International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Medical Diagnosis by AI: Disease Prediction through Symptom-Based Machine Learning

Tushar Soni[1], Shivam Yadav[2], Shreyas Banchhor[3], Shivam Dansena[4]

[1]Dept. of CSE, SSTC,
Author guide: - Prof. Durgeshwari Sahu
Coordinator: - Prof. Aakanksha Choubey
[1]roll no. – 301402221096, [2]roll no. – 301402221196, [3]roll no. – 301402221176, [4]roll no. - 301402221151

Abstract

Artificial Intelligence (AI) is transforming healthcare through predictive models that enhance diagnostic accuracy and efficiency. This study presents a symptom-based diagnosis system using a Decision Tree Classifier to predict diseases from user-input symptoms. Developed in Python with a Tkinter GUI, the system is accessible to non-technical users. Trained on a dataset containing over 130 symptoms across 41 diseases, it achieved 95.7% accuracy. The paper details the system's design, implementation, and potential for use in resource-limited healthcare settings.

I. Introduction

A. Background

The integration of AI in healthcare is revolutionizing diagnostics, especially in environments lacking resources. Traditional methods rely heavily on clinical expertise and extensive testing, which are not always available. AI-driven systems offer fast, cost-effective initial assessments.

B. Research Question

Can a machine learning model reliably predict diseases from user-provided symptoms? How can such a system be made accessible through a GUI?

C. Significance

The proposed system can democratize healthcare access, serving as an early screening tool and potentially reducing the burden on clinical services.

II. Literature Review

A. Related Work

Previous studies demonstrate success in using AI for image-based diagnostics and clinical data. Rajkomar et al. (2019) and Esteva et al. (2017) showed AI's potential in predicting patient outcomes and detecting skin cancer, respectively.

B. Theoretical Framework

This research builds on supervised learning principles, using decision tree algorithms based on entropy and Gini index for classification.

C. Research Gaps

Few existing works focus on user-accessible, symptom-based AI diagnostics, especially for use outside clinical environments.

III. Methodology

A. Design

The study uses an experimental design implemented in Python. The GUI simulates real-life usage for symptom input and diagnosis output.

B. Dataset

Sourced from public medical repositories, the dataset includes binary symptom features and disease labels. Preprocessing included normalization, imputation, and feature selection.

C. Sampling

Stratified sampling ensured balanced training and testing (80:20), preserving class distribution and minimizing bias.

D. Analysis Techniques

A Decision Tree Classifier was tuned using Scikit-learn. Evaluation metrics included accuracy, precision, recall, F1-score, and confusion matrix analysis. Brief comparisons with Random Forest and Naive Bayes models were made.

IV. Results

A. Findings

The model achieved 95.7% accuracy. It performed particularly well on diseases with distinct symptom profiles. GUI response times remained under 2 seconds.

B. Interpretation

Strong precision/recall indicates the model's reliability. Misclassifications occurred mostly with diseases sharing similar symptoms (e.g., flu vs. cold).

C. Research Question Validation

The results support the hypothesis that decision tree-based models can effectively predict disease from symptoms and offer usable diagnostic outputs.

V. Discussion

A. Result Implications

Decision trees offer interpretability and efficiency, making them suitable for home and clinical use. Real-time performance and ease of use are promising for broader deployment.

B. Literature Comparison

Unlike complex deep learning models, this work emphasizes interpretability and non-clinical usability, contributing uniquely to existing literature.

C. Limitations

The system covers a limited set of diseases and depends on user-reported symptoms. It does not handle symptom progression or comorbidities.

VI. Conclusion

A. Summary

This research delivers a high-accuracy, user-friendly AI tool for early-stage disease screening using symptoms.

B. Contributions

It introduces a practical, explainable, and accessible diagnostic model suited for non-specialist users, contributing to digital health equity.

C. Future Work

Expanding symptom and disease databases, clinical validation, mobile deployment, and integration with wearables are recommended.

References

1.**Shrey182. (n.d.). Medical Diagnosis using AI**. GitHub. https://github.com/shrey182/Medical-diagnonis-using---AI

2.**Rajkomar, A., Dean, J., & Kohane, I. (2019).** Machine Learning in Medicine. NEJM, 380(14), 1347–1358.

3.Esteva, A., et al. (2017). Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. Nature, 542, 115–118.

4.**Chaurasia, V., & Pal, S. (2014).** A Novel Approach for Breast Cancer Detection Using Data Mining Techniques. IJIRCCE.

5.Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future: Big Data and Clinical Medicine. NEJM, 375, 1216–1219.

6.Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial Intelligence in Healthcare. Nature Biomedical Engineering, 2(10), 719–731.

7.**Nguyen, P., et al. (2020). Explainable AI:** Interpreting, Explaining, and Visualizing Deep Learning. Springer.

8.**Ribeiro, M. T., Singh, S., & Guestrin, C**. (2016). Why Should I Trust You? Explaining Classifier Predictions. ACM SIGKDD.

9.**Dreiseitl, S., & Binder, M. (2005).** Do Physicians Value Decision Support? AI in Medicine, 33(1), 25–30.

10.**Kononenko, I. (2001)**. Machine Learning for Medical Diagnosis: History, State of the Art, and Perspective. Artificial Intelligence in Medicine, 23(1), 89–109.