Cure AI - AI Powered Care

Ashish Vishwakarma
Btech student, Department of IT
Inderprastha Engineering College, Ghaziabad Uttar Pradesh, India

Somay Sharma
Btech student, Department of IT
Inderprastha Engineering College, Ghaziabad Uttar Pradesh, India

Mayank Kumar Jha
Btech student, Department of IT
Inderprastha Engineering College, Ghaziabad Uttar Pradesh, India

Govind Sharma
Btech student, Department of IT
Inderprastha Engineering College, Ghaziabad Uttar Pradesh, India

ABSTRACT

CureAI is a machine learning-based disease prediction system designed to enhance early diagnosis and medical decision-making through the power of data-driven insights [1][10]. The model utilizes a Random Forest classifier to predict potential diseases based on user-input symptoms, offering a reliable and scalable solution to support healthcare professionals and patients alike [8][9]. In developing CureAI, a carefully curated dataset of symptoms and corresponding diseases was used, emphasizing data preprocessing, feature engineering, and model evaluation to ensure accuracy and generalizability [6][12].

The model achieved high classification performance by leveraging the ensemble learning nature of Random Forest, which reduces overfitting and handles high-dimensional symptom data effectively [1][10]. CureAI is deployed through a Flask-based REST API, enabling easy integration with web or mobile frontends. This architecture ensures real-time predictions with low latency and broad accessibility [13][14].

The motivation behind CureAI lies in addressing the gap between symptom occurrence and timely medical consultation, particularly in resource-limited settings. By offering immediate, preliminary diagnostic insights, CureAI empowers individuals to seek professional care proactively, potentially leading to earlier treatment and better outcomes [3][4].

The system is designed with scalability and interpretability in mind, making it suitable for extension into more complex health prediction frameworks or integration with electronic health records [2][15]. With continued improvements in dataset quality and user interface design, CureAI has the potential to serve as a vital tool in preventive healthcare, disease surveillance, and health education [14][16].

INTRODUCTION

In recent years, the integration of Artificial Intelligence (AI) in the healthcare domain has shown immense potential in improving diagnostic accuracy, speeding up decision-making, and enhancing patient outcomes [10][14]. Among various AI branches, Machine Learning (ML) and Natural Language Processing (NLP) have emerged as powerful tools for automating disease prediction and extracting meaningful insights from unstructured medical data [1][6]. As the volume of health-related data continues to grow—from patient-generated symptom descriptions to electronic health records—there is a critical need for intelligent systems that can interpret this information and assist in timely medical decisions [12][13].

Disease prediction based on symptoms is a complex task, often requiring significant clinical expertise. However, many individuals initially turn to online sources or digital tools to self-assess their conditions. This behavior highlights the demand for intelligent platforms that can bridge the gap between early symptom recognition and professional medical advice [3][4]. In this context, CureAI has been developed as a machine learning-based disease prediction system that utilizes natural language symptom descriptions to identify likely medical conditions. Unlike traditional systems that require manual symptom selection, CureAI leverages a rule-based NLP approach to extract symptoms from plain English sentences, converting them into structured inputs suitable for a Random Forest classification model [8][9].

CureAI represents a novel convergence of frontend usability, backend intelligence, and medical knowledge. Its architecture involves multiple layers, including data preprocessing, symptom extraction, vectorization, and disease classification. The model has demonstrated high predictive performance, with an accuracy of 98%, making it a promising tool for early disease detection and public health awareness [1][10].

This review paper aims to explore the development and impact of CureAI, while also situating it within the broader landscape of AI-driven disease prediction systems. The paper discusses the methodologies, existing literature, and technological underpinnings that inform the model. By comparing CureAI with other models and highlighting its unique contributions, this work provides valuable insights for researchers and practitioners looking to build or improve intelligent healthcare applications [2][15]. Moreover, it outlines the future scope and potential improvements to ensure that such tools can be effectively integrated into real-world healthcare ecosystems [14][16].

LITERATURE SURVEY

Title : **AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes**
Author : V. Jackins, S. Vimal, M. Kaliappan & Mi Young Lee

This study applies data mining and AI to predict diseases, focusing on diabetes, coronary heart disease, and cancer datasets. Using Anaconda for data analysis, Naive Bayes and Random Forest classifiers were implemented, with Random Forest demonstrating higher accuracy (up to 92.4% for cancer) after pre-processing data. This model, superior to K-means and DBSCAN, is intended for early diagnosis but is limited by high data volume processing time. Future work will test other AI algorithms to improve real-time accuracy.

Title : **Medical Chatbot: AI Based Infectious Disease Prediction Model Author :** Srujana Biradar, Dr. Swaroopa Shastri

A medical chatbot leveraging AI and NLP provides healthcare information and guidance in a conversational format. Trained using SVM on labeled symptom data, the chatbot achieved a 97.4% prediction accuracy for infectious diseases. The model's high reliability and ease of access aim to increase public access to credible medical information, showcasing the chatbot's potential to deliver essential healthcare guidance.

Title : **An AI-Based Medical Chatbot Model for Infectious Disease Prediction**
Author : Sanjay Chakraborty1, Hrithik Paul2, Sayani Ghatak2, Saroj Kumar Pandey3, Ankit Kumar3, Kamred Udham Singh4, Mohd Asif Shah5

This paper discusses developing a medical chatbot to promote healthcare awareness and provide solutions for infectious diseases. By using a deep feedforward multilayer perceptron model, the chatbot aims to enhance human interaction and lifestyle improvement. Achieving 94.32% accuracy, the model addresses the lack of guidelines for medical chatbots. With minimal loss and practical applications, this research highlights the technology's potential during health crises like pandemics, emphasizing its role in preventive care and public health awareness.

Title : **Towards a Chatbot for Medical Diagnosis Based on Patient Symptoms Author :** Kisito KABOREa, Omar SAWADOGOa,1, Yaya TRAOREa and Julie THIOMBIANO b

Focusing on LLMs, this paper develops a medical diagnostic chatbot combining algorithms like RandomForestClassifier and LogisticRegression to diagnose diseases. Inspired by LLaMA, it

generates personalized reports based on patient data. The model shows promise in enhancing healthcare accessibility, with future improvements planned through association rules to refine symptom-disease correlations. This approach aims to increase diagnostic precision and improve patient outcomes via personalized chatbot interactions.

Title : **Predictive Health Assistant: AI-Driven Disease Projection Tool**

Author : J. V. Anchitaalagammai; S. Kavitha; S. Murali; K. Lochen Gururaj; J.P. Pranav Kiruthik

This study explores AI-powered chatbots in healthcare, particularly in managing infectious diseases and supporting remote check-ups. Chatbots, operating on computer platforms, offer a solution for patients facing long doctor wait times, especially for chronic illnesses like heart disease. The paper advocates for chatbots as an efficient healthcare resource, emphasizing the potential to raise awareness and facilitate proactive health monitoring, ultimately reducing hospital visits and easing access to healthcare.

Title : **Identifying diseases symptoms and general rules using supervised and unsupervised machine learning**

Author : Fatemeh Sogandi

This study explores the application of machine learning to detect disease symptoms early, aiding in effective treatment management. Using patient data from Kaggle, the study preprocesses and analyzes data with the Apriori algorithm to identify frequent symptom patterns and rules. Various machine learning models, including stepwise regression, SVM, and boosted methods, were applied for disease prediction. Stepwise regression performed best in cross-validation. Significant decision rules were also extracted, highlighting symptom-disease relationships, which can streamline clinical applications and aid non-expert users. The proposed approach supports both healthcare professionals and patients by providing cost-effective, efficient predictive tools for diagnosing and managing diseases.

Title : **Explainable Artificial Intelligence-Based Disease Prediction with Symptoms Using Machine Learning Models**

Author : Gayatri Sanjana Sannala, K. V. G. Rohith, Aashutosh G. Vyas & C. R. Kavitha

Artificial intelligence (AI) offers transformative potential in healthcare by automating tasks and improving diagnosis and treatment efficiency. However, a key challenge is ensuring explainability, as AI decisions directly impact patient care. In disease classification, AI leverages machine learning algorithms to analyze symptoms and classify conditions, yet traditional “black box” models lack transparency, creating barriers to trust and understanding. To address this, techniques like feature importance analysis and Explainable AI (XAI) are employed, providing insights into the decision-making process. XAI enhances transparency, building trust and fostering better acceptance of AI systems among healthcare providers and patients.

Title : **Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach**

Author : Talasila Bhanuteja, Kilaru Venkata Narendra Kumar, Kolli Sai Poornachand, Chennupati Ashish, Poonati Anudeep.

Healthcare data is often stored in unstructured ways, either as paper records or consuming digital space, which limits its accessibility and usefulness. Effective data mining techniques can transform this data into valuable insights, enabling quicker, more accurate clinical decisions. This paper focuses on using models like Random Forest, Decision Tree, and LightGBM to classify diseases and identify key risk factors, offering actionable guidance for healthcare providers. Key objectives include disease classification using multiple algorithms, identifying impactful risk factors, comparing classification methods, and exploring how risk factors (e.g., diabetes influenced by hypertension) interact. These methods support proactive healthcare decisions, enhancing patient outcomes and optimizing healthcare services.

Title : **A Smart Web Application for Symptom-Based Disease Detection and Prediction Using State-of-the-Art ML and ANN Models**

Author : Parvej Reja Saleh, Eeshankur Saikia

Detecting diseases through symptoms becomes complex when symptoms vary or increase in number, making diagnosis challenging. This study utilizes a symptom-disease database, derived from patient discharge summaries, covering 149 common diseases. We developed a Machine Learning model using classification techniques like Multinomial Naïve Bayes, Logistic Regression, Decision Tree, K-nearest Neighbor, Support Vector Machine, and Random Forest for symptom-based disease detection. Additionally, a Feedforward Neural Network using MLP was implemented for precise predictions. Our smart, web-based application integrates this model, enabling users to predict diseases based on symptoms, aiming to lower diagnosis costs and save valuable time.

METHODOLOGY

CureAI is designed as an intelligent disease prediction system that processes natural language symptom inputs and returns accurate diagnostic predictions using a supervised machine learning model. The methodology consists of a multi-stage pipeline that includes data collection, natural language preprocessing, symptom extraction, vectorization, classification using a Random Forest model, and results rendering via a Flask API. The complete system emphasizes modularity, efficiency, and interpretability.

1. Data Collection and Preprocessing:

The model is trained on a curated dataset that maps symptoms to their corresponding diseases. The dataset is cleaned by removing duplicates, standardizing symptom nomenclature, and handling missing values. Label encoding is used to convert categorical disease labels into numerical format. Feature scaling is not required since the symptom data is binary (0/1 indicating absence/presence).

2. Natural Language Processing (NLP):

User inputs are typically unstructured sentences describing their health concerns. A custom-built rule-based NLP engine parses this input using Python's nltk and re libraries. The pipeline includes tokenization, lowercasing, stop word removal, and lemmatization. Keyword matching is then applied to identify known symptoms from a predefined symptom dictionary. This approach helps map diverse user expressions to standardized medical symptom terms, increasing accuracy in symptom detection.

3. Symptom Vectorization:

After symptom extraction, a binary vector of length n (where n = total number of known symptoms) is created. Each vector index corresponds to a particular symptom, marked as 1 if present in the user input and 0 otherwise. This structured format ensures compatibility with the ML model.

4. Machine Learning Model – Random Forest Classifier:

A Random Forest classifier, built using scikit-learn, is trained on the binary symptom vectors. The model leverages an ensemble of decision trees to reduce overfitting and increase prediction robustness. It outputs the most probable disease along with a confidence score. The choice of Random Forest was made due to its interpretability, high accuracy, and resilience to noisy data.

5. Backend Integration with Flask API:

A lightweight Flask API handles the interaction between the user interface and the model. It receives symptom inputs, invokes the NLP engine and classifier, and returns the predicted result to the user interface in real time.

6. Frontend Interface:

The web interface is designed using HTML, CSS, and JavaScript, providing a simple form for users to enter symptoms in plain English. The results are displayed with a clean, user-friendly layout.

This technical stack and workflow allow CureAI to function as a fast, accurate, and easy-to-use disease prediction system that can be deployed for public health applications or clinical pre-screening.

EXPECTED OUTCOME

CureAI is expected to provide an intelligent, accessible, and efficient tool for preliminary disease prediction based on user-reported symptoms in natural language. The primary outcome is to assist users in understanding potential medical conditions before consulting a healthcare professional, thereby promoting early diagnosis and health awareness [10][12].

One of the key expected outcomes is high prediction accuracy. CureAI's Random Forest model, trained on a comprehensive symptom-disease dataset, has demonstrated an accuracy of approximately 98% [1][9]. The system also delivers reliable performance in terms of precision, recall, and F1-score, indicating its ability to reduce false positives and false negatives in medical predictions [6][10].

Another major outcome is intelligent symptom understanding using NLP. The rule-based NLP system can extract meaningful symptom entities from complex or informal user input. This capability makes CureAI suitable for real-world applications, where users may not follow medical terminology but still need effective diagnosis [3][11].

CureAI is also expected to simplify access to health insights by offering a seamless web interface. Users can enter their symptoms in plain English and instantly receive feedback in the form of predicted diseases, increasing health literacy and enabling informed decisions [14][15].

Additionally, the system is designed to be lightweight and deployable, with a backend built on Flask and a clean frontend that can be embedded into healthcare portals or mobile applications. This flexibility ensures that CureAI can be used in various environments—educational, clinical, or even rural settings with limited access to doctors [13].

Overall, CureAI is not a replacement for medical professionals but serves as a supportive tool to empower users with knowledge, improve health-seeking behavior, and reduce delays in seeking professional care [2][4].

CONCLUSION

CureAI presents an innovative and practical approach to disease prediction by combining natural language processing with machine learning. The system is capable of interpreting user-inputted health complaints written in plain English, extracting relevant symptoms, and predicting possible diseases using a trained Random Forest model, as demonstrated in previous studies on AI-based disease prediction systems [1][8]. This blend of rule-based NLP and robust classification techniques enables CureAI to bridge the gap between non-technical users and complex medical diagnostic systems.

The platform has demonstrated promising performance in terms of prediction accuracy, scalability, and user interaction. Its lightweight Flask-based backend and web-friendly frontend make it highly adaptable for integration into various healthcare tools or standalone applications [13]. The symptom vectorization technique and ensemble learning strategy provide a high level of reliability even in cases where symptom descriptions vary among users, which aligns with findings from recent research on machine learning models for disease prediction [6][10].

While CureAI is not intended to replace professional medical advice, it serves as a powerful pre-diagnostic tool for awareness, early screening, and educational purposes [3][12]. With further enhancements such as multilingual support, voice-based interfaces, and larger datasets, CureAI can evolve into a more comprehensive digital health assistant. Ultimately, CureAI contributes to the growing field of AI in healthcare by promoting accessible, intelligent, and user-centric health solutions [14][15].

REFERENCE

[1] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes."

[2] S. Biradar and Dr. S. Shastri, "Medical Chatbot: AI Based Infectious Disease Prediction Model."

[3] S. Chakraborty, H. Paul, S. Ghatak, S. K. Pandey, A. Kumar, K.U. Singh, and M. A. Shah, "An AI-Based Medical Chatbot Model for Infectious Disease Prediction."

[4] K. Kabore, O. Sawadogo, Y. Traore, and J. Thiombiano, "Towards a Chatbot for Medical Diagnosis Based on Patient Symptoms."

[5] J. V. Anchitadagammai, S. Kavitha, S. Murali, L. Gururaj, and J. P. Pranav Kiruthik, "Predictive Health Assistant: AI-Driven Disease Projection Tool."

[6] G. S. Sannala, K. V. G. Rohith, A. G. Vyas, and C. R. Kavitha, "Explain- able Artificial Intelligence-Based Disease Prediction with Symptoms Using Machine Learning Models."

[7] F. Sogandi, "Identifying diseases symptoms and general rules using supervised and unsupervised machine learning."

[8] T. Bhanuteja, K. V. N. Kumar, K. S. Poornachand, C. Ashish, and P. Anudeep, "Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach."

[9] S. Dhanka and S. Maini, "Random Forest for Heart Disease De- tection: A Classification Approach," in Proc. IEEE Int. Conf. on Power, Energy, Environment and Research (ICEPEER), Dec. 2021. DOI: 10.1109/ICEPES52894.2021.9699506

[10] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," Healthcare, vol. 10, no. 3, p. 541, Mar. 2022. DOI: 10.3390/healthcare10030541

[11] X. Luo et al., "A Deep Language Model for Symptom Extraction from Clinical Text and Its Application to Extract COVID-19 Symptoms from Social Media," IEEE J. Biomed. Health Inform., vol. 26, no. 4, pp. 1737–1748, Apr. 2022. DOI: 10.1109/JBHI.2021.3123192

[12] E. Sezgin et al., "Extracting Medical Information From Free- Text and Unstructured Patient-Generated Health Data Using Natural Language Processing Methods: Feasibility Study With Real-World Data," JMIR Formative Res., vol. 7, 2023. DOI: 10.2196/43014

[13] K. V. Rao et al., "Leveraging Flask API and Machine Learning to Forecast Multiple Diseases," Commun. Appl. Nonlinear Anal., vol. 32, no. 1S, Oct. 2024. DOI: 10.52783/cana.v32.2145

[14] L. Tang, J. Li, and S. Fantus, "Medical Artificial Intelligence Ethics: A Systematic Review of Empirical Studies," Digit. Health, vol. 9, 2023. DOI: 10.1177/20552076231186064

[15] I. D. Mienye, "A Survey of Bias and Fairness in Healthcare AI," in Proc. IEEE Int. Conf. Healthcare Informatics (ICHI), Jun. 2024. DOI: 10.1109/ICHI61247.2024.00103

[16] N. Lalwani et al., "Survey on Ethical Challenges of Implement- ing AI in Healthcare," in Proc. IEEE Int. Conf. on Pervasive Ar- tificial Intelligence and Radio Intelligence (PraI), Aug. 2024. DOI: 10.1109/PRAI62207.2024.10827739