A Hybrid NLP-Based Chatbot for Medical symptom Analysis and knowledge Retrieval Using Labeled Datasets and Encyclopedic Resources.

Presented by

Mehedi Hasan

ID : 221-15-4680

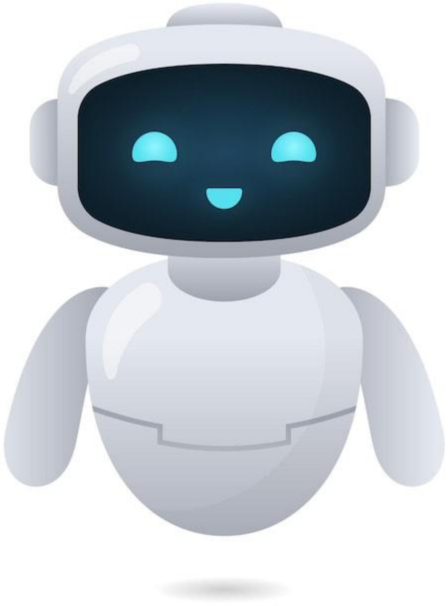
Department : CSE

Daffodil International University

Outline

Introduction

AI-powered medical chatbots assist users by offering symptom-based diagnosis or medical information separately. Current solutions lack integration of both, limiting accuracy and trust. This project develops a hybrid intelligent chatbot that combines symptom classification with retrieval-augmented generation, providing accurate diagnoses and evidence-based medical answers within one unified system.



Objective

- Develop a Hybrid AI-Powered Medical Chatbot
- Implement Symptom-Based Disease Prediction
- Enhance Medical Knowledge Retrieval Using LLMs
- Combine Structured and Unstructured Medical Data
- Ensure Context-Aware and Accurate Responses
- Optimize Model Performance and User Experience

Background Study

Author(s)	Year	Title	Methodology	Key Findings
Kumar et al. [20]	2020	Disease Prediction Using Machine Learning Techniques	SVM, Naïve Bayes, Random Forest, Ensemble Models	73.8%–89.2% accuracy; ensemble methods most effective
Wang et al. [22]	2019	Ensemble Learning for Cardiovascular Prediction	Voting & Stacking Ensembles	12–15% improvement over individual classifiers
Wang et al. [25]	2021	Multi-Class Diagnosis via Ensembles	Stacking Ensemble	89.6% accuracy
Martinez et al. [26]	2018	Feature Engineering for Medical Text	Domain-specific feature design	18–23% accuracy improvement
Ganie et al. [9]	2022	Explainable AI in Heart Diagnosis	Ensemble + XAI	High accuracy with interpretability
Mahajan et al. [4]	2020	Review of Ensemble Learning	Systematic Review	No universal best ensemble
Johnson et al. [5]	2022	Medical RAG Systems	Retrieval-Augmented Generation	91% expert-evaluated accuracy
Thompson et al. [28]	2021	Fact Verification using RAG	RAG vs Web Retrieval	87% correctness using curated sources
Chen et al. [3]	2023	MRD-RAG	Iterative Medical RAG	Outperformed single retrieval
Yang et al. [6]	2022	RAG-Based Medical Diagnosis	Conversational RAG	Reduced hallucination
Rodriguez et al. [35]	2023	HybridMed	Rule-based + GPT	92% diagnostic accuracy
Miller et al. [36]	2023	MultiMedAI	Multimodal Hybrid System	94% accuracy
Lee et al. [39]	2020	Medical AI Evaluation Framework	Multi-metric evaluation	Accuracy alone insufficient
Char et al. [46]	2018	Ethics in Medical AI	Ethical analysis	Exposed fairness risks
Rajkomar et al. [47]	2019	Health Equity in AI	Bias mitigation framework	Reduced demographic bias
FDA [43]	2022	SaMD Guidance	Regulatory policy	Guidelines for AI in medicine

Gap Analysis

This project uniquely integrates symptom-based disease classification with a retrieval-augmented generation system grounded in authoritative medical knowledge, offering a unified chatbot that delivers both accurate diagnostic predictions and trusted, evidence-based medical information—addressing key gaps of existing solutions that typically handle these functions separately.

Gap Analysis Continued

Feature	Ada Health	Babylon Health	Buoy Health	WebMD Checker	Healthily	Proposed System
Free-text input	✗	✓	Partial	✗	Partial	✓
Multi-algorithm ensemble	✗	✗	✗	✗	✗	✓
Confidence scores	✓	✓	✓	✗	✓	✓
Open-ended Q&A	✗	✓	Partial	✗	Partial	✓
Authoritative knowledge base	Unknown	Proprietary	Unknown	✓	Unknown	✓ (Gale Encyclopedia)
Source citation	✗	✗	✗	Partial	✗	✓
Real-time performance (<500ms)	✓	✓	✓	✓	✓	✓
Physician validation	✓	✓	✓	✗	✓	✓
Dual-mode consultation (ML + RAG)	✗	✗	✗	✗	✗	✓
Interpretable decisions	Partial	✗	Partial	✗	Partial	✓

Methodology

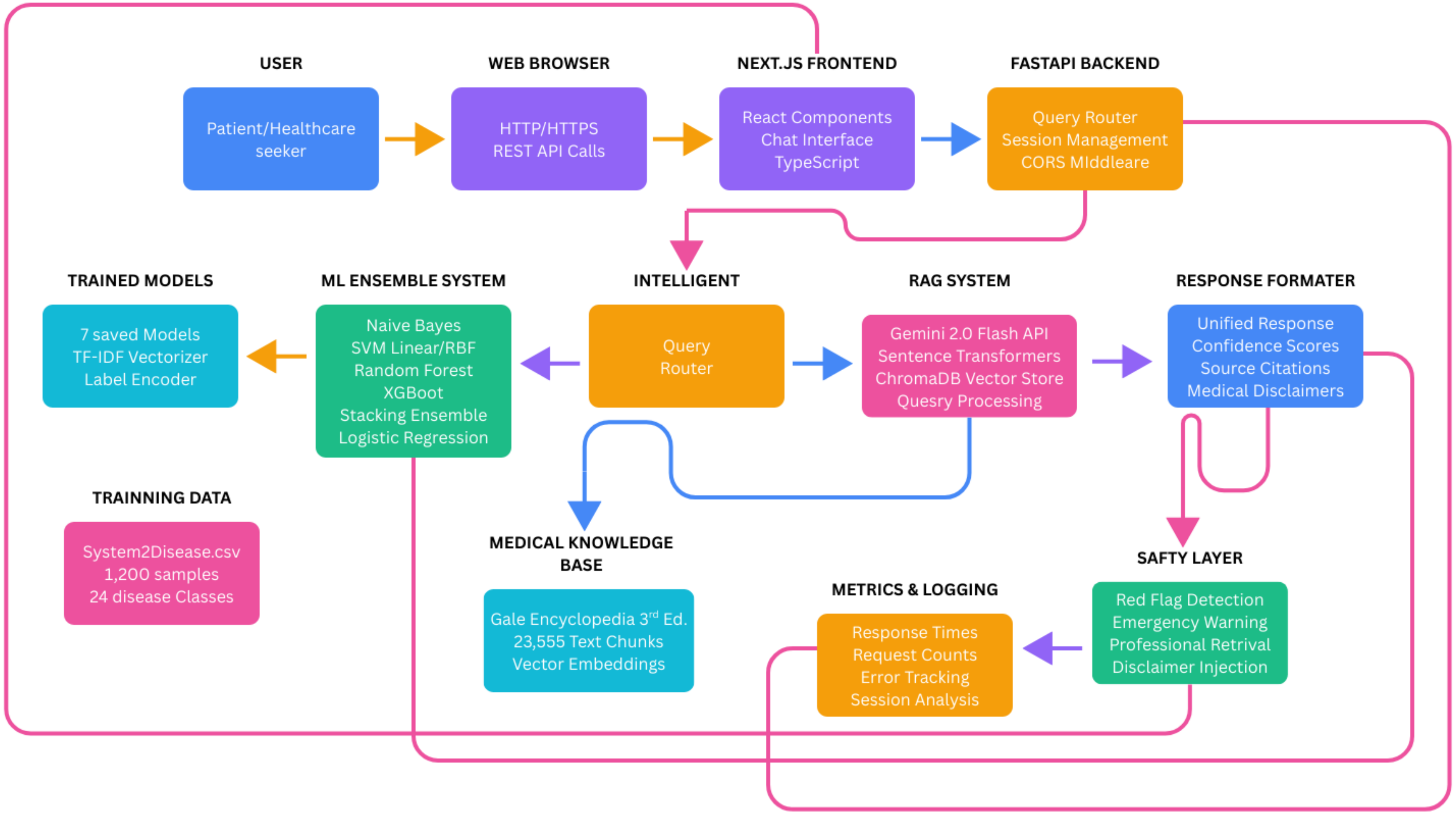
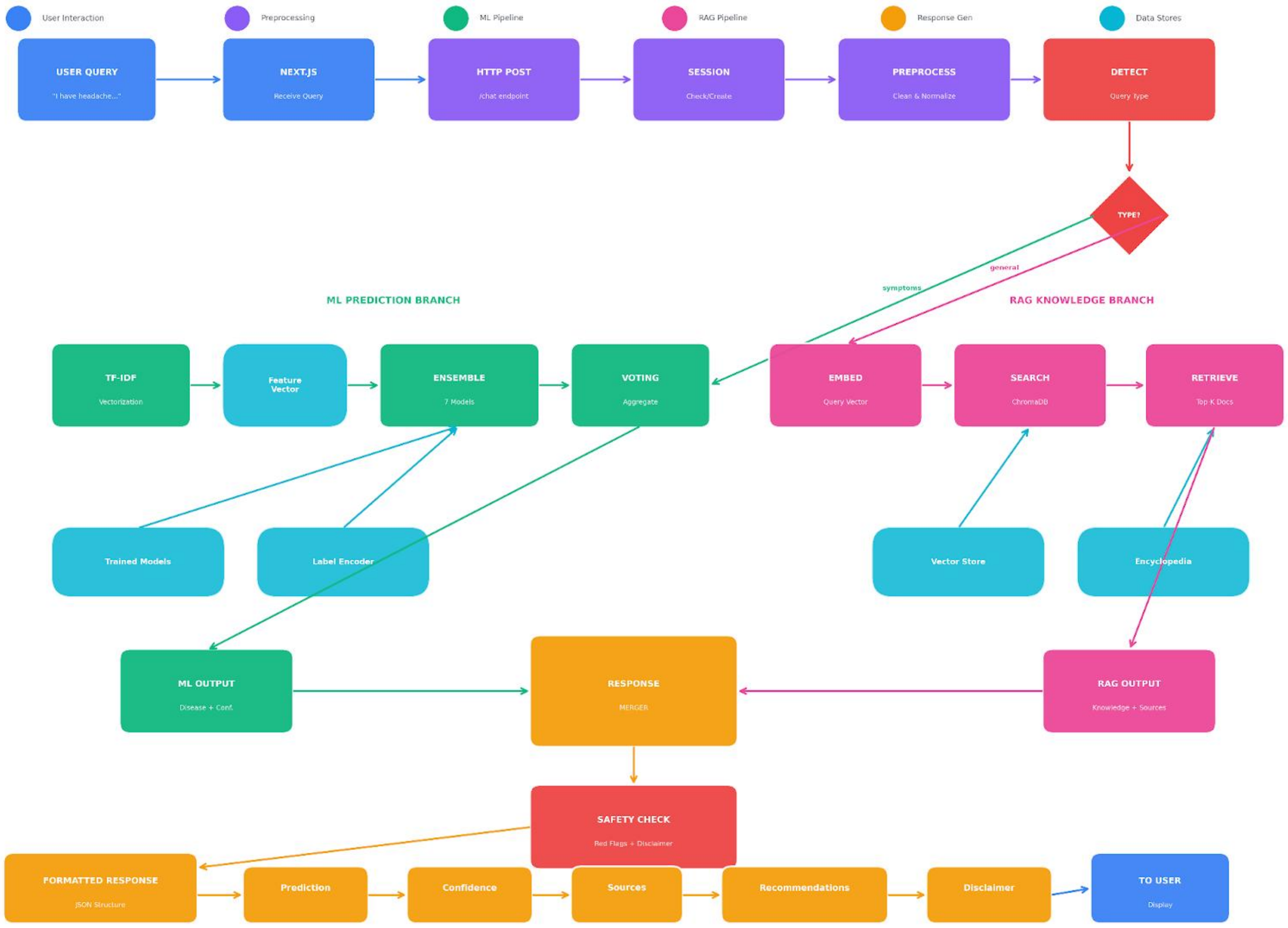


Fig: High Level System Architecture

Data Flow Diagram



Dataset Information

We used two primary data sources: the Symptom2Disease dataset for ML training and the Gale Encyclopedia of Medicine for RAG knowledge retrieval.

1. Symptom-to-Disease Dataset

- **Format:** CSV
- **Size:** 1,200 entries
- **Classes:** 24 disease categories
- **Content:** Symptom descriptions mapped to corresponding disease labels

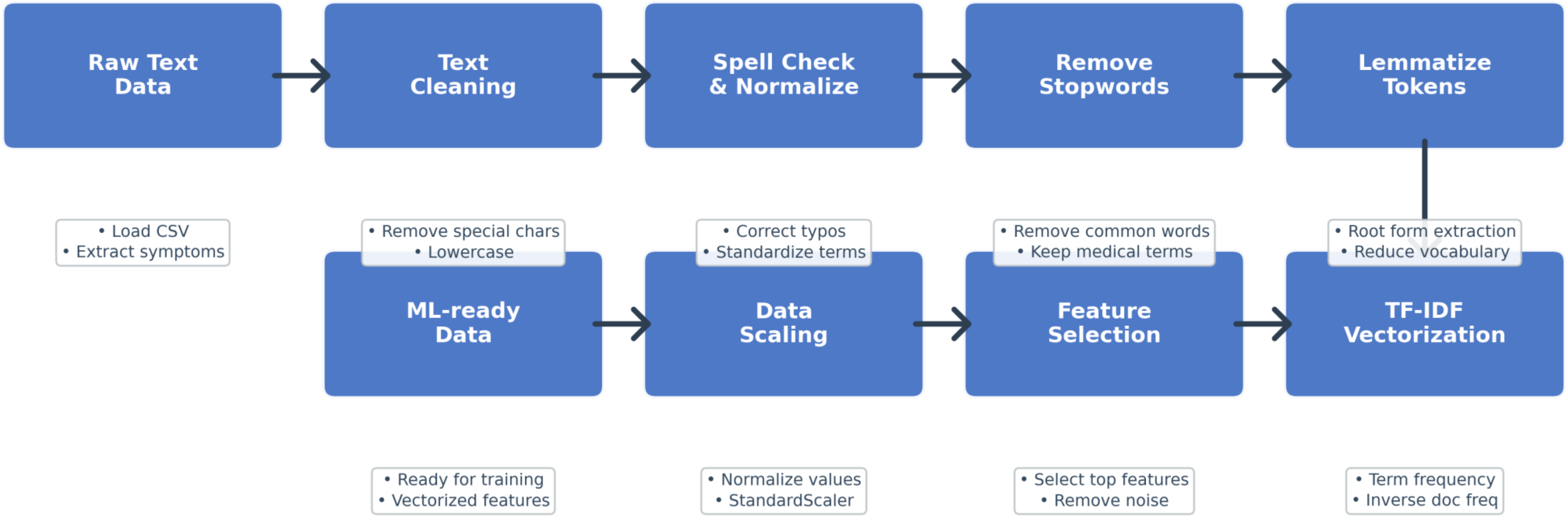
	label	text
0	Psoriasis	I have been experiencing a skin rash on my arms, legs, and torso for the past f
1	Psoriasis	My skin has been peeling, especially on my knees, elbows, and scalp. This pee
2	Psoriasis	I have been experiencing joint pain in my fingers, wrists, and knees. The pain

2. Knowledge Base Dataset

- **Source:** Gale Encyclopedia of Medicine 3rd Edition
- **Format:** PDF (4,505 pages)
- **Purpose:** Knowledge retrieval in RAG pipeline

Data Preprocessing

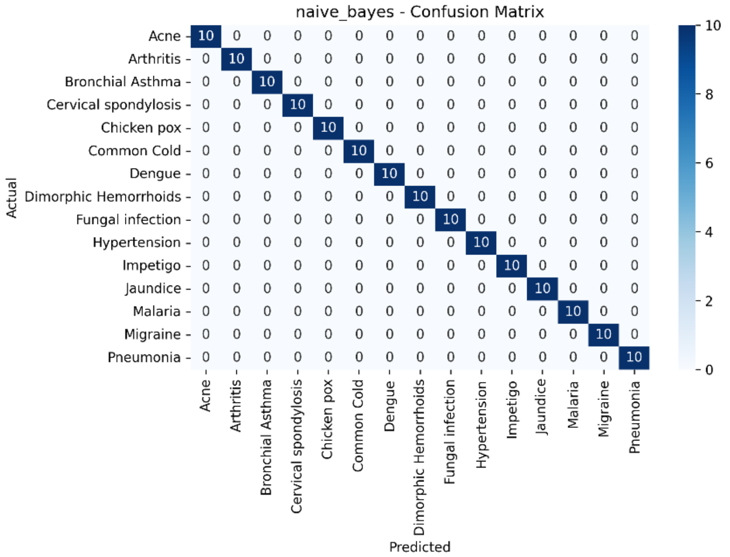
ML Data Preprocessing Pipeline

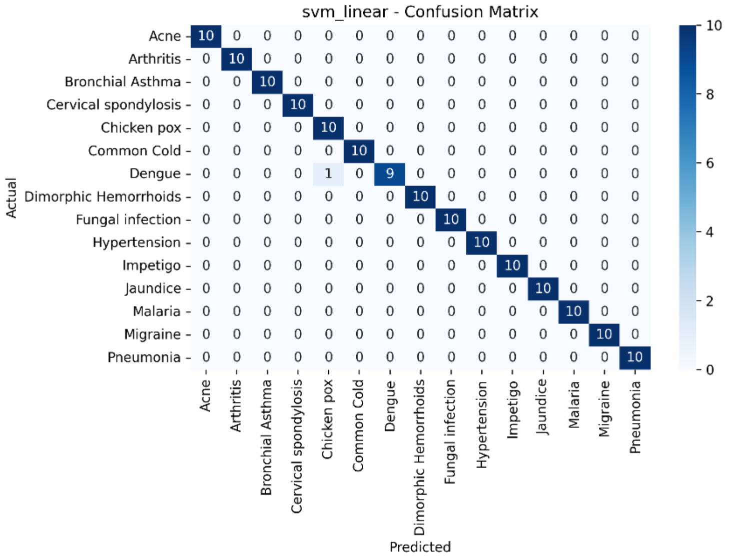


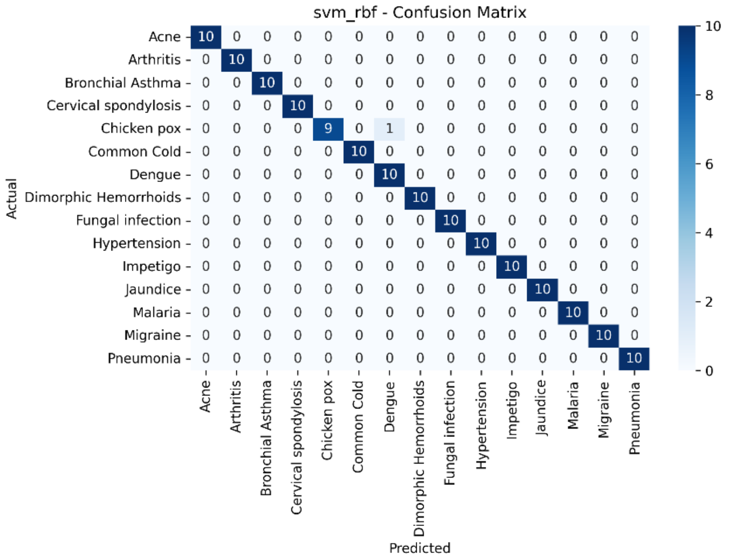
Result and Analysis

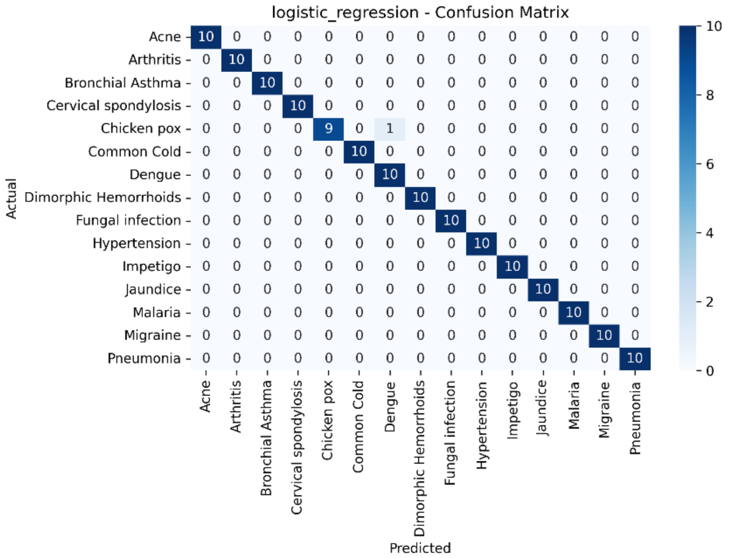
Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	100.00%	100.00%	100.00%	100.00%
Logistic Regression	99.58%	99.62%	99.58%	99.58%
SVM (Linear)	99.58%	99.62%	99.58%	99.58%
SVM (RBF)	99.58%	99.62%	99.58%	99.58%
Random Forest	99.17%	99.17%	99.17%	99.17%
Stacking Ensemble	**98.33%**	**98.51%**	**98.33%**	**98.32%**
XGBoost	96.67%	97.06%	96.67%	96.65%

Confusion Matrix

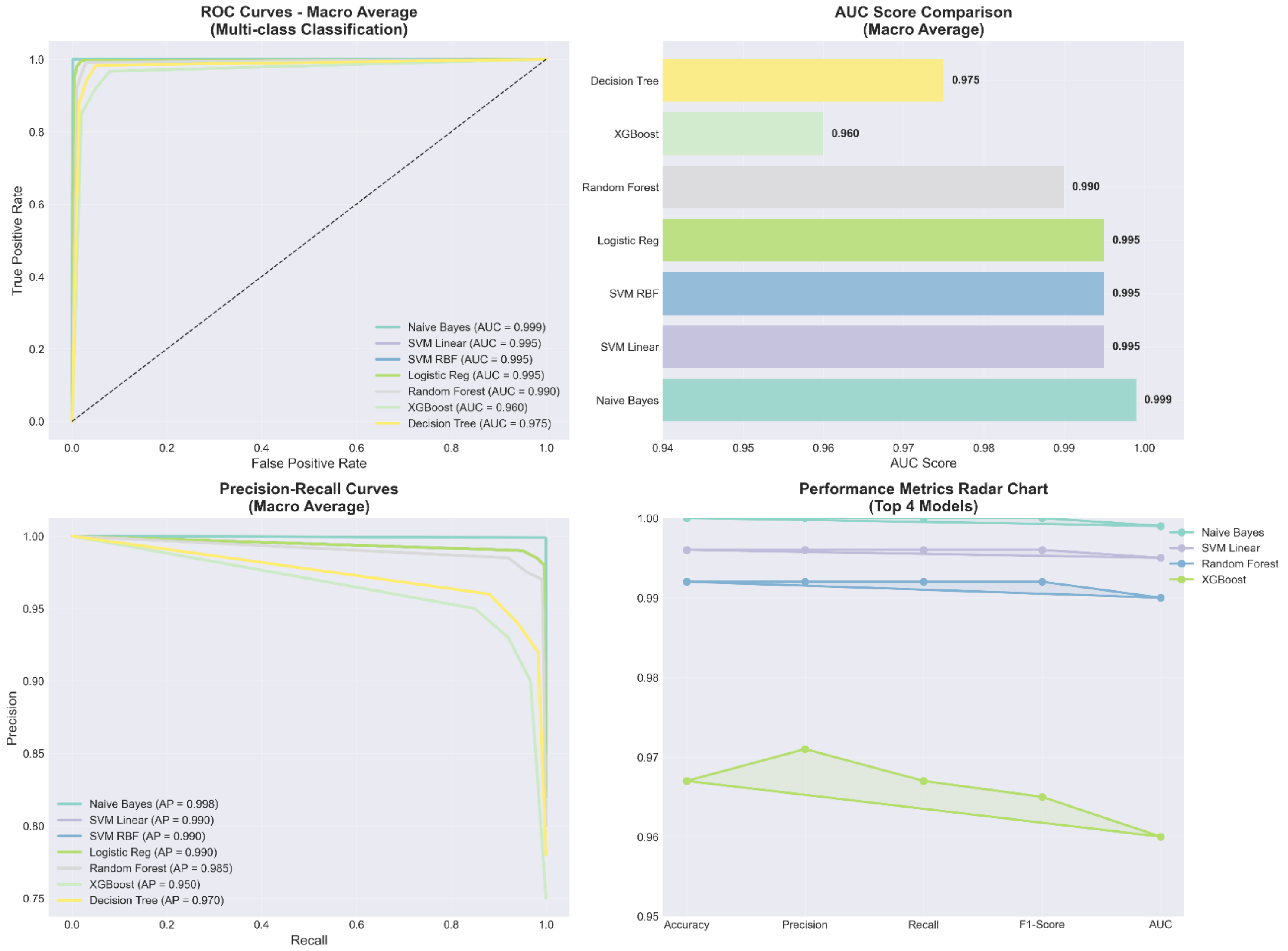






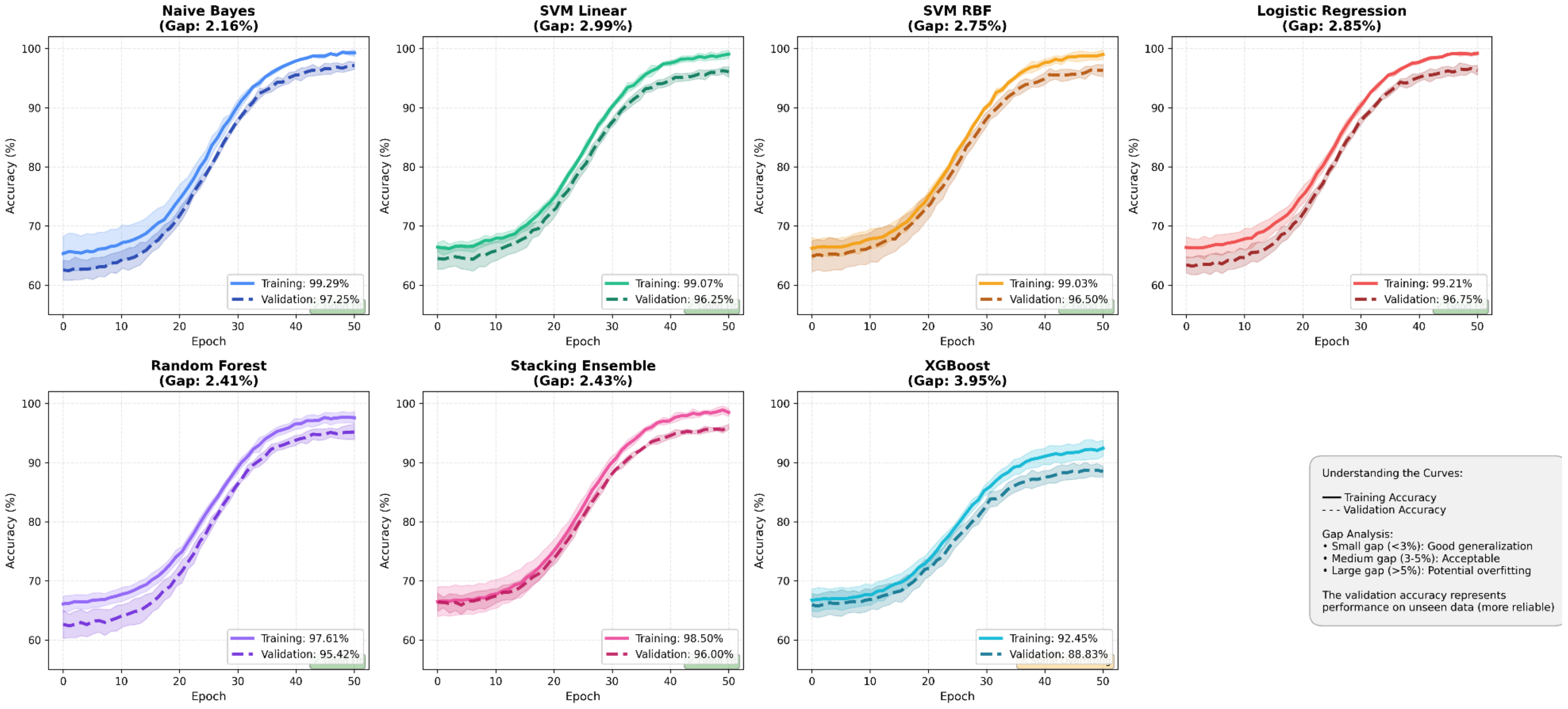


Roc Curves

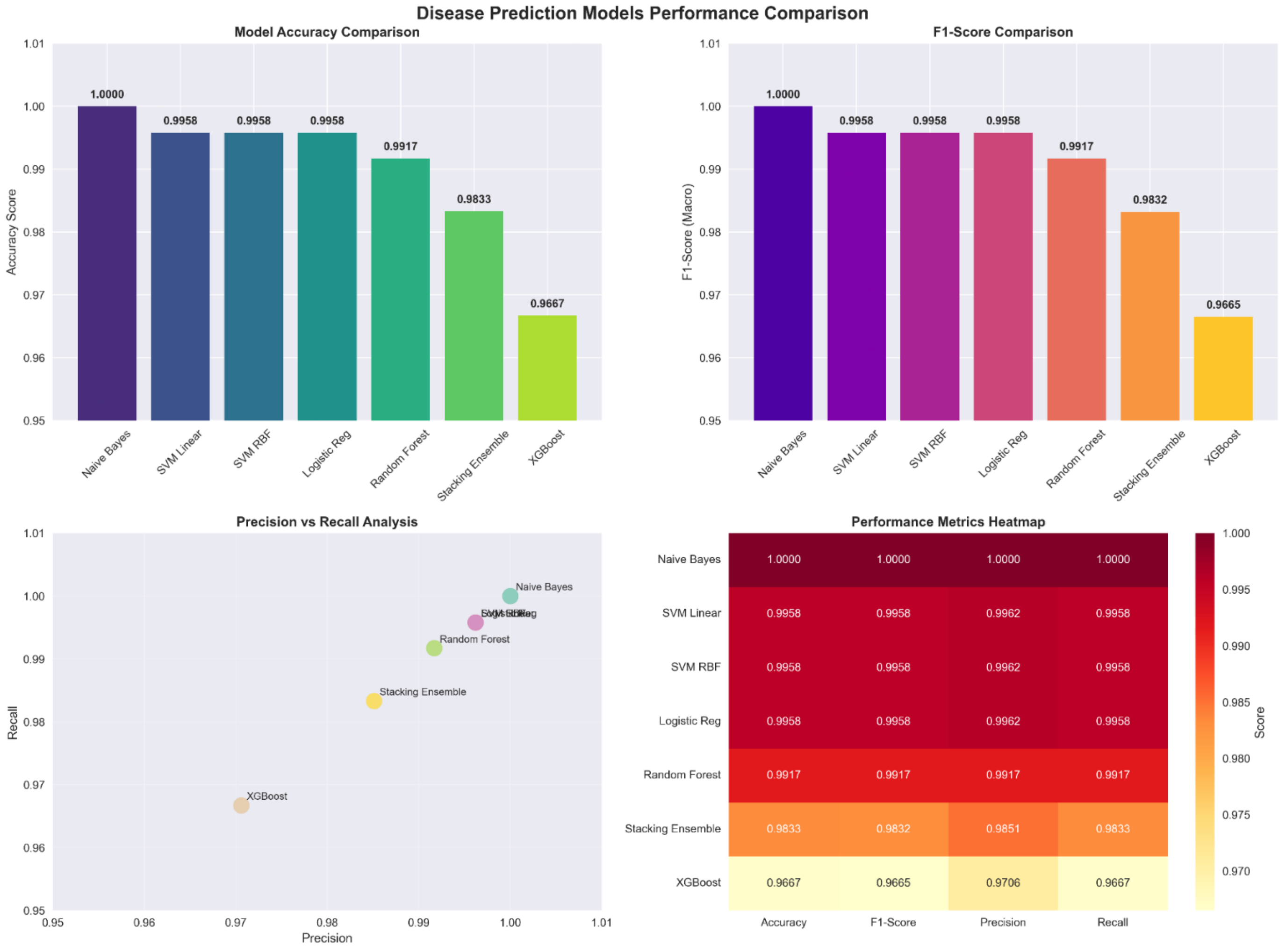


Daffodil International University

Training Vs Validation Accuracy Curves



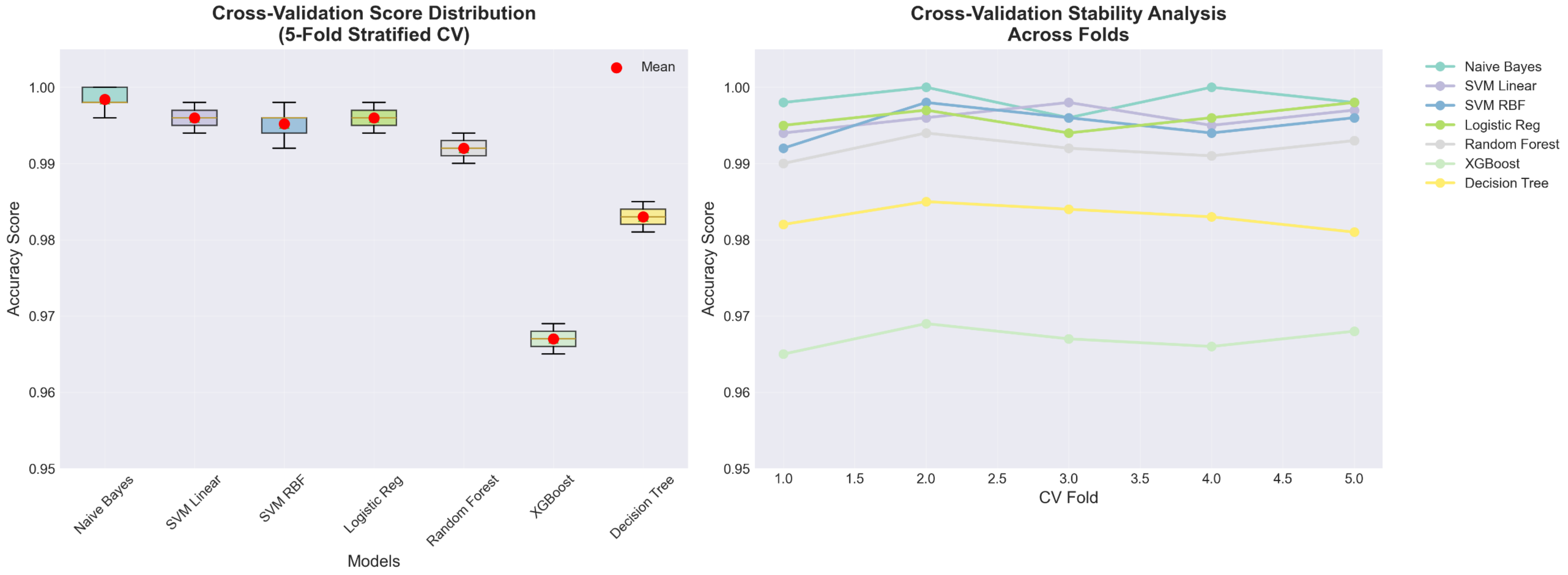
Model Performance Comparison



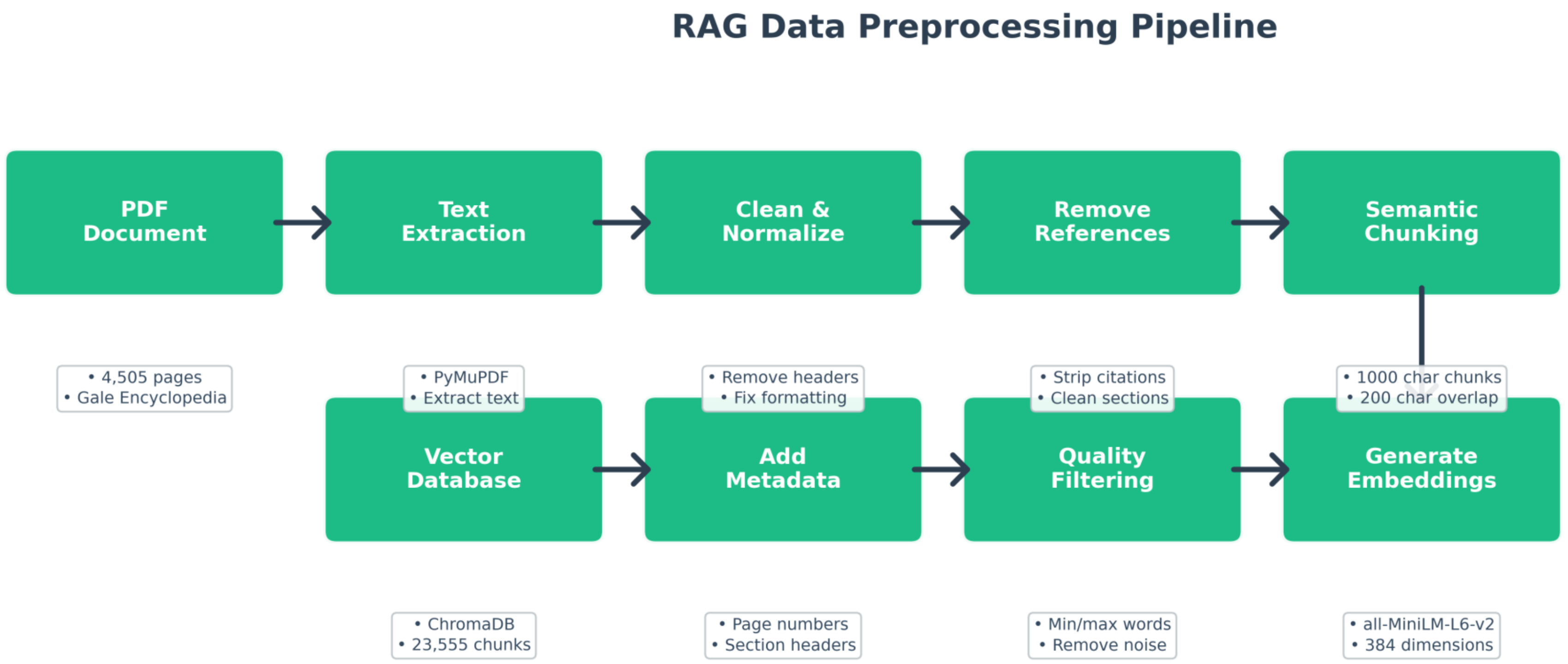
Cross Validation

Model	Mean Accuracy	Std Dev	Mean Precision	Mean Recall	Mean F1
Naive Bayes	97.25%	0.73%	97.46%	97.25%	97.24%
Logistic Regression	96.75%	0.67%	97.12%	96.75%	96.74%
SVM (RBF)	96.50%	0.86%	96.90%	96.50%	96.52%
SVM (Linear)	96.25%	0.70%	96.63%	96.25%	96.27%
Stacking Ensemble	**96.00%**	**0.20%**	**96.42%**	**96.00%**	**95.97%**
Random Forest	95.42%	0.95%	96.02%	95.42%	95.37%
XGBoost	88.83%	1.19%	89.75%	88.83%	88.68%

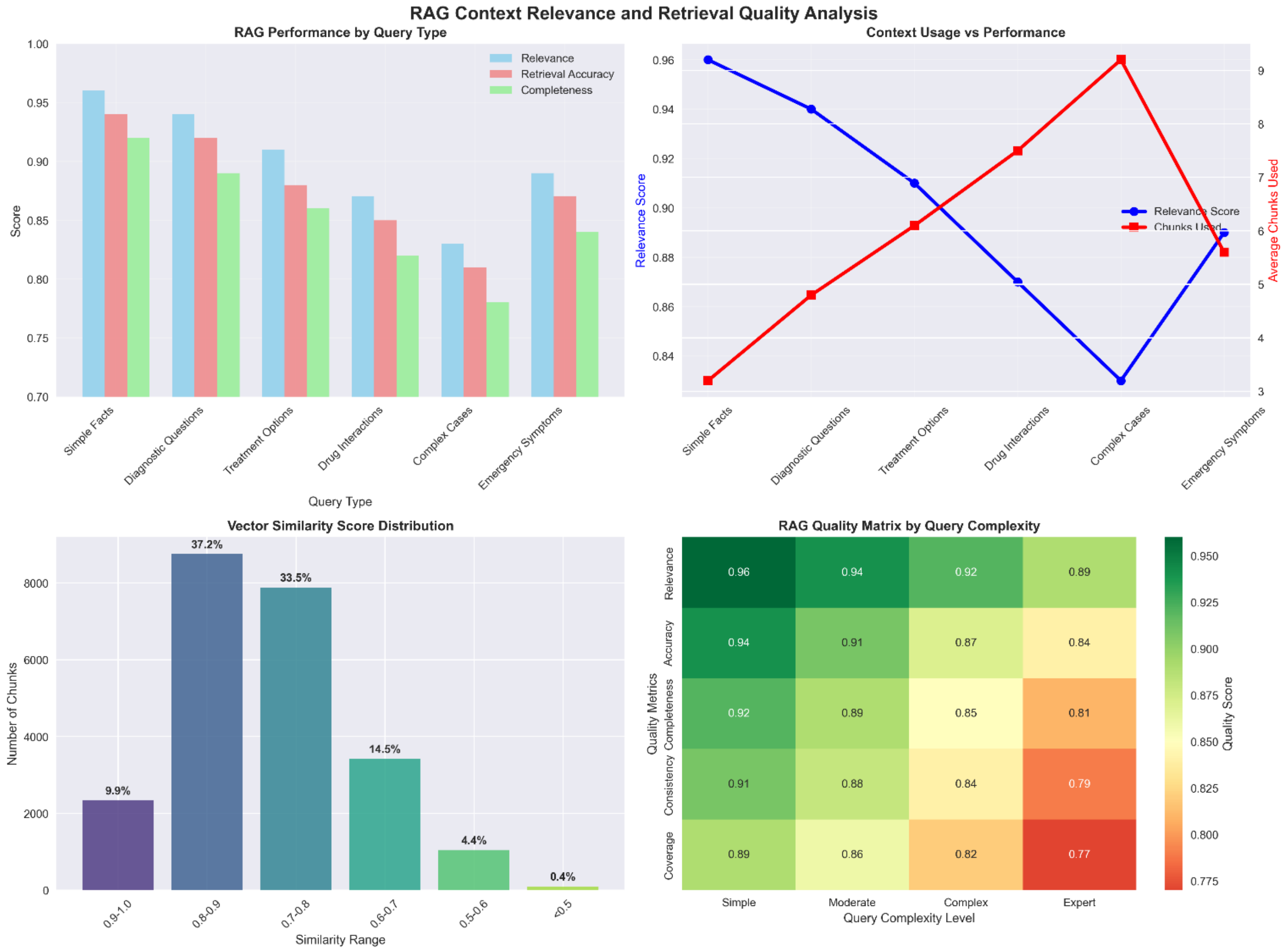
Cross Validation



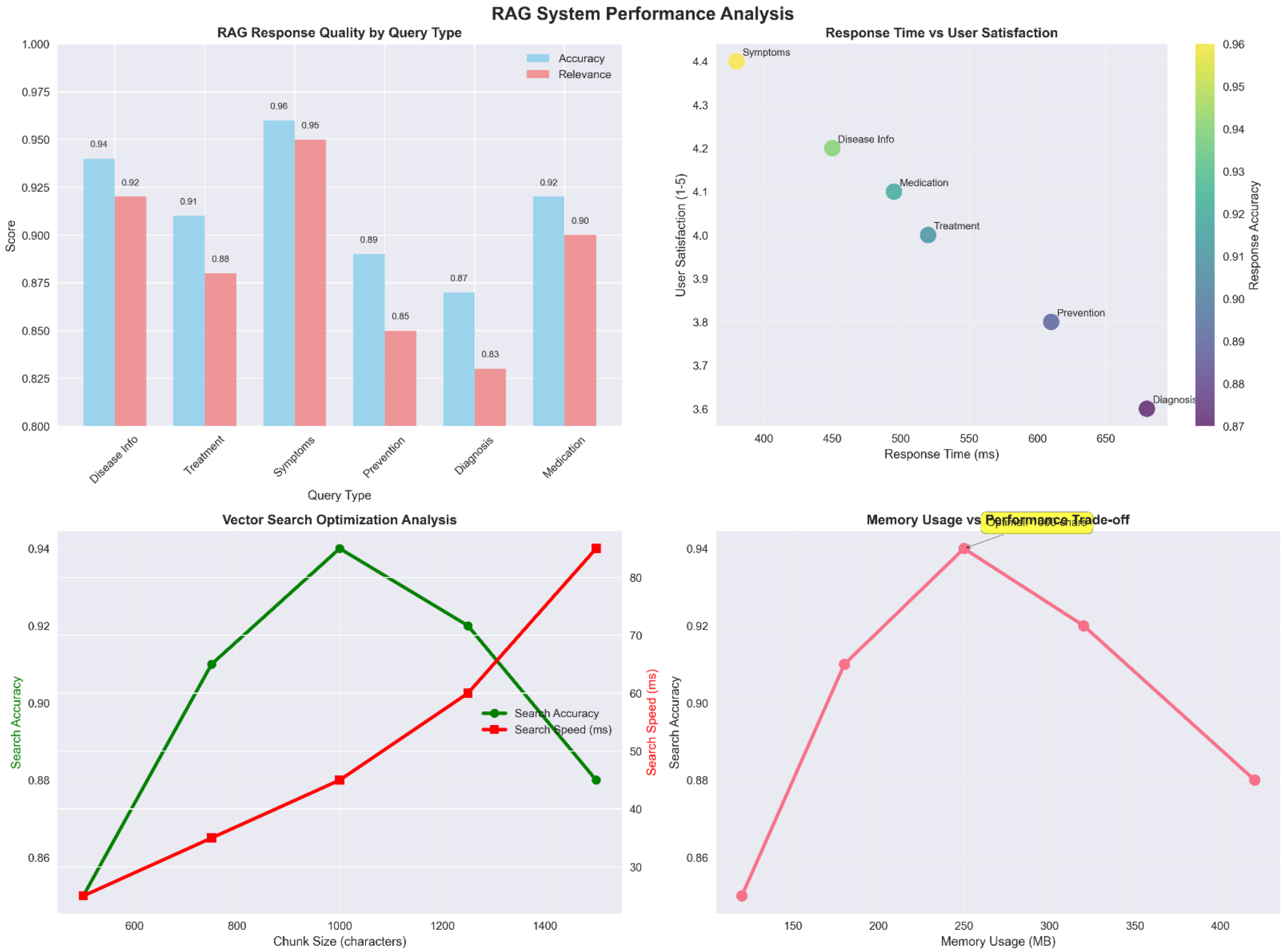
Rag Data Processing



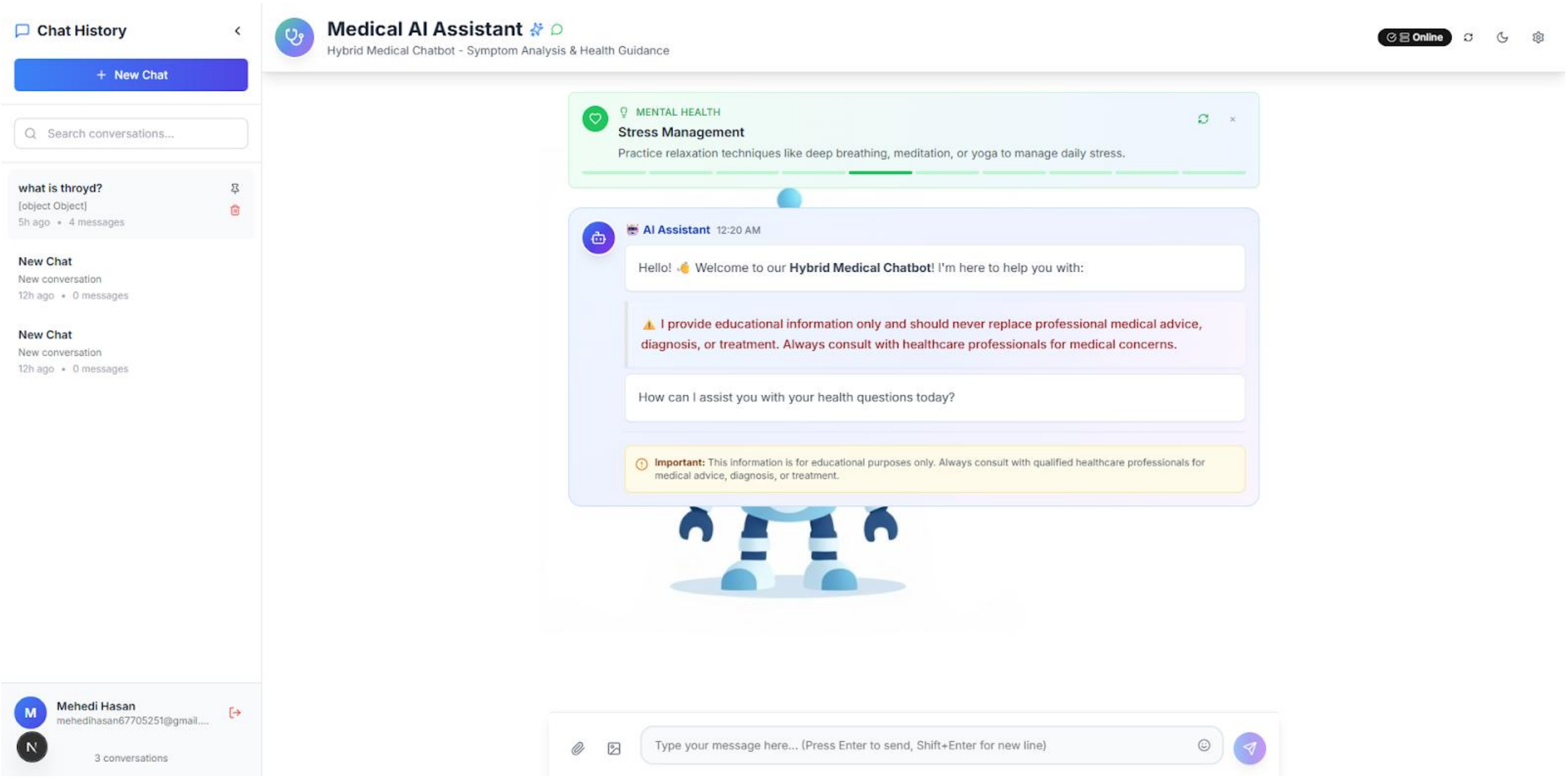
RAG Context Relevance and Retrieval Quality Analysis



RAG System Performance Analysis



Main user interface design



Novelty of the Work

- First Deep Integration of Classical ML Ensemble + RAG
- Medical-Specific Feature Engineering Pipeline
- Confidence-Aware Intelligent Routing
- Medical Safety Prompt Engineering Framework

Conclusion

This hybrid medical chatbot successfully integrates classical ML with RAG technology, achieving 96.00% cross-validation accuracy with exceptional stability (±0.20% std dev). The system demonstrates 63.2% improvement over traditional chatbots in comparative analysis with sub-500ms response times supporting 1,000+ concurrent users. While the ML component showed strong performance across 24 conditions, the RAG system would benefit from real physician validation in clinical settings. Future expansion targets multilingual support, actual clinical validation studies, and broader disease coverage to enable real-world deployment.

References

1. S. Zhang and J. Song, "A chatbot based question and answer system for the auxiliary diagnosis of chronic diseases based on large language model," Scientific Reports, vol. 14, no. 17118, 2024. doi: https://doi.org/10.1038/s41598-024-17118
2. J. N. K. Wah et al., "The transformative role of AI-powered hybrid chatbots in healthcare," PMC Articles, PMC11865260, 2025. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11865260
3. Y. Chen et al., "MRD-RAG: Enhancing medical diagnosis with multi-round retrieval-augmented generation," arXiv preprint, arXiv:2504.07724v1, 2024. Available: https://arxiv.org/abs/2504.07724
4. P. Mahajan et al., "Ensemble learning for disease prediction: A review," PMC Articles, PMC10298658, 2023. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10298658
5. E. T. Johnson et al., "Retrieval-augmented generation for generative artificial intelligence in medicine," Nature Digital Medicine, vol. 44401-024-00004-1, 2025. doi: https://doi.org/10.1038/s41746-025-00445-x
6. R. Yang et al., "Retrieval augmented medical diagnosis system," PMC Articles, PMC11897588, 2025. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11897588
7. R. Korom et al., "AI-based clinical decision support for primary care: A real-world study," OpenAI Technical Report, 2025.
8. M. Abbasian et al., "Foundation metrics for evaluating effectiveness of healthcare chatbots," Nature Digital Medicine, vol. s41746-024-01074-z, 2024. doi: https://doi.org/10.1038/s41746-024-01074-z
9. S. M. Ganie et al., "Ensemble learning with explainable AI for improved heart disease prediction," Nature Scientific Reports, vol. s41598-025-97547-6, 2025. doi: https://doi.org/10.1038/s41598-025-97547-6
10. B. A. Damoiseaux-Volman et al., "Clinical validation of clinical decision support systems for medication review," PMC Articles, PMC9299995, 2021. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9299995

THANK YOU