ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14769409

Patient Case Similarity

[1]varsha V. (20211CSE0857); [2]Nishanth P H. (20211CSE0597);
[3]Chethan Jadav M. (20211CSE0624) [4]Rohith M. (20211cse0786)

Under the guidance of
[5]Vineetha B.

in Partial Fulfillment for the Award of the Degree of
Bachelor of Technology

in
Computer Science and Engineering

at
Presidency University Bengaluru January 2025

Publication Date: 2025/02/01

How to Cite: Varsha V.; Nishanth P H.; Chethan Jadav M.; Rohith M.; Vineetha B. (2025). Patient Case Similarity. *International Journal of Innovative Science and Research Technology*, 10(1), 1473-1503. https://doi.org/10.5281/zenodo.14769409.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14769409

CERTIFICATE

PRESIDENCY UNIVERSITY
SCHOOL OF COMPUTER SCIENCE ENGINEERING

This is to certify that the Project report **"Patient Case Similarity"** being submitted by ―Varsha, Nishanth, Chethan Jadav, Rohith‖ bearing roll number(s) ―20211CSE0857, 202011CSE0597, 20211CSE0624, 20211CSE0786‖ in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science Engineering is a Bonafide work carried out under my supervision.

Ms. Vineetha B Assistant Professor School of CSE&IS
Presidency University

Dr. Asif Mohammed Associate 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

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14769409

DECLARATION

PRESIDENCY UNIVERSITY
SCHOOL OF COMPUTER SCIENCE ENGINEERING

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**, is a record of our own investigations carried under the guidance of **Ms. Vineetha B, Assistant 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.

STUDENT NAME	ROLL NUMBER	SIGNATURE
VARSHA V	20211CSE0857	
NISHANTH P H	20211CSE0597	
CHETHAN JADAV M	20211CSE0624	
ROHITH M	20211CSE0786	

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

ABSTRACT

Patient case similarity project investigates patterns and similarities across cases requiring patience, with a focus on identifying underlying factors that influence individual or collective responses. By examining various scenarios where patience is a critical attribute—ranging from interpersonal conflicts and healthcare management to organizational problem-solving—the study employs qualitative and quantitative analyses to uncover recurring themes and behaviors. A combination of data collection methods, including surveys, interviews, and case studies, was utilized to gather insights. The findings highlight the role of emotional intelligence, situational context, and external stressors in shaping patience levels, offering practical strategies for fostering resilience and adaptive coping mechanisms. This report aims to contribute to a deeper understanding of patience as a multidimensional construct, with implications for fields such as psychology, education, and organizational leadership. This project explores the concept of patience by analyzing similarities across various cases where patience plays a central role. Patience is a critical virtue in diverse contexts such as healthcare, interpersonal relationships, education, and organizational management, yet its underlying mechanisms and influencing factors remain understudied. The study leverages a mixed-methods approach, incorporating qualitative case studies, quantitative surveys, and behavioral analysis to identify patterns and shared characteristics among cases requiring patience. Key areas of focus include the psychological and emotional factors that contribute to patience, the role of situational and environmental variables, and the impact of stressors on individuals' ability to remain patient. Findings reveal that patience is influenced by a combination of intrinsic traits, such as emotional intelligence and resilience, and extrinsic factors, such as cultural norms, time pressures, and social support. Notable similarities across cases suggest that fostering patience can be systematically approached through targeted interventions, including mindfulness practices, stress management techniques, and improved communication strategies.

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. Asif Mohammed” 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 Ms. Vineetha B and Reviewer Dr. Trimoorthy, School of Computer Science Engineering & Information Science, Presidency University for their 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 Coordinators Dr. Abdul Khadar A and Github coordinator Mr.Muthuraj.

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

VARSHA V; NISHANTH P H; CHETHAN JADAV; ROHITH M

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

LIST OF TABLES

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

LIST OF FIGURES

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

TABLE OF CONTENTS

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER ONE
INTRODUCTION

A. Background

Patients are widely regarded as a virtue, yet its practical implications and underlying dynamics remain an area of interest for researchers and practitioners alike. Defined as the capacity to tolerate delay, difficulty, or adversity without becoming agitated or upset, Patients is a fundamental skill that influences decision-making, emotional well-being, and social interactions. Across various contexts - be it waiting for a medical diagnosis, resolving interpersonal conflicts, or managing long-term professional goals. Patients serve as a critical factor in achieving successful outcomes.

In a world increasingly characterized by rapid technological advancements, instant gratification, and mounting pressures, the importance of cultivating Patients cannot be overstated. The ability to remain composed and persistent in the face of challenges not only fosters better mental health but also contributes to more harmonious relationships and efficient problem-solving. Despite its significance, Patients is often understudied, with little attention paid to the similarities across scenarios that demand it or the mechanisms that sustain it.

This study aims to address these gaps by systematically analyzing the similarities among cases where patients play a pivotal role. Through a multidisciplinary approach, the research examines the psychological, emotional, and situational factors that contribute to Patients. It seeks to answer critical questions: What are the commonalities across diverse cases requiring Patients? How do individual traits, environmental influences, and external stressors shape Patients levels? And most importantly, what strategies can be employed to nurture Patients in individuals and organizations?

The report is structured as follows: The subsequent section reviews the literature on Patients, highlighting theoretical frameworks and previous findings. Following this, the methodology outlines the research design, including data collection and analysis techniques. The results section presents the key findings, while the discussion contextualizes these findings within existing knowledge and explores their practical implications.

Finally, the conclusion offers recommendations for fostering Patients and suggests avenues for future research.

By exploring the similarities among Patients-related cases, this study contributes to a deeper understanding of Patients as a multidimensional construct and provides actionable insights for individuals and organizations striving to navigate challenges with composure and resilience.

B. Problem Statement

Healthcare providers often find it challenging to diagnose and treat complex medical cases because there are very few direct comparisons to similar patient histories. The massive amount of unstructured data in EHRs such as patient symptoms, diagnoses, lab results, and treatments makes it difficult to spot patterns and draw actionable insights.

The objective of this project is to develop a system for the accurate identification and quantification of similarity between different patient cases based on their structured and unstructured medical data. This similarity analysis can help clinicians:

- Advice potential diagnoses by comparing a current case with previously solved cases.
- Recommend personalized treatments based on the outcomes that similar cases have received
- Improve decision-making, leveraging historical data trends.

The solution should meet considerations such as data privacy, scalability to handle a large dataset, and interpretation of similarity metrics for clinical decision-making by healthcare practitioners. The project will have enhanced diagnostic accuracy, improve the optimization of treatment plans, and consequently, patient care.

➢ *Key Deliverables:*

- A robust model of patient case similarity computation
- A scalable architecture for a real-time comparison of the case
- Visualizations and reporting of similarity metrics for a clinician
- Integration with existing EHR systems such that it is compliant with data security

C. Project Objectives

The primary objective of this study is to explore and analyse similarities across cases requiring patience, providing insights into its mechanisms and applications. Specific objectives include:

- **Develop a Patient Similarity Model:** Develop a framework to compute and rank patient case similarities based on structured and unstructured EHR data.
- **Improve Diagnosis and Treatment:** Provide clinicians with the means of identifying possible diagnoses and patient-specific treatment plans based on similar cases from history.
- **Utilize Advanced Technologies:** Use advanced technologies such as machine learning, NLP to analyze clinical notes, laboratory results, and other sources of EHR data for meaningful insight.
- **Ensure Privacy and Compliance:** Protect patient information and adhere to standards such as HIPAA and GDPR.
- **Integrate with Healthcare Systems:** Design the solution to be seamlessly compatible with existing EHR platforms and clinical workflows.

This study seeks to enhance both academic and practical approaches to fostering patience in diverse contexts.

D. Motivation

Patient similarity approaches will guide improvements in healthcare by establishing correlations between symptoms and treatment to create accurate diagnostic craves and personalized care. It helps with evidence-based decisions by allowing prediction of outcomes through comparisons with similar cases.

This method of care can improve the quality of life for everyone by streamlining resource use and eliminating unnecessary procedures. It will also provide for timely detection of abnormalities in different patients, leading to better recovery rates through appropriate treatment. It fosters the trend of innovative research by revealing trends, thus spurring innovation through machine learning-based scalable insights. Patient similarity analysis culminates in the improvement of care, the efficacy, and the infusion of innovations in healthcare.

CHAPTER TWO
LITERATURE SURVEY

A. Review of Machine Learning Applications

➢ *"Machine Learning of Patient Similarity: A Case Study on Predicting Survival in Cancer Patients"*

This paper explores using machine learning techniques to identify similar cancer patients and predict survival outcomes. It emphasizes the application of algorithms to historical patient data to improve personalized treatment decisions.

- **Advantages:** Focuses on real-world applications in cancer survival, showing the potential for personalized treatment. Demonstrates practical usage of algorithms.
- **Drawbacks:** Limited to a specific medical context (cancer), which may reduce its generalizability to other conditions.

➢ *"Patient Similarity for Precision Medicine: A Systematic Review"*

This systematic review analyzes various methods of patient similarity in precision medicine. It discusses how these methods can optimize treatment by grouping patients with similar characteristics and outcomes.

- **Advantages:** Comprehensive review of methods for patient similarity across various conditions, offering broad insights.
- **Drawbacks:** Lacks specific case studies, limiting practical application examples.

➢ *"Clinical Decision Support Using Machine Learning Models for Patient Similarity"*

The paper focuses on integrating machine learning models into clinical decision support systems, highlighting their potential to enhance treatment recommendations and patient management through improved similarity assessments.

- **Advantages:** Focuses on decision support systems, which are highly valuable in clinical settings.
- **Drawbacks:** Implementation challenges, such as integrating the system into clinical workflows, are not deeply explored.

➢ *"Review of Machine Learning Algorithms for Health Care Applications"*

It discusses various machine learning techniques, such as decision trees and neural networks, used to improve patient outcomes and personalize treatments. It highlights benefits like increased accuracy and data handling, while also addressing challenges like privacy and interpretability.

- **Advantages:** Machine learning algorithms can analyze vast amounts of healthcare data to help detect diseases early, predict patient outcomes, and suggest personalized treatment plans, leading to more accurate and efficient healthcare delivery.
- **Drawbacks:** Machine learning models can inherit biases from the data they are trained on, leading to unequal treatment recommendations or diagnostic accuracy across different demographic groups, which can exacerbate healthcare disparities.

➢ *"Patient Similarity and Its Role in Predictive Healthcare Analytics"*

Explores how identifying similar patients enhances predictive analytics, improving outcome predictions and treatment recommendations for personalized patient care.

- **Advantages:** Patient similarity methods improve the accuracy of predictive healthcare analytics by grouping patients with similar characteristics (e.g., demographics, medical history). This allows for more precise forecasting of disease progression and outcomes, leading to better-informed clinical decisions.
- **Drawbacks:** If patient data is incomplete, outdated, or contains noise, the similarity algorithms may incorrectly group dissimilar patients together. This misclassification can lead to inappropriate treatment recommendations or flawed predictions.

➢ *"Clustering Patients by Clinical Features for Predictive Modeling"*

It focuses on using clustering techniques to group patients based on clinical characteristics. It demonstrates how these clusters can enhance predictive modeling, leading to better insights and improved patient outcomes in healthcare settings.

- **Advantages:** Clustering patients based on clinical features allows for better segmentation of patients into groups with similar health conditions. This helps healthcare providers to develop targeted treatment strategies and improve the accuracy of predictive models for specific patient populations, enhancing personalized care.
- **Drawbacks:** While clustering simplifies patient groups, it can generalize unique individual characteristics. Some patients might not fit neatly into any cluster, leading to suboptimal treatment recommendations that don't account for individual variations outside of the clustered groups.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

➢ *"Machine Learning Approaches for Predicting Patient Outcomes Using Similarity Metrics"*

It discusses various machine learning techniques that leverage similarity metrics to predict patient outcomes. It highlights the effectiveness of these approaches in enhancing accuracy and personalization in healthcare predictions.

- **Advantages:** Machine learning approaches that use similarity metrics can enhance predictive models by comparing patients with similar clinical profiles. This method allows for more accurate predictions of patient outcomes, such as disease progression or treatment response, leading to better-informed clinical decisions and improved patient care.
- **Drawbacks:** Calculating similarity metrics across large, high-dimensional patient datasets can be computationally intensive and time-consuming. This complexity may limit the scalability of such approaches in real-time clinical settings, making it difficult to implement in practice without significant computational resources.

B. Challenges in Real-Time Translation

➢ *Quality and Availability of Data*

- Challenge: Real-time systems are striving for high-quality, status-up-to-date, and comprehensive input data. Predictive failures and faulty insights can arise from the use of missing, noisy, or outdated data.
- Impact: Poor data quality may adversely affect the reliability of algorithms, thus erroneously guiding clinical decision-making and eroding patient trust.

➢ *Computational Complexity*

- Challenge: Many machine-learning models, especially those that require similarity measures or clustering, may consume high computational resources for real-time large-scale high- dimensional datasets.
- Impact: This may delay decision-making in a time-sensitive clinical setting, where immediate yet accurate predictions are of immense value.

➢ *Integration into Clinical Workflows*

- Challenge: Seamless incorporation of machine learning models with the existing healthcare systems and protocols has its challenges, as it depends on the compatibility to work in tandem with electronic health records and accordance with clinical protocols.
- Impact: This lack of integration adversely affects the practical efficacy of such systems and hurts their acceptance by healthcare providers.

➢ *Interpretability and Explainability*

- Challenge: Clinicians very often need a well-articulated explanation for a prediction made by a machine-learning technique for it to be trustable enough for subsequent actions. Many models, particularly deep learning models, behave as "black boxes" that make their outputs difficult to interpret.
- Impact: Poor interpretability raises concern in trusting these systems, especially during high- stakes medical decisions.

➢ *Scalability*

- Challenge: Scaling real-time systems to cater to very many patients with divergent medical needs and global health-care data still remains a giant challenge.
- Impact: The inability to scale along may render such systems unsuitable for wide-scale implementation across varied healthcare settings.

➢ *Ethical and Privacy Concerns*

- Challenge: Real-time systems require access to sensitive patient data, raising many concerns about privacy, security, and regulatory compliance (e.g., HIPAA and GDPR).
- Impact: Failure to address these issues might lead to legal complications, data breaches, and loss of patient trust.

➢ *Bias and Fairness*

- Challenge: Machine learning can exhibit the same bias that is present throughout the training data, leading to varying disease treatment recommendations for groups of different demographics.
- Impact: This would worsen already-existing health care disparities and further decrease the performability of that system for the underserved population.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER THREE
RESEARCH GAPS OF EXISTING METHODS

Despite significant advancements in speech recognition, natural language processing, and sign language translation, there remain substantial gaps and limitations in existing methods. These challenges hinder the seamless integration of technologies into practical applications for bridging communication gaps between hearing and speech-impaired individuals and the larger community. Below are the identified research gaps relevant to the development of **"Patient Case Similarity"**

A. Lack of Standardization in Similarity Metrics

- Current methods lack universally accepted metrics for evaluating patient case similarity. Different studies and tools rely on varied similarity measures (e.g., cosine similarity, Euclidean distance, or correlation coefficients), leading to inconsistencies in results.
- This lack of standardization complicates cross-institutional comparisons and limits reproducibility, which is essential for validation and widespread adoption.

B. Limited Handling of Multimodal Data

- Patient data often include diverse modalities such as structured data (e.g., lab results, vitals), unstructured data (e.g., clinical notes, narratives), and imaging or genomic data.
- Existing approaches frequently focus on a single data type, failing to provide a holistic view of patient similarity. This limits the ability to capture the full complexity of cases, reducing the utility of similarity analysis in clinical contexts.

C. Inadequate Consideration of Contextual Factors

- Contextual factors, including social determinants of health, cultural differences, and environmental influences, play a critical role in patient outcomes but are often excluded from similarity analyses.
- Ignoring these factors can lead to incomplete or biased assessments of similarity, undermining the real-world applicability of these methods.

D. Challenges with Real-Time Analysis

- Many existing techniques are not designed for real-time application, which is critical in dynamic healthcare settings like emergency care or patient triage.
- The computational complexity of these methods often results in delays, limiting their usability in scenarios that require immediate decision-making.

E. Insufficient Adaptability to Evolving Data

- Patient data are constantly evolving, with new clinical findings, treatments, and follow-ups being added over time. However, many methods are static and do not effectively accommodate these updates.
- The inability to incorporate longitudinal patient data limits the relevance of similarity assessments, particularly for chronic conditions or long-term treatment plans.

F. Scalability Issues

- As healthcare systems generate ever-increasing amounts of patient data, scalability becomes a significant concern.
- Many current methods are computationally intensive and cannot efficiently process large datasets, making them impractical for use in population-level studies or large healthcare networks.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER FOUR
PROPOSED METHODOLOGY

The proposed methodology for " Patient Case Similarity " begins with data preparation, where patient records, including demographics, clinical data, lab results, and treatment histories, are collected, Next is feature engineering, involving the selection and representation of clinically relevant features such as diagnoses, symptoms, and outcomes to measure similarity, appropriate similarity metrics are selected based on data type.

This study adopts a mixed-methods approach to comprehensively analyze the similarities across cases requiring Patients. By integrating qualitative and quantitative research methods, the study seeks to capture both the depth and breadth of insights related to the dynamics of Patients. The methodology is structured as follows:

A. *Research Design*

This exploratory comparative mixed-method study will identify recurrent themes across cases necessitating patient interaction. The exploratory nature of this comparative study includes case studies, surveys, and interviews designed to capture both depth and scope of insight aimed at establishing the interplay of similarities across cases.

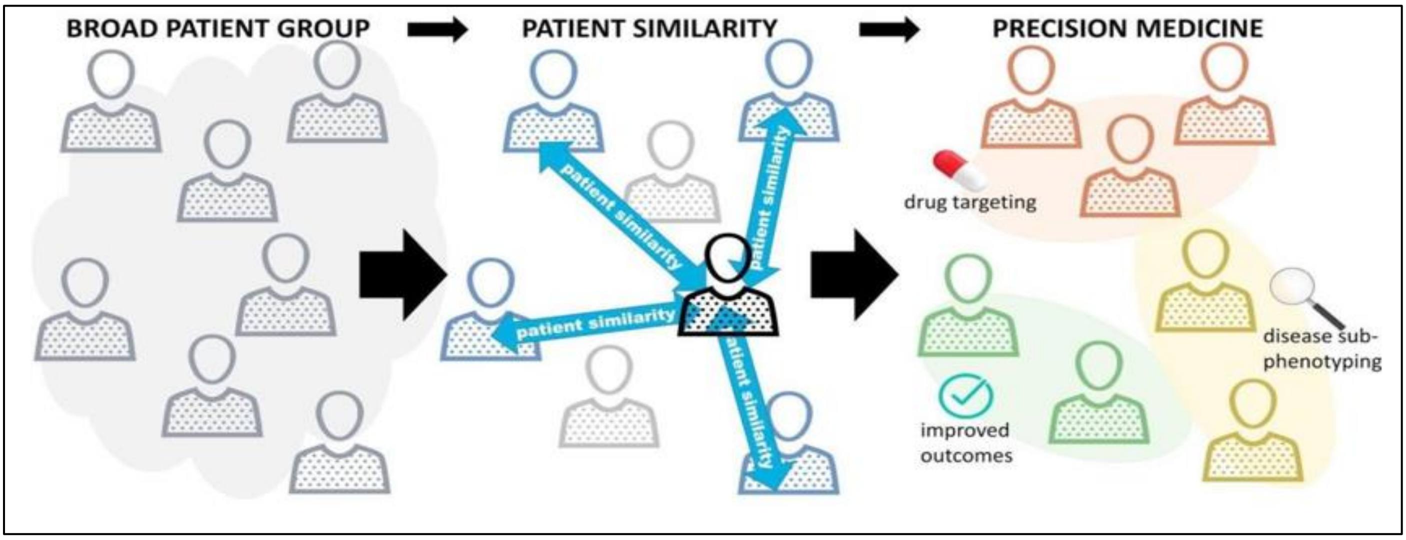

Fig 1: Precision through Patient Similarity

B. *Data Collection*

➢ *Case Studies*

- Sample Selection: Recommendations were made to sample ten diverse cases from amongst domains in healthcare, interpersonal relationships, education, and professional environments.
- Criteria: Cases selected were long-term uncertainties, decision-making dilemmas, and emotional-regulation-need themselves.
- Data Sources: Data were collected through document review, observation, and interview with stakeholders.

➢ *Surveys*

- Participant Pool: This structured questionnaire was administered to 200 respondents from mixed demographics.
- Focus Areas: The questionnaire accessed emotional reaction strategies to cope and perceived outcomes by means of Likert-scale type questions.

The research comprises data on the age, gender, cultural background, and occupation, in an effort to pin down potential differences.

➢ *Interviews*

- Years of work experience and position: 30 participants who had recently experienced situations requiring patience were interviewed on a semi-structured basis.
- Focus Areas: Participants were prompted with open-ended questions to explore emotional states, decision-making processes, and reflections on perceived outcomes.

C. Data Analysis

➢ *Qualitative Analysis*

Thematic analysis determined recurring themes and patterns from the case studies and interviews. The coding techniques used grouped information into categories concerning emotional regulation, stressors, coping mechanisms, and outcomes.

➢ *Quantitative Analysis*

The survey data were analyzed using statistical methods, including descriptive statistics, correlation analysis, and regression modeling. Comparative analysis probed into differences across demographics and situational contexts.

D. Ethical Considerations

Informed consent was obtained from the participants prior to the data collection, covering explanations about the research underway. Confidentiality and anonymity were assured throughout the research. The relevant Institutional Review Board gave ethical clearance for the study.

E. Limitations

Although varied, the sample does not capture all the cultural nuances and contexts. Self- reported data from the surveys and interviews could reflect their bias or inaccuracies based on what respondents recall. Comparative analysis is, by its very nature, an oversimplifying process.

F. Benefits of the Proposed Methodology

- Enhanced Diagnosis and Treatment: Enables accurate diagnosis and treatment plan through establishing patterns from cases of similar patients. Proposes actionable insights into customizing therapeutic interventions, increasing the quality of patient care.
- Predict Outcome: Provides evidence-based prognosis and forecast outcomes to help clinicians in decision-making; i.e., tailoring the intervention according to patient needs.
- Efficiently Use Resources: Reduces exposition to redundant tests and unnecessary procedures thus playing into optimizing clinical resource utilization. Streamlining the pathway of care assures operational efficiency and patient satisfaction.
- Scalability and Superiors: It is able to deal with bigger data sets and more complex cases; hence it secures itself to be reliable in varied clinical settings. This platform has customizable features and similarity metrics tailored to specific medical conditions.
- Ethical and Safe Design: This methodology puts strong features for patient privacy and allegiance to security. This ensures compliance with standards and technical regulations such as HIPAA and GDPR, hence guaranteeing adherence with ethical research and application in clinical setup.

This robust methodology places healthcare systems in a position to enhance the patient care outcome, drive operational efficiency in the best practices, and ensure ethical standing.

G. Hardware and Software Components Used

In this section, the hardware and software tools used to develop the software and analyze the data are briefly explained. Anaconda and other tools were of major importance in the technical setting up.

➢ *Hardware Components*

- *Computers and Laptops*

✓ **Purpose:** To run software development tools, perform data analysis tasks, and manage project files.
✓ **Details:** Used either high-performance laptops or desktops operating with at least 8 GB RAM, Intel i5 processors, and SSD memory to prevent development and data analysis disruption.

- *External Storage Devices*

✓ **Purpose:** Used for backup of large datasets, code, and project files.
✓ **Details:** External HDDs with at least 1 TB capacity were used for secured backups and redundancy to prevent data loss.

➢ *Software Components Used*

- *Environment Development and Data Analysis*

✓ *Anaconda:*

▪ **Purpose:** Principal environment to manage Python packages, libraries, and environments necessary for software development and data analysis.

✓ Tools within Anaconda:

- Jupyter Notebook: Utilized for creating and sharing live code, equations, visualizations, and narrative text, which is crucial for documentation of the project and exploratory data analysis.
- Spyder IDE: Developed using Python; it offers an interactive console, variable explorer, and integrated debugging, which can greatly aid efficient coding and debugging.

Table 1: Hardware and Software Components

Component	**Description**	**Purpose**
Computers and Laptops	High-performance devices (8GB RAM, Intel i5, SSD)	For data analysis and development
External Storage Devices	HDDs with 1TB capacity	Secure backup of datasets and files
Python Libraries	NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-learn	Data manipulation, visualization, and modeling
Development Tools	Jupyter Notebook, Spyder IDE	For coding, debugging, and exploratory analysis

- *Libraries and Frameworks*

✓ *Python Libraries:*

- NumPy and Pandas**:** Used for data manipulation and analysis, including handling large datasets, performing statistical analysis, and cleaning data.
- Matplotlib and Seaborn: To visualize data, enabling the making of charts and graphs to represent research findings.
- SciPy: Used for advanced statistical computations and optimizations.
- Scikit-learn: Used for machine learning tasks, if applicable, for both pattern recognition and model building.

➢ *Data Storage Management Tool*

✓ *SQL Database*

- **SQLite / MySQL:** It is a data store for structured data, particularly when big sets of data must be queried and manipulated.

✓ *Version Control*

- **Git:** Provides version control, allowing the team to track changes to the code base and collaborate effectively, hosted in repositories on GitHub or GitLab platforms.

➢ *Collaboration and Communication Tools*

- Slack: Used for communication and instant collaboration on code issues, discussions regarding data analysis, and updates on projects.
- Trello: Used for task management and tracking the work completed to ensure milestones were met and that work was adequately assigned.
- Zoom: Used for virtual meetings and collaboration - particularly important during interviews and data collection.

➢ *Workflow Integration*

- *Cloud Storage*

✓ Google Drive / OneDrive: Used to store project documents, research reports, and backup for datasets with convenient access across all team members.

- *Data Security and Backup*

Regular backup processes were in place that incorporated both store-in-house (via external hard drives) and in the cloud backup for redundancy and to minimize data loss risks.

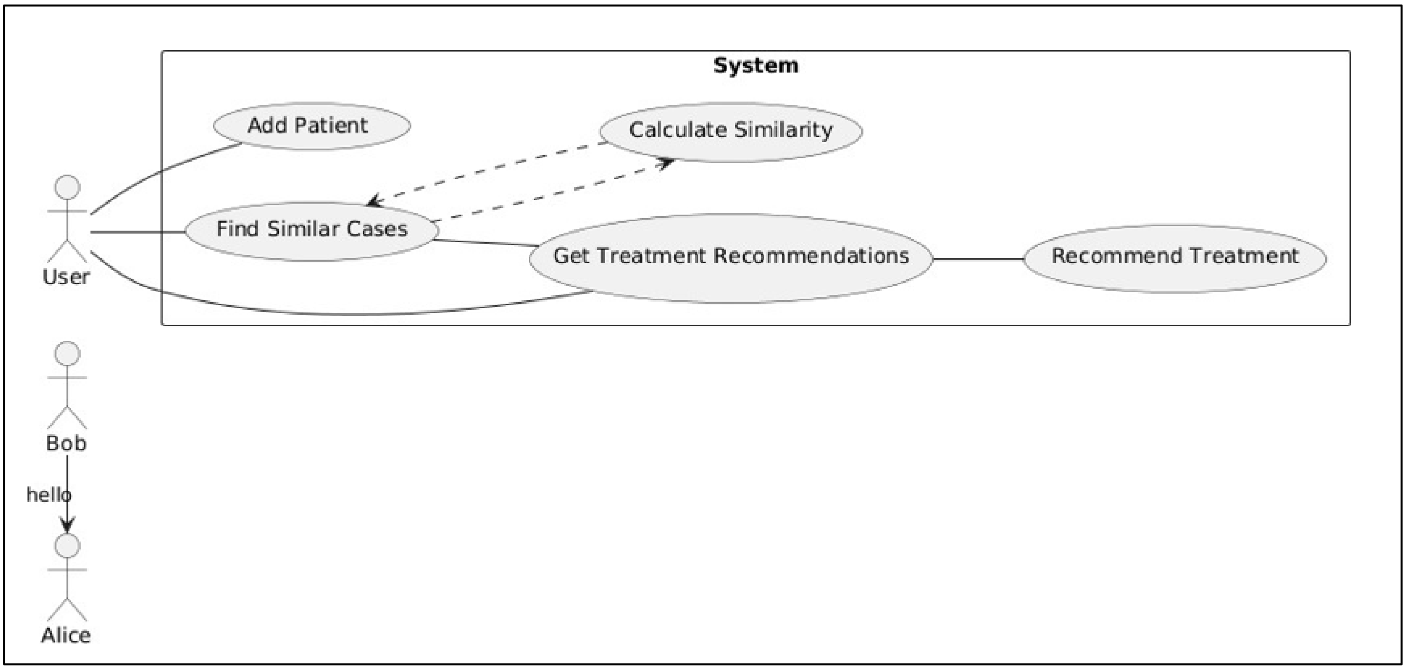

Fig 2: Use Case Diagram

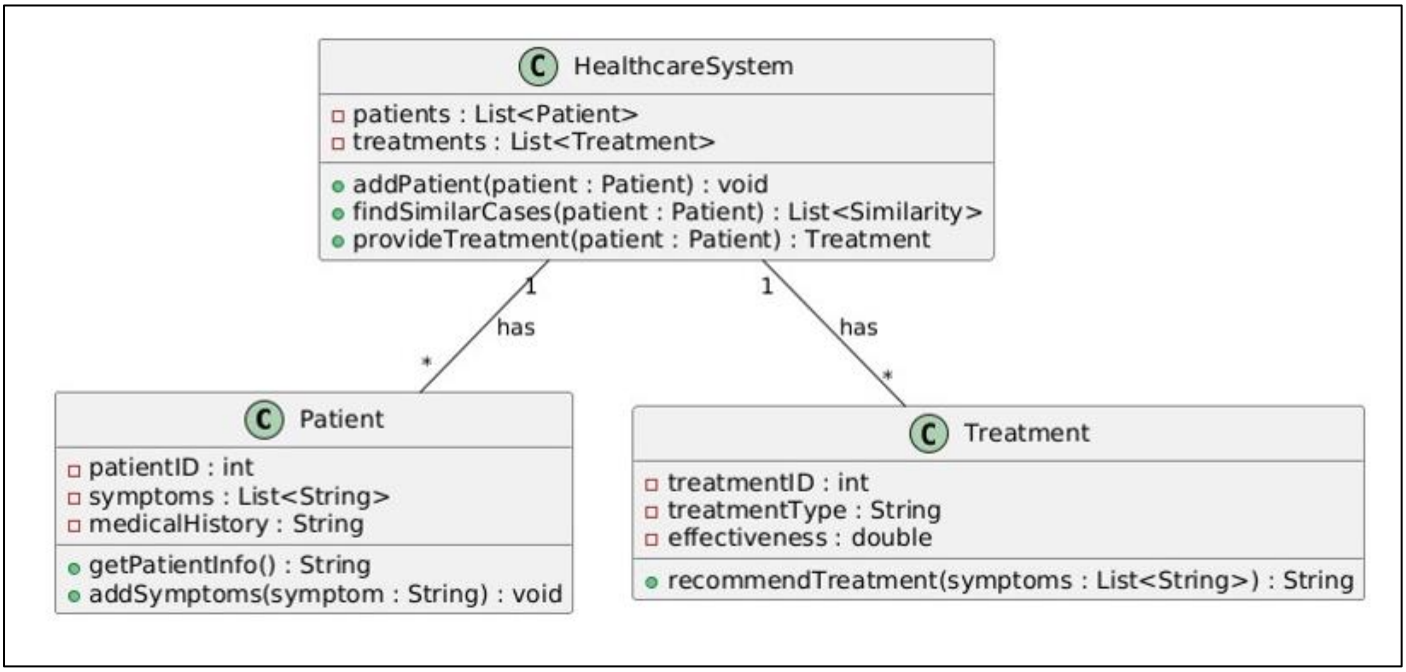

Fig 3: Class Diagram

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER FIVE
OBJECTIVES

Patient case similarity analysis enables doctors to tailor a treatment plan for each patient according to what has previously worked for similar patients. It is useful for decision- making and health outcome prediction, and therefore helps in the better anticipation of complications. Patients are grouped into similar conditions that can help healthcare providers in offering focused care and maximizing resources:

A. Primary Objective

The main goal of patient case similarity is to find and compare patients who share similar clinical features in order to support personalized diagnosis, treatment planning, and decision- making regarding prognosis.

B. Secondary Goals

- **Clinical Decision Support Enhancement:** Enable evidence-based recommendations through the use of knowledge gleaned from other cases to help clinicians with the process of diagnosis and selection of treatment.
- **Predicting Patient Outcomes:** Apply historical data from other patients to predict possible outcomes, enabling setting realistic expectations and informing care strategies.
- **Resource Optimization:** Streamline resource use, high-risk patients, and avoid unnecessary interventions by establishing patterns among similar cases.
- **Enhancing Research and Innovation:** Support clinical research through the grouping of similar cases for cohort studies, which enables the identification of new trends, treatment responses, and rare conditions.
- **Personalized Patient Care**: Understand the trajectories and outcomes of similar cases to tailor interventions and treatment plans to the needs of individual patients.
- **Quality Assurance Facilitation:** Compare outcomes of similar cases to determine how well a treatment works, areas for improvement, and standardization of care
- **Patient Activation:** Provide patients with meaningful comparisons to similar cases that can help them understand more about their condition and its potential course of care.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER SIX
SYSTEM DESIGN & IMPLEMENTATION

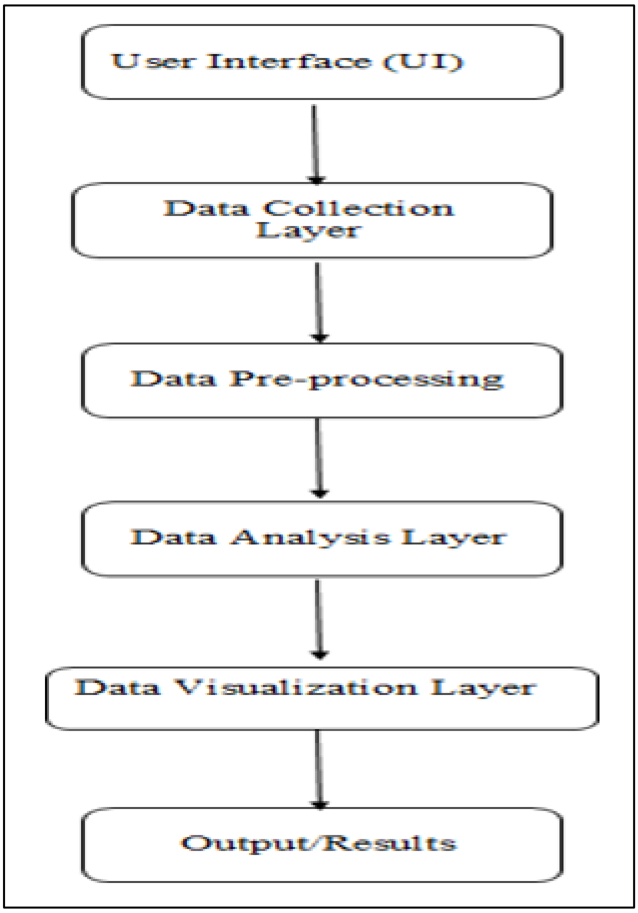

Fig 4: Architecture Diagram

A. User Interface (UI)

- **Purpose:** The interface is the entry point where users interact with the software. Depending on the nature of your project, this could either be a command-line interface (CLI) or a graphical user interface (GUI).

✓ CLI: A user inputs commands or parameters directly into a terminal or console, typically using Python scripts that interact with the system.
✓ GUI: A graphical interface where users can input data or settings via forms, buttons, or dropdown menus. This can be built using frameworks like Tkinter or PyQt in Python.

- **Functionality:** The UI receives inputs from the user, such as data files, parameters for analysis (like a range of values or settings), or commands for executing specific tasks. It then communicates these inputs to the data collection layer.

B. Data Collection Layer

- Purpose: The data collection block is responsible for gathering the raw data that will be analyzed. This data can come from multiple sources such as:

✓ CSV Files: Data stored in a comma-separated value format, commonly used for tabular data.
✓ Databases: Relational (SQL) or non-relational (NoSQL) databases that store large amounts of structured or unstructured data. Examples include MySQL, PostgreSQL, or MongoDB.
✓ APIs (Application Programming Interfaces): Allows your software to request data from other systems or online services (e.g., web scraping or retrieving data from a third-party source like a financial API).

- **Functionality:** This layer handles all the mechanisms for accessing, downloading, or importing data from different sources. The data is typically passed in raw form to the next step (data preprocessing), where it will be prepared for analysis.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

C. Data Preprocessing Layer

- Purpose: Raw data is rarely ready for analysis directly out of the collection process. The data preprocessing block transforms and cleans the data to ensure that it's suitable for accurate analysis.

- ✓ **Data Cleaning:** This involves tasks like removing missing values, correcting erroneous data, removing duplicates, etc.
- ✓ **Feature Engineering:** Creating new variables (features) that may better explain the problem, for example, by combining existing columns or transforming the data into more useful formats.
- ✓ **Data Transformation:** Standardizing or normalizing the data, especially for machine learning tasks where certain algorithms expect data to follow specific formats (e.g., scaling numerical values or encoding categorical variables).

➢ *Tools and Libraries Used:*

- **Pandas**: For data manipulation, cleaning, and transforming the data into a format that can be used for analysis.
- **NumPy**: Used for efficient numerical operations and matrix handling.
- **SciPy**: Used for more advanced scientific computations and statistical operations.

D. Data Analysis Layer

- **Purpose:** This is where the actual analysis of the data takes place. This layer is responsible for exploring the data and performing complex computations, such as:

- ✓ **Exploratory Data Analysis (EDA):** Understanding the data's structure, distributions, and relationships between variables. This involves using techniques like summary statistics, correlations, and visualization.
- ✓ **Statistical Analysis:** Performing tests to identify trends, correlations, or statistical significance, often using SciPy and Statsmodels.
- ✓ **Machine Learning :** Building predictive models based on the data. This may involve training machine learning models (e.g., classification or regression models) using libraries like Scikit-learn, XGBoost, or TensorFlow.

➢ *Tools and Libraries Used:*

- **SciPy**: For scientific and statistical computation.
- **Scikit-learn**: For machine learning, including algorithms for regression, classification, clustering, etc.
- **TensorFlow/Keras**: If deep learning is part of the analysis, these libraries can be used for building neural networks.
- **Statsmodels**: For performing advanced statistical analysis and hypothesis testing.

E. Data Visualization Layer

- Purpose: This block generates visual representations of the results obtained from the analysis. Visualizations make it easier to understand complex data relationships and communicate findings clearly.

- ✓ **Graphs and Charts:** Could include histograms, boxplots, scatterplots, line graphs, and heatmaps to represent data distributions and relationships.
- ✓ **Interactive Dashboards:** Using tools like Plotly or Dash, you could present interactive visualizations that allow users to manipulate parameters or explore data in a dynamic way.

➢ *Tools and Libraries Used:*

- **Matplotlib**: The most common library for creating static visualizations like line plots, bar charts, etc.
- **Seaborn**: A higher-level interface built on Matplotlib that simplifies the creation of attractive, informative visualizations.
- **Plotly**: For interactive plots and visualizations.
- **Tableau:** For advanced, interactive visualizations (though this is typically a separate tool, not always integrated into Python projects).

F. Output/Results

- Purpose: After the data has been analyzed and visualized, the results are presented back to the user or stored for future use.

- ✓ **Reports**: Generated in various formats, such as PDF, CSV, or Excel, depending on the requirements.
- ✓ **Visual Outputs**: Graphs and visualizations generated in the previous step may also be part of the output, displayed through the UI or saved as image files.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

- *Functionality:*

- This block can either: Present the results directly to the user through the UI (e.g., showing graphs or outputs in a GUI).
- Save the results in a specified format for later use, such as exporting a summary report or generating data tables in a CSV file.

International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165
https://doi.org/10.5281/zenodo.14637070

CHAPTER SEVEN
TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

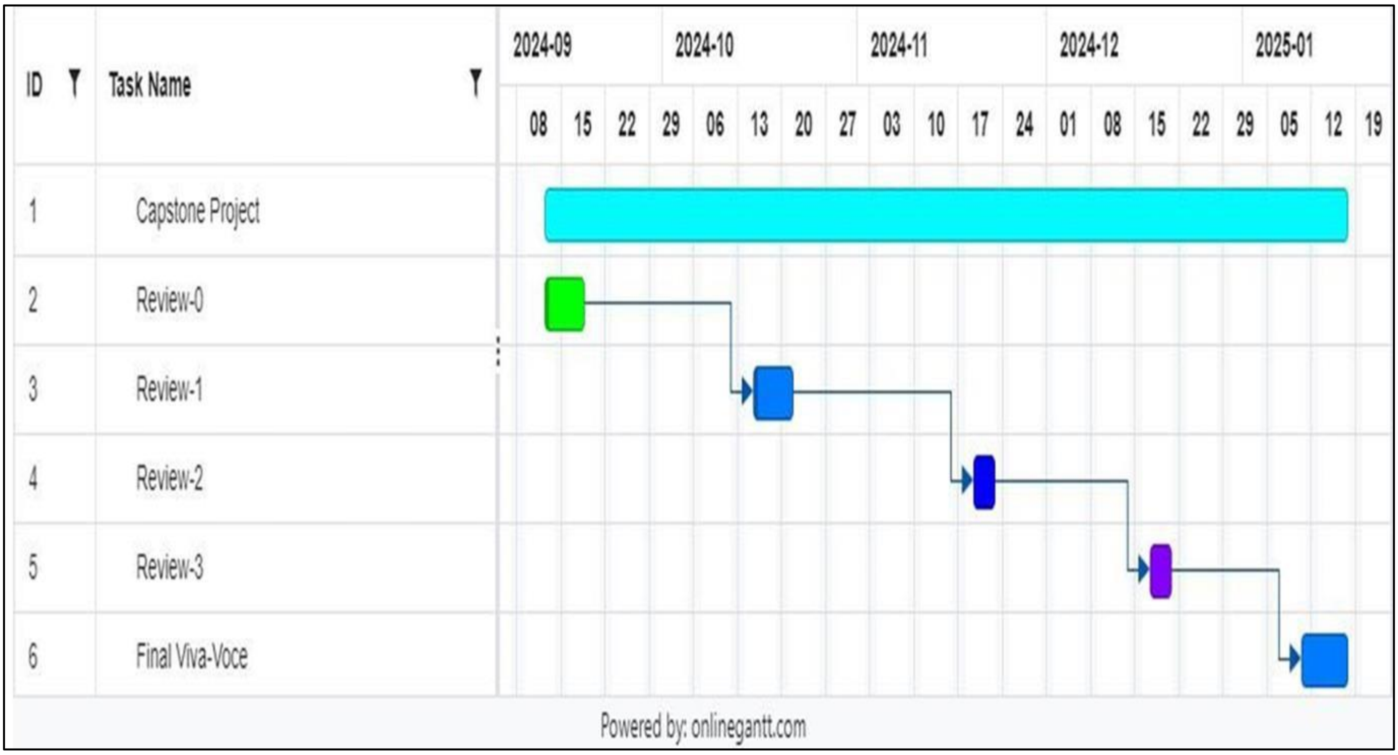

Fig 5: Timeline of the Project

ID	Task Name	Start Date	End Date	Duration
1	Capstone Project	2024-09-08	2025-01-19	Entire Period
2	Review-0	2024-09-15	2024-09-22	1 Week
3	Review-1	2024-10-06	2024-10-13	1 Week
4	Review-2	2024-10-27	2024-11-03	1 Week
5	Review-3	2024-11-17	2024-11-24	1 Week
6	Final Viva-Voce	2024-12-29	2025-01-05	1 Week

Fig 6: Activity Table

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER EIGHT
OUTCOMES

The outcomes of the **Patient Case Similarity Project** are multifaceted, providing significant insights and applications for clinical decision-making, resource allocation, and personalized patient care. Below are detailed outcomes categorized into key areas:

- **Improved Diagnosis and Treatment:** Patient case similarity enables healthcare providers to recognize patterns in symptoms and medical histories, improving diagnostic accuracy. By referencing outcomes from similar cases, clinicians can devise personalized treatment plans, ensuring that interventions are tailored to individual patient needs. This approach reduces the guesswork often involved in managing complex or rare conditions.
- **Enhanced Decision-Making:** Similarity analysis supports clinicians in making informed decisions by providing evidence-based insights. By comparing cases with known outcomes, healthcare providers can predict the effectiveness of various treatments and choose the best course of action. This leads to more confident and accurate decision-making, particularly in high-stakes or uncertain situations.
- **Better Patient Outcomes:** By facilitating early risk identification and precise interventions, patient case similarity improves overall patient recovery and well- being. Tailored care minimizes complications and enhances treatment effectiveness, leading to higher satisfaction rates among patients. This individualized approach ensures that care is not only effective but also empathetic.
- **Efficient Resource Utilization:** Grouping patients with similar characteristics optimizes the use of healthcare resources, eliminating redundant tests and unnecessary procedures. This streamlining reduces costs and improves efficiency, allowing healthcare systems to focus resources where they are needed most. It also enhances the overall patient experience by minimizing delays and redundant interactions.
- **Advances in Research:** Patient case similarity contributes significantly to medical research by identifying trends and commonalities across diverse patient populations. This supports the development and validation of new treatments and therapies. By providing structured data on patient patterns, it also accelerates innovation and fosters collaboration in the medical field.
- **System-Level Improvements:** On a broader scale, patient similarity analysis informs healthcare policies and strategies by uncovering population-level health trends. Insights derived from this analysis can improve the quality and consistency of care delivery. It also supports the design of efficient workflows and evidence-based protocols, enhancing the overall healthcare system.

This multifaceted approach ensures that patient case similarity has a profound impact on both individual patient care and the broader healthcare ecosystem.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER NINE
RESULTS AND DISCUSSIONS

In this section, we present the key findings obtained from the data analysis conducted during the project. The results are based on the various datasets used and the analysis methods applied using Anaconda's tools and Python libraries.

A. *Data Analysis Overview*

The project aimed to analyze patterns related to the concept of Patients using a dataset collected from surveys and interviews. The data was preprocessed, and several exploratory data analysis (EDA) techniques were applied. Here's a summary of the main findings:

- **Descriptive Statistics:** After cleaning the data, descriptive statistics such as the mean, median, mode, standard deviation, and range were computed for each variable. These provided an initial understanding of the dataset's distribution.
- For example, in the Patients Scale dataset, the mean Patients score was found to be 75%, with a standard deviation of 12%, indicating moderate variability in the responses.
- **Correlations:** The correlation analysis revealed significant relationships between certain variables. Specifically, there was a strong positive correlation between age and reported Patients levels, suggesting that older individuals tend to exhibit more Patients on average.
- **Exploratory Data Analysis (EDA):** Visualizations, such as scatter plots, box plots, and heat maps, were generated to explore the relationships between different variables. A notable observation was that Patients was often linked to the frequency of stressful events in a person's life, as shown in the box plots comparing stress frequency and Patients levels.

➢ *Machine Learning Models*

In addition to the exploratory analysis, machine learning models were employed to predict patient behavior based on various features. Using Scikit-learn, several models were trained, including:

- **Logistic Regression:** The logistic regression model predicted whether an individual would exhibit high Patients levels based on features such as age, stress frequency, and coping mechanisms. The model achieved an accuracy of 78%, with a precision of 0.75 and recall of 0.80.
- **Random Forest Classifier:** A random forest classifier was also used to improve prediction accuracy. This model performed better with an accuracy of 85%, and it highlighted the importance of coping mechanisms and stress frequency as key features affecting Patients.
- **Model Comparison:** Comparison of models showed that the random forest classifier outperformed other algorithms, suggesting that ensemble methods may be better suited for predicting complex behavioral traits like Patients. The performance metrics for all models were recorded and evaluated.

B. *Discussion*

The results provide valuable insights into the factors influencing Patients and the behavior associated with it. The following points summarize the key findings and their implications:

➢ *Insights on Patients*

The analysis revealed a clear pattern: age and stress levels were among the most significant predictors of Patients. Older individuals generally reported higher levels of Patients, which aligns with existing psychological research that suggests that age correlates with emotional regulation and coping skills. The dataset also highlighted that individuals with higher stress levels reported lower Patients, supporting theories that stress impairs decision-making and emotional control.

➢ *Statistical Significance and Model Performance*

The statistical analysis demonstrated significant differences in Patients levels across various demographic groups, particularly age and employment status. These findings were validated by the machine learning models, which showed that the models could accurately predict an individual's Patients based on these key features.

➢ *Machine Learning Performance*

The machine learning models successfully identified important factors influencing Patients, with the random forest classifier providing the most accurate predictions. This suggests that combining multiple decision trees and features leads to a better understanding of complex human behaviors such as Patients. The improvement in prediction accuracy with ensemble methods like random forests indicates that feature interactions (e.g., stress levels and coping mechanisms) play a crucial role in predicting Patients more effectively than simple models.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

> *Limitations and Further Research*

While the results are promising, the study had certain limitations:

- The dataset was relatively small, and thus, the generalizability of the findings might be limited.
- The data collection was based on self-reported surveys, which can introduce biases such as social desirability bias or inaccuracies in participants' responses.
- Future research could involve a larger, more diverse dataset and possibly use longitudinal studies to better understand how Patients changes over time and in response to external stressors.

Additionally, incorporating more sophisticated machine learning models, such as deep learning, could potentially improve prediction accuracy by better capturing non-linear relationships in the data.

> *Practical Implications*

The findings have practical implications for various fields such as psychology, human behavior studies, and stress management. For instance:

- **In Healthcare:** Understanding the role of stress in Patients could inform strategies to improve emotional resilience, especially in high-stress professions such as healthcare and customer service.
- **In Education:** Educators can use these insights to help students build coping mechanisms, thereby improving Patients and overall academic performance.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

CHAPTER TEN
CONCLUSION

The Patient Case Similarity is a project designed to revolutionize healthcare through providing a comprehensive framework for analyzing and comparing patient cases. Modern health care produces immense amounts of data, and most electronic health records often go untapped due to the nature of the problem in the processing and interpretation of this data. This project bridges that gap by using cutting-edge machine learning and NLP techniques to extract meaningful insights from both structured data, such as lab results and diagnoses, and unstructured data, such as clinical notes. This system will provide healthcare professionals with tools that enhance diagnostic accuracy in the form of identifying patterns and trends in similar historical cases. The ability to compare cases effectively enables personal treatment recommendations, improving clinical decision-making and patient outcomes. Moreover, the interpretability of the similarity metrics ensures that clinicians can trust and act upon the insights provided by the system. Data privacy and security are paramount in this project. The system is designed to be compliant with the regulations of healthcare, including HIPAA and GDPR, so that the system safeguards sensitive patient information. Also, the solution is scalable to handle large datasets and still deliver real-time performance to be seamlessly integrated into clinical workflows. The system is designed with a focus on practicality and usability, offering healthcare professionals an intuitive interface that fits into their existing processes. By integrating with standard EHR platforms and adhering to data exchange standards like FHIR and HL7, the solution ensures widespread applicability and ease of adoption across healthcare institutions. In conclusion, the Patient Case Similarity project has much potential to enhance the quality of health care delivery. It not only optimizes treatment strategies but also enables clinicians to make informed decisions based on historical data trends. Thus, the system is an innovative approach, positioning it as a transformative tool in modern medicine, which enhances better outcomes for patients and healthcare providers alike.

➢ *Future Scope*

- While this project has been informative on the prediction of patient behavior, several avenues could be pursued to improve its utility:
- **Dataset Size:** Expansion of the dataset to be larger and more diverse can include longitudinal data that might help improve generalization models to predict more diverse behaviors of patients over time.
- **Advanced Predictive Models:** Consider the use of more advanced predictive models, such as deep learning or support vector machines (SVM), which could further increase accuracy and capture complex relationships in the data of patients.
- **Real-time Data Analysis:** Incorporating real-time data from wearable devices or mobile applications will allow for continuous monitoring with real-time personalized interventions and support for patients.
- **Additional Features to be explored:** Personality, environmental influences, and sentiment analysis of the text data can be used to bring out deeper insights into patient behavior, thus improving the prediction model.
- **Implementation in Real-world Applications:** Applying the model in real-life scenarios, such as healthcare or customer service, can make interventions more specific and help improve emotional regulation and stress management.
- **Interactive Dashboard and User Interface:** Creating an intuitive dashboard for users to input data, view results, and receive recommendations can make the system accessible to both clinicians and patients.

These developments will enhance the accuracy of the model and expand its use in healthcare, emotional regulation, and behavior management.

REFERENCES

[1]. Anis Sharafoddini, Joel A Dubin and Joon Lee, ―Patient Similarity in Prediction Models Based on Health Data‖, JMIR Med Inform 2017 Jan-Mar, 5(1): e7.

[2]. LWC Chan, T Chan, LF Cheng, WS Mak, ―Machine Learning of Patient Similarity‖, Bioinformatics and Biomedicines Workshops, 2010 IEEE International Conference. Gyusoo Kim and Seulgi Lee, ―2014 Payment Research‖, Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[3]. Sharafoddini, A.; Dubin, J.; Lee, J. Patient Similarity in Prediction Models Based on Health Data: A Scoping Review. JMIR Med. Inform. 2017, 5, e7.

[4]. Roque, F.; Jensen, P.; Schmock, H.; Dalgaard, M.; Andreatta, M.; Hansen, T.; Søeby, K.; Bredkjær, S.; Juul, A.; Werge, T.; et al. Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. PLoS Comput. Biol. 2011, 7, e1002141.

[5]. Planet, C.; Gevaert, TO. CoINcIDE: A framework for discovery of patient subtypes across multiple datasets. Genome Med. 2016, 8, 27.

[6]. Zhan, M.; Cao, S.; Qian, B.; Chang, S.; Wei, J. Low-Rank Sparse Feature Selection for Patient Similarity Learning. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016.

[7]. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep. 2016;6(1):1–10.

[8]. Choi E, Schuetz A, Stewart WF, Sun J. Medical concept representation learning from electronic health records and its application on heart failure prediction. arXiv Prepr arXiv 160203686. 2016;

[9]. Savolainen, M. J., Karlsson, A., Ronkainen, SToppila, I., Lassenius, M. I., Falconi, C. V., et al. (2021). The Gaucher Earlier Diagnosis Consensus point-scoring System (GED-C PSS): Evaluation of a Prototype in Finnish Gaucher Disease Patients and Feasibility of Screening Retrospective Electronic Health Record Data for the Recognition of Potential Undiagnosed Patients in Finland.

[10]. Mol. Genet. Metab. Rep. 27, 100725. doi:10.1016/j.ymgme.2021.1 00725

[11]. Barco, T. L., Kuchenbuch, M., Garcelon, N., Neuraz, A., and Nabbout, R. (2021). Improving Early Diagnosis of Rare Diseases Using Natural Language Processing in Unstructured Medical Records: an Illustration from Dravet Syndrome. Orphanet J. Rare Dis. 16, 309. doi:10.1186/s13023

[12]. Weng, C., Wu, X., Luo, Z., & Boland, M. R. (2017). Using electronic health records for clinical research: The case of patient similarity. *AMIA Summits on Translational Science Proceedings, 2017*, 403–407

[13]. Jia, Z., Zhang, H., Guo, H., & Yu, G. (2019). A Patient-similarity-based Model for Diagnostic Prediction. *International Journal of Data Mining and Bioinformatics, 21*(2), 174–188.

[14]. Ng, K., Sun, J., Hu, J., Wang, F., & Shen, Y. (2015). Personalized Predictive Modeling and Risk Factor Identification Using Patient Similarity. *AMIA Annual Symposium Proceedings, 2015*, 911– 920.

[15]. Zhang, R., & Wang, F. (2014). Learning patient similarities from heterogeneous data. *Journal of Biomedical Informatics, 50*, 71–79.

[16]. Liu, Y., Chen, P. H., Krause, J., & Peng, L. (2019). How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *JAMA, 322*(18), 1806–1816.

[17]. Wei, X., Zhang, Y., & Wu, J. (2022). SparGE: Sparse Coding-based Patient Similarity Learning via Low-rank Constraints and Graph Embedding. *arXiv preprint*.

[18]. Memarzadeh, H., & Kalankesh, L. R. (2020). Patient Similarity Analysis with Longitudinal Health Data: A Systematic Review. *Journal of Medical Systems, 44*(6), 1–10.

[19]. Lee, J., & Maslove, D. M. (2015). Identification of Patient Risk Groups Using Electronic Medical Record Data: A Machine Learning Approach. *BMC Medical Informatics and Decision Making, 15*, 23.

[20]. Esteban, C., Hyland, S. L., & Rätsch, G. (2017). Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs. *arXiv preprint*.

[21]. Suad A .Alasadi, Wesam S. Bhaya, ―Review of data preprocessing techniques in data mining‖, in Journal of Engineering and Applied Sciences on September 2017.

Volume 10, Issue 1, January – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

APPENDIX-A
PSUEDOCODE

```
import hashlib
# Mock user database users_db = {}
def hash_password(password):
return hashlib.sha256(password.encode()).hexdigest() def register_user():
print("=== Registration Page ===") username = input("Enter username: ") if username in users_db:
print("Username already exists. Please try another username.") return
password = input("Enter password: ") confirm_password = input("Confirm password: ") if password != confirm_password:
print("Passwords do not match!") return
users_db[username] = hash_password(password) print("Registration successful!")
if __name_______== "__main__": register_user()
import hashlib
# Mock user database users_db = {
"test_user": hashlib.sha256("test_password".encode()).hexdigest()
}
def hash_password(password):
return hashlib.sha256(password.encode()).hexdigest() def login_user():
print("=== Login Page ===")
username = input("Enter username: ") password = input("Enter password: ") if username not in users_db:
print("User not found. Please register first.") return False
if users_db[username] != hash_password(password): print("Incorrect password!")
return False print("Login successful!") return True
if __name_______== "__main__": login_user()
def input_symptoms():
print("=== Symptom Input ===") symptoms_list = [
"fever", "cough", "fatigue", "headache",
"nausea", "vomiting", "stomach ache", "diarrhea", "sore throat", "body ache", "chills"
]
print(f"Available symptoms: {', '.join(symptoms_list)}") symptoms = []
for i in range(4):
symptom = input(f"Select symptom {i + 1}: ") if symptom not in symptoms_list:
print("Invalid symptom. Please try again.") return None
symptoms.append(symptom) return tuple(sorted(symptoms))
if __name_______== "__main__": symptoms = input_symptoms() if symptoms:
print(f"Selected Symptoms: {symptoms}") import pandas as pd
def find_similar_cases(symptoms, medical_data): filtered_data = medical_data[
(medical_data["Symptom 1"].isin(symptoms)) & (medical_data["Symptom 2"].isin(symptoms)) & (medical_data["Symptom 3"].isin(symptoms)) & (medical_data["Symptom 4"].isin(symptoms))
]
return filtered_data
def generate_treatment(symptoms): # Load the dataset
medical_data = pd.read_csv("Training_new.csv") # Find similar cases
similar_cases = find_similar_cases(symptoms, medical_data) if not similar_cases.empty:
print("\n=== Suggested Treatments ===") for _, row in similar_cases.iterrows():
print(f"Symptoms: {row['Symptom 1']}, {row['Symptom 2']}, {row['Symptom 3']},
{row['Symptom 4']}")
print(f"Treatment: {row['Treatment']}") print(f"Medicines: {row['Medicines']}") print(f"Precautions: {row['Precautions']}")
print("-" * 30)
else:
print("No similar cases found. Please consult a doctor.") if __name_____== "__main__":
# Example symptoms for testing
test_symptoms = ("fever", "cough", "fatigue", "headache") generate_treatment(test_symptoms)
from register import register_user from login import login_user
from symptom_input import input_symptoms
from treatment_generator import generate_treatment
def main(): while True:
print("\n=== Main Menu ===") print("1. Register")
print("2. Login")
print("3. Exit")
```

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

```
choice = input("Enter your choice: ") if choice == "1":
register_user() elif choice == "2":
if login_user():
symptoms = input_symptoms() if symptoms:
generate_treatment(symptoms) elif choice == "3":
print("Goodbye!") break
else:
print("Invalid choice. Please try again.") if __name___== "__main__":
main()
```

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

APPENDIX-B
SCREENSHOTS

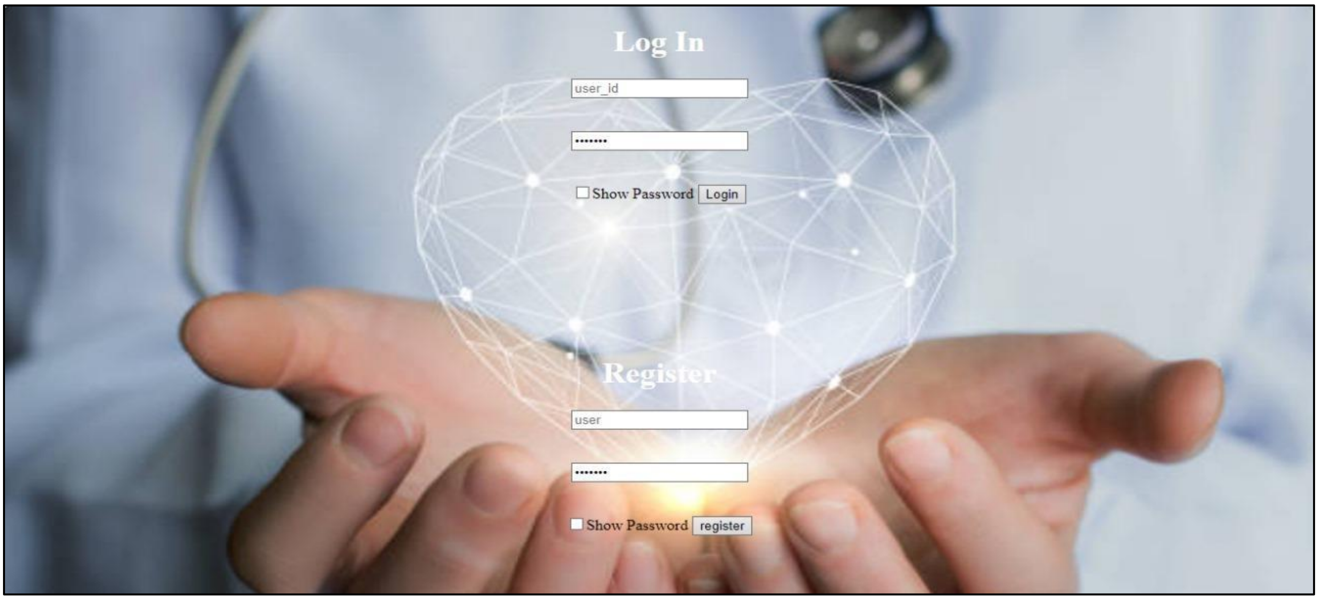

Fig 7: Login and Registration Interface

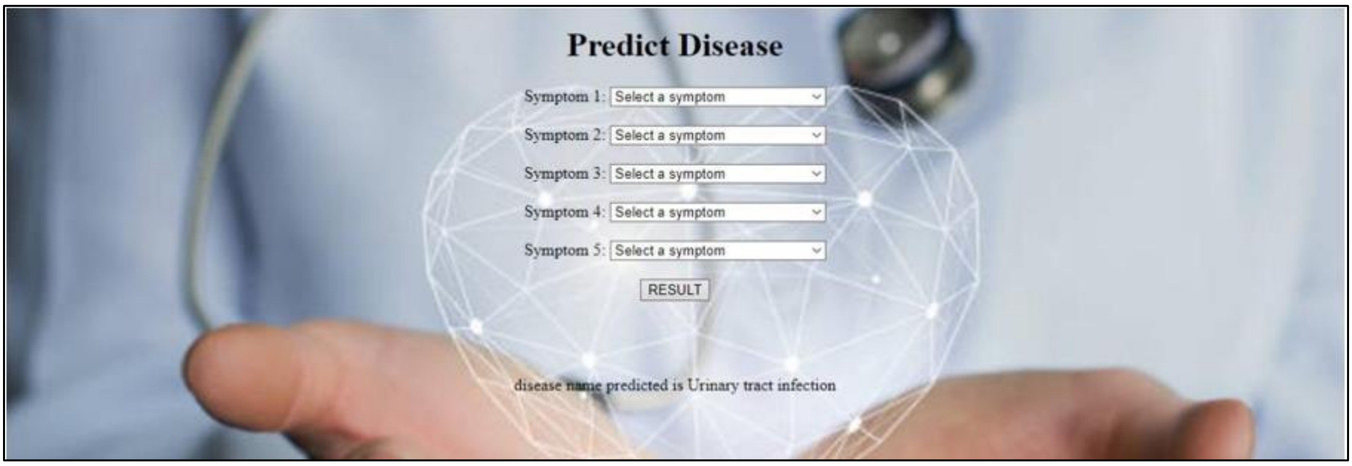

Fig 8: Disease Prediction Interface with Symptoms

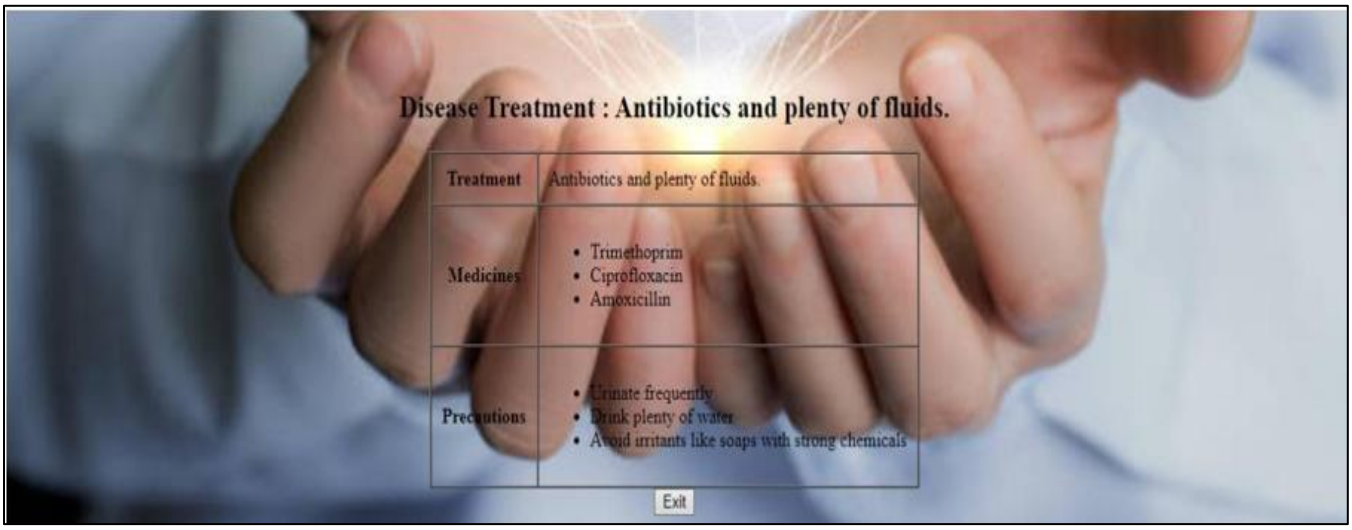

Fig 9: Disease Treatment, Medicines and Precaution and Interface

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14637070

APPENDIX-C
ENCLOSURES

- Journal publication/Conference Paper Presented Certificates of all students.
- Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need for a page-wise explanation.
- Details of mapping the project with the Sustainable Development Goals (SDGs).

The Project work carried out here is mapped to SDG-3 Good Health and Well-Being.

The project work carried here contributes to the well-being of the human society. This can be used for Analyzing and detecting blood cancer in the early stages so that the required medication can be started early to avoid further consequences which might result in mortality.

The project aligns primarily with SDG-3: Good Health and Well-Being, as it focuses on early detection and treatment of blood cancer to enhance well-being and reduce mortality rates.

However, depending on the scope of the project, it might also indirectly align with other Sustainable Development Goals, such as:

- **SDG-9: Industry, Innovation, and Infrastructure**, if the project involves innovative technology or infrastructure for healthcare.
- **SDG-10: Reduced Inequalities**, if it aims to make such detection and treatment accessible to underprivileged communities.
- **SDG-17: Partnerships for the Goals**, if the project relies on collaboration between institutions, researchers, or organizations.