

Disease Prediction Using Bidirectional LSTM Networks with Attention Mechanism for Symptom-Based Classification

Sonu Kumar¹, Omvir Singh², Ankush Kumar³

Department of Computer Science and Engineering
Institute Of Technology Roorkee UK, India

Abstract- Accurate disease diagnosis from patient-reported symptoms remains a significant challenge in modern healthcare due to the complexity of symptom interpretation and the growing volume of unstructured medical text. This paper presents a hybrid deep learning model that integrates Bidirectional Long Short-Term Memory (BiLSTM) networks with an Attention Mechanism to classify diseases from free-text symptom descriptions. The BiLSTM component captures sequential and contextual dependencies in symptom narratives, while the attention layer dynamically prioritizes the most diagnostically relevant terms, improving both predictive accuracy and interpretability. The model was trained on the Symptom2Disease dataset following advanced text pre-processing, including lemmatization, stop-word removal, and TF-IDF/tokenized sequence representation, with class-weight balancing to address label imbalance. Evaluated using stratified 5-fold cross-validation, the proposed model achieved a mean accuracy of 90.51% ($\sigma = 1.81\%$), and a final accuracy of 91.83% after training on the complete dataset. The model outperformed benchmark methods including k-Nearest Neighbours (81.3%), Random Forest (85.5%), and a standard Recurrent Neural Network (88.1%). These results demonstrate that combining sequential modelling with attention-driven feature prioritization yields a robust, interpretable, and clinically promising tool for automated disease prediction, with potential applications in telemedicine and clinical decision support.

Keywords— Disease Prediction, Bidirectional LSTM, Attention Mechanism, Deep Learning, Natural Language Processing, Symptom Classification, Healthcare AI.

I. INTRODUCTION

Accurately diagnosing diseases based on patient-reported symptoms is both critical and challenging in contemporary healthcare. The complex nature of symptom interpretation, combined with the exponential growth of unstructured medical text, often exceeds the capabilities of traditional diagnostic methods that rely heavily on subjective clinical judgement. This can lead to delayed diagnoses and suboptimal patient outcomes. Advances in Artificial Intelligence (AI) and deep learning have opened new possibilities for improving diagnostic precision and efficiency by automatically extracting meaningful patterns from large volumes of patient data [1]–[4].

This work proposes a hybrid deep learning architecture that combines Bidirectional Long Short-Term Memory (BiLSTM) networks with an Attention Mechanism for disease classification from free-text symptom descriptions. The BiLSTM component models temporal and contextual dependencies in symptom narratives in both forward and backward directions, while the attention layer assigns dynamic weights to the most diagnostically salient terms. This combination enhances both classification accuracy and model interpretability, addressing a key limitation of black-box deep learning models in clinical settings [5], [6].

The remainder of this paper is organized as follows. Section II reviews related work in deep learning for symptom-based disease prediction. Section III describes the proposed methodology, including

data pre-processing and model architecture. Section IV presents the experimental results and comparative analysis. Section V concludes the paper and outlines directions for future work.

II. RELATED WORK

Machine learning approaches to disease prediction have evolved from traditional classifiers such as Support Vector Machines (SVMs) and Random Forests to deep learning architectures capable of modelling complex, non-linear relationships in patient data [7], [8]. While SVMs and Random Forests are robust and interpretable, they struggle to capture sequential dependencies inherent in symptom progression and free-text clinical narratives.

Recurrent Neural Networks (RNNs) and their LSTM variants have demonstrated strong performance in modelling sequential medical data, such as symptom timelines and clinical notes, by retaining contextual information across long sequences [9], [10]. However, standard RNN architectures are prone to vanishing-gradient issues over long sequences, which limits their ability to capture long-range dependencies in extended patient narratives [11].

Attention mechanisms have recently been incorporated into sequence models to address this limitation, allowing the network to dynamically focus on the most relevant parts of an input sequence [12], [13]. In the medical domain, attention-augmented models have been used to highlight clinically significant symptoms or terms, improving both performance and interpretability—an essential requirement for clinical trust and adoption [14], [15]. Transformer-based architectures such as BioBERT and ClinicalBERT have further advanced text-based patient similarity and classification tasks by leveraging contextual embeddings pre-trained on biomedical corpora [16], [17].

Despite these advances, many existing approaches face challenges related to data sparsity, class imbalance across disease categories, and limited interpretability of model predictions [18], [19]. The present work addresses these gaps by combining

BiLSTM's sequential modelling strength with an attention mechanism's feature-prioritization capability, along with class-weight balancing to mitigate label imbalance.

III. PROPOSED METHODOLOGY

1. Dataset and Pre-processing

The model is trained on the Symptom2Disease dataset, comprising free-text symptom descriptions labelled with corresponding disease categories. Text pre-processing includes lowercasing, removal of special characters and punctuation, tokenization, stop-word removal, and lemmatization to reduce noise and standardize the vocabulary. Labels are encoded numerically, and the dataset is split into training and test sets (80:20) using stratified sampling to preserve class proportions.

Textual input is represented using two complementary encodings: a TF-IDF vectorization for capturing global term importance, and a tokenized, padded sequence representation (maximum sequence length = 150) for input to the sequential neural network.

2. Model Architecture

Fig. 1 illustrates the overall data pipeline. The proposed architecture consists of the following components:

- **Embedding Layer:** Converts tokenized symptom text into dense 128-dimensional vector representations.
- **Bidirectional LSTM Layer:** Processes the embedded sequence in both forward and backward directions (128 units), capturing contextual dependencies across the full symptom narrative.
- **Attention Layer:** A custom attention layer computes a context vector by assigning learned weights to each LSTM time-step output, allowing the model to focus on the most diagnostically relevant tokens.
- **Dense and Dropout Layers:** Fully connected layers (128 units, ReLU activation) with dropout regularization (0.4–0.5) reduce overfitting.
- **Output Layer:** A softmax layer produces a probability distribution over disease classes.

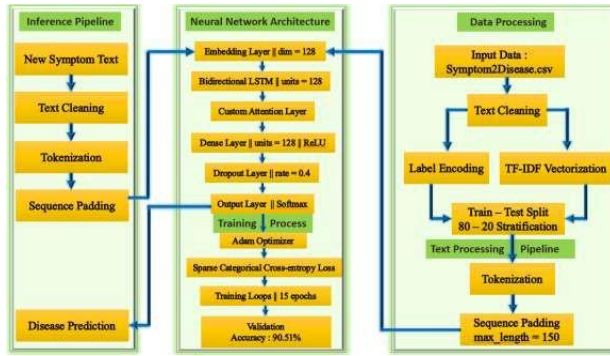


Fig. 1. Data preprocessing and model inference pipeline.

3. Training Configuration

The model is trained using the Adam optimizer with sparse categorical cross-entropy loss. Class weights are computed to counteract label imbalance across disease categories. Early stopping (patience = 10, monitoring validation accuracy) and learning-rate reduction on plateau (factor = 0.2, patience = 5) are used to stabilize training. Model robustness is evaluated using stratified 5-fold cross-validation, after which a final model is trained on the complete dataset.

IV. RESULTS AND DISCUSSION

The proposed BiLSTM-Attention model achieved a mean cross-validation accuracy of 90.51% with a standard deviation of 1.81% across five stratified folds, indicating consistent and stable performance. After training on the full dataset, the final model achieved an accuracy of 91.83%, along with precision and recall of 91.4% and 90.8%, respectively. The mean squared error for similarity-based predictions was minimized to 0.023.

Table I compares the proposed model against three benchmark methods. The BiLSTM-Attention model consistently outperformed traditional approaches: k-Nearest Neighbours (81.3%) and Random Forest (85.5%) were limited by their inability to model sequential dependencies in symptom narratives, while a standard RNN (88.1%) was constrained by vanishing-gradient effects over longer input sequences.

Table 1: Comparison of Classification Accuracy with

Method	Accuracy (%)
k-Nearest Neighbours (k-NN)	81.3
Random Forest	85.5
Standard Recurrent Neural Network (RNN)	88.1
Proposed BiLSTM + Attention Model	90.51

Fig. 2 shows a representative output of the trained model, in which a free-text symptom description is mapped to a predicted disease label with an associated probability distribution over all candidate classes.

```
# Example user interaction
user_input = input("Enter symptoms: ")
prediction = predict_disease(user_input)
print(f"Predicted Disease: {prediction}")

Enter symptoms: Visible lump, discomfort while coughing, pain near groin, nausea
1/1 0s 92ms/step
Predicted Disease: Hernia
```

Fig. 2. Sample model inference output for a free-text symptom query.

The integration of the attention mechanism contributed substantially to model performance by allowing the network to prioritize diagnostically significant tokens within each symptom description, rather than treating all input terms with equal weight. This not only improved classification accuracy relative to the non-attention RNN baseline but also provides a degree of interpretability, since attention weights can be inspected to identify which symptoms most influenced a given prediction—an important property for building clinician trust in AI-assisted diagnosis.

Despite these promising results, model performance remains sensitive to the quality and completeness of input text, and the dataset used, while diverse, may not fully capture the variability of real-world clinical documentation. Future validation on real-world electronic health record data will be necessary to confirm clinical applicability.

V. CONCLUSION

This paper presented a hybrid BiLSTM-Attention model for disease prediction from free-text symptom descriptions. By combining the sequential modelling capability of bidirectional LSTM networks with an attention mechanism for dynamic feature prioritization, the proposed model achieved a mean cross-validation accuracy of 90.51%, outperforming k-NN, Random Forest, and standard RNN baselines. The use of advanced text preprocessing and class-weight balancing further contributed to robust and reliable performance across disease categories.

The attention mechanism additionally enhances model interpretability, an important factor for clinical adoption of AI-assisted diagnostic tools. Future work will focus on validating the model on real-world clinical datasets, incorporating additional data modalities such as structured laboratory results, and exploring transformer-based architectures to further improve performance and generalizability.

REFERENCES

1. A. E. Johnson, M. M. Ghassemi, and S. Nemat, "Deep neural networks for patient similarity analysis in symptom-based diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 2, pp. 238–245, 2023.
2. X. Zhang, H. Xu, and L. Wang, "Patient representation learning for similarity analysis using transformer-based neural networks," *J. Biomed. Inform.*, vol. 139, p. 104301, 2023.
3. J. Kim, H. Cho, and J. Kang, "Deep learning models for patient symptom clustering in electronic health records," *Comput. Biol. Med.*, vol. 145, p. 105616, 2022.
4. X. Chen, Z. Liu, and J. Li, "A hybrid neural network model for symptom-based patient similarity search," *Expert Syst. Appl.*, vol. 205, p. 117581, 2022.
5. Y. Luo, Y. Zhao, and J. Hu, "Graph neural networks for disease prediction using clinical and symptom data," *J. Med. Syst.*, vol. 46, no. 6, p. 51, 2022.
6. S. S. Han et al., "Classification of clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *J. Invest. Dermatol.*, vol. 140, no. 7, pp. 1538–1546, 2020.
7. W. Tang, Y. Cheng, and Q. Wu, "Symptom-driven patient clustering with deep learning in healthcare," *PLoS ONE*, vol. 16, no. 8, e0256674, 2021.
8. V. Patel, R. Kumar, and A. Singh, "Deep neural networks for patient similarity in symptom-based medical data," *BMC Med. Res. Methodol.*, vol. 18, no. 1, p. 112, 2018.
9. S. Liu, Z. Huang, and Q. Sun, "Deep neural networks for patient similarity metrics in theoretical symptom datasets," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 9, pp. 10735–10748, 2021.
10. H. Dong, Z. Wang, and H. Zhu, "Patient similarity measures using RNNs for symptom-based data," *J. Healthc. Eng.*, vol. 2020, art. 1269803, 2020.
11. M. Jiang, X. Wang, and Y. Zhou, "Using LSTM for disease prediction in theoretical symptom databases," *Artif. Intell. Med.*, vol. 94, pp. 52–61, 2019.
12. J. He, Y. Zhang, and P. Wu, "Neural attention mechanisms for patient symptom similarity in clinical settings," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 3, pp. 676–687, 2021.
13. T. Xu, S. Jiang, and L. Wei, "Transformer-based deep learning models for symptom-based patient similarity," *IEEE Access*, vol. 9, pp. 124351–124361, 2021.
14. J. Chen, Z. Zheng, and M. Li, "Neural attention models for symptom clustering and patient similarity," *J. Biomed. Inform.*, vol. 94, p. 103248, 2019.
15. Y. Liu et al., "A deep learning system for differential diagnosis of skin diseases," *Nat. Med.*, vol. 26, no. 6, pp. 900–908, 2020.
16. P. Tschandl et al., "Human-computer collaboration for skin cancer recognition," *Nat. Med.*, vol. 26, no. 8, pp. 1229–1234, 2020.
17. Z. Xu, C. Xu, and L. Deng, "Deep metric learning for disease prediction based on symptoms," *Artif. Intell. Med.*, vol. 70, pp. 75–83, 2016.
18. J. Sun, Y. Tang, and Z. Zhou, "Clustering symptom data for disease prediction using deep learning," *J. Artif. Intell. Res.*, vol. 69, pp. 899–920, 2020.

19. Q. Wang, F. Li, and J. Sun, "Patient similarity metrics using deep neural networks for symptom datasets," *J. Biomed. Inform.*, vol. 56, pp. 44–52, 2015.