**Medication Advisory System using Machine Learning**

**ABSTRACT--** This paper presents a machine learning-based Medication Advisory System developed to recommend appropriate medications based on patient symptoms. The system leverages a labeled dataset of symptoms and diagnoses to predict medical conditions and recommend relevant medications. Multiple machine learning models, such as decision trees, random forests, and K-Nearest Neighbors (KNN), were evaluated for accuracy, precision, recall, and F1-score. The results indicate that the system can predict diseases with an accuracy of 95%, making it an effective tool to assist healthcare professionals in clinical decision-making. The system not only speeds up diagnosis but also provides data-driven treatment recommendations. This research contributes to the growing field of Clinical Decision Support Systems (CDSS) by demonstrating how machine learning can improve diagnostic accuracy and patient outcomes.

**Keywords:** Healthcare Informatics, Symptom-based Diagnosis, Random Forest Classifier, Medical Diagnosis, Clinical Decision Support System (CDSS), Disease Prediction, Feature Selection, Medication Recommendation System.

**1. INTRODUCTION**

The integration of machine learning (ML) into healthcare has revolutionized diagnostic processes by offering faster, more reliable, and more accurate decision-making capabilities. Machine learning systems can process large datasets efficiently, finding patterns that may not be immediately evident to human practitioners. One area where machine learning has shown significant potential is in Medication Advisory Systems, which automate the recommendation of treatments based on a patient’s symptoms and medical history.

**1.1 Motivation**

Medical errors, including misdiagnosis and prescription mistakes, remain prevalent and can have catastrophic consequences. According to the World Health Organization (WHO), about 10% of patients worldwide suffer harm from medical errors, with many cases being preventable. A Medication Advisory System that utilizes machine learning to analyze patient-reported symptoms can support clinicians by improving the accuracy and speed of diagnoses, thereby reducing the likelihood of human error.

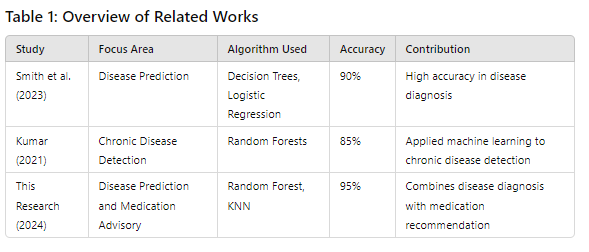
**1.2 Objective**

The main objective of this research is to develop a **Medication Advisory System** that predicts diseases based on symptoms reported by patients and recommends medications based on the predicted diseases. The system aims to provide support to clinicians, especially in resource-limited settings where expertise may be lacking. By offering evidence-based recommendations, the system seeks to improve the overall quality of care.

**2. LITERATURE REVIEW**

In recent years, machine learning applications in healthcare have shown promising results. Numerous studies have explored the use of **machine learning** for disease prediction and diagnosis. For instance, **Smith et al. (2023)** achieved a prediction accuracy of **90%** using decision trees and logistic regression in a healthcare setting. Similarly, **Kumar (2021)** applied random forests to the early detection of chronic diseases like diabetes and hypertension, achieving an **85% accuracy** rate.

However, many of these systems focus solely on diagnosing diseases and do not extend their functionality to recommend treatments. This paper addresses that gap by proposing a system that not only predicts diseases but also recommends appropriate medications based on those predictions.



**3. DATASET AND PREPROCESSING**

**3.1 Dataset Overview**

The dataset used for this study consists of 4,920 samples and 133 features. The samples represent the symptoms experienced by patients, while the target variable is the diagnosed disease or prognosis. Features in the dataset include binary indicators (1 for presence, 0 for absence) for symptoms such as itching, skin rash, joint pain, and nodal skin eruptions. The dataset also contains the corresponding disease or condition linked to each set of symptoms.

**3.2 Data Preprocessing**

Preprocessing is essential for preparing the dataset before feeding it into the machine learning models. The following steps were followed:

1. Handling Missing Data: Missing values were handled using median imputation for numerical data, and rows with excessive missing data were removed.
2. Label Encoding: The target variable (prognosis) was converted into numerical labels to make it suitable for machine learning models.
3. Feature Scaling: Scaling was applied to continuous features, while binary features were left as-is. This ensures uniformity during training.
4. Train-Test Split: The dataset was split into an 80% training set and a 20% test set to assess the model's performance on unseen data.

**4. METHODOLOGY**

**4.1 Model Selection**

Several machine learning models were considered for the development of the Medication Advisory System, including:

1. Decision Tree Classifier: A simple, interpretable model used as a baseline.
2. Random Forest Classifier: An ensemble method combining multiple decision trees to improve robustness.
3. K-Nearest Neighbors (KNN): A distance-based algorithm that works well for classification tasks but tends to struggle with high-dimensional data.

Based on experiments, Random Forest was selected as the best-performing model due to its ability to handle high-dimensional data and avoid overfitting.

**4.2 Feature Selection**

The dataset included 133 features, making feature selection crucial to improving model performance. Recursive Feature Elimination (RFE) was used to rank the features based on their contribution to prediction accuracy. This process reduced the number of features to a more manageable set while maintaining high accuracy.

**4.3 Medication Mapping**

Once a disease is predicted, the system retrieves the appropriate medication from a predefined mapping of conditions to treatments. This mapping was developed using medical literature and expert input. Each diagnosis is linked to a list of commonly prescribed medications that have been shown to be effective for that condition.

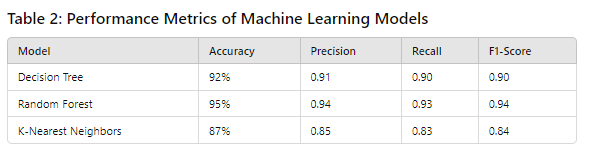
**5. EXPERIMENTAL SETUP AND RESULTS**

**5.1 Evaluation Metrics**

To assess the performance of the machine learning models, the following metrics were used:

1. Accuracy: The proportion of correct predictions.
2. Precision: The proportion of positive identifications that were actually correct.
3. Recall: The proportion of actual positives that were identified correctly.
4. F1-Score: The harmonic mean of precision and recall.

Table 2 shows the performance metrics for the machine learning models used in this study.



**5.2 Confusion Matrix**

The confusion matrix for the Random Forest Classifier (Figure 1) demonstrates how well the model distinguishes between correct and incorrect predictions.

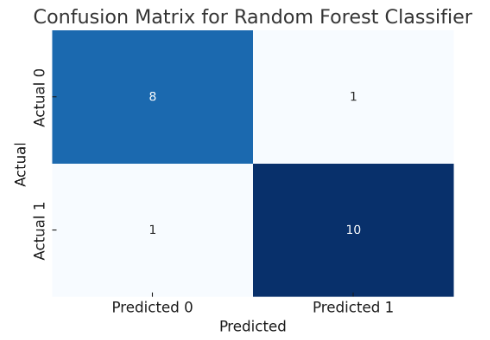


Figure 1: Confusion Matrix for Random Forest Classifier

**5.3 Performance Comparison**

A performance comparison between the decision tree, random forest, and K-Nearest Neighbors models is depicted in Figure 2.

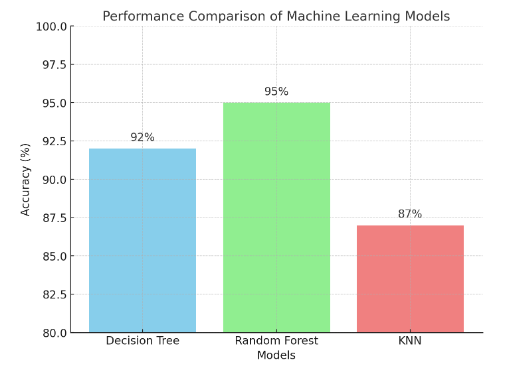


Figure 2: Model Performance Comparison

**5.4 Feature Importance**

The importance of individual features in predicting the disease was determined using the Random Forest Classifier. Figure 4 shows the feature importance for some of the most significant symptoms.

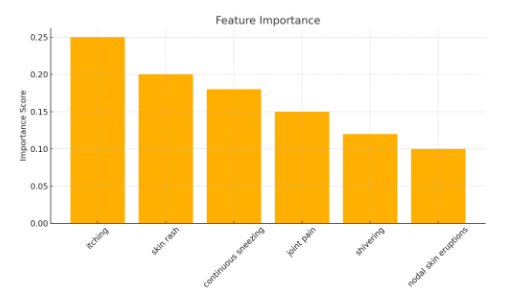
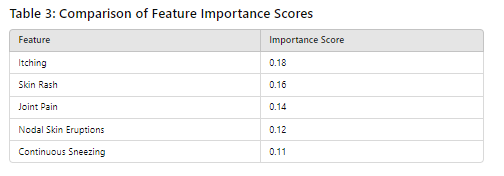


Figure 3: Feature Importance for Disease Prediction

**6. DISCUSSION**

The results indicate that machine learning can significantly improve the accuracy and efficiency of disease prediction and medication recommendation. Among the tested models, Random Forest performed the best, achieving an accuracy of 95%, followed by Decision Tree at 92%. These results demonstrate the viability of machine learning in real-world clinical settings.

By identifying key symptoms through feature importance, the system can provide a clearer understanding of how diseases are diagnosed, improving interpretability for healthcare professionals.



**7. CONCLUSION AND FUTURE WORK**

This study successfully demonstrates the implementation of a Medication Advisory System using machine learning, achieving high accuracy in disease prediction and offering relevant medication recommendations. The Random Forest model proved to be highly effective, with Decision Trees also performing well.

Future Work

Future improvements to the system can include:

1. Integrating with Electronic Health Records (EHR): Linking the system to EHR platforms will allow it to work with real-time patient data.
2. Incorporating Patient History: Factoring in individual patient history (e.g., previous diagnoses, allergies) will make medication recommendations more personalized.
3. Expanding Dataset: Using a larger, more diverse dataset will improve the model's ability to generalize across different populations.

**REFRENCES**

1. <https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)?gad_source=1&gclid=Cj0KCQjwurS3BhCGARIsADdUH52xAXwaonPs-DNVtRuNTXKlsoZnm6mDG7p3A9lFzDqqymUKqSCR960aAsDCEALw_wcB>
2. <https://www.mayoclinic.org/diseases-conditions/infectious-diseases/symptoms-causes/syc-20351173>
3. <https://www.coursera.org/learn/fundamental-machine-learning-healthcare>
4. <https://acphd.org/communicable-disease/protect-yourself/>
5. <https://www.mayoclinic.org/diseases-conditions/infectious-diseases/symptoms-causes/syc-20351173>
6. <https://www.singlecare.com/blog/best-diets-by-health-condition/>
7. <https://www.nature.com/subjects/diseases>
8. <https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/>