International Journal of Science, Strategic Management and Technology
Volume **02** Issue **07** July-2026 | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Deep Learning for Symptom-to-Disease Triage to Improve Diagnosis in Rural and Underserved Communities

Tasneem Kagzi[1*], Dr.Urvashi Makwana[2] and Husen Kagdi[1]

[1*]School of Engineering, P P Savani University, Dhamdod, Kosamba, 394125, Gujarat, India.
[2]P P SAVANI HOMOEOPATHIC MEDICAL COLLEGE AND HOSPITAL, P P Savani University, Dhamdod, Kosamba, 394125, Gujarat, India.

*Corresponding author(s). E-mail(s): tasneem.kagzi@ppsu.ac.in; Contributing authors: urvashi.makwana@ppsu.ac.in; husen.kagdi@ppsu.ac.in;

https://doi.org/10.55041/ijsmt.v2i7.075

Cite this Article: Kagzi, T., Makwana, U. & Kagdi, H. (2026). Deep Learning for Symptom-to-Disease Triage to Improve Diagnosis in Rural and Underserved Communities. International Journal of Science, Strategic Management and Technology, *02*(7), 1-9. https://doi.org/10.55041/ijsmt.v2i7.075

Abstract

Many people in rural and underdeveloped places continue to face significant chal-lenges in accessing dependable, qualified medical advice. When professional help is unavailable, patients may be forced to rely on traditional home remedies or local health myths, delaying the prompt and precise diagnosis required for effec-tive treatment. We aimed to close this essential diagnostic gap by developing a practical, machine learning-driven system for symptom-to-disease prediction. Our approach is simple; it takes a user's reported symptoms and immediately generates a prioritized list of the top five most probable matching illnesses. We rigorously tested the solution using a variety of machine learning and deep learning techniques, including Random Forest, Decision Tree, Support Vector Machines (SVM), and a Deep Neural Network. Our final, optimized model achieved a robust 90% accuracy on a public dataset of disease and symptom specifications. Crucially, a clinical expert in homeopathy reviewed our model's output and validated its accuracy for delivering basic, initial medical guidance and supporting early triage decisions for patients.

Keywords: Deep Neural Network (DNN), Multi-Class Classification, Machine Learning, Symptom Analysis, Tabular Data, Feature Engineering, Symptom Profile, Homeopathy, Clinical Correlation, Disease Diagnosis

1 Introduction

People today are not as healthy as their predecessors were, thus it is crucial to treat these issues and receive a timely diagnosis. With unhealthy eating habits and consum-ing processed food, also to meet market demands the use of chemicals in agriculture have created an urgent need for timely and accurate medical diagnosis, guided by established clinical standards. [1, 2] Unfortunately, access to expert medical advice is still limited for residents of underserved and rural areas. They are frequently forced to rely on home cures and old health myths due to this crucial diagnostic gap, inevitably delaying proper treatment. While massive health awareness campaigns [3] and sig-nificant global investment [4] are positive steps, the physical challenge of delivering medical facilities to remote communities remains a persistent barrier.[5] As a result, creating dependable, easily accessible digital tools is not only beneficial but also cru-cial for early intervention, a requirement that has been brought to light by researchers who are concentrating on offering check-up models for people who cannot afford expen-sive consultation fees.[6] This context has driven extensive research into AI solutions, which range from early virtual assistants [8] to highly accurate classifiers for specific conditions like skin diseases.[9] While studies have shown promising results, such as the 89.3% accuracy achieved by Setegan et. al.,[10] the medical community strongly agrees that models must offer more than just high performance; the outputs must be transparent and explainable (XAI) so both practitioners and patients can trust the recommendation.[11] Our research directly contributes to this effort by introducing a robust deep learning-based system designed to close this diagnostic gap. We devel-oped a highly accurate model that accepts user symptoms and generates a prioritized list of the five most probable diseases. By leveraging a Deep Neural Network ensemble alongside traditional models (Random Forest, Decision Tree, and SVM), we achieved a verified 90% predictive accuracy. Crucially, a clinical expert reviewed and validated the model, confirming its practical utility for facilitating initial, basic medical triage where professional help is scarce.

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 07 July-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

2 Materials and Methods

2.1 Public Dataset and Preprocessing

Proposed model training and testing work is carried out on dataset posted by Behzad hassan which is available at Kaggle entitled as SympScan.[12] It provides 231 symp-toms details in binary format for around 100 unique diseases with around 96000 instances. It also provides post- diagnosis resources to provide details of disease, its medication and suggestions for diets, precautions and workouts. In Figure-1 shows fre-quency of each disease in dataset. Dataset was split in 80-20 % train-test split. Also Label Encoding is performed on data to convert string fields into categorical data.

2.2 Traditional Machine Learning Models

Before experimenting through Deep Learning Models, we first evaluated performance using traditional Machine learning models like Na¨ıve Bayes, Decision Tree, Random Forest, K-NN, AdaBoost and SVM Classifier were also used. All these ML models are

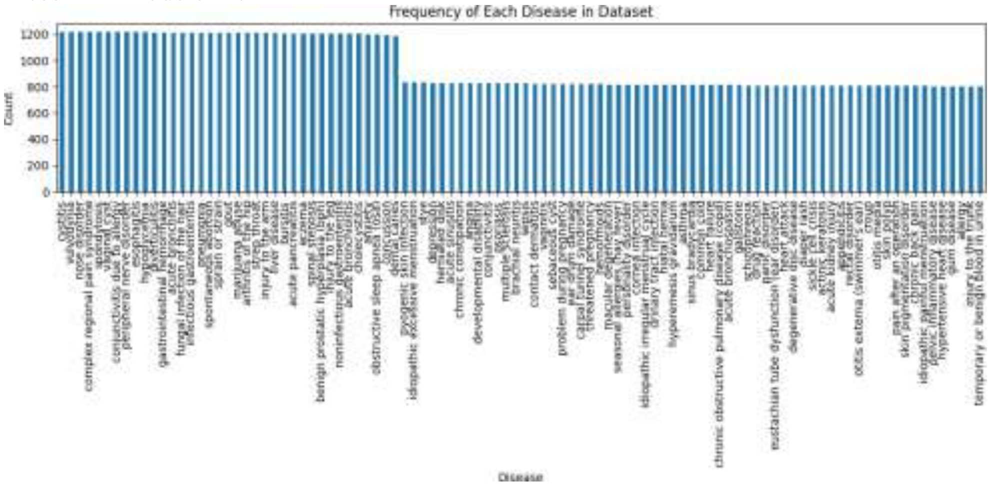


Fig. 1: Your figure caption here

most reliable and takes less time to provide results. However, there performance varies based on complexity of dataset.

2.3 Deep Neural Network

The proposed deep learning model follows a deep feed-forward neural network archi-tecture designed for multi-class disease classification. Beginning with Input layer connected to dense layer of 512 neurons with ReLU activation, followed by a 20% dropout to reduce overfitting. The network then progressively decreases in size through subsequent dense layers of 256, 128, 64, and 32 neurons, each using ReLU activa-tion. This gradual reduction in layer size enables effective feature compression while maintaining learning depth. Finally, the output layer consists of number of classes neu-rons with softmax activation, producing a probability distribution across all disease categories for accurate multi-class prediction.

3 Results

As shown in Figure-3a , the training and validation curves demonstrate excellent model performance. The Deep Neural Network (DNN) achieved rapid learning, and achieved 90% accuracy. The training and validation curves align closely which confirms that the model has strong generalization capabilities and is highly resistant to overfitting. The model depicts fast learning from beginning, curve spikes from 70% to approximately 88%. Even train and validation accuracy are stable on 88-90% The validation loss as depicted in Figure-3b stays consistently low and does not rise relative to the training loss. This confirms that the model is not overfitting and that training beyond 30 epochs would not show major improvements which indicates an efficient and optimal stopping point was reached.

Figure-2 shows Accuracy Comparison of various Machine Learning Models and Deep Learning Models. As experimented lowest accuracy is achieved by AdaBoost which is just 8.9%. Other ML based models achieved considerably good accuracy between 84 to 89%

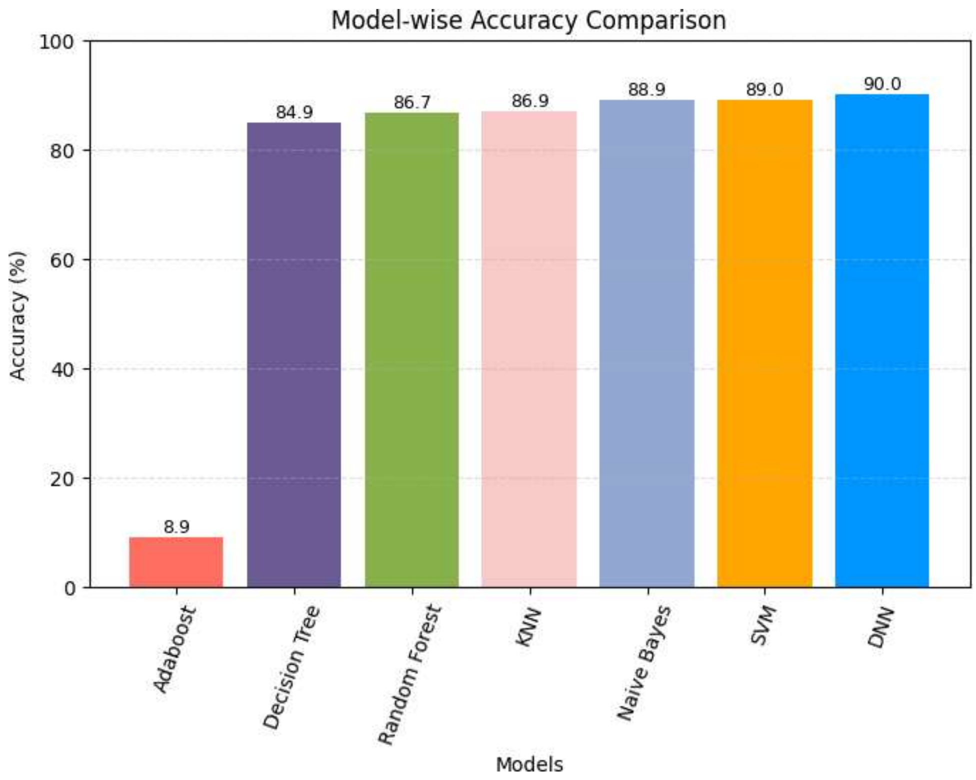
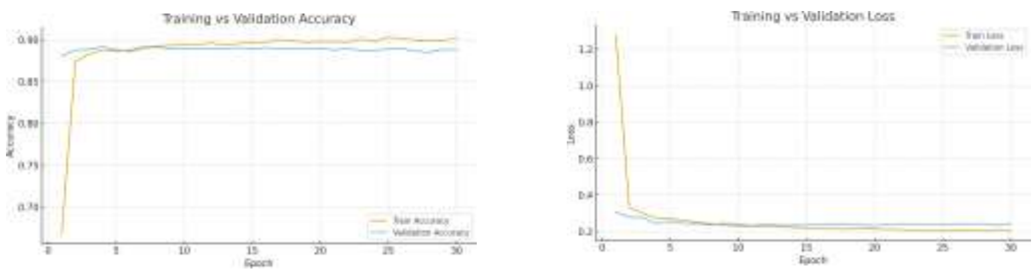


Fig. 2: Model-wise Accuracy Comparision

4 Discussion and Future Work

4.1 Performance Analysis and Data Limitations

Proposed model definitely achieved 90% accuracy. However, more dense layers and max pooling layers can be added to capture more complex non-linear relationships within the symptom data. Also, publicly available datasets have limited symptoms details and mapping between symptoms and disease states can be ambiguous or questionable. So, it is advisable to prepare a dataset by collecting data from hospitals. However, the ethical challenge of patient privacy presents a significant challenge to this data collection process.



(a) Training vs. Validation Accuracy

(b) Training vs. Validation Loss

Fig. 3: Performance evaluation curves of the proposed DNN-based Model

4.2 Clinical Relevance and homeopathic perspective

The model's clinical relevance was assessed by expert homeopathic practitioner. And as per the results they stated that from inputted symptoms, model is able to list 5 possible diseases but the correct, most probable disease was frequently found at a lower position in the list, often associated with low probability score. So, model clearly needs improvement in terms of training as well as authentic and clinically verified dataset is required for better results. Clinician has also shown their concern regarding binary values 1 and 0 to represent symptom detection. They believe that intensity of symptoms will be more prominent factor to determine disease. So clearly this states requirement for curating a dataset from scratch by referring standard practice books of medical science. As disease symptoms evolve and new pathologies emerge, regular updating of dataset is also important factor. The model can be advanced to suggest medications, precautions, diet etc. And it will be even helpful to medical practitioners if it can provide details of various blood or radiological reports required for further assessment.

5 Conclusion

This research addresses the critical diagnostic gap in rural and underserved commu-nities by developing a Deep Neural Network (DNN) system for symptom-to-disease triage. The model achieved accuracy of 90.0%, which outperforming all machine learning models. This performance validates the system's feasibility as a depend-able and easily accessible tool for initial medical triage in areas with limited access to professional care. Although, the model was unable to consistently rank the most probable disease first. The results can be directly tied to the current reliance on binary (0/1) symptom inputs. To elevate the model to a true clinical standard, future work must prioritize the ethical creation of a new, proprietary dataset that cap-tures the intensity (ordinal values) of patient symptoms. Moving forward, the system will be expanded into a comprehensive decision-support tool, offering suggestions for necessary diagnostic reports, preventative advice, and medication recommendations.

References

[1] Harrison's Principles of Internal Medicine, 22nd Edition, Vol. 1 and 2. McGraw-Hill Education, 2025.

[2] *Davidson's Principles and Practice of Medicine*, 23rd Edition.

[3] Vernon, E., Gottesman, Z., & Warren, R. (2021). The value of health awareness days, weeks and months: A systematic review. Social Science & Medicine, 268, 113553. https://doi.org/10.1016/j.socscimed.2020.113553

[4] Rakhimov, M., Akhmadjonov, R., & Javliev, S. (2022, October). Artificial intel-ligence in medicine for chronic disease classification using machine learning. In2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT)(pp. 1-6). IEEE.

[5] Khan, L., Shahreen, M., Qazi, A., Jamil Ahmed Shah, S., Hussain, S., & Chang, H.

[6] T. (2024). Migraine headache (MH) classification using machine learning methods with data augmentation. Scientific Reports,14(1), 5180.

[7] Sogandi, F. (2024). Identifying diseases symptoms and general rules using supervised and unsupervised machine learning. Scientific Reports, 14(1), 17956.

[8] Khan, M. R., Haider, Z. M., Hussain, J., Malik, F. H., Talib, I., & Abdul-lah, S. (2024). Comprehensive Analysis of Cardiovascular Diseases: Symptoms, Diagnosis, and AI Innovations.Bioengineering,11(12), 1239.

[9] Tiwari, A., Saha, T., Saha, S., Bhattacharyya, P., Begum, S., Dhar, M., & Tiwari, S. (2022). Symptoms are known by their companies: towards association guided disease diagnosis assistant.BMC bioinformatics,23(1), 556.

[10] Mohan, J., Sivasubramanian, A., & Ravi, V. (2025). Enhancing skin disease classi-fication leveraging transformer-based deep learning architectures and explainable ai.Computers in Biology and Medicine,190, 110007.

[11] Setegn, G. M., & Dejene, B. E. (2025). Explainable AI for Symptom-Based Detec-tion of Monkeypox: a machine learning approach.BMC Infectious Diseases,25(1), 419.

[12] Hassan, E., Abd El-Hafeez, T., & Shams, M. Y. (2024). Optimizing classification of diseases through language model analysis of symptoms.Scientific reports,14(1), 1507.

[13] Behzad Hassan. SympScan — Symptoms to Disease Dataset. Kaggle, 2022. Accessed from https://www.kaggle.com/datasets/behzadhassan/sympscan-symptomps-to-disease