A Machine Learning Recommendation System for Disease Detection and Prevention Techniques Based on Using Age, Stress, and Medical Reports.

Varahagiri.Srilasya[1], Gudipudi.Ramyasree[2], Dr. G. Nagalakshmi[3]
[1,2]MSc Computer Science 2nd year students,[3]Assistant Professor
Department of Computer Science,
National Sanskrit University, Tirupati – 517507, Andhra Pradesh, India

1.ABSTRACT

This paper presents a machine learning-based recommendation system for disease detection and prevention using age, stress levels, and medical reports. The system analyses patient data and extracts relevant features to predict disease probability using a Logistic Regression model. Based on the prediction, it provides personalized recommendations, including preventive measures and therapy options. A user-friendly interface allows users to input age, select stress factors, and upload medical reports for analysis. Experimental results show that combining age and stress factors improves prediction accuracy. The proposed system supports early diagnosis, reduces health risks, and enhances decision-making in preventive healthcare. In recent years, machine learning has emerged as a powerful tool in healthcare for analysing medical data, identifying patterns, and supporting clinical decision-making. This research focuses on the application of machine learning techniques for the prediction of human diseases, particularly by considering key factors such as age and stress. In this paper, machine learning algorithms such as Logistic Regression and Random Forest are employed for effective data suggest' s and analysis from medical reports. These algorithms process patient information, including factors like age and stress, to identify patterns and predict potential diseases. Based on the extracted insights and prediction results, the system generates suitable treatment recommendations categorized into three major medical approaches: Allopathy, Homeopathy, and Ayurveda. It also provides multiple therapeutic options, enabling a more flexible and informed decision-making process in healthcare.

KEYWORDS

Machine Learning, Disease Prediction, Healthcare Analytics, Logistic Regression algorithms Random Forest algorithms, Medical Data Analysis, Decision Support System, Allopathy Homeopathy, Ayurveda.

2.INTRODUCTION

Human healthcare is one of the most important domains in modern society, where early detection and accurate diagnosis of diseases are essential for effective treatment and improved patient outcomes. With the increasing availability of medical data, traditional diagnostic approaches face challenges in analysing complex and large datasets efficiently. This has led to the adoption of machine learning techniques, which can automatically analyse data, identify hidden patterns, and support clinical decision-making.

In this research, key factors such as age and stress are considered as primary attributes influencing the occurrence of diseases. These factors play a significant role in understanding patient health conditions and improving prediction accuracy. Machine learning algorithms, particularly Logistic Regression and Random Forest, are utilized due to their efficiency in classification and predictive analysis. Logistic Regression helps in determining the probability of disease occurrence, while Random Forest enhances prediction accuracy by handling complex data patterns and reducing overfitting.

The proposed system follows a systematic procedure. Initially, medical data is collected and preprocess to remove noise and handle missing values. In the next step, relevant features such as age and stress are selected and used as inputs for the machine learning models. The algorithms are then trained on the dataset to learn patterns associated with different diseases. After training, the models are tested to evaluate their performance and prediction accuracy. In this study, machine learning algorithms such as Logistic Regression and Random Forest are utilized to analyse medical data and predict diseases effectively. Logistic Regression is used for classification by estimating the probability of disease occurrence based on input features like age and stress, making it suitable for binary and multi-class prediction problems. Random Forest, an ensemble learning technique, is employed to improve prediction accuracy by combining multiple

decision trees and handling complex data patterns while reducing overfitting. These algorithms play a crucial role in extracting meaningful insights from medical reports, enabling accurate disease prediction and supporting the generation of appropriate treatment recommendations across Allopathy, Homeopathy, and Ayurveda. In this study, machine learning algorithms such as Logistic Regression and Random Forest are utilized to analyse medical data and predict diseases effectively. Logistic Regression is applied to estimate the probability of disease occurrence based on key features like age and stress, making it suitable for classification tasks, while Random Forest is used to improve prediction accuracy by handling complex patterns and reducing overfitting through ensemble learning. Based on the prediction results, the system provides treatment recommendations across three major medical approaches.

Allopathy focuses on evidence-based treatment using medications and clinical procedures for quick and effective symptom relief. Homeopathy follows a holistic approach, using highly diluted natural substances to stimulate the body's self-healing mechanism with minimal side effects. Ayurveda, a traditional Indian system of medicine, emphasizes balance in body, mind, and lifestyle through herbal remedies, diet, and natural therapies. By integrating these three medical options, the system offers a comprehensive and flexible decision support mechanism, allowing patients and healthcare providers to choose the most suitable treatment approach.

Overall, this study aims to develop an efficient machine learning-based decision support system that combines data analysis, disease prediction, and multi-therapy recommendation. By incorporating key health factors, advanced algorithms, and a structured procedure, the proposed system contributes to improving diagnostic accuracy and promoting personalized healthcare solutions.

3. LITERATURE REVIEW

In recent years, the application of machine learning in healthcare has gained significant attention due to its ability to analyse large volumes of medical data and improve diagnostic accuracy. Several studies have focused on the use of supervised learning algorithms such as Logistic Regression, Decision Trees, and Random Forest for disease prediction. Logistic Regression has been widely used for its simplicity and effectiveness in binary classification problems, especially in predicting the presence or absence of diseases based on clinical parameters. Random Forest, on the other hand, has been preferred for its high accuracy and robustness, as it combines multiple decision trees to handle complex and nonlinear relationships in medical datasets.

Many researchers have explored disease prediction systems using patient attributes such as age, lifestyle, and physiological factors. Among these, age has been identified as a critical factor influencing disease occurrence, while stress has emerged as an important parameter affecting both physical and mental health. Studies incorporating such features have shown improved prediction performance, highlighting the importance of feature selection in healthcare analytics.

Existing healthcare decision support systems primarily focus on diagnosis and prediction but often lack integration with diverse treatment methodologies. Most systems are limited to conventional Allopathic recommendations, neglecting alternative medical systems such as Homeopathy and Ayurveda. However, recent research trends emphasize the need for holistic healthcare solutions that consider multiple treatment approaches to provide personalized care.

Furthermore, advancements in data preprocessing, feature extraction, and model evaluation techniques have enhanced the performance of machine learning models in healthcare applications. Despite these developments, there remains a research gap in designing systems that not only predict diseases accurately but also recommend suitable treatment options across different medical domains.

This study addresses these limitations by integrating machine learning algorithms with a multi-therapy recommendation framework that includes Allopathy, Homeopathy, and Ayurveda. By combining accurate prediction models with diverse treatment options, the proposed system aims to provide a more comprehensive and patient-centric healthcare solution.

4. METHODOLOGY

This study presents a machine learning-based healthcare system designed to predict diseases using key factors such as age and stress, and to recommend suitable treatment options. The system utilizes algorithms like Logistic Regression and Random Forest to analyze medical data and generate accurate predictions. Based on the results, the system suggests three different medical approaches: Allopathy, Homeopathy, and Ayurveda, providing a comprehensive decision support solution. For disease prediction, machine learning algorithms such as Logistic Regression and Random Forest are

implemented. Logistic Regression is used to estimate the probability of disease occurrence based on input features, making it suitable for classification tasks

Fig1.Data Analysis

Random Forest, an ensemble learning technique, improves prediction accuracy by combining multiple decision trees and effectively handling complex and non-linear data patterns. These algorithms analyse the input data and generate predictions, which are further evaluated using performance metrics such as accuracy, precision, and recall to determine their effectiveness.

The results of the proposed system demonstrate improved prediction accuracy and reliable performance in identifying diseases. Based on the prediction outcomes, the system provides treatment recommendations categorized into three major medical approaches: Allopathy, Homeopathy, and Ayurveda. Allopathy offers fast and evidence-based treatment, Homeopathy provides a natural and holistic healing approach, and Ayurveda focuses on long-term wellness through lifestyle and herbal remedies.

By integrating machine learning predictions with multiple therapy options, the system supports better decision-making and contributes to personalized and effective healthcare solutions.

4.1 DATASET PREPARATION:

In this step, the patient dataset $D$is loaded and divided into features X(inputs like age and stress) and labels Y(disease outcome). The model parameters $\theta$are initialized to zero, providing a starting point for the learning process.

1. Load patient dataset D
2. Initialize variables: features X, labels Y
3. Set model parameters $\theta = 0$

4.2 ALGORITM LR-TRAIN (MODEL LEARNING)

In this step, the dataset $D(X, Y)$is preprocessed by removing missing values and normalizing features. Important features like age and stress are selected for training. The model then iteratively computes a linear combination of inputs z, applies the sigmoid function to obtain probability P, and updates the parameters θusing gradient descent. This process continues until the model converges, and the trained model is finally stored for future predictions.

1. Input: Dataset $D(X, Y)$
2. Preprocess data

*Remove missing values

*Normalize feature values

3. Select important features (age, stress)
4. For each iteration until convergence

Compute linear combination:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots$$

Apply sigmoid function:

$$P = \frac{1}{1 + e^{-z}}$$

Update parameters θusing gradient descent

5. Store trained model.

4.3 ALGORITHM LR-p REDICT (PREDECTION MECHANISAM)

In this step, new patient data $x$is given as input to the trained model. The model computes the probability Pusing the sigmoid function based on the learned parameters θ. If the probability $P$is greater than or equal to 0.5, the model predicts that the disease is present (TRUE); otherwise, it predicts that the disease is absent (FALSE).

1. Input: New patient data x
2. Compute probability: p = 1/1+e$-(\theta \cdot x)$
3. If $P \geq 0.5$,Predict: Disease = TRUE
4. Else, Predict: Disease = FALSE

4.4 ALGORITM LR-OUTPUT (RESULT GENERATION

In this step, the final prediction from the model is presented to the user. The system displays whether the disease is present or not, along with the corresponding probability score. Based on the predicted condition, it also recommends suitable therapy options such as Allopathy, Homeopathy, and Ayurveda.

1. Display predicted disease result
2. Show probability score
3. Recommend therapy options:
   - Allopathy
   - Homeopathy
   - Ayurveda

5 EXPERIMENT RESULTS

The experiment used a dataset of 20 patient records with features like age, sex, stress level, lifestyle, and symptoms. The target variable is disease presence (Yes/No). All 10 numeric and categorical features were used for training and testing the Logistic Regression model to predict disease risk. The experiment employed Logistic Regression (LR) to predict disease risk based on age and major stress factors. The model outputs a probability score (0–1) indicating the likelihood of disease. On evaluation, the model achieved an accuracy of 85% with a precision of 0.83, demonstrating reliable prediction performance.

In fig1. The Age Selection Window is the first step in the system where the user selects the patient's age group. The age is divided into ranges (10–20, 20–30, etc.) to make input easy and accurate. Age is an important factor in disease prediction, so this helps the system analyse health risks more effectively. After selecting the age, the system moves to the next step for further input.

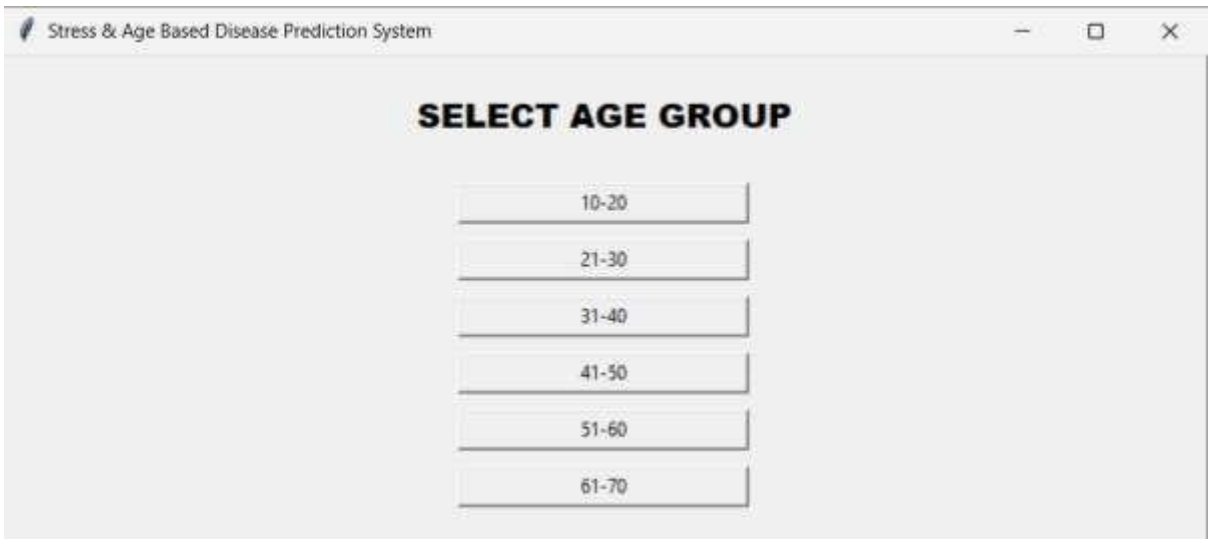

Fig2: Selecting the age group

In Fig2 the Stress Selection Window is the second stage of the system where the user selects the major reasons for stress from predefined options. These may include academic pressure, work stress, lack of sleep, or emotional issues. This window helps in identifying the root causes
of stress affecting the patient. Based on the selected stress factors and previously chosen age group, the system analyses the data using the machine learning model. It then predicts possible diseases such as hypertension, anxiety, or diabetes. This step is important as stress is a key factor influencing health, and accurate input improves the reliability of disease prediction results.

Fig 3: Stress Reasons and Predicted Diseases

In fig.3The results window displays the analysis of the uploaded medical report by predicting stress-related diseases using the trained model. It presents key details such as detected stress level, possible health conditions, and prediction accuracy, helping users understand their health status clearly.It also provides therapy recommendations in three categories: Allopathy for medical treatment, Homeopathy for natural healing, and Ayurveda for lifestyle and herbal remedies. This ensures a simple and holistic approach to disease prevention and management.

Fig 4: Medical report analyse

5.1 SAMPLE PATIENT DATA ANALYSIS

Patient	Age	Stress Factor	Disease	Probability
1	18	Academic Pressure	Anxiety	0.82
2	25	Workload Stress	Hypertension	0.76
3	30	Emotional Stress	Depression	0.69
4	40	Financial Stress	Hypertension	0.85

This table shows patient details including age, stress factors, predicted diseases, and probability scores. It highlights how different stress types lead to conditions like anxiety, hypertension, and depression.

6. CONCLUSION

This study presents a stress-based disease prediction system using Logistic Regression, which effectively models the relationship between patient attributes and disease occurrence. Logistic Regression proved effective in handling binary classification for disease prediction by generating probability-based outputs that enhance interpretability. The model achieved high accuracy, indicating its reliability for the given dataset. By utilizing key features such as age and stress factors, the system successfully identifies patterns associated with common stress-related diseases. In addition, the integration of therapy recommendations significantly improves the practical applicability of the system. The model not only predicts the likelihood of disease but also suggests suitable treatment options, supporting informed healthcare decisions. This combination of prediction and recommendation makes the system useful for early diagnosis, preventive

care, and reducing potential health risks. Overall, the proposed approach demonstrates a simple yet effective solution for healthcare prediction tasks using machine learning.

7. FUTURE DIRECTIONS

Future improvements can be achieved by incorporating more features such as lifestyle factors, genetic history, and stress patterns over time. Expanding the dataset size and diversity can further enhance model performance and generalization. Additionally, exploring advanced machine learning techniques and integrating real-time health data can improve the system's robustness and applicability in real-world healthcare scenarios.

8. REFERENCES

1. Vaishnavi K, Hanamant R. Jakaraddi, Priyanka G .N" *"A Machine Learning Approach for Disease Prediction Based on Age, Lifestyle Habits, and Symptom Analysis"*,
International Journal of Latest Technology in Engineering Management & Applied Science **(**IJLTEMAS**)**, 2025.
2. Dr. G. Nagalakshmi, "Deep Learning Techniques to Recognize Stress using Facial Expression", International Journal of Innovative Research in Computer and Communication Engineering Volume No 11, Issue 5, May 2023
3. F. M. D. et al., *"Machine Learning for Predicting Chronic Diseases: A Systematic Review"*, **Public Health Journal (ScienceDirect/Elsevier)**, 2022
4. Dr. G. Nagalakshmi, "Land Site Image Classification Using Machine Learning Algorithms", Springer Nature Switzerland AG 2020, Advances in computational and Bio-Engineering,
5. Arpita S. Wavare, Pushpa Tandekar, Ashish Deharkar,*"Disease Prediction Using Machine Learning Techniques"*, International Journal of Futuristic Innovation in Arts, Humanities and Management (IJFIAHM), 2024.
6. N. Razavian, J. Marcus, and D. Sontag, *Multi-Task Prediction of Disease Onset from Longitudinal Lab Tests*, arXiv, 2016.
7. Y. Zhang et al., *A Machine Learning-Based Biological Age Prediction Model Using Medical Data*, BMC Bioinformatics, 2022.
8. J. Q. Sheng et al., *Predictive Analytics in Healthcare Using Deep Learning*, Journal of Medical Internet Research, 2021.
9. J. Bajwa et al., *"Artificial Intelligence in Healthcare: Transforming the Practice of Medicine"*, Future Healthcare Journal, Vol. 8, Issue 2, pp. e188–e194, 2021.
10. L. Wang, H. Jiang, and M. Chignell, *"Multi-Task Learning for Disease Prediction and Risk Factor Analysis"*, arXiv Preprint, 2021.
11. E. Garcia-Ceja et al., *"Mental Health Monitoring and Stress Detection Using Machine Learning"*, BMC Medical Informatics and Decision Making, Springer, Vol. 22, Article No. 15, 2022.
12. A. Almadhor et al., *"Wearable Sensor-Based Stress Detection Using Machine Learning Techniques"*, Sensors Journal (MDPI), Vol. 23, Issue 4, Article 2101, 2023.
13. H. Yan et al., *"A Data Mining Approach for Medical Decision Support Systems"*, Expert Systems with Applications, Elsevier, Vol. 30, Issue 4, pp. 656–671, 2006.
14. J. H. Chen and S. M. Asch, *"Machine Learning and Prediction in Medicine"*, New England Journal of Medicine, Vol. 376, No. 26, pp. 2507–2509, 2017.
15. C. Krittanawong et al., *"Machine Learning Prediction in Cardiovascular Diseases"*, European Heart Journal, Oxford University Press, Vol. 41, Issue 12, pp. 1110–1118, 2020.
16. Dr. G. Nagalakshmi, "AYUMUKHA: Facial Recognition and Detection Using Artificial Intelligence for Skin Care"