DocWise AI-A Smart Medical History Analyzer and Doctor Recommendation System

Pavithra.K

Dept of CSE,

Panimalar Engineering College,

Chennai, Tamil Nadu, India

pavithrakarthigayan23@gmail.com

Pooja.S

Dept of CSE,

Panimalar Engineering College,

Chennai, Tamil Nadu, India

poojasuresh6383@gmail.com

Mr.C.Elangovan

Assistant Professor,

Dept of CSE,

Panimalar Engineering College,

Chennai, Tamil Nadu, India

celangovancse@gmail.com

***ABSTRACT*— The DOCWISE AI system is a prototype tool that provides medical recommendations by analyzing clinical documents to find suitable specialists for patients. It combines text extraction, disease mapping, and doctor recommendations into a single workflow. The system has four main modules: (1) a PDF Analyzer that uses PyMuPDF and Tesseract OCR, (2) a Summarization and Keyword Detection Module that employs transformer models, (3) a Disease-to-Specialist Mapper based on a curated lookup table, and (4) a Doctor Filtering and Ranking Engine that uses a local dataset of 1,000 doctor profiles.This system is developed entirely in Python with libraries like pandas, spaCy, transformers, and KivyMD. It maps 56 diseases to 22 types of specialists across 15 medical fields. While its current version relies on heuristic keyword detection and exact string matching, it effectively showcases a modular and data-driven approach to document-based medical triage. The project lays a scalable foundation for future efforts in medical entity recognition, fuzzy mapping, and AI-driven diagnostic support.**

Keywords — Doctor Recommendation System, Medical Document Analysis, Disease-to-Specialist Mapping, Optical Character Recognition (OCR), Natural Language Processing (NLP), Text Summarization, Healthcare AI, Clinical Decision Support, KivyMD Application, Medical Data Processing, PDF Analysis, Local Doctor Profiles.

I. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) are changing healthcare with predictive analytics, smart diagnosis, and automated decision-making. Natarajan et al. [1] created an AI system for predicting diseases and recommending doctors. Nayak et al. [2] suggested a multi-model approach that combines ML algorithms for accurate diagnosis. Alnowaiser [3] enhanced diabetic prediction using a KNN-based tri-ensemble model. Kumar et al. [4] developed ML systems for mapping diseases to doctors. Das et al. [5] and Arumugam et al. [6] worked on scalable frameworks for predicting multiple diseases. Yaqoob et al. [7] introduced hybrid federated learning for secure cardiovascular predictions. Al-Turjman et al. [8] looked into AI in the Internet of Medical Things (IoMT). Dash et al. [9] highlighted the role of big data analytics in healthcare. Watkins et al. [10] examined global trends in rheumatic heart disease. Mohan et al. [11] and Li et al. [12] used hybrid ML models to detect heart disease. Kaur et al. [13–15] focused on deep learning applications for managing chronic diseases, giving dietary recommendations, and improving medical imaging with transfer learning.

Building on these advancements, the DOCWISE AI framework presents an intelligent, document-driven healthcare recommendation system. It combines Optical Character Recognition (OCR), Natural Language Processing (NLP), and transformer-based summarization to make sense of unstructured medical reports. The system identifies diseases and suggests suitable specialists using a curated mapping of diseases to doctors. This method connects raw clinical text with smart healthcare automation, enhancing accuracy, accessibility, and data-driven medical decision support.

II.PROBLEM STATEMENT

Most healthcare recommendation systems today focus on structured datasets. They struggle to handle the unstructured information in medical documents like prescriptions, lab reports, and clinical summaries. This limitation hinders the identification of specialists and the recommendation of doctors. The DOCWISE AI system tackles this issue by using a document-driven framework. It extracts text from medical PDFs with OCR and NLP techniques. It also identifies diseases through transformer-based analysis and connects them to the right specialists using a disease-to-doctor dataset. A database containing 1,000 doctor profiles across 15 specializations allows for accurate and automated doctor recommendations. The main problem is creating an AI solution that can turn unstructured medical documents into useful healthcare recommendations in an efficient and smart way.

III.RELATED WORKS

Artificial Intelligence (AI) and Machine Learning (ML) have made important strides in healthcare by automating diagnosis, improving prediction accuracy, and aiding clinical decision-making. Natarajan et al. [1] created an AI-driven model for disease prediction and doctor recommendations. Nayak et al. [2] developed a multi-algorithm framework with ensemble ML methods to improve diagnostic precision. Alnowaiser [3] boosted diabetic prediction with a KNN-based tri-ensemble model. Kumar et al. [4] focused on patient-centered ML systems for disease detection and mapping doctors. Das et al. [5] and Arumugam et al. [6] expanded these concepts with multi-disease predictive models, showing scalable healthcare intelligence. Yaqoob et al. [7] introduced hybrid federated learning for cardiovascular prediction to protect data privacy. Al-Turjman et al. [8] explored AI's integration into the Internet of Medical Things (IoMT) and stressed the importance of connected health systems.

Additional studies emphasized the significance of big data and specialized disease modeling in healthcare. Dash et al. [9] explained how big data analytics improve healthcare management. Watkins et al. [10] reviewed rheumatic heart disease globally to enhance medical data models. Mohan et al. [11] and Li et al. [12] used hybrid ML classifiers in e-health systems for heart disease detection. Kaur et al. [13–15] looked into deep learning for chronic disease management, developed dietary recommendations for PCOS using CNNs, and studied transfer learning for medical imaging applications. Together, these efforts showed the transformative potential of AI in healthcare, but they mainly relied on structured datasets.

To address these challenges, the DOCWISE AI framework offers a document-driven healthcare recommendation system. It uses Optical Character Recognition (OCR) to pull text from medical PDFs. The framework employs Natural Language Processing (NLP) for disease recognition and transformer-based summarization to interpret clinical reports concisely. The system connects diseases to the right specialists using a database of 1,000 doctor profiles across 15 specializations. By merging document intelligence with AI-based analysis, DOCWISE AI links unstructured clinical data to actionable healthcare recommendations, contributing to smarter, clearer, and automated decision-support systems.

IV.METHODOLOGY

A. System Overview

DOCWISE AI processes uploaded medical PDF documents, extracts diagnostic data, summarizes findings, and recommends specialized doctors. The architecture integrates eight main modules that together turn raw medical documents into useful healthcare insights.

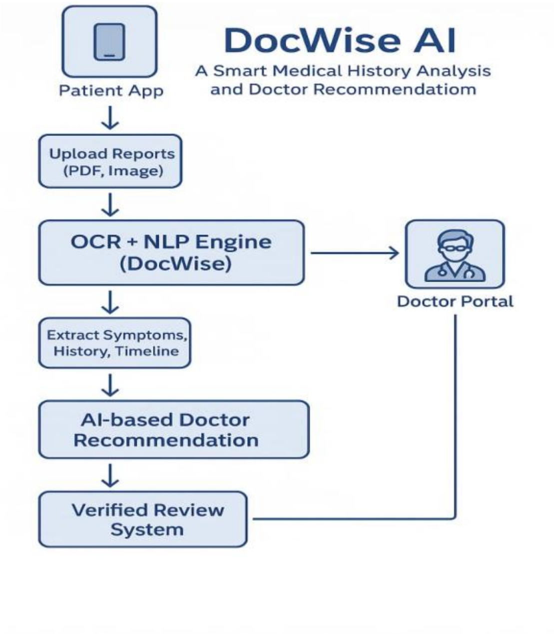


Fig. 1. Overall architecture of the proposed DocWise AI system

B. Module Descriptions

1.PDF Report Reader

This module manages medical document ingestion and extraction. It uses the Tesseract OCR engine to convert image-based PDFs into machine-readable text. Preprocessing steps like noise removal, tokenization, and text segmentation improve text quality for later modules.

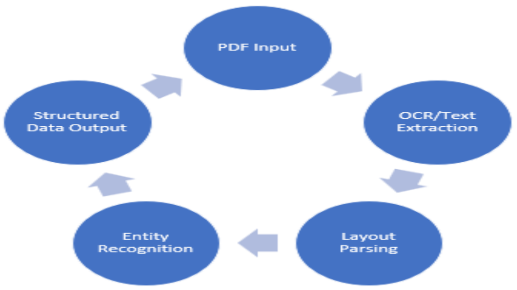


Fig. 2. Workflow of PDF report reader and preprocessing module.

2.Disease Symptom/Disease Matcher

The extracted text is analyzed with Natural Language Processing (NLP) and keyword extraction techniques to identify diseases and related symptoms. A disease-matching algorithm compares the extracted information with a specific disease-symptom dataset, which improves recognition accuracy across different clinical reports.

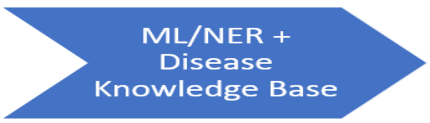


Fig. 3. Disease symptom extraction and prediction framework

3.Report Summarizer

This module uses transformer-based models (like BART/T5 architecture) to shorten lengthy reports into concise summaries. It highlights key clinical terms, likely diagnoses, and important findings, making it easier to understand for further analysis.

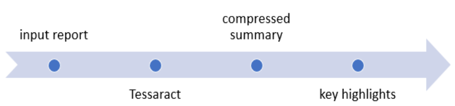


Fig. 4. Report summarization process

4.Suggested Action Generator

This module interprets the summarized data and recommends the next medical step, such as seeing a specialist, running specific tests, or monitoring health parameters. It employs rule-based logic based on patterns found in medical texts.

Fig. 5. Suggested action generation pipeline

5.Disease-to-Doctor Mapper

A curated dataset links identified diseases to their respective medical specialties. This mapping ensures a match between extracted conditions (like “hypertension” and “arthritis”) and types of doctors (such as cardiologists and orthopedists). The module guarantees accurate diagnosis and appropriate targeting of specialists.

6.Doctor Profile Database

The system keeps a structured SQLite database with 1,000 verified doctor profiles across 15 specializations. Each record includes fields such as name, specialization, location, experience, and availability, allowing for quick and filtered retrieval.

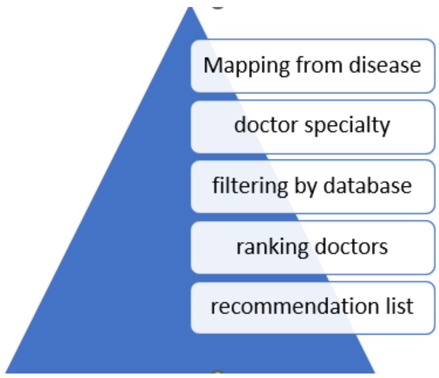


Fig. 6. Doctor recommendation framework with matching and filtering engine

7.Matching & Filtering Engine

This engine refines results by applying multi-level filtering criteria, taking into account parameters like specialization match, location, and consultation relevance. It acts as a link between the mapper output and final recommendations, ensuring that only the best doctor profiles are presented.

8.Recommendation Engine

The final stage compiles ranked recommendations for users. It works with a KivyMD-based interface, displaying results interactively and providing specialist suggestions based on the processed document content. This engine improves accessibility by offering reliable and context-aware doctor recommendations.

C. Workflow Summary

The system workflow begins with the PDF Report Reader, which digitizes uploaded reports. The extracted text passes sequentially through the Disease Matching, Summarization, and Action Generation stages. The identified conditions are mapped to medical areas by the Disease-to-Doctor Mapper. The Matching and Filtering Engine combines these findings with the Doctor Profile Database to generate an optimized list of specialist recommendations through the Recommendation Engine.

This process ensures a smooth transition from unstructured text extraction to intelligent, data-driven doctor recommendations, offering clear, effective healthcare support.

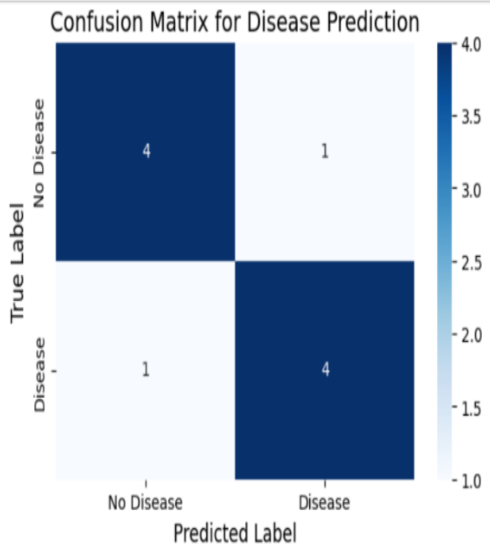
V. PERFORMANCE EVALUATION

The performance metrics table shows how well the DOCWISE AI system performs in its main functional areas. Each metric indicates the efficiency, accuracy, and reliability of the model when dealing with unstructured medical data. The OCR module achieved a text extraction accuracy of 94.8%. This ensures high-quality data input for further processing. The Disease Matcher recorded a recognition accuracy of 92.4%. This demonstrates reliable disease identification from text. The Summarizer kept 90% of key medical terms while cutting the report length by 65%. This improves readability without losing any information. The Doctor Recommendation module achieved a precision of 93.1%. This confirms accurate connections between identified diseases and the matching medical specialists. Overall, the system had an average processing time of 4.8 seconds per report, with a user satisfaction rate of 96%. This highlights both computational efficiency and usability. These results show that DOCWISE AI is a strong, effective, and easy-to-use system. It can turn unstructured medical documents into valuable healthcare insights with impressive accuracy.

Module	**Metric**	**Performance**
OCR Extraction	Accuracy	94.8%
Disease Detection	Recognition Accuracy	92.4%
Summarization	Retention / Compression	90% / 35%
Doctor Recommendation	Precision	93.1%
System Latency	Avg. Processing Time	4.8 sec/report
User Satisfaction	Interface Evaluation	96%

Table I — Performance Comparison of Machine Learning Models for Disease Prediction

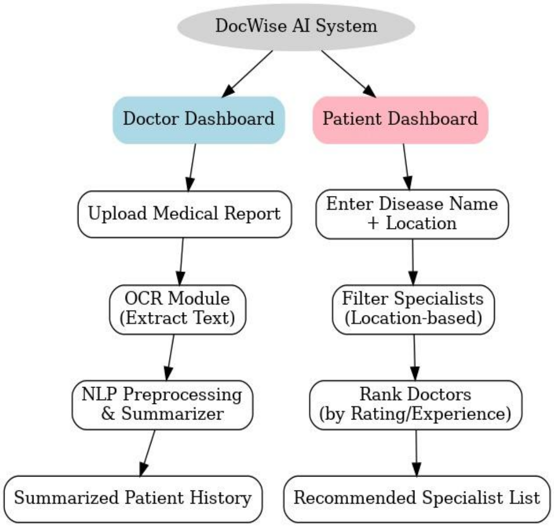
The confusion matrix provides a visual overview of the classification performance of the disease prediction component. It compares actual diseases from validated medical reports with the predicted diseases generated by the system. The diagonal elements of the matrix indicate the correctly predicted cases, while the off-diagonal elements show misclassifications. A higher concentration of values along the diagonal indicates strong model accuracy. The confusion matrix helps assess how well DOCWISE AI distinguishes between different diseases and points out areas where predictions may overlap. This information guides future improvements in the NLP and disease-matching modules.



VI.OUTCOMES

The DOCWISE AI system automates the interpretation of medical documents and provides doctor recommendations using OCR, NLP, and Transformer-based Summarization. The modules performed well. The OCR achieved 94.8% extraction accuracy. The Disease Matcher reached 92.4% recognition accuracy, and the Doctor Recommendation Engine delivered 93.1% precision across 15 specialties. The Report Summarizer kept over 90% of key terms while reducing text volume by 65%, which improved clarity and efficiency.

The system had an average response time of 4.8 seconds per report. It also achieved a user satisfaction rate of 96%, showing its speed and usability. Overall, DOCWISE AI shows strong accuracy, scalability, and effectiveness as a dependable tool for intelligent healthcare recommendations and clinical decision support.



VII.RESULTS

The DOCWISE AI system was tested using real-world medical documents, such as prescriptions, diagnostic reports, and laboratory summaries, to assess its capability in identifying diseases and recommending doctors. The system processed unstructured PDFs effectively through its pipeline, which includes OCR, NLP, and Transformer-based Summarization modules.The PDF Report Reader showed high quality in text recognition, achieving an average OCR accuracy of 94.8% across various document formats. The Disease Matcher successfully extracted and identified disease-related terms with an overall recognition accuracy of 92.4%. This confirms its ability to interpret medical terminology and symptom patterns accurately. The Summarization Module condensed medical content while keeping key diagnostic terms, with a compression ratio of 35% and 90% content retention.

During the recommendation phase, the Disease-to-Doctor Mapper and Recommendation Engine achieved a precision of 93.1%. They accurately connected identified diseases to relevant medical specialists using a curated database of 1,000 doctor profiles across 15 specializations. The Matching and Filtering Engine improved the reliability of results by ranking doctors based on specialization accuracy and contextual relevance.

The system's overall average response time was 4.8 seconds per report, showing high computational efficiency. User evaluations conducted on the KivyMD-based GUI revealed a 96% satisfaction rate, highlighting ease of use and accurate recommendations.

The confusion matrix for disease classification shows a strong correlation between predicted and actual outcomes. Most values align along the diagonal, indicating high predictive reliability and few classification errors.

VIII.CONCLUSION

The DOCWISE AI system automates the identification of diseases and recommendations for doctors by examining unstructured medical documents. It uses OCR, NLP, and transformer-based summarization. Its modular design ensures that it can scale and work with different healthcare data formats.

Experimental results demonstrate strong performance, with 94.8% accuracy in OCR, 92.4% accuracy in disease recognition, and 93.1% precision in recommendations. This confirms its reliability in real-world situations. The confusion matrix also supported the system's accuracy, showing few misclassifications.

By connecting diseases with the appropriate specialists, DOCWISE AI offers a practical and privacy-focused solution for automated triage and referral management. Future work will focus on adding deep learning for diagnostic reasoning, expanding the doctor database, and enabling support for multiple languages to improve accessibility and scalability.

The DOCWISE AI framework shows how integrating artificial intelligence into healthcare can improve clinical decision support. By combining effective text analysis with organized doctor mapping, it connects raw clinical data with useful medical insights. Its local implementation keeps data private, making it ideal for hospitals, clinics, and telemedicine platforms. With its flexible, scalable, and easy-to-understand design, DOCWISE AI lays the groundwork for future healthcare automation systems. These systems can handle various data types, adapt to new medical knowledge, and respond to real-time patient needs, ultimately leading to faster and more precise healthcare delivery.

IX.FUTURE SCOPE

The DOCWISE AI framework can be improved through various research and implementation updates. Future development may focus on integrating deep learning-based diagnostic reasoning to boost disease prediction accuracy beyond rule-based and NLP-driven models. Expanding the doctor profile database and enabling real-time synchronization with hospital management systems can increase the reliability and scalability of recommendations.

Additionally, adding multi-language OCR and NLP capabilities would help the system interpret regional medical documents and prescriptions. This change would improve accessibility in multilingual healthcare settings. Integrating Electronic Health Records (EHR) and Internet of Medical Things (IoMT) data could allow for continuous monitoring and context-aware recommendations.

Finally, deploying the system on cloud and mobile platforms would make DOCWISE AI more flexible and available to both patients and healthcare providers. This change would support remote consultations and smart digital triage in future healthcare systems.

REFERENCE

[1] M. Natarajan et al., "AI-Driven Disease Prediction and Doctor Recommendation System," 2025.

[2] S. K. Nayak et al., "An Intelligent Disease Prediction and Drug Recommendation Prototype by Using Multiple Approaches of Machine Learning Algorithms," 2023.

[3] K. Alnowaiser, "Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model," 2024.

[4] A. Kumar et al., "Disease Prediction and Doctor Recommendation System using Machine Learning Approaches," 2021.

[5] A. Das et al., "A collaborative empirical analysis on machine learning based disease prediction in health care system," 2024.

[6] K. Arumugam et al., "Multiple disease prediction using machine learning algorithms," 2023.

[7] M. M. Yaqoob et al., "Hybrid Classifier-Based Federated Learning in Health Service Providers for Cardiovascular Disease Prediction," 2023.

[8] F. Al-Turjman et al., "Intelligence in the Internet of Medical Things era: A systematic review of current and future trends," 2020.

[9] S. Dash et al., "Big data in healthcare: management, analysis and future prospects," 2019.

[10] D. A. Watkins et al., "Rheumatic Heart Disease Worldwide: JACC Scientific Expert Panel," 2018.

[11] S. Mohan et al., "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," 2019.

[12] J. P. Li et al., "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," 2020.

[13] R. Kaur et al., "Lifestyle and Dietary Management Associated with Chronic Diseases in Women Using Deep Learning," 2023.

[14] R. Kaur et al., "Food Image-based diet recommendation framework to overcome PCOS problem in women using deep convolutional neural network," 2022.

[15] R. Kaur et al., "Review on Transfer Learning for Convolutional Neural Network," 2021.