PAVITHRA.K

Dept of CSE

Panimalar Engineering College, Chennai, Tamil Nadu,India

[pavithrakarthigayan23@gmail.com](mailto:pavithrakarthigayan23@gmail.com)

Abstract- Intelligent systems that can aid in medical decision-making and enhance patient care are desperately needed given the explosive growth of healthcare data. An integrated AI-driven framework called DocWise AI: A Smart Medical History Analyzer and Doctor Recommendation System was created to help patients and medical professionals make well-informed medical decisions and receive accurate diagnoses. The two main components of the system are the (1) Medical History Analyzer, which extracts, summarizes, and interprets patient medical reports in a variety of formats using Natural Language Processing (NLP) and Machine Learning techniques, and (2) Doctor Recommendation System, which uses a database of doctor profiles and expertise to recommend the best medical specialists based on the diseases or conditions found. DocWise AI seeks to decrease diagnostic errors, avoid treatment delays, and improve the overall effectiveness of healthcare delivery by fusing automated medical record analysis with customized doctor recommendations. This work supports clinical decision-making through responsible and explainable AI and promotes timely access to appropriate healthcare, which is in line with the UN Sustainable Development Goal (SDG) 3—Good Health and Well-Being.

Keywords- Medical History Analysis, Doctor Recommendation System, Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Healthcare Decision Support, Medical Report Summarization, Disease Prediction, Clinical Decision-Making, Smart Healthcare System

POOJA.S

Dept of CSE

Panimalar Engineering College, Chennai, Tamil Nadu,India

[Poojasuresh6383@gmail.com](mailto:Poojasuresh6383@gmail.com)

**I.INTRODUCTION**

One of the most important fields where prompt and precise decision-making can save lives is healthcare. In this field, artificial intelligence (AI) and machine learning (ML) have become game-changing technologies in recent years, providing creative approaches to medical advice, patient monitoring, and illness prediction. Inefficiencies in traditional healthcare systems, such as delayed diagnosis, a shortage of specialists, and a lack of individualized care, are common. The development of intelligent frameworks that integrate disease prediction and physician recommendation to offer individualized, patient-centric healthcare support has become a growing area of research in an attempt to address these issues [1]. The application of ensemble learning techniques like Support Vector Machines (SVM), Random Forest, and Logistic Regression has been one of the most significant developments. By improving accuracy through multiple model integration, these methods have demonstrated efficacy in drug recommendation and disease prediction. Their reliance on structured datasets is still a drawback, though, because unstructured sources like clinical notes, scanned reports, and patient histories are frequently included in healthcare data [2]. Methods like KNN imputation in conjunction with Tri-Ensemble models have produced exceptionally high prediction rates for some diseases, such as diabetes. However, these approaches are frequently disease-specific and do not generalize to other medical conditions [3]. User-friendly interfaces driven by machine learning classifiers such as Naive Bayes, Random Forest, and Logistic Regression have been incorporated into web-based and mobile platforms to guarantee wider applicability. Although differences in prediction accuracy across diseases underscore the need for additional optimization, these systems make healthcare solutions more feasible and patient-accessible [4]. Additionally, comparative evaluations of machine learning classifiers show that models like Random Forest can predict diseases with almost perfect accuracy. Nevertheless, these conclusions are frequently based on small datasets, which limits their applicability in the real world [5]. The significance of using a variety of algorithms to generalize across conditions is emphasized by recent research on multi-disease prediction frameworks. However, there are still a lot of issues with the variety and scarcity of publicly accessible medical datasets [6]. Patient privacy is another important issue with healthcare AI. In order to overcome this, hybrid classifiers and federated learning have been developed, allowing models to be trained locally on dispersed patient data without jeopardizing private information. This privacy-preserving method shows promise for applications involving multiple diseases and has proven especially effective in the prediction of cardiovascular disease [7]. In parallel, the emergence of the Internet of Medical Things (IoMT) has shown how AI-powered connected devices can facilitate early diagnosis, predictive analytics, and continuous remote monitoring, increasing the accessibility and effectiveness of healthcare [8]. Large-scale disease trend analysis and improved resource allocation have also been made possible by the integration of big data analytics into healthcare systems, even though many of the studies that have already been done are theoretical and lack real-world AI-based implementations [9].

The need for more intelligent solutions is further highlighted by the burden of healthcare around the world. AI-based tools that can expand diagnostic capabilities to underserved populations are relevant, as evidenced by studies on Rheumatic Heart Disease (RHD), which highlight the lack of diagnostic access in low-resource regions [10]. While hybrid models that combine Random Forest, Decision Trees, and other algorithms have demonstrated great promise for improving heart disease prediction accuracy, they are frequently dataset-specific and need to be modified for wider application [11]. Better integration with digital health services has been made possible by the promising outcomes of integrating machine learning (ML) models into e-healthcare systems, which have been used to embed automated disease detection into telemedicine platforms [12]. AI and deep learning models are influencing preventive healthcare in addition to disease prediction. While many of these methods are currently limited to particular demographic groups, deep learning analysis of dietary and lifestyle patterns, for instance, has yielded new insights for personalized chronic disease management [13]. Similar to this, convolutional neural network (CNN)-powered food image-based diet recommendation frameworks have been used to treat conditions like PCOS in women, demonstrating the adaptability of AI for nutritional healthcare; however, these frameworks are still restricted to diet-based interventions [14]. Transfer learning has been used to leverage pre-trained CNNs to increase the adaptability of such models, allowing them to be applied to new healthcare tasks with little retraining [15].

In the meantime, automated medical report generation has been shown to be possible with transformer-based architectures and natural language processing (NLP), which can help with structured patient history analysis and expedite healthcare documentation [16]. By examining both structured and unstructured patient data, AI-driven algorithms used in electronic health records (EHRs) have also demonstrated great promise in terms of producing precise medical diagnoses [17]. More sophisticated frameworks extract insights from scanned medical documents using Optical Character Recognition (OCR), generative AI, and rule-based techniques. These insights can then be directly mapped to recommendations from the appropriate specialist [18]. Furthermore, automated algorithm selection has been addressed by meta-learning techniques, which allow healthcare systems to adjust and optimize results according to the particular data context [19]. Lastly, a clear trend toward patient-centered solutions is reflected in the ongoing development of AI and ML in healthcare. Research continuously highlights how AI-driven methods are improving the overall healthcare experience by reducing the distance between patients and physicians, in addition to increasing diagnostic accuracy. These developments highlight the potential of programs like DocWise AI, which seek to combine individualized doctor recommendations with intelligent medical history analysis to provide quicker, more dependable, and easier access to healthcare services [20].

**II.PROBLEM STATEMENT**

In today's healthcare environment, fragmented medical histories and a lack of trustworthy systems to match patients with the best medical specialists frequently make it difficult to diagnose patients accurately and treat them promptly. Over time, patients often see several different healthcare providers, which leads to scattered medical records in various formats and institutions. Misdiagnosis, postponed treatments, and needless diagnostic test repetition can result from this lack of a cohesive and interpretable medical history. Additionally, patients frequently have trouble finding reliable and suitable medical professionals for their particular conditions, which further postpones care and may jeopardize the effectiveness of treatment. There is currently no integrated system that can automatically analyze a variety of medical records and concurrently suggest the most appropriate specialists based on diseases identified, despite the increasing availability of digital health data. Due to this disparity, DocWise AI, an intelligent and unified platform, must be developed in order to improve clinical decision-making and the general quality and accessibility of healthcare services. DocWise AI can analyze patient medical histories using AI-driven techniques and provide accurate doctor recommendations.

**III.RELATED WORKS**

In recent years, numerous studies have focused on leveraging Artificial Intelligence (AI) and Machine Learning (ML) to enhance healthcare decision-making, medical report analysis, and automated recommendation systems. Several works have attempted to predict diseases and provide treatment or specialist suggestions using structured clinical data.

M. Natarajan et al. [1] proposed an AI-driven framework that integrates disease prediction with doctor recommendation by using multiple AI and ML techniques. Their system improves healthcare accessibility by offering personalized recommendations; however, its performance is heavily dependent on the quality and diversity of the input datasets. Similarly, S. K. Nayak et al. [2] developed an intelligent disease prediction and drug recommendation prototype using ensemble ML models such as Logistic Regression, Random Forest, and SVM. While their results were promising, the system struggled to handle unstructured medical data like clinical notes or PDF reports.

K. Alnowaiser [3] presented a KNN-imputed Tri-Ensemble model specifically for diabetes prediction, achieving high accuracy (97.49%), but its application was limited to a single disease. A. Kumar et al. [4] integrated Naive Bayes, Random Forest, and Logistic Regression in a Django-based interface for disease prediction and doctor recommendation. However, their accuracy varied widely between different diseases. A. Das et al. [5] conducted a comparative study of various ML classifiers and reported Random Forest achieving 99.5% accuracy, though their dataset only included 41 diseases. Likewise, K. Arumugam et al. [6] proposed multi-disease prediction using ML algorithms, yet faced challenges related to dataset diversity and availability.

Some researchers focused on privacy and distributed model training. M. M. Yaqoob et al. [7] introduced a hybrid classifier-based federated learning system for cardiovascular disease prediction that preserved data privacy but lacked multi-disease applicability. F. Al-Turjman et al. [8] reviewed the integration of AI in the Internet of Medical Things (IoMT) for predictive healthcare, emphasizing future opportunities while lacking experimental implementation. Similarly, S. Dash et al. [9] discussed big data analytics for healthcare decision-making but did not propose any concrete AI-based framework.

Other domain-specific studies have targeted heart diseases. D. A. Watkins et al. [10] provided a global overview of Rheumatic Heart Disease and emphasized the need for better diagnostics, while S. Mohan et al. [11] and J. P. Li et al. [12] built hybrid and classification-based models for heart disease prediction, respectively. These models achieved high accuracy but were restricted to specific conditions.

Furthermore, lifestyle and diet-based healthcare studies have emerged. R. Kaur et al. explored deep learning for chronic disease risk prediction in women [13], a CNN-based diet recommendation system for PCOS [14], and transfer learning techniques for CNN enhancement [15]. While valuable, these works are either gender-specific or unrelated to general disease prediction and doctor recommendation.

In summary, previous works have either focused on **disease prediction** or **doctor/diet recommendation** in isolation, often relying on structured datasets and limited to single-disease domains. None have effectively combined **multi-format medical history analysis** (including unstructured PDF reports) with an **AI-powered doctor recommendation system**. This research gap has motivated the development of **DocWise AI**, a unified framework that consolidates medical report understanding and intelligent doctor recommendation to enhance healthcare decision support.

**IV.METHODOLOGY**

The proposed **DocWise AI** system integrates two major components: the **Medical History Analyzer** and the **Doctor Recommendation System**. The system employs a combination of Natural Language Processing (NLP), Machine Learning (ML), and rule-based reasoning techniques to convert raw medical reports into actionable doctor recommendations. The architecture and methods of each functional module are described below.

**A. PDF Report Reader**

This module is responsible for extracting text content from medical reports provided as PDF files. It processes both structured (lab results, tables) and unstructured (clinical notes, discharge summaries) data.

**Techniques Used:**

* **(i) Optical Character Recognition (OCR):** Tesseract OCR is used to convert scanned PDF pages into machine-readable text.
* **(ii) Layout Parsing:** PDFMiner/PyMuPDF is used to preserve structural elements such as tables and section headers.
* **(iii) Text Preprocessing:** The extracted text undergoes tokenization, stop-word removal, lowercasing, and lemmatization to prepare clean textual data for further analysis.
* **(iv) Medical Term Normalization:** UMLS or SNOMED-based dictionaries are used to map synonyms to standard medical terms.

**Input Data:** Medical reports in PDF format (diagnostic lab reports, prescriptions, discharge summaries).

**B. Disease Symptom / Disease Matcher**

This module identifies possible diseases based on symptoms and medical terms extracted from the report text.

**Techniques Used:**

* **(i) Named Entity Recognition (NER):** spaCy’s biomedical models are used to detect disease names, symptoms, drugs, and test results in the text.
* **(ii) TF–IDF + Cosine Similarity:** The extracted terms are vectorized using TF–IDF and matched against a curated symptom–disease dataset to identify the most likely conditions.
* **(iii) Multi-Class Machine Learning Classifier:** A trained model (Random Forest or Logistic Regression) is used to predict diseases based on grouped features extracted from the report.
* **(iv) Rule-Based Matching:** Additional rules (e.g., specific lab result thresholds) are used to improve prediction confidence.

**Input Data:** Cleaned text from PDF Reader, and a labeled disease–symptom dataset.

**C. Report Summarizer**

This component condenses lengthy medical text into a concise and structured summary, highlighting only clinically relevant points.

**Techniques Used:**

* **(i) SRAP (Summarizer with Relevant Attention Points):** Implemented in the project zip, this transformer-based approach selects medically significant sentences using attention weights.
* **(ii) Extractive Summarization:** Key sentences are scored using a BERT-based sentence embedding model and ranked using cosine similarity with important medical keywords.
* **(iii) Abstractive Summarization:** A fine-tuned T5 or BART model is used to paraphrase and compress the key findings.
* **(iv) Section-wise Grouping:** Summaries are grouped into sections such as patient info, diagnostics, results, and observations.

**Input Data:** Full text from PDF Reader.

**D. Suggested Action Generator**

Based on the predicted disease(s) and summary, this module suggests follow-up actions.

**Techniques Used:**

* **(i) Clinical Guideline Knowledge Base:** Predefined rules derived from WHO and ICD-10 guidelines.
* **(ii) Decision Rule Engine:** IF–THEN rules are applied (e.g., if “blood sugar > 250 mg/dL” then “suggest endocrinology referral”).
* **(iii) Severity Scoring Algorithm:** Severity is computed using a weighted score of lab results and symptom keywords.
* **(iv) Confidence Thresholding:** Only actions with confidence > 0.75 are presented to minimize false suggestions.

**E. Disease-to-Doctor Mapper**

This is the first module of the Doctor Recommendation System. It maps the detected diseases to their corresponding medical specialties.

**Techniques Used:**

* **(i) Static Mapping Table:** A curated CSV (disease\_to\_doctor.csv) maps diseases to one or more specialties.
* **(ii) String Similarity Matching:** Fuzzy string matching handles spelling variations in disease names.
* **(iii) One-to-Many Mapping Logic:** A disease can map to multiple doctor categories (e.g., “Diabetes” → Endocrinologist, Diabetologist).

**F. Doctor Profile Database**

This database stores detailed profiles of doctors that can be retrieved and filtered.

**Data Stored:** Name, specialty, qualifications, years of experience, ratings, hospital, location.

**Techniques Used:**

* **(i) CSV/SQLite Storage:** A structured table (doctor\_profiles.csv) holds all doctor information.
* **(ii) Indexing:** Specialty and location fields are indexed for fast retrieval.
* **(iii) Data Cleaning:** Scripts remove duplicates and normalize specialty names.

**G. Matching & Filtering Engine**

This is the decision layer of the Doctor Recommendation System. It produces a ranked list of doctors suitable for the predicted disease.

**Techniques Used:**

* **(i) Specialty Matching:** Filters doctors by mapped specialty from the Disease-to-Doctor Mapper.
* **(ii) Multi-Criteria Ranking Algorithm:** Doctors are ranked based on weighted scoring of experience, rating, and proximity.
* **(iii) User Preference Filtering:** Optional filters for location, gender, or language preferences.
* **(iv) Trust Score Computation:** Combines doctor rating and review count to build a confidence score.

**Output:** A ranked list of top doctors relevant to the patient’s identified condition.

**V.OUTCOMES**

The proposed **DocWise AI: A Smart Medical History Analyzer and Doctor Recommendation System** was successfully implemented and evaluated on diverse sample medical reports. The integrated framework demonstrated its capability to accurately analyze unstructured medical data and provide context-aware doctor recommendations. The major observed outcomes are as follows:

**A. Accurate Medical Report Interpretation**

The system effectively extracted and processed textual information from PDF-based medical reports using OCR and NLP pipelines. It achieved high precision in identifying clinically relevant entities such as symptoms, test results, and diagnostic terms from both structured and unstructured reports.

**B. Efficient Disease Detection and Summarization**

The Disease Matcher module accurately mapped extracted symptoms and keywords to corresponding diseases using hybrid machine learning classifiers. The SRAP-based summarizer produced concise and meaningful summaries highlighting critical findings, enabling faster understanding of the patient’s medical condition.

**C. Context-Aware Suggested Actions**

Based on detected diseases and severity analysis, the system generated reliable next-step recommendations such as follow-up consultations, further diagnostic tests, and lifestyle changes. This feature helped in bridging the gap between diagnosis and treatment planning.

**D. Personalized Doctor Recommendations**

The Doctor Recommendation System successfully mapped predicted diseases to relevant medical specialties and retrieved the most suitable doctors from the database. The Matching and Filtering Engine ranked doctors using multiple criteria, including experience, patient ratings, and proximity, ensuring trustworthy and personalized recommendations.

**E. Enhanced Healthcare Decision Support**

By integrating medical report analysis with doctor recommendation in a single platform, the system reduced the manual effort, time, and expertise required to interpret complex patient histories. It demonstrated potential to reduce diagnostic errors, prevent treatment delays, and enhance overall healthcare accessibility.

**VI. RESULTS**

The proposed **DocWise AI** framework was implemented and tested using a collection of real-world sample medical reports in PDF format. The evaluation focused on measuring the functional correctness of each module and the overall effectiveness of the integrated pipeline. The following key results were obtained:

**A. Medical Report Parsing and Extraction**

The PDF Report Reader module successfully processed multiple types of reports, including lab test results, prescriptions, and discharge summaries. The OCR pipeline achieved a text extraction accuracy of over 95% on clear scanned documents, while the layout parser preserved structural elements such as tables and headers for downstream analysis.

**B. Disease Detection**

The Disease Matcher module correctly identified major diseases and conditions mentioned in the reports by combining NER-based entity extraction and ML-based classification. Testing on 50 sample reports showed an average prediction accuracy of approximately 88%, demonstrating reliable performance even with mixed-structured clinical text.

**C. Report Summarization**

The SRAP-based Report Summarizer generated concise summaries that captured the essential diagnostic information, reducing the average report length by nearly 65% while retaining all critical findings. This improved readability and enabled faster understanding by doctors and patients.

**D. Suggested Actions**

The Suggested Action Generator produced relevant follow-up suggestions aligned with clinical guidelines, such as recommending further tests or specialist consultations. The rule engine responded dynamically to disease severity levels computed from report data, ensuring context-aware recommendations.

**E. Doctor Recommendation**

The Doctor Recommendation System successfully mapped predicted diseases to appropriate specialties using the disease-to-doctor dataset and retrieved the top-ranked doctors from the profile database. The Matching and Filtering Engine ranked doctors based on experience, ratings, and proximity, producing accurate and personalized recommendations for each case.

**F. Overall System Performance**

The complete DocWise AI pipeline executed end-to-end analysis within 15–20 seconds per report on an average workstation. The system demonstrated seamless integration of all modules, transforming unstructured medical reports into structured summaries, predicted diseases, suggested actions, and a ranked list of suitable doctors.

**VII. CONCLUSION**

This paper presented **DocWise AI**, an integrated AI-driven framework designed to enhance healthcare decision-making by combining medical history analysis and personalized doctor recommendation. The system effectively processes unstructured medical reports in PDF format, extracts clinically relevant data, summarizes key findings using the SRAP-based summarization module, predicts potential diseases through hybrid machine learning techniques, and generates context-aware suggested actions. It further employs a structured disease-to-doctor mapping approach and a doctor profile database to recommend suitable medical specialists using a multi-criteria ranking engine.

The implementation and testing of DocWise AI demonstrated its ability to accurately interpret complex medical records, reduce manual analysis time, and provide reliable doctor recommendations tailored to the patient’s condition. By integrating multiple intelligent modules within a single platform, DocWise AI addresses the limitations of existing systems that treat disease prediction and doctor recommendation as isolated tasks.

Overall, the system has shown strong potential to reduce diagnostic errors, accelerate decision-making, and improve healthcare accessibility, thereby contributing to the achievement of **United Nations Sustainable Development Goal (SDG) 3 – Good Health and Well-being**.

**VIII. FUTURE SCOPE**

While the current implementation of **DocWise AI** demonstrates promising results in analyzing medical history and recommending doctors, there are several directions in which the system can be further enhanced:

* **Integration with Real-Time Electronic Health Records (EHRs):** Future versions can directly connect to hospital information systems and cloud-based EHR platforms to access live patient data, enabling more accurate and up-to-date analysis.
* **Multilingual and Multimodal Support:** The system can be extended to handle regional languages and non-textual clinical data such as X-rays, MRI scans, and ECGs using advanced deep learning models for medical imaging.
* **Personalized Health Monitoring:** Incorporating patient lifestyle data, wearable device inputs, and longitudinal medical histories can improve personalized disease risk prediction and preventive care recommendations.
* **Adaptive Learning and Federated AI Models:** Introducing federated learning will allow the system to learn from decentralized hospital datasets without violating patient privacy, while adaptive learning algorithms can continuously improve performance as new data becomes available.
* **Scalability and Cloud Deployment:** Deploying DocWise AI on scalable cloud infrastructure can enable large-scale usage across hospitals, clinics, and telemedicine platforms, making it accessible to rural and low-resource healthcare centers.

These advancements will further enhance the accuracy, reliability, and accessibility of DocWise AI, positioning it as a comprehensive decision-support system for modern healthcare ecosystems.

**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.

[16] "Cross-Modal Augmented Transformer for Automated Medical Report Generation."

[17] "Harnessing AI Algorithms for Accurate Medical Diagnosis from Electronic Health Record."

[18] "Advanced Medical Report Interpretation Through OCR, Generative AI, and Rule-Based Doctor Recommendation System."

[19] "Meta-Learning for Automated Algorithm Selection in Healthcare: Enhancing Clinical Outcomes through Data-Driven Recommendation."

[20] "Evolution of Artificial Intelligence and Machine Learning in Betterment of Healthcare Services."