International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 01 | Jan 2025 www.irjet.net p-ISSN: 2395-0072

DIAGNOSIS OF ACUTE DISEASES IN VILLAGES AND SMALLER TOWNS USING AI

Chandrashekhar K S [1], Praveen P [2], Bhuvaneshwar C [3], Manasa H A [4], Radhika Sreedharan [5]

[1234] UG student, Dept. of Computer Science & Technology, Presidency University, Bengaluru.
[5] Assistant Professor, Dept. of Computer Science & Engineering, Presidency University, Bengaluru.

---***---

Abstract - *The rapid growth in AI-driven healthcare solutions has paved the way for advanced diagnostic tools, especially in resource-constrained environments like villages and smaller towns. Acute diseases often require timely intervention, and delays in diagnosis can have severe health consequences. This project addresses these challenges by developing an AI-based system capable of diagnosing acute diseases in under-served areas. The system leverages machine learning models trained on diverse medical datasets and offers a cost-effective, scalable solution to support healthcare providers in rural areas. The development process of this AI-based diagnostic system involves several key phases, including data collection, preprocessing, model training, system integration, and performance evaluation. A large dataset comprising patient records and disease-related data has been used to train and validate the models. This project aims to bridge the healthcare accessibility gap in underdeveloped areas, providing timely, accurate, and cost-effective diagnostic support. The AI-driven system offers a scalable solution that can be extended to diagnose a wider range of diseases in the future. By leveraging AI, the system empowers healthcare workers and reduces the burden on overstretched medical resources, ultimately leading to better health outcomes for rural populations.*

Key Words: Artificial Intelligence (AI) in Healthcare, Acute Disease Diagnosis, AI for Rural Healthcare ,Machine Learning in Disease Prediction, Natural Language Processing (NLP) for Symptom Analysis, Deep Learning in Medical Imaging, Predictive Analytics in Health, AI-based Recommendation Systems ,Symptom Mapping and Feature Engineering ,AI in Resource-Constrained Environments, Healthcare Accessibility in Rural Areas, AI-Driven Healthcare Innovations ,Explainability in AI Models, Real-Time Disease Diagnosis, AI for Healthcare Workers Support ,AI and Data Privacy in Healthcare ,Scalable Healthcare Solutions, Modular System Design in AI ,Disease Prediction Accuracy, Community Health Monitoring with AI.

1. INTRODUCTION

AI has established itself as one of the key drivers of transformation in the 21st century and is reshaping industries by providing new and efficient and data-driven solutions to some of the biggest and challenging problems global. Every industry that rely on lot of amount of data, from medicinal, finance, energy, logistical right the way to even car manufacturing has benefited from the use of AI for its data processing, pattern recognition and predicting abilities.

1.1 The Role of AI in Solving Complex Challenges

In light of this, with the steady increase in sizes and sophistication of datasets that is readily available, Machine learning models have never failed to amaze every solver of challenges that one could think could not be solved. Traditionally uses of inpaints such as pattern recognition, predictive analytics, anomaly detection, optimization, and the likes have since improved through the incorporation of AI. But to successfully apply these technologies in solving problems within a given domain, one has to get to a refined method.

For instance:
In the context of healthcare, AI models are now applied for the purpose of augmenting the accuracy of the current disease diagnosis, assessing patient prognosis and for defining selecting customized treatment plans. SDKs still are an issue for rural and underprivileged locations, leaving poor diagnosis as an area where AI can excel.

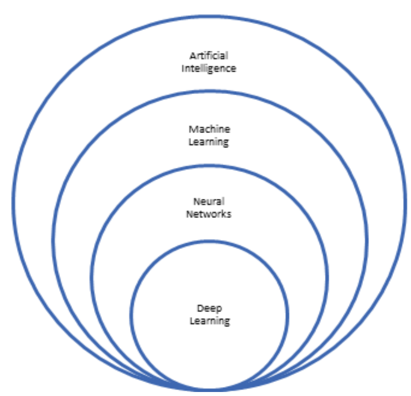

Fig. 1.1 Overview of Artificial Intelligence

The project's objectives focus on multiple aspects. First, a robust data preprocessing pipeline will eliminate inefficiencies by cleaning datasets and applying normalization techniques for optimal model performance.

Accurate disease prediction models will be developed using machine learning algorithms like Random Forest, SVM, and Gradient Boosting, with iterative improvements based on performance metrics. User-friendly web interface, accessible features include screen reader compatibility and adjustable font sizes. Performance will be optimized using predictive analytics with time series analysis and ensemble techniques to accurately predict disease trends. Secure, scalable data storage with encryption and access control for user and dataset growth. A real-time recommendation engine will give personalized health advice based on symptoms, in line with the latest medical guidelines. Administrative management will be simplified through a dashboard offering insights into system performance, user engagement, and data trends, along with features for updating datasets and monitoring activities. Early disease detection will be encouraged through regular symptom tracking and user education on recognizing initial symptoms. The model will be validated by using cross-validation techniques and analysis of the ROC curve to be used for accuracy and updates based on population health trends. This module includes a health awareness section, which provides users with educational content-articles and interactive tools-designed to guide informed choices by the user.

2. LITERATURE SURVEY

The development in Artificial Intelligence is vast and its application in healthcare sector has a vast opportunity of automated diagnosis, predictive analysis and most importantly a treatment plan. Health care remains a challenge in the rural areas, especially in the village and small town added up with problems like poor facility, insufficient staff, and late diagnose.

This paper is aimed to overcome the aforementioned challenges by means of using machine learning and technologies available at present. It is evident that structured dataset, symptom-based diagnosis and smart recommendation systems, can help to fill the health divide in rural areas. Subsequently, the study discusses the contributions, methodologies, and limitations of prior work to relate them to the project.

2.1 Existing Methods

Existing approaches in exploiting AI and mobile technology in health care are solutions that help break through diagnostic issues in remote or less privileged locations. mHealth utilizes mobile equipment to capture in real-time patient symptoms and diagnosis information. Therefore, health care can be brought to the periphery of most locations at little cost. This technology, however, has drawbacks such as depending on the networks of the mobiles and equipment capability. It will be underutilized when there is bad connectivity. AI-Based Telemedicine Systems allow remote diagnosis by analyzing patient symptoms and medical data using AI models, offering expert-level diagnosis cost-effectively while reducing the strain on healthcare infrastructure. Nevertheless, these systems require internet access and raise concerns about data privacy.

AI is used in Symptom-Based Diagnostic Systems to effectively function in places with poor health infrastructure by analyzing patient symptoms, though their accuracy depends on the quality of input data and cannot replace in-person evaluations. The AI for Rural Health Screening models are mainly for detecting common diseases such as tuberculosis and malaria among underserved populations, providing scalable and fast solutions. However, they do not perform well in complex or rare conditions because datasets are limited. Similarly, Disease Prediction via Machine Learning uses demographic and health data to predict disease trends and outbreaks, enhancing preventive care but demanding large training data for accuracy.

AI-Driven Medical Decision Support Systems reduce diagnostic errors and speed up the diagnostic process for medical professionals. They are especially useful in rural areas where there is less expertise but demand specific infrastructure and pose integration issues. Deep Learning Models for Medical Diagnosis provide precise, complex diagnosis using large sets of structured and unstructured data. However, they require much computational power and hence are challenging to deploy in low-resource settings.

In addition, AI-Based Symptom Checkers are mobile or web-based user-friendly tools that allow rural populations to self-diagnose. However, though scalable, their accuracy is very much dependent on the quality of input from users in their cases, which may lead to incorrect diagnosis in complicated cases. Lastly, Multi-modal Disease Prediction Using AI integrates different data types like text, images, and audio for complete diagnostic insight. While this approach increases accuracy, it involves complex data integration and requires large, diverse datasets for effective training.

3. PROPOSED METHODOLOGY

3.1 Flow chart

This disease diagnosis flow begins when the user initiates this process by feeding in their symptoms, either through a questionnaire, text input, or other methods. It then proceeds to evaluate the validity of those symptoms. If the symptoms are recognized and matched with the knowledge base of the system, then the flow moves to the "Predict" step, where the system uses algorithms, machine learning models, or a medical knowledge base to predict a potential disease. If the symptoms are invalid or insufficient, the process ends at the "Stop" node.

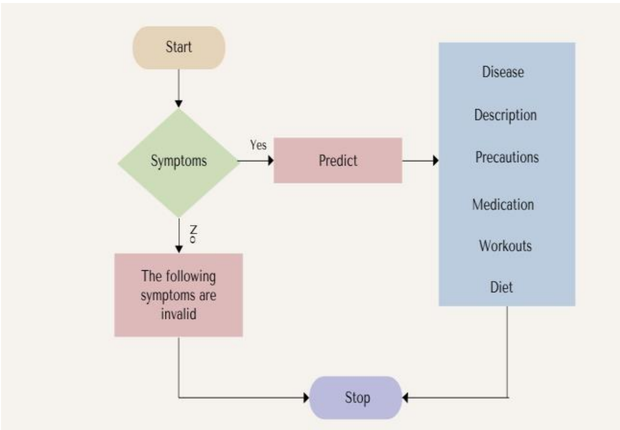

Fig. 3.1: Flow chart

If the symptoms are recognized and matched with the knowledge base of the system, then the flow moves to the "Predict" step, where the system uses algorithms, machine learning models, or a medical knowledge base to predict a potential disease. If the symptoms are invalid or insufficient, the process ends at the "Stop" node. If a disease is predicted, the system gives detailed information including a general description of the disease, its causes, precautions, recommended medications, exercises, and dietary suggestions. The process ends at the "Stop" node, signaling the completion of the diagnosis.

3.2 Data Collection and Preprocessing

This paper aims to design and develop a Personalized Medical Recommendation System that predicts the potential diseases based on user-provided symptoms. The project integrates machine learning, data science, and web development to provide a trustworthy and user-friendly healthcare tool. This system will enable users to input their symptoms and a trained machine learning model will predict the most probable disease(s) with high accuracy.

The explanation of each phase will be done in detail, focusing on model selection, working mechanisms, and evaluation processes.

3.1.1 Data Collection

The first step is gathering the right medical datasets that provide a structured mapping of symptoms to diseases, forming the foundation of the machine learning model. Sources include public medical datasets such as Kaggle's Disease Prediction Datasets and the UCI Machine Learning

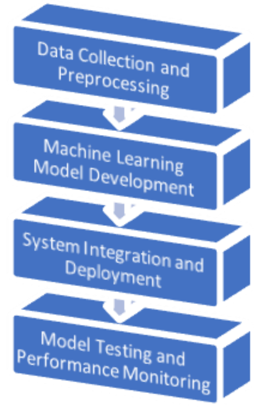

Fig. 3.2: Phases of Proposed Methodology

Repository, as well as symptom-disease relationships obtained from medical journals and published literature. In addition, manual curation involves developing symptom-disease pairs using expert knowledge to enhance the comprehensiveness of the dataset.

3.1.2 Data Preprocessing

Preprocessing ensures the data is clean, uniform, and ready for learning. Missing values are dealt with by either replacing them with "unknown" or the most frequent value, known as the mode. Also, duplicate rows are removed so that the model is not skewed, and outlier detection is made by filtering anomalous symptom-disease mappings including irrelevant symptoms.

3.1.3 Feature Engineering

Symptom encoding is the process of converting text symptom names into numerical features using methods such as label encoding, which maps a unique integer to each symptom, and one-hot encoding, which represents symptoms as binary vectors. Severity mapping categorizes diseases into levels such as low, moderate, and high. Additionally, input representation transforms symptoms into vectors that serve as input features for the machine learning model.

3.1.4 Data Splitting

The dataset is divided into training and testing sets, with 80% used to train the machine learning model and 20% reserved for validation. For consistency in performance, k-fold cross-validation with k=5 is applied, where the dataset is divided into five subsets, and the model is trained and validated iteratively on all subsets.