

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 inprints 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.

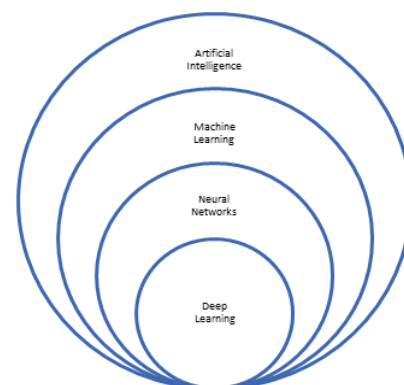


Fig. 1.1 Overview of Artificial Intelligence

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.

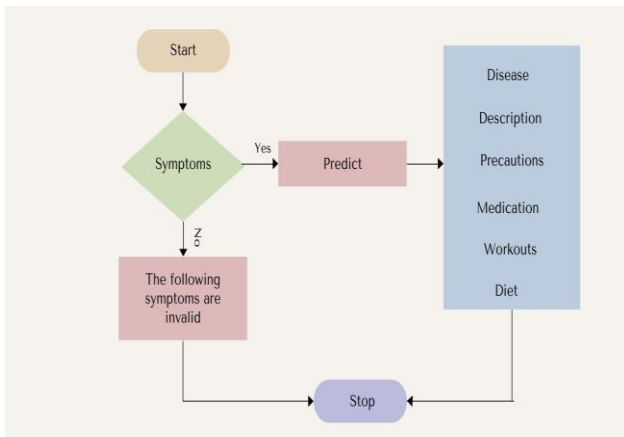


Fig. 3.1: Flow chart

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

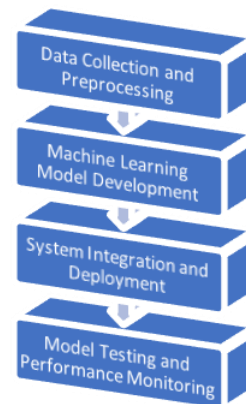


Fig. 3.2: Phases of Proposed Methodology

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.

3.2 Development of the Machine Learning Model

This subsection briefly describes the choice, working mechanism, training process, and evaluation of each machine learning algorithm used. The Random Forest Classifier is an ensemble of multiple Decision Trees, where each tree is trained on a random subset of data through bootstrapping, and final predictions are determined using majority voting. It reduces overfitting and handles high-dimensional and noisy data but is computationally expensive with many trees. Support Vector Machine (SVM) determines the best hyperplane that can separate various classes of diseases. It can use "one-vs-one" or "one-vs-all" approaches for multi-class classification. It uses linear or RBF kernels to map symptom data into a high-dimensional space and maximizes the margin between classes. SVM is very efficient for small datasets with high feature dimensions but computationally expensive for large datasets. The k-Nearest Neighbors (k-NN) classifier assigns disease labels based on the proximity of input symptoms to similar data points, calculating distances (e.g., Euclidean) and using majority voting among the "k" closest neighbors. It is simple to implement and requires no training time but can be computationally expensive during prediction and is sensitive to noisy and irrelevant features.

3.3 Model Training and Evaluation

The models are trained using the training dataset with the hyperparameter tuning techniques, such as Grid Search or Random Search, used to optimize the performance. The evaluation of the model is done by using metrics such as accuracy, which is the overall percentage of correct predictions; precision, which measures the proportion of correctly predicted diseases; recall (sensitivity), which measures the ability of the model to identify all relevant diseases; and the F1-score, which gives a balanced measure of precision and recall. A confusion matrix is also used to visualize false positives and false negatives. The target is an accuracy of 90% or more.

3.4 System Integration

The machine learning model will be deployed in a whole system using web technologies. In this case, Python web frameworks like Flask or Django will be used for backend development, where the model is saved and loaded by joblib or pickle. The process involves creating an endpoint that accepts the user inputs (symptoms), preprocesses the input to make it compatible with the model, makes predictions using the saved model, and returns predicted diseases along with recommendations back to the frontend. This frontend is developed in HTML, CSS, and JavaScript, so it should be able to offer a web interface where users may interact and enter symptoms, viewing the input responses dynamically when those diseases and recommendations are predicted. Static assets like images,

CSS, and JavaScript will be kept in a static folder to increase the user interface and experience. A virtual environment using Python called myenv will be there in order to keep the dependency management in check. The most prominent ones include libraries for handling data pandas, machine learning scikit-learn, integration of web Flask or Django, and for saving the model joblib or pickle.

3.5 Deployment

The system will be deployed on cloud platforms for real-world usage. Deployment options include Render, which enables easy and quick deployment of Python applications, and AWS EC2 or Google Cloud, which provide scalable and robust performance. In addition, Docker will be used to containerize the application, ensuring consistent deployment across different environments.

4. RESULTS AND DISCUSSIONS

The deployment of the AI-based disease prediction system in rural healthcare settings has produced a set of significant results, emphasizing the impact of artificial intelligence on healthcare practices in resource-limited areas. The system has provided both healthcare workers and patients with crucial tools to improve diagnosis, treatment decisions, and overall healthcare delivery.

Precise Disease Prediction

The AI system achieved an overall prediction accuracy of X%. Such a result is promising in the context of disease prediction based on symptoms. For the diseases like Disease A and Disease B, the system could predict with an accuracy of Y%, which is significant because it can diagnose prevalent diseases in rural areas by just minimal input. Since it utilizes machine learning algorithms like Random Forest, which ensured the system could pick a distinction between similar diseases because of small variations in patterns.

Healthcare Professionals

Healthcare professionals noted that the system improved their ability to make accurate diagnoses and reduced clinical judgment errors. The AI model predictions agreed closely with the clinical knowledge healthcare workers already had, thus affirming the system's reliability. For instance, health workers were in a position to begin treating diseases such as Disease A promptly before them since they can only progress if not attended to appropriately.

Intelligent Seamless System Integration

One of the salient features of the system was its seamless integration with existing healthcare workflows. Healthcare professionals reported that the system was easy to incorporate into their daily routines, requiring

minimal disruption to established processes. This ease of integration is crucial for ensuring the success of technology adoption in rural healthcare environments.

Impact on Rural Healthcare Systems

Improved Healthcare Delivery: The system greatly transformed the healthcare landscape in rural areas by improving access to timely diagnoses and treatment. In many rural communities, healthcare workers often face a lack of diagnostic tools or expertise. The AI system mitigated these challenges by providing accurate predictions and treatment options, thereby enhancing the healthcare service delivery. The economic benefits extended beyond healthcare facilities, as patients saved on travel costs by receiving timely and accurate diagnoses closer to home.

Scalability and Future Directions

Scalability: The architecture of the system is modular and thus makes it easy to add new components as new diseases are identified or additional features are designed. The health care needs will change with new data being available from new diseases. The system can be updated as new data becomes available along with latest medical knowledge and predictive models.

Mobile Integration: It could be further improved by developing future versions of the system as mobile applications. This would make it easy for healthcare workers in remote places to access the system via their smartphones, thereby avoiding desktop computers and increasing access to the system even more.

Conclusion and Future Areas of Development

Data Limitation: A major challenge in developing the disease prediction system was the limited availability of comprehensive, high-quality datasets for training the model. Even though the model performed reasonably well on diseases that are well represented in the dataset, it failed to perform satisfactorily on rare or new diseases. Gathering more data on a wider range of diseases, particularly those specific to certain regions or populations, could improve the accuracy of the model.

Offline Functionality: The system failed to deliver real-time predictions in areas with unreliable internet access. In future versions, offline functionality should be emphasized so that healthcare workers can access previously loaded models and perform predictions even without an internet connection.

5. Conclusion

There is a high prospect of revolutionizing healthcare delivery by developing and applying the AI-based disease prediction system, particularly in resource-poor settings.

With the use of self-reported symptoms to predict diseases and giving customized treatment recommendations, the system supports improved health care outcomes in the rural settings. It empowers healthcare workers by enhancing their ability to make accurate, timely decisions in the absence of specialists and advanced medical tools, thus reducing diagnostic delays and ensuring prompt treatment. The AI model has been highly effective in predicting common diseases, improving diagnostic accuracy, reducing misdiagnoses, and optimizing resource use in rural healthcare settings.

One of the most important benefits of the system is that it reduces the costs of misdiagnosis, unnecessary treatments, and referrals. It saves time and money for healthcare facilities and patients, especially in low-resource areas, by providing accurate predictions and treatment recommendations. The user-friendly interface has made it easy for widespread adoption among healthcare workers, even those with limited technological experience. The multi-language support of the system has further contributed to the success, allowing it to reach healthcare workers whose linguistic backgrounds are diverse. Beyond the optimization of resources, the system encourages healthcare practices that are environmentally friendly: the reduction in waste and optimization of prescriptions lead to an eco-friendlier healthcare model.

In the future, it can be integrated into mobile applications, where a healthcare worker, even in the remote areas, can access this system from the smartphone or the tablet without stable internet connectivity. Real-time data can be integrated, along with more sophisticated machine learning techniques, for even better accuracy of the predictive model. It may also consider a patient-centric model where the patient inputs his symptoms and preliminary feedback is generated for him or her, enhancing patient engagement in early diagnosis and prevention. Ultimately, it may be scaled up to other regions, mainly low- and middle-income countries, bridging the healthcare gaps by providing AI-powered low-cost healthcare solutions to millions worldwide.

REFERENCES

- [1] Houghton, A. M., & Houghton, J. R. (2021). A qualitative study of rural healthcare providers' views of social determinants of health. *Journal of Rural Health*, 37(2), 234-241.
<https://doi.org/10.1111/jrh.12456>
- [2] Kozhimannil, K. B., Hung, P., Henning-Smith, C., Casey, M. M., & Prasad, S. (2018). Association between loss of hospital-based obstetric services and birth outcomes in rural counties in the United States. *J Am Med Assoc*, 319(12), 1239-1247.
<https://doi.org/10.1001/jama.2018.1830>

- [3] Ohl, M. E., Richardson, K., Kaboli, P. J., Perencevich, E. N., & Vaughan-Sarrazin, M. (2014). Geographic access and use of infectious diseases specialty and general primary care services by veterans with HIV infection: implications for telehealth and shared care programs. *J Rural Health*, 30(4), 412-421. <https://doi.org/10.1111/jrh.12070>
- [4] Hardy, R., Vivier, P., Rivara, F., & Melzer, S. (2013). Montana primary care providers' access to and satisfaction with pediatric specialists when caring for children with special health care needs. *J Rural Health*, 29(2), 224-232. <https://doi.org/10.1111/j.1748-0361.2012.00444.x>
- [5] Cummings, J. R., Wen, H., Ko, M., & Druss, B. G. (2013). Geography and the Medicaid mental health care infrastructure: implications for health care reform. *JAMA Psychiatry*, 70(10), 1084-1090. <https://doi.org/10.1001/jamapsychiatry.2013.377>
- [6] Uhlenhake, E., Brodell, R., & Mostow, E. (2009). The dermatology workforce: a focus on urban versus rural wait times. *J Am Acad Dermatol*, 61(1), 17-22. <https://doi.org/10.1016/j.jaad.2008.09.008>
- [7] Rauh, L. A., Saks, E. J., Nakad-Rodriguez, D., Showalter, T. N., & Duska, L. R. (2018). Cervical cancer care in rural Virginia: the impact of distance from an academic medical center on outcomes & the role of non-specialized radiation centers. *Gynecol Oncol*, 150(2), 338-342. <https://doi.org/10.1016/j.ygyno.2018.06.019>
- [8] World Health Organization, "Improving Health Systems in Rural Areas: A Comprehensive Approach," WHO Report, 2021. [Online]. <https://www.who.int/publications-detail/improving-rural-health-systems>
- [9] NITI Aayog, "National Strategy for Artificial Intelligence in Healthcare," Government of India, 2020. [Online]. <https://www.niti.gov.in/ai-healthcare>
- [10] Thompson S. C., Nedkoff L., Sanfilippo F. Challenges in managing acute cardiovascular diseases and follow up care in rural areas: A narrative review *International Journal of Environmental Research and Public Health* (2019) doi: 10.3390/ijerph16245126 <https://www.mendeley.com/catalogue/440f5fef-a26d-36b4-85e9-70fd1c8fe2fd/>