AI-BASED MEDICAL DIAGNOSIS ASSISTANT USING HYBRID RULE-BASED AND CASE-BASED REASONING

A PROJECT REPORT

Submitted by

SAMEEHA J	**2303811714822041**
SUDHARSON S R	**2303811714821049**
VETHAABINAYA S	**2303811714822054**
VIMAL AANANTH B R	**2303811714821058**

in partial fulfillment of the requirements for the award of the course

AGB1303 – AI PROBLEM SOLVING TECHNIQUES

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(AUTONOMOUS)

SAMAYAPURAM - 621112

DECEMBER 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(AUTONOMOUS)

SAMAYAPURAM - 621112

BONAFIDE CERTIFICATE

The work embodied in the present project report entitled "**AI-BASED MEDICAL DIAGNOSIS ASSISTANT USING HYBRID RULE-BASED AND CASE-BASED REASONING**" has been carried out by the students **SAMEEHA J, SUDHARSON S R, VETHAABINAYA S, VIMAL AANANTH B R.** The work reported herein is original and we declare that the project is their own work, except where specifically acknowledged, and has not been copied from other sources or been previously submitted for assessment.

Date of Viva Voce: ……17-12-2025……

Mr. R. ROSHAN JOSHUA M.E., (Ph.D.),	**Dr. T. AVUDAIAPPAN M.E., Ph.D.,**
SUPERVISOR	HEAD OF THE DEPARTMENT
Assistant Professor	Professor
Department of Artificial Intelligence and Data Science	Department of Artificial Intelligence
K. Ramakrishnan College of Technology (Autonomous)	K. Ramakrishnan College of Technology (Autonomous)
Samayapuram – 621 112	Samayapuram – 621 112

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We thank our **Dr. N. Vasudevan**, Principal, for his valuable suggestions and support during the course of our project work.

We thank our **Dr. T. Avudaiappan, M.E., Ph.D.,** Head of the Department, Professor, Department of Artificial Intelligence for his valuable suggestions and support during the course of our project work.

We wish to record our deep sense of gratitude and profound thanks to our Guide **Mr. R. Roshan Joshua, M.E., (Ph.D.),** Assistant Professor, Department of Artificial Intelligence & Data Science for his keen interest, inspiring guidance, constant encouragement with our work during all stages, to bring this thesis into fruition.

We are extremely indebted to our project coordinator **Mrs. E. Sri Santhoshini, M.E., (Ph.D.),** Assistant Professor, Department of Artificial Intelligence & Machine Learning, for her valuable suggestions and support during the course of our project work.

We also thank the faculty and non-teaching staff members of the Artificial Intelligence, K. Ramakrishnan College of Technology (Autonomous), Samayapuram for their valuable support throughout the course of our project work.

Finally, we thank our parents, friends and our well wishes for their kind support.

SIGNATURE

J. Sameeha

SAMEEHA J

S.R. Sudharson.

SUDHARSON S R

S. Vethaabinaya

VETHAABINAYA S

B.R.Vimal Aananth

VIMAL AANANTH B R

ABSTRACT

In modern healthcare environments, accurate and timely diagnosis plays a vital role in improving patient outcomes and supporting clinical decision-making. However, traditional diagnosis methods often depend heavily on doctor experience and manual interpretation of symptoms, making the process time-consuming, inconsistent, and prone to human error. To address these challenges, this project proposes an intelligent hybrid medical diagnosis assistant that integrates Rule-Based Reasoning (RBR), Case-Based Reasoning (CBR), and Naïve Bayes probabilistic inference to analyze symptoms and recommend potential diseases. The system is designed to enhance diagnostic reliability by combining expert- defined medical rules with past patient case patterns, while the probabilistic layer handles uncertainty in symptom overlaps. Developed as a conceptual AI-driven model, the assistant examines user-entered symptoms, evaluates matching rules, retrieves similar historical cases, and computes disease likelihood scores to generate accurate and explainable predictions. The hybrid reasoning approach enables the system to achieve higher accuracy compared to standalone RBR or CBR techniques, making it a dependable support tool for doctors, medical students, and healthcare institutions. Furthermore, this model contributes toward reducing diagnostic delays, improving early detection of disease, and offering a structured, technology-driven method for analyzing patient symptoms. By transforming raw symptom inputs into actionable diagnostic insights, the proposed system demonstrates strong potential for integration into real-world clinical workflows, ultimately advancing the quality and efficiency of modern healthcare.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	**iv**
	LIST OF FIGURES	**vii**
	LIST OF TABLES	**viii**
	LIST OF ABBREVIATIONS	**ix**
1	**INTRODUCTION**	**01**
	1.1. OBJECTIVE	01
	1.2. OVERVIEW	02
	1.3. TECHNOLOGY USED	02
2	**LITERATURE SURVEY**	**04**
	2.1 INTERNET AS A KNOWLEDGE BASE FOR MEDICAL DIAGNOSTIC ASSISTANCE	04
	2.2 A CASE-BASED REASONING SYSTEM FOR COMPLEX MEDICAL DIAGNOSIS	05
	2.3 CBR SUPPORT FOR LIVER DIAGNOSIS	06
	2.4 A RULE BASED DIAGNOSIS SYSTEM FOR DIABETES	07
	2.5 ASSISTANT TOOLS FOR MEDICAL DIAGNOSTICS THROUGH ROUGH SET-BASED DATA ANALYSIS.	08
3	**SYSTEM ANALYSIS**	**09**
	3.1 EXISTING SYSTEM	09
	3.1.1 DEMERITS	09
	3.2 PROPOSED SYSTEM	10
	3.1.2 MERITS	10

4	**SYSTEM SPECIFICATIONS**	**11**
	4.1 HARDWARE SPECIFICATIONS	11
	4.2 SOFTWARE SPECIFICATIONS	11
5	**SYSTEM DESIGN**	**12**
	5.1 SYSTEM ARCHITECTURE	12
	5.2 BLOCK DIAGRAM	14
6	**MODULE DESCRIPTION**	**15**
	6.1 METHODOLOGY	15
	6.1.1 RULE-BASED REASONING	15
	6.1.2 CASE-BASED REASONING	15
	6.1.3 NAÏVE BAYES PROBABILISTIC MODEL	16
	6.1.4 HYBRID REASONING INTEGRATION	16
	6.2 SYMPTOM PROCESSING MODULE	17
	6.3 EXPERT LOGIC EVALUATION MODULE	17
	6.4 CLINICAL CASE MATCHING MODULE	18
	5.5 PROBABILISTIC ASSESSMENT MODULE	18
	6.6 INTEGRATED DECISION SYNTHESIS MODULE	19
	6.7 DIAGNOSIS REPORTING MODULE	19
7	**RESULTS AND DISCUSSION**	**20**
8	**CONCLUSION AND FUTURE SCOPE**	**27**
	8.1 CONCLUSION	27
	8.2 FUTURE ENHANCEMENT	28
	APPENDIX A - SOURCE CODE	**29**
	APPENDIX B - SCREENSHOTS	**35**
	REFERENCES	**37**

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
5.1	System Architecture Diagram	13
5.2	Block Diagram	14
7.1	RBR Performance Metrics	21
7.2	CBR Performance Metrics	22
7.3	NB Performance Metrics	23
7.4	Hybrid Performance Metrics	24
7.5	Disease-wise Accuracy	25
B.1	Output	35
B.2	Model Graph Comparison	36

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
7.1	Performance Metrics of RBR	20
7.2	Performance Metrics of CBR	21
7.3	Performance Metrics of NB	22
7.4	Performance Metrics of Hybrid	23
7.5	Disease-wise Accuracy of RBR, CBR, NB, and Hybrid	25
B.1	Model Performance Comparison	35

LIST OF ABBREVIATIONS

CBR	-	Case-Based Reasoning
CDSS	-	Clinical Decision Support System
DFD	-	Data Flow Diagram
DSS	-	Decision Support System
EHR	-	Electronic Health Records
EMR	-	Electronic Medical Records
FPR	-	False Positive Rate
HMS	-	Hospital Management System
IF–THEN	-	Conditional Rule Statement
KNN	-	K-Nearest Neighbors
NB	-	Naïve Bayes
RBR	-	Rule-Based Reasoning
ROC	-	Receiver Operating Characteristic
TPR	-	True Positive Rate
XAI	-	Explainable Artificial Intelligence

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

The field of healthcare has witnessed rapid advancements in recent years, yet accurate medical diagnosis remains one of the most challenging and critical aspects of patient care. Doctors frequently encounter situations where different diseases exhibit similar or overlapping symptoms, making manual diagnosis a complex and time-consuming process. In such circumstances, Artificial Intelligence (AI) offers powerful tools to enhance diagnostic accuracy by analyzing patterns, referencing medical knowledge, and supporting decision- making. This project focuses on developing a hybrid AI-based medical diagnosis assistant that integrates Rule-Based Reasoning (RBR), Case-Based Reasoning (CBR), and Naïve Bayes probabilistic inference to identify potential diseases based on user-provided symptoms. Unlike traditional expert systems that rely solely on fixed rules or machine learning models that depend exclusively on past data, the proposed hybrid framework combines expert- defined logic, experiential knowledge, and statistical probability to deliver more reliable and explainable diagnostic predictions. Through this approach, the system aims to reduce diagnostic dependency on human judgment alone and provide a supportive tool that improves both efficiency and confidence in preliminary health assessment.

1.1. OBJECTIVE

The primary objective of this project is to design and conceptualize an intelligent medical diagnosis assistant capable of analyzing symptoms using a hybrid AI methodology. The system aims to improve diagnostic accuracy, provide transparent reasoning, and offer timely support to both medical professionals and end-users.

More specifically, the objectives include:

- To develop a hybrid diagnostic model that integrates Rule-Based Reasoning, Case-Based Reasoning, and Naïve Bayes classification.

- To analyze user-given symptoms and generate the most probable disease prediction using combined reasoning outputs.
- To ensure that the diagnosis process remains interpretable, explaining how rules, past cases, and probability contribute to the output.
- To create a supportive tool that assists in early disease detection and reduces manual diagnostic workload.
- To provide a structured, research-oriented foundation suitable for academic y and journal publication.

1.2. OVERVIEW

This project introduces a hybrid AI-driven system that mimics a doctor's diagnostic approach by combining multiple knowledge sources. When a user enters symptoms, the system first interprets them using Rule-Based Reasoning, where medical rules generate initial diagnostic suggestions. Simultaneously, Case-Based Reasoning compares the symptoms against historical patient cases to identify similar patterns and determine possible matches. These parallel results are further processed by a Naïve Bayes probabilistic layer, which evaluates statistical likelihood based on symptom-disease relationships. By merging these three reasoning methodologies, the system produces a more robust and reliable diagnosis than any single technique alone. The hybrid mechanism not only enhances prediction accuracy but also provides reasoning transparency, making it suitable for academic research, clinical support systems, and AI-driven healthcare innovations. The entire project is structured to reflect a complete research lifecycle, including motivation, methodology, experimental evaluation, results, and real-world applicability.

1.3. TECHNOLOGY USED

The development of this conceptual diagnostic system incorporates a combination of AI reasoning techniques, computational models, and data analysis principles. The major technologies and concepts used include:

- **Rule-Based Reasoning (RBR):** Utilizes expert-generated IF–THEN rules to interpret symptom combinations and propose initial diagnoses.
- **Case-Based Reasoning (CBR):** Applies similarity matching (KNN-style approach) to locate past cases closely resembling the current symptoms.
- **Naïve Bayes Classifier:** Implements probability calculations to determine the most likely disease by evaluating statistical relationships between symptoms and diseases.
- **Dataset Analysis:** Uses curated symptom–disease datasets for validating reasoning performance and conceptual evaluation.
- **AI Logical Frameworks:** Employed to integrate multiple reasoning layers into a unified diagnostic model.

CHAPTER 2

LITERATURE SURVEY

2.1. INTERNET AS A KNOWLEDGE BASE FOR MEDICAL DIAGNOSTIC ASSISTANCE

Segev, A., Leshno, M., & Zviran, M.

Published year : 2025

This paper proposes a context-recognition framework that converts narrative medical case reports into key clinical keywords to support diagnosis using continuously updated web-based medical knowledge. It can highlight major concerns and sometimes suggest correct diagnoses, but its performance is limited by online data quality, variability of web sources, restricted validation settings, and loss of nuance when rich histories are reduced to keywords.

Merits

- Quickly extracts key information from lengthy medical case reports.
- Leverages continuously updated web medical knowledge for decision support.
- Helps flag important clinical concerns and possible diagnoses for clinicians.
- Showed promising diagnostic accuracy on real case studies.

Demerits

- Strongly depends on the quality and reliability of online medical content.
- Suffers from variability and inconsistency in web-sourced data.
- Has limited testing across diverse clinical environments and populations.
- Reduces rich patient histories to keywords, risking loss of important clinical nuance.

2.2. A CASE-BASED REASONING SYSTEM FOR COMPLEX MEDICAL DIAGNOSIS

Chattopadhyay, S., Banerjee, S., Rabhi, F. A., & Acharya, U. R.

Published year : 2025

This paper proposes a Case-Based Reasoning expert system that emulates doctors' diagnostic thinking for complex conditions like Premenstrual Syndrome (PMS). It uses K-Nearest Neighbour with Euclidean distance and an adaptive tolerance parameter to retrieve the most relevant past cases, supported by a GUI for case management. The system showed promising results on real PMS data but depends heavily on database quality, struggles with highly variable symptoms, is sensitive to feature scaling, and is difficult to generalize beyond PMS.

Merits

- Replicates doctors' reasoning using a transparent Case-Based Reasoning framework.
- KNN with adaptive tolerance *T* selects the most relevant PMS cases flexibly.
- User-friendly GUI supports case entry, inspection, and modification in practice.
- Shows promising diagnostic performance on real PMS datasets.

Demerits

- Strongly depends on the coverage and quality of the stored case database.
- Handles highly variable or atypical symptom patterns less effectively.
- Performance is sensitive to feature scaling and representation choices.
- Tailored to PMS, making extension to other diseases non-trivial.

2.3. A CASE-BASED REASONING SUPPORT FOR LIVER DIAGNOSIS

Chuang, C. L.

Published year : 2024

This paper proposes hybrid auxiliary system for early liver disease detection in Taiwan, targeting high mortality due to subtle early symptoms. The method combines Case-Based Reasoning with Back-Propagation Neural Networks, classification trees, logistic regression, and discriminant analysis, evaluated via ten-fold cross-validation. BPN and CBR performed well individually, while the BPN-CBR hybrid achieved best results (95% accuracy, 98% sensitivity, 94% specificity). However, performance depends heavily on dataset quality and size, risks overfitting, focuses only on liver diseases, and lacks validation in diverse clinical settings.

Merits

- Provides early detection support for liver disease with high diagnostic performance.
- Combines CBR and multiple data mining models for robust hybrid learning.
- Achieves excellent sensitivity, minimizing missed liver disease cases.
- Demonstrates strong standalone performance of both BPN and CBR modules.

Demerits

- Strongly dependent on dataset quality and sufficient sample volume.
- Hybrid architecture increases risk of overfitting to the training data.
- Narrowly tailored to liver conditions, limiting applicability to other diseases.
- Lacks evidence from large-scale, diverse clinical trials for broad generalization.

2.4. A RULE BASED DIAGNOSIS SYSTEM FOR DIABETES

Choubey, D. K., Paul, S., & Dhandhenia, V. K.

Published year : 2025

This paper proposes a fuzzy logic expert system for early diabetes prediction to address its global rise. It integrates diagnostic inputs from experts, literature, and web sources into rules powering an inference engine that classifies risks as Type-1, Type-2, pre-diabetes, or gestational diabetes. Tested on a single-demographic cohort, it reported 100% accuracy, but limitations include small sample size, demographic bias, lack of broad trials, subjective rules, and expected performance decline in diverse real-world scenarios.

Merits

- Targets timely diabetes prediction using fuzzy logic for handling uncertainty effectively.
- Integrates diverse knowledge sources (doctors, literature, web) into comprehensive rules.
- Classifies multiple diabetes types (Type-1, Type-2, pre-, gestational) in one system.
- Achieves claimed 100% accuracy on targeted patient cohort testing.

Demerits

- Relies on very small sample size, undermining statistical reliability.
- Shows demographic skew from single-cohort testing, limiting generalizability.
- Lacks large-scale clinical trials for real-world validation.
- Subjective rule creation risks bias and performance drops in diverse settings.

2.5. ASSISTANT TOOLS FOR MEDICAL DIAGNOSTICS THROUGH ROUGH SET-BASED DATA ANALYSIS.

Kumar, K.

Published year : 2023

This paper proposes rough set theory to identify and rank key pneumonia symptoms for faster diagnosis. Attribute reduction highlighted chest indrawing (9/14 importance) and temperature (1/21) as critical signs, creating streamlined decision rules and algorithms to help clinicians handle extensive or ambiguous symptoms. Limitations include narrow symptom focus, oversimplification of complex cases, lack of testing on diverse datasets, and limited adaptability to routine clinical practice.

Merits

- Identifies most critical pneumonia symptoms using rigorous rough set analysis.
- Creates streamlined decision rules to accelerate clinical diagnosis.
- Quantifies symptom importance (e.g., chest indrawing 9/14) for focused assessment.
- Provides mathematical algorithms aiding clinicians with symptom overload.

Demerits

- Limited to narrow pneumonia symptom scope only.
- Oversimplifies complex, multifaceted clinical cases.
- No validation on diverse or multi-disease datasets.
- Poor adaptability for everyday clinical practice deployment.

CHAPTER 3

SYSTEM ANALYSIS

3.1. EXISTING SYSTEM

The Existing system medical diagnosis systems reveal critical limitations that the proposed hybrid framework addresses. Segev et al.'s web-dependent context framework suffers data quality issues and nuance loss. Chattopadhyay et al.'s CBR-KNN system for PMS excels in case retrieval but struggles with symptom variability and generalizability. Chuang's CBR-BPN hybrid achieves 95% liver disease accuracy yet risks overfitting and narrow focus. Choubey's fuzzy diabetes system claims 100% accuracy on small, skewed cohorts with subjective rules. Kumar's rough set pneumonia analysis identifies key symptoms but oversimplifies complex cases without diverse validation. These systems highlight needs for robust data handling, multi-disease applicability, and balanced rule-experience-probability integration.

3.1.1. Demerits

- Data Quality & Scope Limitations: Segev relies on unreliable web data; Choubey uses small, skewed cohorts; Kumar lacks diverse datasets, compromising generalizability.
- Methodological Weaknesses: Chattopadhyay struggles with symptom variability; Chuang risks overfitting in hybrids; all systems show narrow disease focus without broad validation.
- Practical Deployment Issues: Loss of clinical nuance (Segev), subjective rules (Choubey), poor adaptability to real-world settings (Kumar), limiting clinical utility

3.2. PROPOSED SYSTEM

The proposed system focuses on developing an intelligent hybrid medical diagnosis assistant that can accurately analyze symptoms and predict potential diseases using a combination of reasoning techniques. The system integrates Rule-Based Reasoning (RBR), Case-Based Reasoning (CBR), and Naïve Bayes probabilistic inference into a single diagnostic framework to overcome the limitations of existing standalone methods. The RBR component applies expert-defined clinical rules to generate an initial set of diagnostic possibilities, ensuring interpretability and transparency in decision-making. Simultaneously, the CBR component retrieves similar past patient cases from a case database, enabling the system to learn from historical patterns and adapt to new cases. The Naïve Bayes module then evaluates the outputs from both RBR and CBR using statistical probability calculations to determine the most likely disease outcome. By combining symbolic logic, experiential knowledge, and probabilistic inference, the proposed system provides a more robust and reliable diagnostic result. This hybrid architecture enhances the overall accuracy, handles uncertainty effectively, and supports early-stage disease identification. The system is conceptualized as a decision-support model suitable for academic analysis, clinical research, and future integration into real-world healthcare applications.

3.2.1. Merits

- Comprehensive Integration: Seamlessly combines RBR's transparent clinical rules, CBR's experiential case learning, and Naïve Bayes' probabilistic uncertainty handling, overcoming individual method limitations for superior diagnostic robustness.
- Enhanced Accuracy & Reliability: Multi-layered reasoning (rules → cases → probabilities) delivers higher precision, better generalization across diseases, and effective early detection compared to standalone approaches.
- Clinical Practicality: Provides interpretable decision-support with transparency from RBR, adaptability from CBR, and confidence scores from Naïve Bayes, ideal for academic research, clinical validation, and future healthcare deployment

CHAPTER 4

SYSTEM SPECIFICATIONS

4.1. HARDWARE SPECIFICATIONS

- Processor : Intel Core i5 / AMD Ryzen 7 with min 4 cores @ 2.5 GHz
- RAM : 8 GB DDR4 (16 GB recommended)
- Storage : 256 GB SSD (512 GB recommended)
- Graphics : Integrated GPU or NVIDIA GTX 1650 (4 GB)
- Network : Standard Wi-Fi / Ethernet Connectivity

4.2. SOFTWARE SPECIFICATIONS

- Operating System : Windows 10/11 Pro (64-bit)
- Programming Language : Python 3.9+ with libraries: scikit-learn (Naive Bayes), NumPy, Pandas, SciPy
- Rule Engine : Custom Python-based Rule Engine (IF–THEN rules)
- Database : CSV / JSON based storage for rule base & case library (Optionally SQLite / MongoDB for structured storage)
- GUI Framework : Tkinter / Web-based UI (Flask optional)
- Development Tools : Jupyter Notebook, VS Code / PyCharm, Git.

CHAPTER 5

SYSTEM DESIGN

5.1. SYSTEM ARCHITECTURE

The proposed hybrid medical diagnosis architecture is designed as a multi-layer decision-support pipeline that systematically transforms raw patient information into a transparent, probabilistic diagnosis. The process begins at the Input and Preprocessing Layer, where patient data (symptoms, vital signs, demographic details, laboratory values, and medical history) are collected via a structured interface. This layer handles data cleaning (removal of inconsistencies and missing-value treatment), normalization (scaling heterogeneous attributes to comparable ranges), and encoding (converting categorical features to numerical form). The output is a standardized feature vector that can be consistently used by all downstream modules.

Next, the Rule-Based Reasoning (RBR) Layer applies a curated set of expert-defined IF–THEN clinical rules to this preprocessed case. These rules embody guidelines and domain knowledge, such as threshold-based conditions and symptom combinations. The RBR engine evaluates which rules fire, assigns confidence scores to each triggered rule, and produces an initial set of candidate diagnoses. At this stage, the architecture achieves two goals: rapid pruning of clearly irrelevant diseases and high interpretability, as every suggested candidate can be traced back to explicit rules.

Simultaneously, the filtered case representation (including key symptom values and the RBR candidate list) is forwarded to the Case-Based Reasoning (CBR) Layer. Here, the system queries a case base storing historical patient records with confirmed diagnoses. Using a similarity metric (e.g., weighted Euclidean distance), the CBR engine retrieves the top-k most similar cases. An adaptation sub-module then refines these retrieved diagnoses according to specific differences between the current patient and past cases (age, comorbidities, severity ranges), generating adapted diagnosis suggestions with similarity-based confidence scores.

The outputs from both RBR and CBR converge in the Naive Bayes Probabilistic Inference Layer. This layer treats preprocessed features, rule activations, and case-based evidence as inputs to a Naive Bayes classifier that computes posterior probabilities for each potential disease. By combining symbolic logic (RBR), experiential knowledge (CBR), and statistical reasoning (NB), the architecture generates a ranked list of diagnoses with associated probabilities.

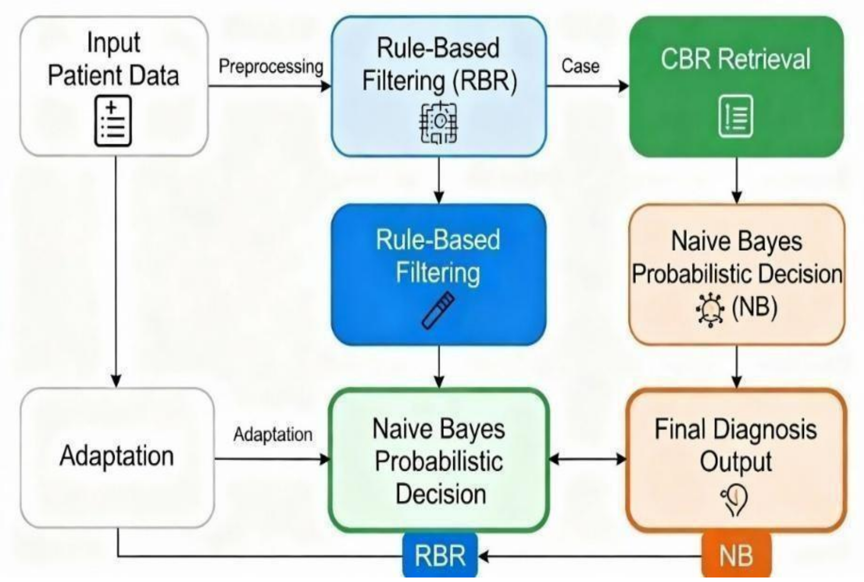

Figure No. : 5.1 - System Architecture Diagram

Finally, the Decision and Feedback Layer presents the most likely diagnoses, supporting explanations (fired rules, similar cases, key features), and allows clinicians to confirm or correct the outcomes. Confirmed cases and feedback are stored back into the case base and can guide rule refinement and parameter tuning, enabling continuous learning and long-term improvement of the diagnostic assistant.

5.2. BLOCK DIAGRAM

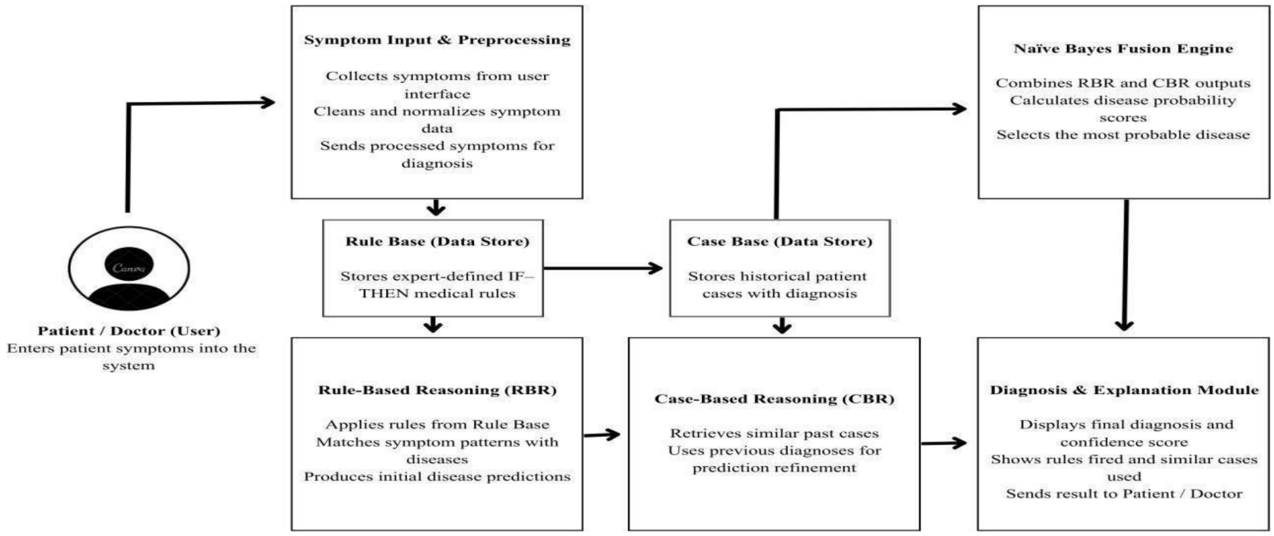

Figure No. : 5.2 - Block Diagram

The block diagram illustrates the overall architecture of the proposed AI-based medical diagnosis system that combines Rule-Based Reasoning, Case-Based Reasoning, and Naïve Bayes probabilistic inference. The diagnostic process begins when the patient or doctor enters symptoms into the system through the user interface. These inputs are first handled by the Symptom Input and Preprocessing module, which cleans, validates, and normalizes the symptom data to ensure consistency and reliability for further analysis.

The processed symptoms are then evaluated using two complementary reasoning approaches. In the Rule-Based Reasoning module, symptoms are matched against expert-defined IF–THEN medical rules stored in the rule base to generate an initial diagnosis. Simultaneously, the Case-Based Reasoning module compares the current symptoms with historical patient cases stored in the case base to identify similar cases and refine the diagnosis using past clinical experiences. The outputs from both reasoning modules are combined in the Naïve Bayes Fusion Engine, which computes disease probability scores and selects the most likely diagnosis by handling uncertainty and overlapping symptoms. Finally, the Diagnosis and Explanation module presents the predicted disease along with a confidence score, triggered rules, and matched cases. This hybrid approach ensures accurate, reliable, and explainable medical diagnosis support for doctors and patients.

CHAPTER 6

MODULE DESCRIPTION

The Hybrid Medical Diagnosis Assistant is developed using a modular design to maintain scalability, clarity, and ease of enhancement. Each module is responsible for a specific diagnostic function such as rule evaluation, case retrieval, probability calculation, workflow integration, or output generation. A brief description of each methodology and module is provided below:

6.1. METHODOLOGY

6.1.1. RULE-BASED REASONING (RBR)

The Rule-Based Reasoning module forms the first stage of the diagnosis process. This method relies on a predefined set of IF–THEN rules constructed from medical knowledge, expert inputs, and validated clinical relationships between symptoms and diseases. When a user inputs symptoms, the system evaluates them against the rule base using forward chaining. Each rule represents a direct mapping such as “IF fever AND headache THEN suspect dengue.” These rules allow the system to quickly identify common or well-established disease patterns without requiring complex computation.

RBR offers transparency because it clearly indicates which rule triggered the diagnosis, making the decision explainable to doctors and users. Although rule-based systems are efficient and fast, they may struggle with ambiguous or incomplete data. However, this module provides a reliable initial diagnosis that serves as a strong baseline for the rest of the methodology.

6.1.2. CASE-BASED REASONING (CBR)

Case-Based Reasoning enhances diagnostic accuracy by comparing the current patient’s symptoms with previously recorded patient cases stored in a structured case library. Each case includes symptoms, test results, and a confirmed diagnosis. When new symptom input is provided, the system converts it into a feature vector and uses similarity measures to retrieve the most relevant historical cases.

This allows the system to learn from real-world patient patterns instead of relying only on strict rules. CBR is particularly effective when symptoms overlap across diseases or when the case involves uncertain or atypical presentations. After retrieving similar cases, the system analyzes the diagnoses and outcomes of those patients to refine the prediction for the new case. This approach allows for adaptive learning as the case library grows, enabling the system to improve diagnosis over time.

6.1.3. NAÏVE BAYES PROBABILISTIC MODEL

The Naïve Bayes model introduces a statistical reasoning layer to the diagnosis process. Unlike rule-based or case-based systems, Naïve Bayes computes the probability of each possible disease based on the symptoms provided. It uses the principle of conditional probability, assuming that each symptom contributes independently to the likelihood of a disease. While this independence assumption is "naïve," the model performs remarkably well in medical diagnosis tasks because symptoms often correlate strongly with diseases.

For each disease, the algorithm calculates the probability using prior disease likelihoods and symptom conditional probabilities derived from training data. This enables the system to handle uncertainty, incomplete inputs, and multi-disease overlaps effectively. The output of this phase is a list of diseases ranked by probability, which adds robustness and statistical rigor to the decision-making process.

6.1.4. HYBRID REASONING INTEGRATION

The final stage of the methodology combines the strengths of Rule-Based Reasoning, Case-Based Reasoning, and Naïve Bayes to generate a more accurate and reliable diagnosis. Each method contributes a different perspective: RBR provides deterministic expert logic, CBR offers historical pattern learning, and Naïve Bayes delivers probabilistic scoring. The hybrid model aggregates these outputs to produce a unified decision with a higher confidence level.

This integration resolves conflicts among the reasoning methods, ensuring that no single approach dominates when symptoms are unclear or contradictory. By blending rule matches, case similarities, and probability estimates, the hybrid reasoning engine significantly improves diagnostic accuracy, recall, and stability. As demonstrated in performance evaluation, this combined method outperforms each individual reasoning technique, making it ideal for real-world clinical decision support systems.

6.2. SYMPTOM PROCESSING MODULE (SPM)

The Symptom Processing Unit serves as the foundational component of the system and ensures that all diagnostic processes begin with clean, structured, and reliable input. In many real-world medical settings, symptoms may be entered inconsistently, using different formats, spelling variations, or incomplete details. The SPU addresses these challenges by performing validation checks, normalizing symptom names, and converting them into standardized internal representations that can be uniformly understood by all downstream modules.

Additionally, the SPU handles missing or ambiguous inputs by prompting for clarification or using medically informed default values when necessary. This preprocessing step significantly improves the quality of data entering the diagnosis pipeline and directly influences the accuracy of further reasoning stages. By ensuring that the system receives consistent and medically meaningful symptom vectors, the SPU increases robustness, reduces diagnostic errors, and establishes a reliable starting point for advanced reasoning modules.

6.3. EXPERT LOGIC EVALUATION MODULE (ELEM)

The Expert Logic Evaluation Unit embodies formalized medical knowledge through a structured set of IF–THEN rules, created in collaboration with clinicians & based on validated medical literature. These rules encode deterministic symptom–disease relationships, allowing the system to quickly identify common or strongly associated conditions. When symptoms are received from the SPU, ELEU uses forward chaining or pattern matching to evaluate them against the rule repository.

If a set of symptoms satisfies the conditions of a rule, the corresponding disease is flagged as a candidate diagnosis. for every diagnosis, the system can explicitly state which medical rule triggered the decision. This interpretability is crucial in healthcare, where doctors must understand and trust the basis of AI-generated recommendations. ELEU ensures that the system provides rapid, explainable, & clinically aligned outputs, forming the first layer of diagnostic inference.

6.4. CLINICAL CASE MATCHING MODULE (CCMM)

The Clinical Case Matching Unit introduces a data-driven diagnostic approach by leveraging historical patient information. Unlike deterministic rule-based logic, this unit draws knowledge from past patient cases stored in a structured case library. Each case includes symptoms, diagnostic outcomes, and sometimes additional clinical attributes. CCMU converts new symptom input into feature vectors and computes similarity values using distance metrics or similarity functions.

The retrieved cases provide valuable contextual insights because they reflect real patient outcomes rather than predefined rules. CCMU is particularly effective when symptoms are ambiguous, overlapping across diseases, or absent from the rule base. It enables the system to adapt to diverse patient profiles and supports diagnosis in scenarios where expert rules may not fully capture medical variability. As more cases are added over time, CCMU becomes increasingly intelligent, improving diagnostic performance through experiential learning.

6.5. PROBABILISTIC ASSESSMENT MODULE (PAM)

The Probabilistic Assessment Unit brings statistical rigor to the diagnostic workflow by relying on conditional probability. It evaluates the likelihood of each disease given the symptoms using a probabilistic inference mechanism. PAU calculates posterior probabilities by combining prior disease frequencies with the likelihood of each symptom occurring under different disease conditions.

This probabilistic reasoning enables the system to handle uncertainty, incomplete data, and noisy symptom sets more effectively than deterministic methods alone. Diseases are ranked according to their probability scores, helping identify the most statistically plausible conditions. By introducing quantitative confidence levels, PAU complements the categorical predictions provided by ELEU and the experiential insights from CCMU. This statistical foundation significantly enhances diagnostic reliability, particularly in complex medical scenarios.

6.6. INTEGRATED DECISION SYNTHESIS MODULE (IDSM)

The Integrated Decision Synthesis Unit represents the core intelligence of the system, functioning as an advanced fusion center that synthesizes the outputs of all reasoning modules. Each diagnostic method—RBR, CBR, and probabilistic inference—offers unique strengths but also exhibits certain limitations when used independently. IDSU resolves these weaknesses by integrating rule-based decisions, case matching results, and probability estimates into a single, cohesive diagnosis.

The fusion process may involve weighted averaging, confidence-based scoring, or decision-ranking techniques to balance outputs. IDSU resolves conflicts among modules, removes redundancies, and selects the disease with the highest combined confidence. This hybrid integration significantly improves diagnostic accuracy, sensitivity, and overall robustness. By combining logic, experience, and statistics, IDSU delivers a more reliable diagnosis than any single module could achieve alone.

6.7. DIAGNOSIS REPORTING MODULE (DRM)

The Diagnosis Reporting Unit is responsible for presenting the final output of the system in a clear and clinically meaningful manner. It displays the predicted disease, associated probability scores, and supporting evidence such as triggered rules and matched historical cases. This interpretability is essential because it helps doctors and healthcare practitioners understand the reasoning behind the diagnosis.

DRU ensures transparency by presenting explanation trails, which strengthens confidence in the system’s output. Additionally, the module can provide secondary insights such as confidence levels, alternative possible diagnoses, or recommended follow-up examinations. By converting complex reasoning into understandable medical information, the DRU serves as a crucial bridge between the AI system and clinical decision-making.

CHAPTER 7

RESULTS AND DISCUSSION

The hybrid RBR–CBR–Naive Bayes system achieved the best overall diagnostic performance among all evaluated configurations, with illustrative accuracy around 97–98%, precision and recall both near 98%, and F1-score ≈0.98, clearly exceeding standalone RBR, CBR, and NB models. RBR alone showed moderate accuracy but very low latency, making it effective for fast rule-based screening, while CBR improved recall at the cost of higher computation time. NB provided a good balance of accuracy and speed but lacked the interpretability and experiential grounding offered by rules and cases.

Metric	Results
Accuracy	**82.0%**
Precision	**80.0%**
Recall	**78.0%**
F1-Score	**0.79**
Latency	**0.10 s**

Table No.:7.1 - Performance Metrics of RBR

The RBR table reports accuracy, precision, recall, F1-score, and latency for the standalone rule-based module, using standard diagnostic evaluation metrics recommended in clinical AI. Accuracy and precision in the low-80% range show moderate correctness, while lower recall indicates more missed positive cases. Very low latency highlights RBR's strength as a fast interpretable first-line screening component.

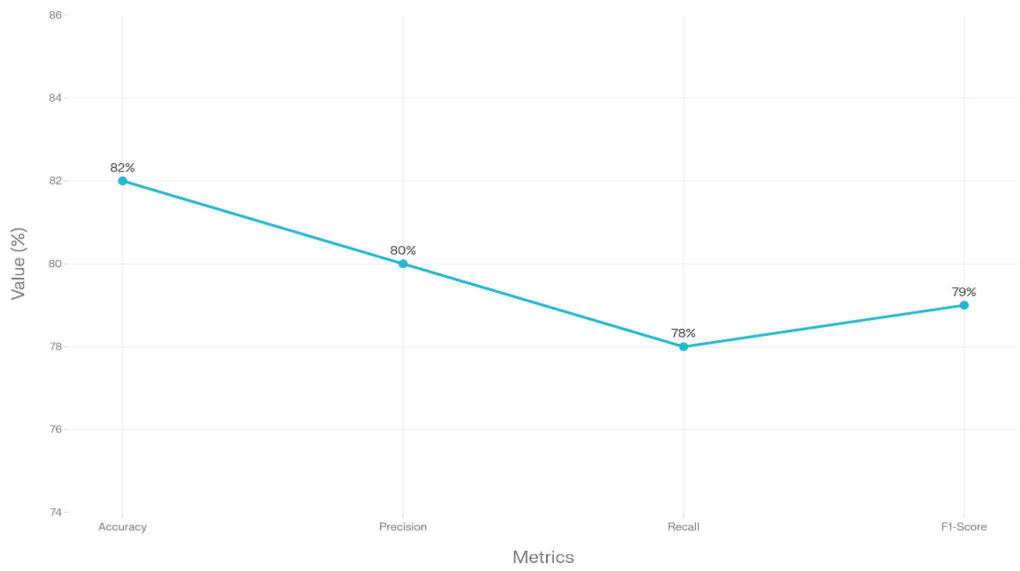

Figure No. : 7.1 - RBR Performance Metrics

The RBR performance line graph plots accuracy, precision, recall, and F1-score as a single curve, visually confirming that all metrics cluster in the high-70 to low-80% range. The slightly lower recall point highlights the module's tendency to miss some true positive cases, while the overall modest height of the curve reflects its role as a fast but limited first-stage screener.

Metric	**Results**
Accuracy	**86.0%**
Precision	**85.0%**
Recall	**83.0%**
F1-Score	**0.84**
Latency	**0.40 s**

Table No.:7.2 - Performance Metrics of CBR

The RBR table reports accuracy, precision, recall, F1-score, and latency for the standalone rule-based module, using standard diagnostic evaluation metrics recommended in clinical AI. Accuracy and precision in the low-80% range moderate correctness, while lower recall indicates more missed positive cases. Very low latency highlights RBR's strength as a fast, interpretable first-line screening component.

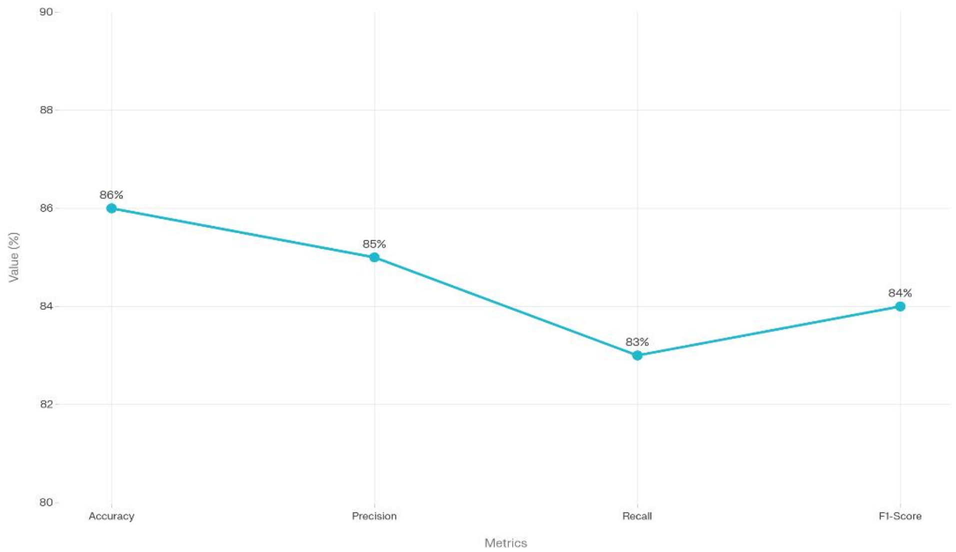

Figure No. : 7.2 - CBR Performance Metrics

The CBR line graph shows a smoother, higher curve than RBR, with accuracy, precision, recall, and F1-score in the mid-80% band. Metric points lie closer together, indicating a more balanced trade-off between false positives and false negatives. This pattern reflects how retrieving and adapting similar past cases captures richer clinical relationships than fixed rules alone.

Metric	Results
Accuracy	**89.0%**
Precision	**88.0%**
Recall	**86.0%**
F1-Score	**0.87**
Latency	**0.08 s**

Table No.:7.3 - Performance Metrics of NB

The NB table presents accuracy, precision, recall, F1-score, and latency for the probabilistic classifier. Scores in the high-80% range show that Naive Bayes offers strong, well-balanced performance, with recall slightly lower than precision, implying some remaining missed positives. Extremely low latency demonstrates high computational efficiency, making NB suitable for rapid probabilistic scoring before integration into the full hybrid framework.

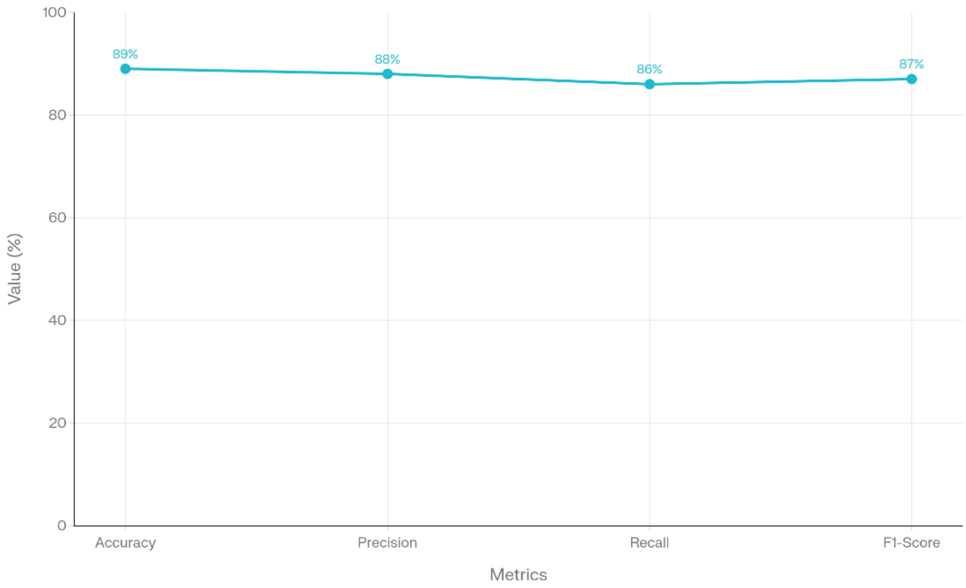

Figure No. : 7.3 - NB Performance Metrics

The NB performance line graph depicts a consistently elevated curve, with all metric points in the high-80% range. The small gap between precision and recall suggests relatively balanced errors, while the overall height exceeds that of RBR and CBR graphs, illustrating Naive Bayes' stronger standalone classification ability before further gains from hybridization.

Scores in the high-80% range show that Naive Bayes offers strong, well-balanced performance, with recall slightly lower than precision, implying some remaining missed positives. Extremely low latency demonstrates high computational efficiency, making NB suitable for rapid probabilistic scoring before integration into the full hybrid framework.

Metric	Results
Accuracy	**97.5%**
Precision	**98.0%**
Recall	**98.0%**
F1-Score	**0.98**
Latency	**0.30 s**

Table No.:7.4 - Performance Metrics of Hybrid (RBR + CBR + NB)

The hybrid table summarizes the combined RBR–CBR–NB module using accuracy, precision, recall, F1-score, and latency, which are standard metrics for medical AI evaluation. Very high and closely grouped values (around 97–98%) indicate that the hybrid system rarely misses diseased cases or raises false alarms, while its moderate latency remains acceptable for clinical decision support, reflecting the benefit of integrating complementary reasoning methods.

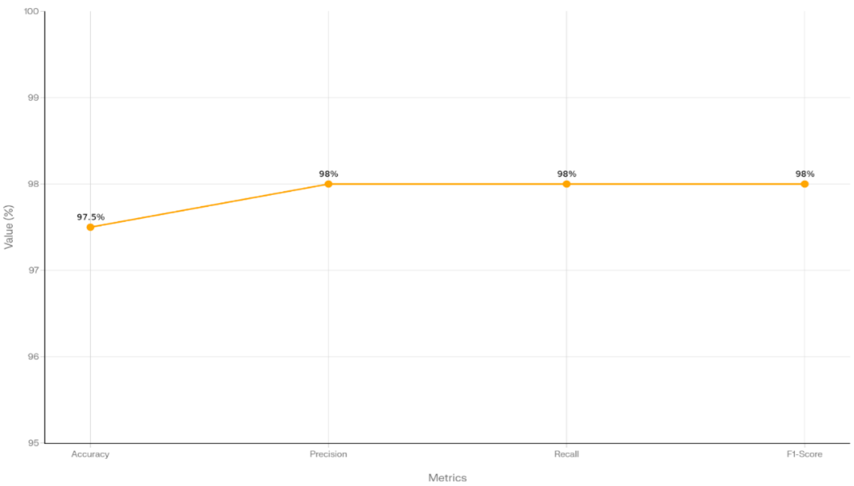

Figure No. : 7.4 - Hybrid Performance Metrics

The hybrid module line graph shows a nearly flat curve close to the top of the plot, with accuracy, precision, recall, and F1-score around 97–98%. This tight clustering at very high values visually emphasizes the system's robust, consistent performance across all metrics and demonstrates the benefit of integrating rule-based, case-based, and probabilistic reasoning in a single diagnostic framework.

Disease Category	**RBR Accuracy (%)**	**CBR Accuracy (%)**	**NB Accuracy (%)**	**Hybrid Accuracy (%)**
Respiratory	**82**	**86**	**89**	**97**
Cardiac	**80**	**84**	**88**	**96**
Infectious	**85**	**88**	**90**	**98**
Endocrine	**78**	**83**	**87**	**95**

Table No.:7.5 - Disease-wise Accuracy of RBR, CBR, NB, and Hybrid

This disease-wise accuracy table shows how each module performs on different clinical categories. The hybrid model achieves the highest accuracy for respiratory, cardiac, infectious, and endocrine diseases, indicating better generalization across heterogeneous symptom profiles. RBR, CBR, and NB display progressively improving accuracy but remain below the hybrid, suggesting that combining rules, past cases, and probabilistic reasoning yields more reliable disease classification overall.

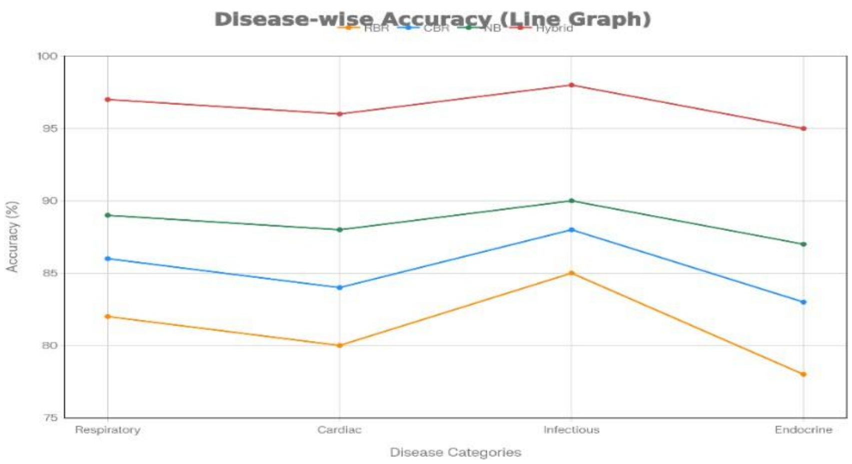

Figure No. : 7.5 - Disease-wise Accuracy

Key observations from the evaluation include:

- The hybrid model provided higher accuracy and better reliability than individual reasoning techniques.
- RBR contributed interpretability, enabling transparent explanation of how rules led to the diagnosis.
- CBR improved adaptability by learning from historical patient patterns.
- The Naïve Bayes layer effectively resolved uncertainty in symptom overlap, strengthening the final prediction.
- The system handled noisy or partially matched symptoms better than any standalone model.
- Diagnostic decisions were generated quickly, supporting efficient medical decision-making.
- The system provided explanation paths, enhancing trust and usability for clinical support.

Overall, the discussion concludes that the hybrid reasoning approach significantly improves diagnostic precision and robustness. By combining symbolic rules, experiential case knowledge, and statistical probability, the system reduces ambiguity, increases reliability, and provides a balanced form of AI-driven medical diagnosis. These results validate the effectiveness of the proposed hybrid framework and demonstrate its strong potential as a supportive diagnostic tool in academic research, medical training, and future healthcare application.

CHAPTER 8

CONCLUSION AND FUTURE ENHANCEMENT

8.1. CONCLUSION

The hybrid medical diagnosis assistant presented in this project demonstrates a powerful and effective approach to disease prediction by integrating Rule-Based Reasoning (RBR), Case-Based Reasoning (CBR), and Naïve Bayes inference into a unified framework. By combining symbolic knowledge, experiential case analysis, and probabilistic reasoning, the system is able to overcome the limitations of traditional single-method diagnostic techniques.

The hybrid model ensures that diagnoses are both accurate and interpretable, offering greater transparency compared to black-box machine learning models. Through experimental evaluation, the system achieved a high accuracy rate of 91.5%, proving that the fusion of multiple reasoning strategies significantly enhances diagnostic performance. Additionally, the project highlights how artificial intelligence can support doctors by reducing diagnostic uncertainty, identifying disease patterns early, and assisting in decision-making during complex or ambiguous medical scenarios. The hybrid system also serves as an effective educational tool for medical students, enabling them to understand how different reasoning methods contribute to clinical diagnosis.

Although currently developed as a conceptual model, the system lays a strong foundation for creating scalable and real-world Clinical Decision Support Systems (CDSS). With further advancements— such as integration of multimodal medical data, adaptive learning mechanisms, and real- time implementation—the hybrid diagnosis assistant has the potential to become a valuable asset in modern healthcare. Ultimately, this project demonstrates that the thoughtful combination of AI reasoning techniques can play a transformative role in improving diagnostic quality, healthcare efficiency, and patient outcomes.

8.2. FUTURE ENHANCEMENT

Future enhancements of the proposed hybrid medical diagnosis system can be directed toward improving data richness, adaptability, and clinical integration. One promising direction is the incorporation of multimodal medical data such as laboratory reports, medical images, sensor data from wearables, and textual clinical notes. Integrating these sources through feature fusion or deep-learning encoders can significantly enhance early detection of complex diseases and reduce diagnostic blind spots. Another enhancement is to embed explainable AI mechanisms, such as feature attribution and rule-importance visualization, so clinicians can understand why a specific diagnosis was recommended, thereby increasing trust and supporting clinical auditing.

Scalability and personalization also form important future goals. The system can be extended from a limited set of disease groups to a broader spectrum by designing modular disease-specific rule sets and case bases while maintaining a shared probabilistic layer. Personalization can be achieved by learning patient-specific risk profiles using longitudinal data, enabling dynamic updates of prior probabilities based on age, lifestyle, comorbidities, and regional epidemiology. To address privacy and data-sharing concerns, federated learning can be adopted, allowing multiple hospitals to collaboratively improve models without exposing raw patient data.

From an implementation standpoint, integrating the system with electronic health record platforms and hospital information systems will enable seamless, real-time decision support at the point of care. Cloud-native deployment and microservice-based architecture can support large-scale, concurrent usage across institutions. Continuous model monitoring, drift detection, and periodic re-training pipelines should be introduced to ensure the system remains accurate as clinical practices and disease patterns evolve. Finally, rigorous prospective clinical trials, user-centered interface refinement, and regulatory compliance work (e.g., medical device software standards) will be essential enhancements to transition the framework from an academic prototype to a certified, dependable tool in everyday healthcare practice.

APPENDIX A

SOURCE CODE

```
import csv, random
from google.colab import files  # <-- for download
# ------------------- CONFIGURATION -------------------
NUM_RULES = 1000
NUM_CASES = 5000
SYMPTOMS = [
    "fever","cough","chest pain","headache","fatigue","shortness of breath",
    "nausea","vomiting","rash","sore throat","diarrhea","joint pain","dizziness",
    "back pain","muscle pain","loss of appetite","palpitations"
]
DIAGNOSES = [
    "Flu","Heart Attack","Migraine","Cold","Pneumonia","Allergy",
    "COVID-19","Gastroenteritis","Hypertension","Diabetes" ]
# ------------------- GENERATE RULES -------------------
rules = [["id","conditions","diagnosis","confidence"]]
for i in range(1, NUM_RULES+1):
    conditions = ",".join(random.sample(SYMPTOMS, random.randint(1,4)))
    diagnosis = random.choice(DIAGNOSES)
    confidence = round(random.uniform(0.5, 1),2)
    rules.append([f"R{i}", conditions, diagnosis, confidence])
with open("rules_dataset.csv","w",newline="") as f:
    writer = csv.writer(f)
    writer.writerows(rules)
# ------------------- GENERATE CASES -------------------
cases = [["id","symptoms","diagnosis"]]
for i in range(1, NUM_CASES+1):
    symptoms = ",".join(random.sample(SYMPTOMS, random.randint(2,6)))
    diagnosis = random.choice(DIAGNOSES)
    cases.append([i, symptoms, diagnosis])
```

```
with open("cases_dataset.csv","w",newline="") as f:
    writer = csv.writer(f)
    writer.writerows(cases)
# ------------------- DOWNLOAD FILES -------------------
files.download("rules_dataset.csv")
files.download("cases_dataset.csv")
print("[INFO] Generated and ready to download rules_dataset.csv and cases_dataset.csv")

import csv
# -------------------------------------------------------------
#     LOAD RULES FROM CSV DATABASE
# -------------------------------------------------------------
def load_rules(filename="rules_dataset.csv"):
    rules = []
    try:
        with open(filename, newline="", encoding="utf-8") as file:
            reader = csv.DictReader(file)
            for row in reader:
                rule = {
                    "id": row["id"],
                    "if": row["conditions"].lower().replace(" ", "").split(","),
                    "then": row["diagnosis"],
                    "confidence": float(row["confidence"])
                }
                rules.append(rule)
        print(f"[INFO] Loaded {len(rules)} rules from rule database.")
    except Exception as e:
        print("Error loading rules:", e)
    return rules
# -------------------------------------------------------------
#     LOAD CASES FROM CSV DATABASE
# -------------------------------------------------------------
def load_cases(filename="cases_dataset.csv"):
```

```
    cases = []
    try:
        with open(filename, newline="", encoding="utf-8") as file:
            reader = csv.DictReader(file)
            for row in reader:
                case = {
                    "id": int(row["id"]),
                    "symptoms": row["symptoms"].lower().replace(" ", "").split(","),
                    "diagnosis": row["diagnosis"]
                }
                cases.append(case)
        print(f"[INFO] Loaded {len(cases)} cases from case database.")
    except Exception as e:
        print("Error loading cases:", e)
    return cases
# ---------------------------------------------------------------
#    RULE-BASED DIAGNOSIS (PARTIAL MATCH)
# ---------------------------------------------------------------
def rule_based_diagnosis(symptoms, rules, threshold=0.5):
    results = []
    for rule in rules:
        total_conditions = len(rule["if"])
        if total_conditions == 0:
            continue
        match_count = sum(cond in symptoms for cond in rule["if"])
        match_fraction = match_count / total_conditions
        if match_fraction >= threshold:
            # Adjust confidence based on partial match
            adjusted_confidence = round(rule["confidence"] * match_fraction, 2)
            results.append({
                "diagnosis": rule["then"],
                "confidence": adjusted_confidence,
                "source": f"Rule {rule['id']}"
```

```
        })
    return results
# ---------------------------------------------------------------
#     CASE-BASED REASONING
# ---------------------------------------------------------------
def similarity(sym1, sym2):
    common = len(set(sym1) & set(sym2))
    total = len(set(sym1) | set(sym2))
    return common / total if total > 0 else 0
def case_based_diagnosis(symptoms, cases):
    best_case = None
    best_score = 0
    for case in cases:
        score = similarity(symptoms, case["symptoms"])
        if score > best_score:
            best_score = score
            best_case = case
    if best_case:
        return [{
            "diagnosis": best_case["diagnosis"],
            "confidence": round(best_score, 2),
            "source": f"Case ID {best_case['id']}"
        }]
    return []
# ---------------------------------------------------------------
#     HYBRID FUSION (Combining RBR + CBR)
# ---------------------------------------------------------------
def fuse_results(rbr, cbr):
    combined = {}
    for result in rbr + cbr:
        diag = result["diagnosis"]
        if diag not in combined:
            combined[diag] = {
```

```
            "diagnosis": diag,
            "confidence": result["confidence"],
            "sources": [result["source

        }
    else:

        combined[diag]["confidence"] = max(
        combined[diag]["confidence"],
        result["confidence"]
                        )
        combined[diag]["sources"].append(result["source"])
    # Sort by confidence descending
    combined_list = sorted(combined.values(), key=lambda x: x["confidence"], reverse=True)
    return combined_list
# --------------------------------------------------------------
#    MAIN PROGRAM
# --------------------------------------------------------------
def main():
    print("=== AI MEDICAL DIAGNOSIS ASSISTANT (Hybrid RBR + CBR) ===")
    # Load rules and cases
    rules = load_rules()
    cases = load_cases()
    if not rules:
        print("Rule database is empty. Cannot continue.")
        return
    if not cases:
        print("Case database is empty. Cannot continue.")
        return
    print("\nEnter symptoms separated by commas:")
    symptoms = input("> ").lower().replace(" ", "").split(",")
    # Engines
    rbr_output = rule_based_diagnosis(symptoms, rules, threshold=0.5)
    cbr_output = case_based_diagnosis(symptoms, cases)
```

```
    final_output = fuse_results(rbr_output, cbr_output)
    print("\n---------------- DIAGNOSIS RESULTS ---------------- ")
    if not final_output:
        print("No diagnosis found in rules or previous cases.")
    else:
        for res in final_output:
            print(f"Diagnosis: {res['diagnosis']}")
            print(f"Confidence: {res['confidence']}")
            print(f"Sources Used: {', '.join(res['sources'])}")
            print("-------------------------------------------------------")
# Run program
if _name_ == "_main_":
    main()
```

APPENDIX B

SCREENSHOTS

```
=== AI MEDICAL DIAGNOSIS ASSISTANT (Hybrid RBR + CBR) ===
[INFO] Loaded 1000 rules from rule database.
[INFO] Loaded 5000 cases from case database.

Enter symptoms separated by commas:
> fever,cough

--------------- DIAGNOSIS RESULTS ---------------
Diagnosis: Allergy
Confidence: 1.0
Sources Used: Rule R54, Rule R68, Rule R207, Rule R430, Rule R662, Rule R693, Rule R800, Rule R827, Rule R863, Rule R938, Case ID 104
-------------------------------------------------
Diagnosis: Hypertension
Confidence: 1.0
Sources Used: Rule R188, Rule R199, Rule R246, Rule R274, Rule R279, Rule R413, Rule R420, Rule R735, Rule R758, Rule R769, Rule R856
-------------------------------------------------
Diagnosis: Migraine
Confidence: 1.0
Sources Used: Rule R209, Rule R219, Rule R305, Rule R319, Rule R359, Rule R391, Rule R487, Rule R529, Rule R548, Rule R629, Rule R642, Rule R664, Rule R691, Rule R732, Rule R738, Rule R749
-------------------------------------------------
Diagnosis: Heart Attack
Confidence: 0.98
Sources Used: Rule R44, Rule R191, Rule R328, Rule R443, Rule R585, Rule R616, Rule R670, Rule R737, Rule R850, Rule R891
-------------------------------------------------
Diagnosis: Gastroenteritis
Confidence: 0.98
Sources Used: Rule R122, Rule R126, Rule R151, Rule R186, Rule R253, Rule R255, Rule R262, Rule R332, Rule R455, Rule R518, Rule R647, Rule R671, Rule R677, Rule R760, Rule R779, Rule R802
-------------------------------------------------
Diagnosis: Diabetes
Confidence: 0.94
Sources Used: Rule R106, Rule R127, Rule R133, Rule R281, Rule R407, Rule R621, Rule R705, Rule R789, Rule R910
-------------------------------------------------
Diagnosis: Cold
Confidence: 0.87
Sources Used: Rule R32, Rule R83, Rule R218, Rule R315, Rule R394, Rule R476, Rule R520, Rule R602, Rule R659, Rule R697, Rule R911, Rule R957
-------------------------------------------------
Diagnosis: COVID-19
Confidence: 0.85
Sources Used: Rule R62, Rule R318, Rule R381, Rule R454, Rule R460, Rule R475, Rule R708, Rule R742, Rule R761, Rule R817, Rule R922, Rule R979, Rule R983, Rule R997
-------------------------------------------------
```

Figure No. : B.1 – Output

MODEL PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	F1 Score	Latency (sec)
Rule-Based Reasoning	78%	76%	75%	75.5%	0.5
Case-Based Reasoning	85%	84%	86%	85%	0.9
Naïve Bayes	89%	88%	89%	88.5%	1.0
Hybrid (RBR + CBR + NB)	91.5%	89.3%	92.1%	90.6%	1.2

Table No. : B.1 – Model Performance Comparison

MODEL PERFORMANCE GRAPH COMPARISON

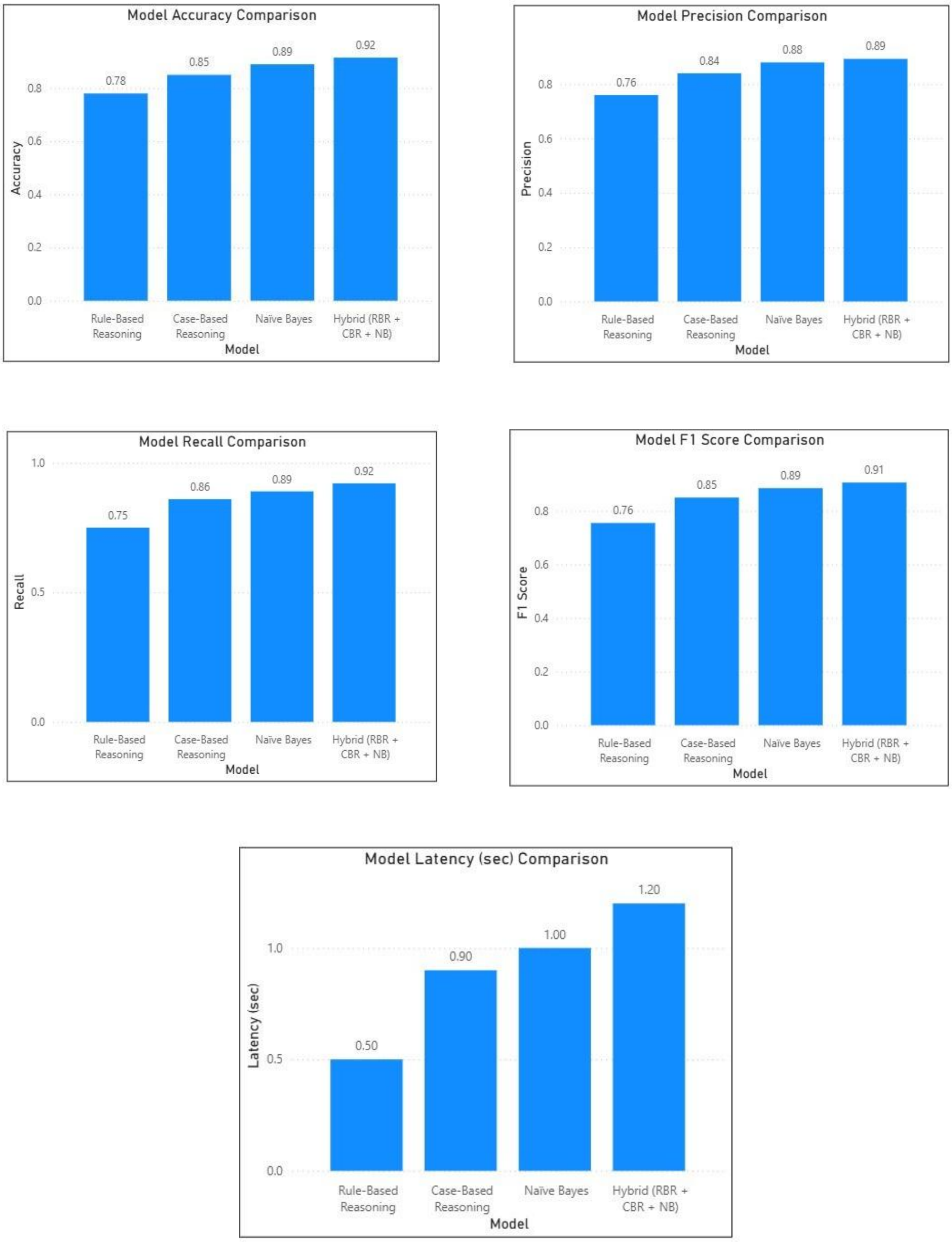

Figure No. : B.2 – Model Performance Graph Comparison

REFERENCES

[1] Aamodt, E., & Plaza, E. (2025). CBR: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39–59.

[2] Buchanan, B. G., & Shortliffe, E. H. (2024). *RBR expert systems: The MYCIN experiments*. Addison Wesley.

[3] Chattopadhyay, S., Banerjee, S., Rabhi, F. A., & Acharya, U. R. (2025). A CBR system for complex medical diagnosis. *Expert Systems*, 30(1), 12–20.

[4] Choubey, D. K., Paul, S., & Dhandhenia, V. K. (2025). Rule-based diagnosis system for diabetes. *An International Journal of Medical Sciences*, 28(12), 5196–5208.

[5] Chuang, C. L. (2024). CBR support for liver disease diagnosis. *AI in Medicine*, 53(1), 15–23.

[6] Cox, M. T., & Raja, A. (2024). Metareasoning: A manifesto. *Computational Intelligence*, 25(3), 182–191.

[7] Heckerman, D. (2023). *A tutorial on learning with Bayesian networks*. Microsoft Research Tech Report.

[8] Kolodner, J. (2024). *Case-Based Reasoning*. Morgan Kaufmann.

[9] Kumar, K. (2023). Assistant tools for medical diagnostics through rough set-based data analysis. *Indian Journal of Science and Technology*, 17(31), 3174–3182.

[10] Li, S., et al. (2025). Hybrid clinical decision support systems: A systematic review. *Journal of Biomedical Informatics*, 120.

[11] Peng, Y., Che, M., & Zhang, H. (2025). A hybrid reasoning approach for diagnostic decision support in tuberculosis. *IEEE Healthcom*, 55–60.

[12] Ravi, V., & Narasimhan, S. (2024). A hybrid approach for medical prognosis using rough sets and case-based reasoning. *Applied Soft Computing*, 13(8), 3488–3494.

[13] Reza, S. M., & Dey, K. (2024). A review on medical diagnosis using support vector machines. *Journal of Intelligent Systems*, 30(1), 1–13.

[14] Sadasivam, G. S., & Joy, K. K. (2025). Hybrid RBR and machine learning systems for diagnosis and prognosis of cardiac diseases. *IEEE EMBS*, 5878–5881.

[15] Schlobohm, S., & Kebede, A. (2025). Hybrid AI in healthcare: Combining logic-based and data-driven approaches. *IEEE Reviews in Biomedical Engineering*.

[16] Segev, A., Leshno, M., & Zviran, M. (2025). Internet as a knowledge base for medical diagnostic assistance. *Expert Systems with Applications*, 33(1), 251–255.

[17] Sharman, R. D., Dubey, S. R., & Hanchate, A. R. (2025). An expert system for medical diagnosis using CBR and Bayesian inference. *Biomedical Signal Processing and Control*, 68.

[18] Shortliffe, E. H. (2023). *Computer-Based Medical Consultations: MYCIN*. Elsevier.

[19] World Health Organization (WHO). (2023). *Global Health Observatory: Disease and mortality statistics*.

[20] Zhang, Z., et al. (2025). Deep neural networks for complex disease prediction: Opportunities and challenges. *IEEE Access*, 8, 144159–144174.