**Human Disease Prediction using Machine Learning Techniques and Real-life Parameters**

**Link:** <https://www.ije.ir/article_169090_5525e34b7bd485c6f9f9cc710f62522f.pdf>

**List of Algorithms Used:**

The research paper focuses on human disease prediction using machine learning techniques and real-life parameters. The model incorporates multiple algorithms to enhance accuracy and efficiency. The primary algorithms used in the proposed system are Random Forest (RF), Long Short-Term Memory (LSTM), and Support Vector Machine (SVM). Random Forest is utilized for symptom classification, LSTM is employed to analyze historical patient data, and SVM serves as the final classifier to verify correlations and ensure accurate predictions.

**Systems Used to Increase Accuracy:**

To increase accuracy, the researchers have employed several techniques. One of the most significant enhancements is data transformation. The dataset, sourced from Kaggle, was modified by assigning weights to rare symptoms to give them higher importance during classification. Additionally, geographical filtering was applied, meaning that diseases that do not commonly occur in a specific location were eliminated from the prediction process. These modifications helped refine the model’s ability to diagnose diseases based on symptoms while considering real-world variations.

**Methods to Improve Accuracy:**

The Random Forest algorithm plays a crucial role in the model. It works by generating multiple decision trees from different subsets of data, with each tree voting on the disease prediction. The final decision is made based on majority voting. Random Forest is preferred due to its ability to handle both categorical and continuous variables, making it highly effective for classification problems. Meanwhile, LSTM enhances the model by analyzing time-series data, allowing it to consider the patient’s historical health records. This makes the predictions more stable and reliable over time. Once both RF and LSTM have provided their predictions, SVM acts as the final decision-making layer, ensuring that the results from RF and LSTM are not only correlated but also causally linked to the input symptoms.

**Limitations of the Model:**

Despite the improvements, the model has some limitations. The execution time is high due to the complexity of the Random Forest and LSTM algorithms, which require significant computational resources. Additionally, overfitting can occur, especially when the model is exposed to noisy data. Another limitation is the lack of extensive time-series data; while LSTM is capable of learning from patient history, its accuracy would improve if more historical data were available.

**Future Scope:**

Looking ahead, there are multiple future applications for this model. It can be further improved by incorporating a larger variety of symptoms and rare diseases, making it applicable to a wider range of medical conditions. The enhancement of time-series data could allow LSTM to provide even more precise predictions by analyzing long-term patient records. Additionally, real-time implementation of this model in healthcare facilities could automate preliminary disease diagnosis, reducing the workload on doctors and making medical services more accessible to patients.

**Conclusion:**

This research presents an effective machine-learning approach to disease prediction, leveraging a combination of Random Forest, LSTM, and SVM. By refining data preprocessing techniques and incorporating historical patient data, the model achieves higher accuracy compared to previous methods. While there are still challenges to address, such as computational efficiency and data availability, the proposed system holds great potential for improving healthcare diagnostics and automation.