefore, both patients as well as healthcare providers can use the advantages of this system. This system also aims to make healthcare more accessible to a wider population by making the total hospital fees cheaper by eliminating unnecessary tests and travelling costs.
- To identify early detection of heart diseases so that better prevention steps can be taken and a better personal plan can be made. The system helps healthcare professionals to identify the health risks early, which allows to develop a tailored preventive plan. These personalized plans can point towards the individual needs and risk factors of the patient. This objective also sheds light upon a much greater goal

which is reducing of heart diseases.

- To improve healthcare efficiency by automating the identification of patient similarities, saving the time and resources of healthcare providers. This allows professionals to focus on providing quality care and treatment while relying on the automated system for accurate and prompt insights. The reliability of healthcare services can further be enhanced by reducing risk of human errors
- Enhancement of clinical research by giving the researchers a tool that helps them in their studies about heart diseases. It helps researches in exploring new dimensions of medical science by providing detailed insights and predictive analysis. Development of innovative therapies and treatments can be done by analyzing patient similarities and predicting outcomes. The project is committed towards providing advancements in medical research and promoting a deeper level of understanding of heart diseases.
- The model aims to create more awareness about strengthening preventive care, taking precautions and making future generations aware of the horrors of heart diseases. It highlights the factors and causes that contribute to heart diseases.
- A user friendly website has been developed which is easy and can be understood by everyone. This website makes sure that patients as well as healthcare professionals can access it easily and understand the insights predicted by the system. It simplifies the complicated medical data and presents it in an understandable format which promotes greater engagement among users.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

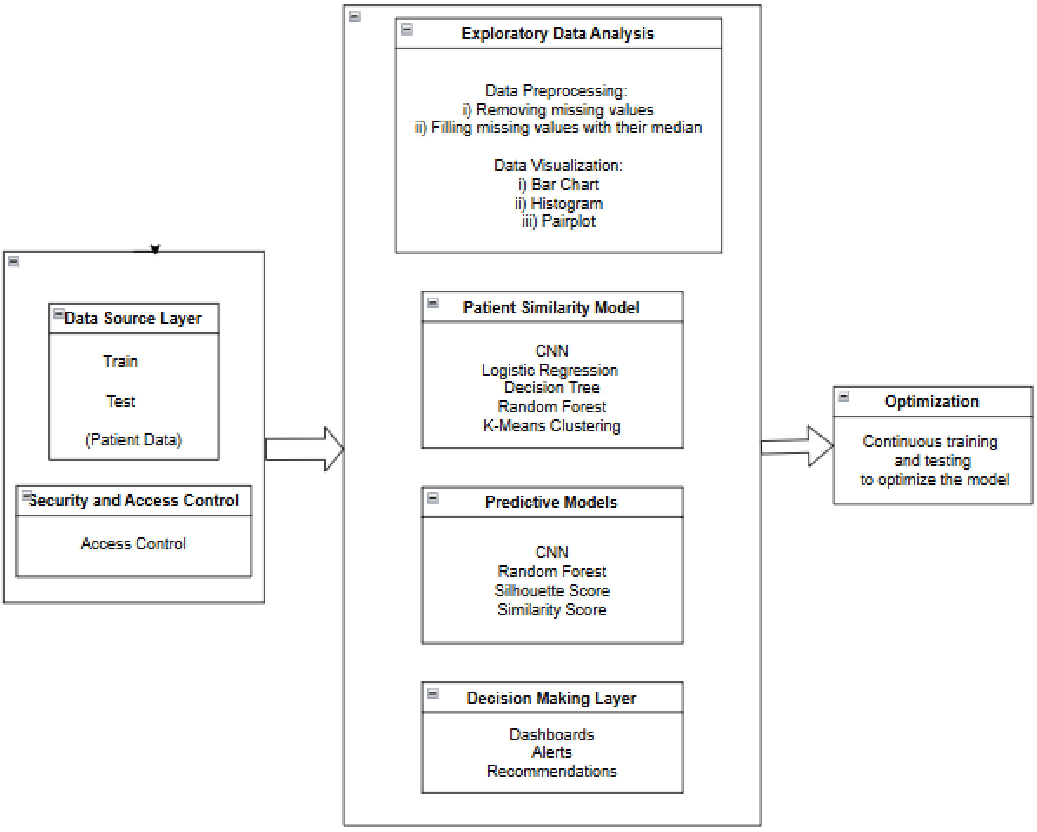


Fig 6.1: System Design

This diagram shows a system designed to analyze patient data and help doctors make better decisions.

i) Data Source Layer:

- It collects patient data and splits it into two parts: one to train the models and another to test how well they work.
- Key feature: Keeps data secure and only allows authorized access.

ii) Exploratory Data Analysis (EDA):

- Prepares the data and finds useful insights.
- Steps of EDA include:-
    - Fixing problems in the data, like missing information.
    - Creating simple charts (like bar charts, histograms, and pair plots) to understand patterns.

iii) Patient Similarity Model:

- It groups patients who are similar to each other.
- Uses tools like:
    - Neural networks (CNN) for analyzing complex data.
    - Simple models (Logistic Regression, Decision Trees).
    - Clustering (K-Means) to create groups.

iv) Predictive Models:

- It predicts what might happen to a patient (e.g., risks or outcomes).
- Uses advanced models like:
    - Neural networks and Random Forests for predictions.
    - Scores (Silhouette and Similarity) to measure accuracy and patient matching.

v) Decision-Making Layer:

- Helps doctors with clear, actionable outputs.
- Easy-to-read visual summaries.
- It notifies about critical events.
- Suggestions for treatment or next steps.

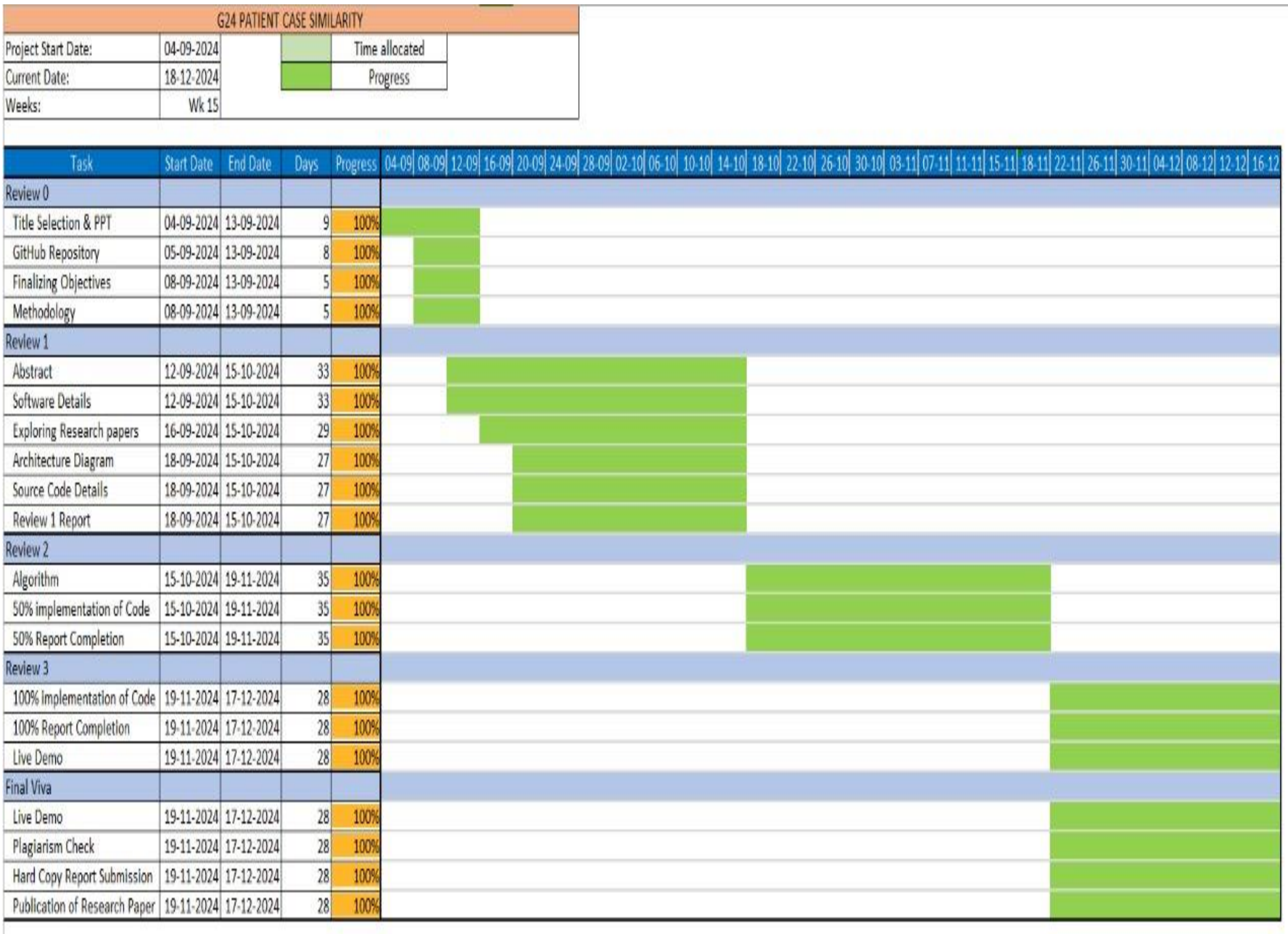
vi) Optimization:

- It continuously improves the models by retraining them with new data to make better predictions over time.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)



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	100%
Review 1 Report	18-09-2024	15-10-2024	27	100%
Review 2				
Algorithm	15-10-2024	19-11-2024	35	100%
50% implementation of Code	15-10-2024	19-11-2024	35	100%
50% Report Completion	15-10-2024	19-11-2024	35	100%
Review 3				
100% Implementation of Code	19-11-2024	17-12-2024	28	100%
100% Report Completion	19-11-2024	17-12-2024	28	100%
Live Demo	19-11-2024	17-12-2024	28	100%
Final Viva				
Live Demo	19-11-2024	17-12-2024	28	100%
Plagiarism Check	19-11-2024	17-12-2024	28	100%
Hard Copy Report Submission	19-11-2024	17-12-2024	28	100%
Publication of Research Paper	19-11-2024	17-12-2024	28	100%

Fig 7.1: Timeline using Gantt Chart

Review Wise Workload:

Review 0:

- Title Selection.
- Define project scope and objectives.
- Create GitHub Repository.
- Come up with methodology to be followed
- Make Review 0 PPT

Review 1:

- Define abstract.
- Define the software details.
- Explore Research Papers.
- Draw Architecture Diagram
- Source Code Details.
- Make Review 1 Report
- Make Review 1 PPT

Review 2:

- Finalize Algorithm details.
- Develop 50% Code.
- Complete 50% Report.

Review 3:

- Develop 100% Code.
- Ensure Model is working properly.
- Complete Project Report
- Plagiarism Check
- Publish Research Paper

CHAPTER-8
OUTCOMES

- Patient Case Similarity Model revolves around the idea of giving a similarity score of a new patient by referring to the historical data of an old or existing patient.
- In our project, we have determined the similarity score between a new patient and the five most similar patients in the historical data.
- This outcome is to bring more substantial improvements in the healthcare management.
- The similarity score determines how closely their cases align, with values approaching 1 representing higher similarity.

8.1 Outcome 1

- With the below outcome, we can conclude that the patients 14 and 21 give exceptionally high similarity scores with the new patient data given by the user.
- These outcomes can support the following in future scenarios:
  - i) To provide personalized treatment plans.
  - ii) To prioritize a case review.

Most Similar Patients:

Patient ID: 10, Similarity: 0.9894803221396218

Patient ID: 11, Similarity: 0.9896825252977868

Patient ID: 14, Similarity: 0.9972892831540058

Patient ID: 21, Similarity: 0.9921180983850746

Patient ID: 23, Similarity: 0.9886220303454638

8.2 Outcome 2

- In the image given below, the similarity scores are lower compared to the previous outcomes. This indicates that the new patient data does not match well with the five patients of historical data.

- The reasons for low similarity scores can be the following:

  i) The new patient's medical data may differ from the historical patients' data.

  ii) The new patient's case may involve rare conditions which may not have been detected in the past.

Most Similar Patients:

Patient ID: 1, Similarity: 0.7798798968758358

Patient ID: 6, Similarity: 0.8224959896114975

Patient ID: 10, Similarity: 0.773295464154435

Patient ID: 14, Similarity: 0.782637663524027

Patient ID: 21, Similarity: 0.7797529557896715

CHAPTER-9
RESULTS AND DISCUSSIONS

9.1 Results

- The dataset “test.csv” and “train.csv” were merged for data analysis and modeling.
- Missing values in columns such as cigsPerDay, BPMeds, totChol, BMI, heartRate, glucose, were filled with their median values.
- Valuable insights were gained from visualization. A bar chart showed the relationship between age of patients and their heart rates. Histogram showed a normal distribution of heart rate values. Pairwise relationships between numeric variables were visualized using pair plot.
- “Logistic Regression” gave an accuracy of 85.84%, “Decision Tree Classifier” gave an accuracy of 75.42% and “Random Forest Classifier” gave an accuracy of 85.05%.
- K-Means clustering grouped patients together based on their features and for these clusters a Silhouette score was determined.
- We have successfully determined the similarity score of a new patient in comparison to five patients from the historical data.

9.2 Discussions

- Improving Model Accuracy:

  i) Hyperparameter Tuning: This involves optimizing the models such as “Logistic Regression, Decision Tree, and Random Forest classifiers”, using techniques such as Grid Search or Random Search. This could lead to enhanced model performances.

  ii) Advanced Models: Models such as Gradient Boosting Machines (GBM), XGBoost, etc., can be explored for better performances.
- Handling missing values is done by median imputation. Median imputation is particularly effective for skewed distributions. However, other imputation strategies such as mean imputation, KNN imputation, etc., can also be considered to get a good accuracy from the model.
- “The Logistic Regression model” gave the best accuracy (85.84%). This suggests that the model was able to capture the linear relationships in the data very effectively and also interpret the data accurately. This model can also be used for binary

classification tasks.

- Different ranges of k-values gave various silhouette scores, some even higher, which suggested better and clearer defined clusters. K-means clustering could also help in identifying distinct patient groups with similar health data which can be used for personalized medical treatment.
- The number of historical patients can be adjusted as desired and the model will still accurately compute the similarity scores. This flexible nature of the model allows further fine-tuning and adds robustness to the model.
- Our project “Patient Case Similarity” could be further personalized by providing healthcare recommendations based on the similarity scores. This would make the project even useful in risk prediction where similar historical cases could inform personalized treatment plans or strategies for preventing the diseases.

CHAPTER-10
CONCLUSION

Advances in “Artificial Intelligence and Machine Learning” have set the stage for big changes in healthcare, with the Patient Case Similarity project being a prime example. The project "Patient Case Similarity" has successfully been used to demonstrate the power of Machine Learning in order to enhance clinical decision support systems. It shows how AI systems can assist doctors in making more accurate diagnoses and predicting health outcomes.

By using historical data, the project has created an easy-to-use web application that looks at patient information, finds patterns, and predicts results by comparing current cases with similar past ones.

The system employs powerful Machine Learning models, including “Logistic Regression, Decision Tree, and Random Forest Classifiers”. Among all these, the Logistic Regression model was the most reliable for predicting healthcare outcomes. The K-Means clustering method grouped patients with similar characteristics, and the results were supported by strong silhouette scores, showing that the system can create clear and accurate clusters. The project also used visual tools like bar charts, histograms, and pair plots to provide useful insights into how different factors are related, improving the analysis. Furthermore, the system demonstrated its precision and adaptability by accurately calculating similarity scores for new patients compared to historical cases.

The project provides immense benefits because it not only reduces the overall healthcare costs by eliminating unnecessary tests but also promotes medical decision-making by improving efficiency. The user-friendly website promotes accessibility for medical professionals and patients improving user interaction and user satisfaction. It also reduces the gap between clinical practices and data driven insights by encouraging personalized medicine and better healthcare.

The Patient Case Similarity project addresses an important problem in healthcare—the need for manual analysis of large datasets, which is time-consuming and prone to mistakes. By

simplifying this process, the system improves efficiency and accuracy, helping healthcare professionals make better decisions.
It encourages preventive care, which helps to improve patient outcomes and overall health.

In the future, the project could be enhanced by expanding the dataset and combining Deep Learning for better predictive accuracy and clustering efficiency.

REFERENCES

[1] Cao, Hui (2008) "Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases.", Journal of biomedical informatics 41.6: 882-888.

[2] Brown, Sherry-Ann. (2016) "Patient similarity: emerging concepts in systems and precision medicine.", Frontiers in physiology 7: 561.

[3] Parimbelli, Enea (2018) "Patient similarity for precision medicine: A systematic review.", Journal of biomedical informatics 83: 87-96.

[4] Jia, Zheng, (2020) "A patient-similarity-based model for diagnostic prediction.", International journal of medical informatics 135: 104073.

[5] Wang, Ni (2019) "Measurement and application of patient similarity in personalized predictive modeling based on electronic medical records.", Biomedical engineering online 18: 1-15.

[6] Zhu, Zihao (2016) "Measuring patient similarities via a deep architecture with medical concept embedding.", 2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE.

[7] Vishakha Sharma, Tannya Mittal, Shrey Shukla, Tushar Babbar, Priyanka Bhardwaj (2020) “Patient-Case Similarity.”, International Journal of Computer Science and Information Technology Research.

[8] Dai, Leyu, He Zhu, and Dianbo Liu. (2020) "Patient similarity: methods and applications.", arXiv preprint arXiv:2012.01976.

[9] Fang, Hao Sen Andrew (2021) "Patient similarity analytics for explainable clinical risk prediction.", BMC medical informatics and decision making 21.1: 207.

[10] Navaz, Alramzana Nujum(2022) "A novel patient similarity network (PSN) framework based on multi-model deep learning for precision medicine.", Journal of Personalized Medicine 12.5 : 768.

[11] Sun, Zhaohong(2022) "Deep dynamic patient similarity analysis: model development and validation in ICU.", Computer Methods and Programs in Biomedicine 225 : 107033.

[12] Mahima, VB Jeevine, Mohammed Salman Khan (2024) “Patient Case Similarity.”, International Research Journal of Modernization in Engineering Technology and Science.

APPENDIX-A

PSUEDOCODE

FRONTEND

Main.html

1. Define the HTML document structure.
   - Specify the DOCTYPE as HTML.
   - Start the HTML tag.
   - Define the HEAD section for metadata and styles.
2. Set metadata and title in the HEAD section.
   - Use `<meta>` tag to set the character encoding to ISO-8859-1.
   - Set the `<title>` tag to "Patient Case Similarity."
3. Define the CSS styles within a `<style>` block.
   - Set an image as a background for the body:
     - Define image path and ensure it covers the full page.
     - Use a sans-serif font for the text.
   - Style the header image:
     - Set it to cover 100% width and specific height (720px).
     - Position it relative for layering with other elements.
   - Style the header text:
     - Use absolute positioning for placement over the header image.
     - Center-align the text and apply color and shadow effects.
   - Define text size and bold formatting for `<h1>` and `<h2>` elements.
   - Style the navigation ribbon:
     - Use a flex container for alignment and spacing.
     - Set a background color, dimensions, and border radius.
   - Style links in the ribbon:
     - Define text alignment, padding, and hover effects.
     - Highlight the active link with a specific background color.
4. Define the BODY content.
   - Add a `<br>` tag for spacing.

- Create a container with relative positioning for the header image and text:
  - Use an `<img>` tag to display the hospital image as the header image.
  - Create a `<div>` with the class `header-text`:
    - Add two headings, `<h1>` and `<h2>`, with the specified content.
  - Add a navigation ribbon (`<div>` with class `rib`):
    - Insert four links (`<a>` tags):
      - Link to pages: Home, About Us, Find a Location, and Patient Portal.
      - Apply hover effects and the active state to specific links.

5. End the BODY and HTML tags.

About.html

1. Define the HTML document structure.
   - Specify the DOCTYPE as HTML.
   - Set the language attribute in the HTML tag to "en".
2. Define the “Head” section.
   - Setting the metadata:
     - Character encoding to UTF-8.
     - Viewport for responsive design.
   - Add the page title: "About Us - Patient Case Similarity System."
   - Include CSS styles within a `<style>` block:
     - Set a background image and size for the body.
     - Define font family and text color for the body.
     - Style headings (`<h1>` and `<h2>`):
       - Use white color and shadow for visibility.
       - Set font size, weight, alignment, and margins.
     - Style the navigation ribbon (`.rib`):
       - Set background color, dimensions, and border radius.
       - Define link styles (`.rib a`):
         - Text color, alignment, hover, and active states.
     - Style the "About Us" section (`.about-container`):
       - Use a semi-transparent black background.
       - Add padding, border radius, and shadow.
       - Ensure text is justified and readable with appropriate font size and line height.

- Highlight specific words using `<strong>` with a unique color.

3. Define the BODY section.
   - Add a HEADER section:
     - Include a `<h1>` tag for the page title.
   - Add a navigation menu (NAV):
     - Create a div with the class `rib`.
     - Center-align links:
       - Include four links: Home, About Us, Find a location, and Patient Portal.
       - Mark the "About Us" link as active.
   - Create the "About Us" section:
     - Use a container div with the class `about-container`:
       - Add a `<h2>` for the section title.
       - Include paragraphs describing the platform's purpose, features, and values.
       - Add a bulleted list of core values with highlights (`<strong>`) such as innovation, compassion, excellence, and integrity.
     - Conclude with a thank-you message.
4. End the BODY and HTML tags.

<u>Locateus.html</u>

1. Define the HTML document structure.
   - Specify the DOCTYPE as HTML.
   - Set the language attribute in the HTML tag to "en".
2. Define the “Head” section.
   - Setting the metadata:
     - Character encoding to UTF-8.
     - Viewport for responsive design.
   - Add the page title: "Hospitals and Doctors for Heart Diseases."
   - Include CSS styles within a `<style>` block:
     - Set a background image and size for the body.
     - Define font family and text color for the body.
     - Style headings (`<h1>` and `<h2>`):
       - Use white color and shadow for visibility.
       - Set font size, weight, alignment, and margins.

- Style the navigation ribbon (`.rib`):
  - Set background color, dimensions, and border radius.
  - Define link styles (`.rib a`):
    - Text color, alignment, hover, and active states.
- Style hospital containers (`.hospital-container`):
  - Use a semi-transparent background with padding and rounded corners.
  - Center-align content and add shadows.
- Style hospital images (`.hospital-img`):
  - Set width, height, object-fit property, and border radius.
- Define text styles for address and doctor lists:
  - Set font size and alignment.
  - Remove list styles for the unordered list.

3. Define the BODY section.
- Add a HEADER section:
  - Include a `<h1>` tag for the page title.
- Add a navigation menu (NAV):
  - Create a div with the class `rib`.
  - Center-align links:
    - Include links: Home, About Us, Find a location, and Patient Portal.
- Create a section for hospital details:
  - For each hospital:
    - Create a container div with the class `hospital-container`.
    - Add an image of the hospital (`<img>` tag with the class `hospital-img`).
    - Add a `<h2>` tag for the hospital name.
    - Include a `<p>` tag for the address with the class `address`.
    - Add a `<h3>` tag for the doctors' section.
    - List doctors using an unordered list (`<ul>`):
      - Include list items (`<li>`) with doctors' names and specialties.
    - Add a `<h3>` tag for the hospital's rating.

4. End the BODY and HTML tags.

<u>Registrationpage.html</u>

1. Define the HTML document structure.

- Specify the DOCTYPE as HTML.
- Set the language attribute in the HTML tag to "en".

2. Define the “Head” section.
   - Set the metadata as following:
     - Character encoding to UTF-8.
     - Viewport for responsive design.
   - Add the page title: "Patient Registration Form."
   - Include CSS styles within a `<style>` block:
     - Set a background image and font family for the body.
     - Style the navigation ribbon (`.rib`):
       - Define dimensions, background color, and border radius.
       - Add hover and active state styles for links.
     - Style headings (`<h1>` and `<h2>`):
       - Center-align text, set color and shadow for readability.
     - Define the container for the registration form:
       - Set width, margin, background color, padding, and border radius.
     - Style form elements:
       - Inputs for text, numbers, and submit button:
         - Define width, padding, borders, and margins.
         - Style the submit button with hover effects.
   - Add JavaScript for form validation:
     - Create a function `validateForm`:
       - Retrieve form input values: Name, ID, Phone, Age, Sex, and other fields.
       - Check if any field is empty. If so, display an alert and return `false`.
       - Ensure ID, Phone, and Age are numeric. If not, display an alert and return `false`.
       - If all validations pass, return `true`.
3. Define the BODY section.
   - Add a center-aligned heading for the page title.
   - Include a navigation menu (`.rib`):
     - Add links: Home, About Us, Find a location, and Patient Portal.
   - Add a center-aligned heading for the registration form.
4. Create a registration form section.
   - Define a container for the form.
   - Add a `<form>` element:

   - Set the form name to "registrationForm."
   - Specify `action` as "registered.html" and `method` as "post."
   - Set the `onsubmit` attribute to call the `validateForm` function.
- Add input fields:
   - For each field:
      - Add a `<label>` for the field.
      - Add an `<input>` with appropriate `type` (text, number, tel).
      - Mark each field as `required`.
- Add a submit button:
   - Define an `<input>` of type `submit` with a label "Submit."

5. End the BODY and HTML tags.

BACKEND

- Our project “Patient Case Similarity” uses a Python-based backend.
- This python-based backend is implemented with Flask.
- The backend handles the patient similarity analysis and processes the API requests and then sends it to the frontend part of the project.
- The below is the pseudo code for our project “Patient Case Similarity”:

1. Importing Libraries:

Importing necessary libraries for backend:-

- Pandas: This python package is used for data manipulation.
- SimpleImputer: It is used to handle missing values.
- Cosine_similarity: It is used for similarity computation of vectors or matrices.
- Flask: This is a web framework which is used for creating web applications. In our project, it is used for creating the API.
- CORS: It is used for Cross-Origin resource sharing. In this code, it is used to process the requests.

2. Initialization of Flask app:

- A Flask application is an instance of the Flask class, that is, everything about this application will be registered with the class. We created a flask application instance for our code.
- After creating the flask application instance we enabled CORS for the app.

3. Preprocessing the data:

- Created a function called “preprocess_data()”.
- Loaded the datasets test.csv and train.csv.
- Combined test and train into a single dataset.
- Dropped unnecessary columns such as “TenYearCHD, is_smoking, education”.
- For some specific columns, filled the missing values with median of the respective column.
- Returned the preprocessed data and the original train dataset.

4. Encoding the data:

- Created another function called “encode_data(X, original_columns)”.
- Converted specific numeric columns to numeric types so that some columns that contain different data types can be converted into numbers.
- Identified categorical columns and encoded them using one-hot encoding.
- SimpleImputer with the "median" strategy is used to handle missing values.
- Returned the encoded dataset as a dataframe.

5. Encoding input data:

- Function called “encode_input_data(input_data, X_encoded)” is created.
- The ‘input_data’ is converted into a dataframe.
- Converted specific columns to numeric types.
- One-hot encoding is applied to categorical columns, aligning with the structure of ‘X_encoded’.
- Missing columns are filled with zeros to match the ‘X_encoded’ structure.
- Returned the encoded input data as a dataframe.

6. Similarity Score of patients:

- Function “find_similar_patients(input_data, X_encoded)” is created in this step.
- Encoded the input data using ‘encode_input_data’.
- Concatenated the encoded input data with the dataset ‘X_encoded’.
- Computed a similarity matrix using ‘cosine_similarity’.
- Extracted the similarity scores for the input patient (new patient) and found the five most similar patients from the historical data.
- Returned the five similar patients and their similarity scores.

7. Defining API Endpoint:

- Route: /predict – this makes the endpoint accessible in the web server.
- Request Type: POST – the client sends data to the server using POST method as it is suitable for structured data like JSON.
- Extract JSON data from the client request.
- Created two dictionaries - input_data and pushing_data. After which it is extracted

and structured for processing.

- The dataset is preprocessed using ‘preprocess_data()’.
- The training data is then encoded using ‘encode_data()’.
- Similarity score is found using ‘find_similar_patients(input_data, X_encoded)’.
- Appended ‘pushing_data’ (new patient record) to train.csv:
- Loaded existing train.csv.
- Appended new data to it.
- Saved the updated dataset back to the CSV file.
- Finally, the code returned the similarity score of similar patients.

8. Error Handling:

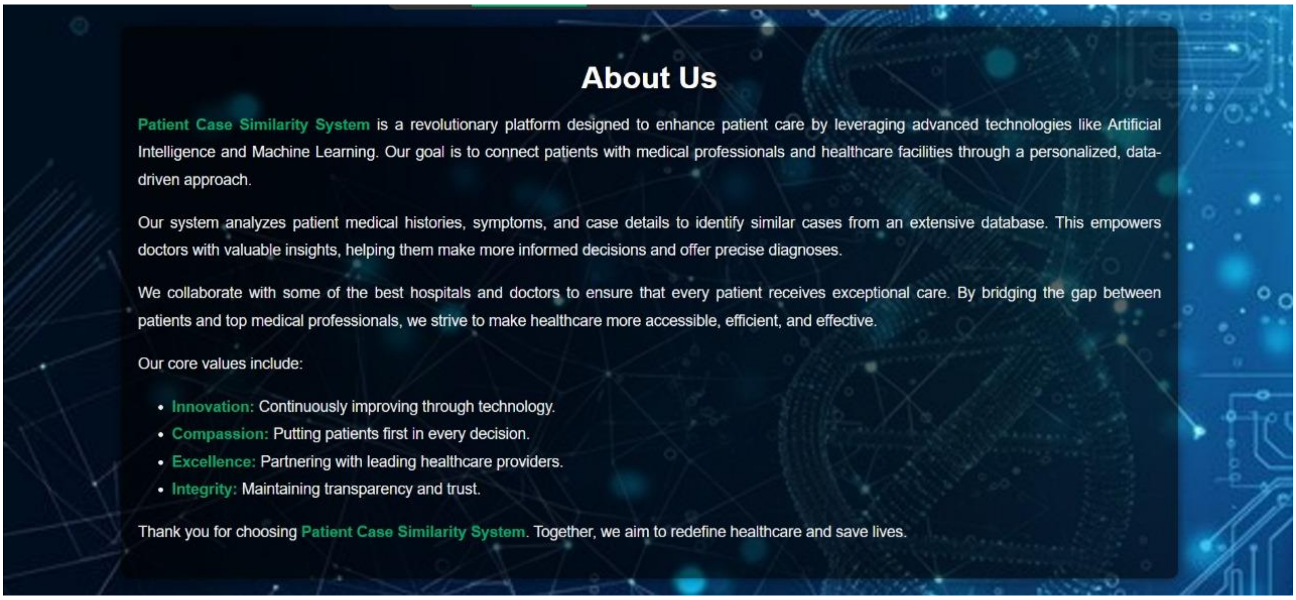
- Error handling is done to handle exceptions during file input/output operations or any other processing errors.
- It returns an error in response in JSON format if an exception occurs.

9. Running the Flask App:

- The Flask app is started in debug mode.

APPENDIX-B

SCREENSHOTS

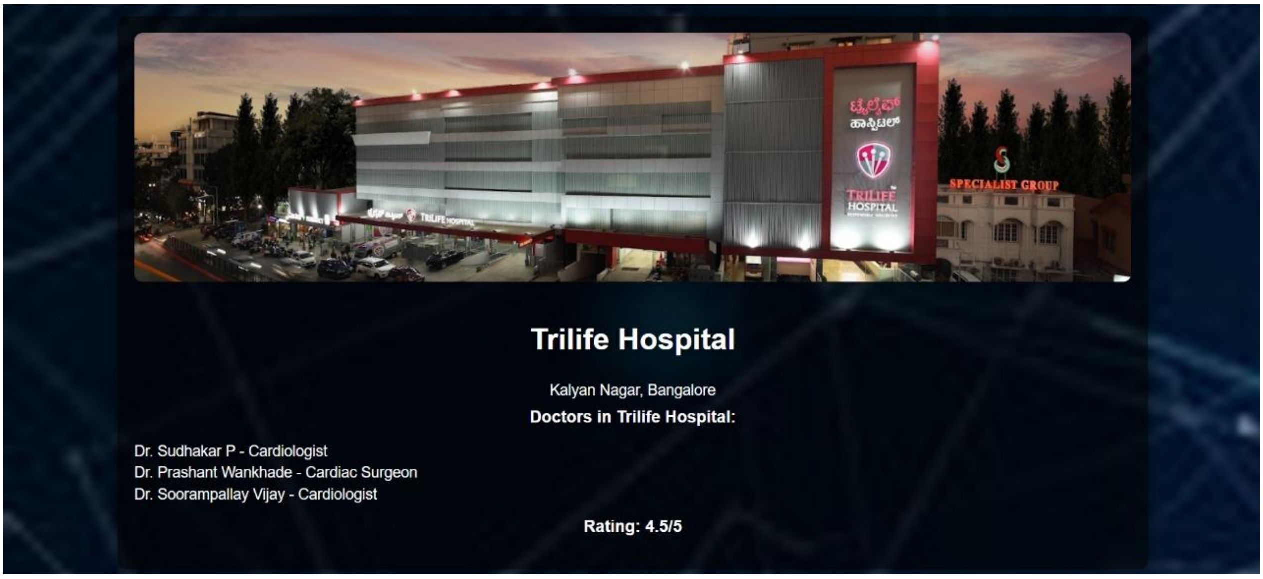

About Us
Patient Case Similarity System is a revolutionary platform designed to enhance patient care by leveraging advanced technologies like Artificial Intelligence and Machine Learning. Our goal is to connect patients with medical professionals and healthcare facilities through a personalized, data-driven approach.
Our system analyzes patient medical histories, symptoms, and case details to identify similar cases from an extensive database. This empowers doctors with valuable insights, helping them make more informed decisions and offer precise diagnoses.
We collaborate with some of the best hospitals and doctors to ensure that every patient receives exceptional care. By bridging the gap between patients and top medical professionals, we strive to make healthcare more accessible, efficient, and effective.
Our core values include:
Innovation: Continuously improving through technology.
Compassion: Putting patients first in every decision.
Excellence: Partnering with leading healthcare providers.
Integrity: Maintaining transparency and trust.
Thank you for choosing Patient Case Similarity System. Together, we aim to redefine healthcare and save lives.


Patient Case Similarity System
Home
About Us
Find a location
Patient Portal
manipalhospitals
LIFE'S ON
Manipal Hospital
Malleswaram West, Bangalore
Doctors in Manipal Hospital:


manipalhospitals
LIFE'S ON
Manipal Hospital
Malleswaram West, Bangalore
Doctors in Manipal Hospital:
Dr. Prabhakara Shetty Heggunje - Cardiologist
Dr. Mohammed Rehan Sayeed - Cardiothoracic Surgeon
Dr. Srikanth Vijaysimha - General Surgeon
Rating: 4.5/5


Aster
CMI HOSPITAL
Aster Cmi Hospital
Sahakara Nagar, Bangalore
Doctors in Aster Cmi Hospital:
Dr. Arul D Furtado - Cardiac Surgeon
Dr. Ganesh Krishnan Iyer - Cardiothoracic Surgeon
Dr. Y M Prashanth - Cardiothoracic and Vascular Surgeon
Rating: 4/5

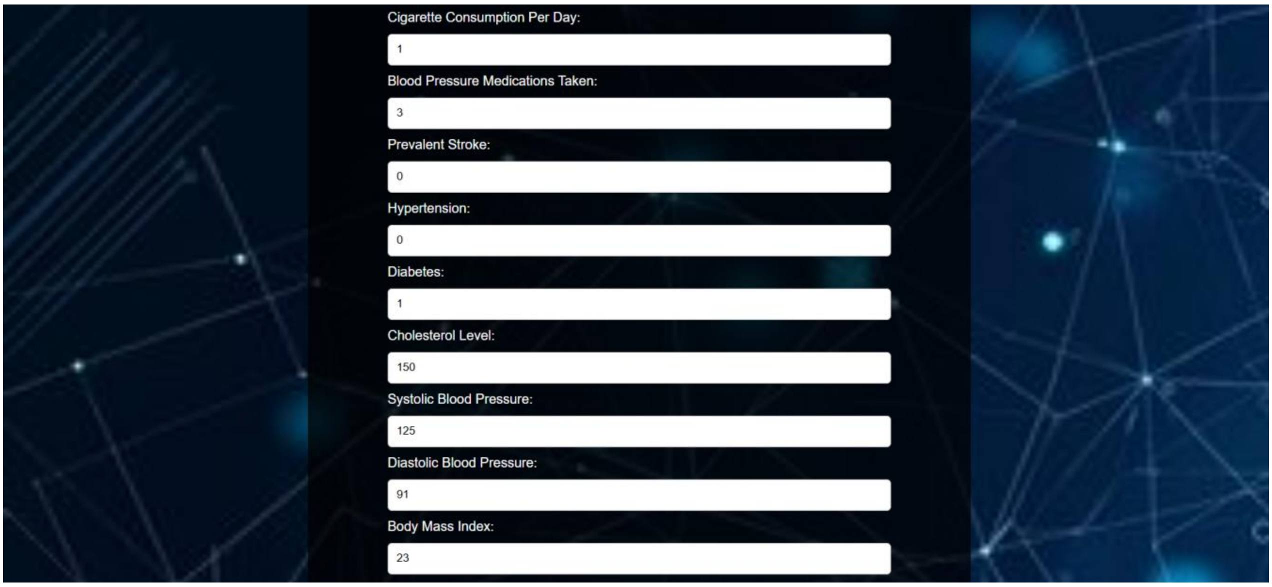

ಅಪೋಲೋ ಆಸ್ಪತ್ರೆ
Apollo Hospitals
Apollo Hospital
Bannerghatta Road, Bangalore
Doctors in Apollo Hospital:
Dr. Raghavendra Chikkatur - Cardiothoracic Surgeon
Dr. Umesh Satish Gheewala - Vascular Surgeon
Rating: 4.5/5

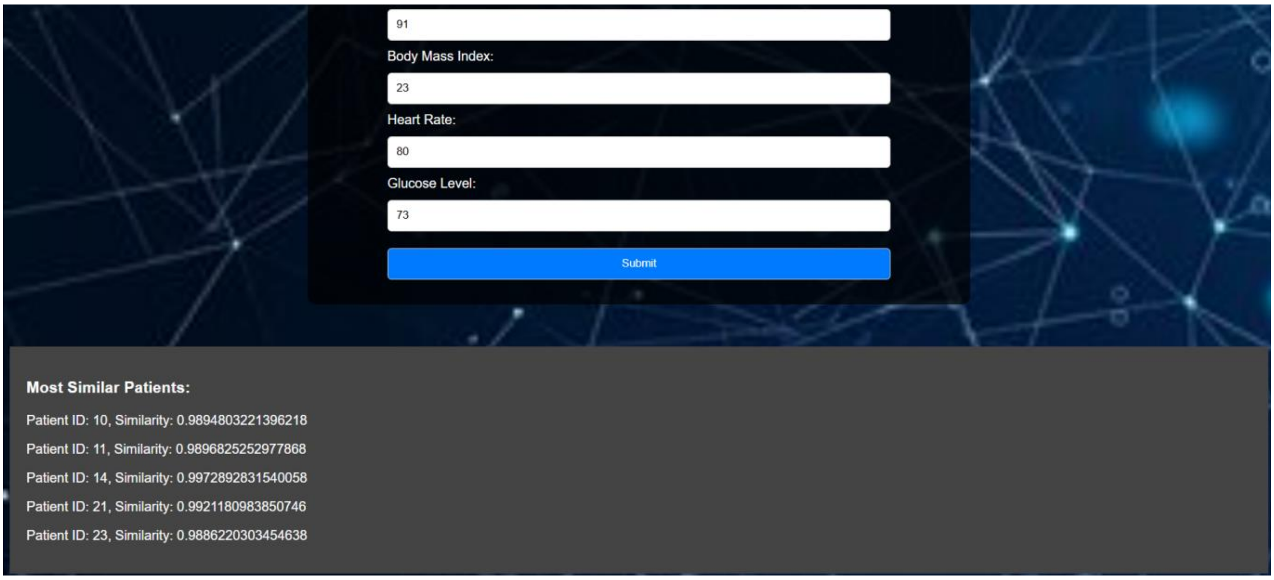

TRILIFE
HOSPITAL
SPECIALIST GROUP
Trilife Hospital
Kalyan Nagar, Bangalore
Doctors in Trilife Hospital:
Dr. Sudhakar P - Cardiologist
Dr. Prashant Wankhade - Cardiac Surgeon
Dr. Soorampallay Vijay - Cardiologist
Rating: 4.5/5


manipalhospitals
Manipal Hospital
Millers Road, Bangalore
Doctors in Manipal Hospital:
Dr. Girish Godbole - Cardiac Surgeon
Dr. Praveen Kumar A V - Cardiac Surgeon
Dr. G Sridhara - Cardiac Surgeon
Rating: 4.5/5


Tathagat
Heart Centre
1st Floor
Tathagat Heart Hospital
Crescent Road, Bangalore
Doctors in Tathagat Heart Hospital:
Dr. Mahantesh R Charantimath - Cardiologist
Rating: 4/5


Patient Case Similarity System
Home
About Us
Find a location
Patient Portal
Registration Form
Patient Name:
N
Patient ID:
78
Phone Number:
8898745632
Age:
50
Sex:
M


Cigarette Consumption Per Day:
1
Blood Pressure Medications Taken:
3
Prevalent Stroke:
0
Hypertension:
0
Diabetes:
1
Cholesterol Level:
150
Systolic Blood Pressure:
125
Diastolic Blood Pressure:
91
Body Mass Index:
23


91
Body Mass Index:
23
Heart Rate:
80
Glucose Level:
73
Submit
Most Similar Patients:
Patient ID: 10, Similarity: 0.9894803221396218
Patient ID: 11, Similarity: 0.9896825252977868
Patient ID: 14, Similarity: 0.9972892831540058
Patient ID: 21, Similarity: 0.9921180983850746
Patient ID: 23, Similarity: 0.9886220303454638

APPENDIX-C

ENCLOSURES

- **Journal publication certificates of all students.**

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, High Impact Factor, Open Access Journal since 2013)

IJIRCCE

Impact Factor 8.625

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

DR. MOHAMMADI AKHEELA KHANUM

Professor, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Patient Case Similarity"

in IJIRCCE, Volume 12, Issue 12, December 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, High Impact Factor, Open Access Journal since 2013)

IJIRCCE

Impact Factor 8.625

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

PRERNA KAKADE

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Patient Case Similarity"

in IJIRCCE, Volume 12, Issue 12, December 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, High Impact Factor, Open Access Journal since 2013)

IJIRCCE

Impact Factor 8.625

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

NIDA AIYMAN

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Patient Case Similarity"

in IJIRCCE, Volume 12, Issue 12, December 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, High Impact Factor, Open Access Journal since 2013)

Impact Factor 8.625

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

BHUVANA V

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Patient Case Similarity"

in IJIRCCE, Volume 12, Issue 12, December 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, High Impact Factor, Open Access Journal since 2013)

IJIRCCE

Impact Factor 8.625

CERTIFICATE

OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

PRANAV GANESH

Student, Department of Computer Science and Engineering, Presidency University, Bengaluru, India

in Recognition of Publication of the Paper Entitled

"Patient Case Similarity"

in IJIRCCE, Volume 12, Issue 12, December 2024

Google scholar · Crossref · Mendeley · doi · INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2320-9801
p-ISSN: 2320-9798

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijircce.com ijircce@gmail.com

- **Plagiarism Check**

Mohammadi Akheela Khanum - Final Report Capstone G24 (1)

ORIGINALITY REPORT

19%	16%	12%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Presidency University Student Paper	11%
2	www2.mdpi.com Internet Source	1%
3	eitca.org Internet Source	1%
4	fastercapital.com Internet Source	1%
5	experts.umn.edu Internet Source	<1%
6	Submitted to Coventry University Student Paper	<1%
7	medinform.jmir.org Internet Source	<1%
8	E. Parimbelli, S. Marini, L. Sacchi, R. Bellazzi. "Patient similarity for precision medicine: a systematic review", Journal of Biomedical Informatics, 2018 Publication	<1%

9	scholarworks.uaeu.ac.ae Internet Source	<1 %
10	Submitted to University of North Texas Student Paper	<1 %
11	www.irjmets.com Internet Source	<1 %
12	www.scribd.com Internet Source	<1 %
13	amds123.github.io Internet Source	<1 %
14	Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, Indra Kishor. "Recent Advances in Sciences, Engineering, Information Technology & Management - Proceedings of the 6th International Conference “Convergence2024” Recent Advances in Sciences, Engineering, Information Technology & Management, April 24–25, 2024, Jaipur, India", CRC Press, 2025 Publication	<1 %
15	business.uc.edu Internet Source	<1 %
16	donut.topology.rocks Internet Source	<1 %
17	www.kluniversity.in Internet Source	<1 %

18	bmcmedinformdecismak.biomedcentral.com Internet Source	<1 %
19	coek.info Internet Source	<1 %
20	assets-eu.researchsquare.com Internet Source	<1 %
21	arxiv.org Internet Source	<1 %
22	www.collegesinstitutes.ca Internet Source	<1 %
23	Poonam Tanwar, Tapas Kumar, K. Kalaiselvi, Haider Raza, Seema Rawat. "Predictive Data Modelling for Biomedical Data and", River Publishers, 2024 Publication	<1 %
24	Submitted to The University of Manchester Student Paper	<1 %
25	Submitted to University of Hertfordshire Student Paper	<1 %
26	Submitted to University of Stirling Student Paper	<1 %
27	towardsdatascience.com Internet Source	<1 %
28	www.openpr.com Internet Source	

<1 %

29	lawyersnjurists.com Internet Source	<1 %
30	Suresh Pokharel, Guido Zuccon, Xue Li, Chandra Prasetyo Utomo, Yu Li. "Temporal Tree Representation for Similarity Computation between Medical Patients", Artificial Intelligence in Medicine, 2020 Publication	<1 %
31	Hui Cao, Genevieve B. Melton, Marianthi Markatou, George Hripcsak. "Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases", Journal of Biomedical Informatics, 2008 Publication	<1 %
32	Rama Devi K, Syed Siraj Ahmed N, Ramchandran P, Parvathi S, Althaf Ali A. "Proactive Detection of Malicious Webpages Using Hybrid Natural Language Processing and Ensemble Learning Techniques", Journal of information and organizational sciences, 2024 Publication	<1 %
33	aquatera.ca Internet Source	<1 %

34	ebin.pub Internet Source	<1 %
35	epub.lib.aalto.fi Internet Source	<1 %
36	mdpi-res.com Internet Source	<1 %
37	"Software Engineering Methods in Systems and Network Systems", Springer Science and Business Media LLC, 2024 Publication	<1 %
38	Syed Nisar Hussain Bukhari. "Data-Driven Farming - Harnessing the Power of AI and Machine Learning in Agriculture", CRC Press, 2024 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

- **Details of mapping the project with the Sustainable Development Goals (SDGs).**

- Our project “Patient Case Similarity” is mapped to “SDG-3”, that is, Good health and Well Being.
- “The United Nations Sustainable Development Goals (SDGs)” are detailed plan to make the world healthier by 2030 and each goal focuses on a specific problem like a doctor treating different health issues in a patient. These goals include:
  - ➢ Giving quality healthcare for everyone like making sure a patient gets the right treatment.
  - ➢ Ensuring clean water and sanitation, like making sure patients have clean and safe facilities.
  - ➢ Providing sustainable energy like making sure patients live in a healthy environment.
  - ➢ Ensuring fairness and good governance like building trust and ethics in healthcare.
  - ➢ Encouraging global teamwork like promoting cooperation among healthcare providers.

- Our project contributes to the improvement of people’s health and enhancing their quality of life.
- By getting the similarity score between historical patients and the new patients, we make it easier to cluster patients based on heart diseases.
- Improving the diagnostic accuracy and predictive models is the basic guarantee of good health and well-being for people of all generations. This has been ensured in our project “Patient Case Similarity”.