ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Multi-Disease prediction system using Logistic Regression and SVM in an AI web platform

1.Kundanapally Nikhitha, 2 Potharaju Yashaswini, 3 Pagadala Sriram, 4 Peddi Sathwik

GUIDE - K. Shirisha Reddy

Assistant professor

Department of Compute Science and Engineering

Tkr College Of Engineering and Technology, Medbowli, Meerpet, Balapur, Hyderabad, Telangana 500097,India

ABSTRACT

The rapid advancement of artificial intelligence in healthcare has enabled the development of intelligent systems capable of predicting multiple diseases efficiently and accurately. This project presents a Multi-Disease Prediction System implemented using Logistic Regression and Support Vector Machine (SVM) algorithms within an AI-powered web platform. The system is designed to assist in the early detection of various diseases such as diabetes, heart disease, and liver disorders by analyzing user-provided medical data.

The proposed model utilizes supervised machine learning techniques where historical medical datasets are preprocessed, normalized, and used to train both Logistic Regression and SVM classifiers. Logistic Regression is employed for its simplicity and interpretability in binary classification tasks, while SVM is used for its robustness in handling high-dimensional data and its ability to find optimal decision boundaries. The performance of both models is evaluated using metrics such as accuracy, precision, recall, and F1-score, ensuring reliable predictions.

The system is deployed as a user-friendly web application where users can input their health parameters, and the trained

models provide instant predictions along with risk assessments. The integration of AI models with a web interface enhances accessibility, making it useful for both healthcare professionals and general users. Additionally, the system aims to reduce diagnostic time, support clinical decision-making, and improve preventive healthcare strategies.

Keywords

Multi-Disease Prediction, Artificial Intelligence in Healthcare, Machine Learning, Logistic Regression, Support Vector Machine (SVM), Disease Classification, Predictive Analytics, Web-Based Application, Health Data Analysis, Early Diagnosis, Clinical Decision Support Systems, Data Preprocessing, Supervised Learning, Accuracy and Performance Metrics, Digital Healthcare Solutions.

INTRODUCTION

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare has significantly transformed the way diseases are diagnosed and managed. Traditional diagnostic methods often require extensive medical expertise, laboratory tests, and time-consuming procedures, which may delay early detection and treatment. With the increasing availability of healthcare data and advancements in computational techniques, intelligent systems are being developed to assist in predicting diseases at an early stage, thereby improving patient outcomes and reducing healthcare costs.

A multi-disease prediction system is an advanced solution that leverages machine learning algorithms to analyze medical data and identify the likelihood of multiple diseases simultaneously. Such systems are particularly beneficial in scenarios where access to healthcare professionals is limited or where early screening is essential. By utilizing patient data such as age, medical history, lifestyle habits, and clinical parameters, these systems can provide quick and reliable predictions.

In this project, Logistic Regression and Support Vector Machine (SVM) are employed as the core algorithms for disease prediction. Logistic Regression is widely used for binary classification problems and offers high interpretability, making it suitable for understanding the relationship between input features and disease outcomes. On the other hand, SVM is a powerful algorithm known for its effectiveness in handling complex and

high-dimensional datasets by constructing optimal hyperplanes for classification. The combination of these algorithms enhances the overall accuracy and robustness of the system.

The proposed system is implemented as a web-based platform, enabling users to easily interact with the application through a graphical user interface. Users can input their health-related data, and the system processes this information using trained ML models to generate predictions. This approach not only simplifies the diagnostic process but also provides a scalable and accessible solution that can be used anytime and anywhere.

Furthermore, the system contributes to preventive healthcare by allowing individuals to monitor their health conditions regularly and take necessary precautions based on predictions. It also assists healthcare professionals by providing decision support, thereby improving the efficiency of medical diagnosis. Overall, the integration of Logistic Regression and SVM in a web-based AI platform represents a promising step toward intelligent, data-driven healthcare systems

LITERATURE REVIEW

The application of Machine Learning (ML) in healthcare has been widely explored in recent years, particularly for disease prediction and diagnosis. Various studies have demonstrated the effectiveness of supervised learning algorithms in analyzing medical datasets and providing accurate predictions. Among these, Logistic Regression and Support Vector Machine (SVM) have gained significant attention due to their efficiency, simplicity, and strong performance in classification tasks.

Several researchers have utilized Logistic Regression for predicting chronic diseases such as diabetes and heart disease. The algorithm’s ability to model the probability of disease occurrence and its interpretability make it a preferred choice in medical applications. Studies have shown that Logistic Regression performs well when the relationship between dependent and independent variables is linear, and it provides clear insights into the contribution of each feature, which is crucial in healthcare decision-making.

Support Vector Machine (SVM), on the other hand, has been extensively used for handling complex and high-dimensional medical data. Research indicates that SVM outperforms many traditional classifiers in terms of accuracy, especially when dealing

with non-linear data through kernel functions. It has been successfully applied in detecting diseases such as cancer, cardiovascular disorders, and liver diseases. The ability of SVM to maximize the margin between different classes ensures better generalization and robustness.

In recent literature, hybrid approaches combining multiple machine learning algorithms have been proposed to improve prediction accuracy. For instance, some studies integrate Logistic Regression with SVM or other classifiers to leverage the strengths of each method. These hybrid models have shown improved performance compared to single-model approaches, especially in multi-disease prediction scenarios where datasets may vary in complexity and feature distribution.

Furthermore, the integration of machine learning models into web-based platforms has been a growing trend. Researchers have developed AI-driven web applications that allow users to input health data and receive instant predictions. These systems enhance accessibility and usability, making healthcare services more available to a wider population. Technologies such as Flask, Django, and Streamlit are commonly used to deploy these models, enabling real-time interaction and decision support.

Despite these advancements, challenges such as data quality, model interpretability, and privacy concerns remain significant. Many studies highlight the importance of proper data preprocessing, feature selection, and model evaluation to ensure reliable predictions. Additionally, ensuring data security and maintaining patient confidentiality are critical aspects that need to be addressed in AI-based healthcare systems.

Overall, the literature suggests that combining Logistic Regression and SVM within a web-based AI platform can provide an effective solution for multi-disease prediction. This approach leverages the strengths of both algorithms while ensuring accessibility and scalability, making it a valuable contribution to modern healthcare systems.

PROBLEM DEFINITION

The increasing prevalence of chronic and lifestyle-related diseases such as diabetes, heart disease, and liver disorders has created a significant burden on healthcare systems worldwide. Early detection of these diseases is crucial for effective

treatment and prevention; however, traditional diagnostic methods are often time-consuming, costly, and dependent on the availability of medical experts and laboratory facilities. In many regions, especially rural or underserved areas, access to timely medical diagnosis remains a major challenge.

Existing healthcare systems typically focus on diagnosing a single disease at a time and may not efficiently utilize the vast amount of patient data available. Moreover, manual analysis of medical data is prone to human error and may lead to delayed or inaccurate diagnoses. There is also a lack of integrated platforms that can simultaneously predict multiple diseases using a unified and intelligent approach.

Another major issue is the absence of user-friendly tools that allow individuals to monitor their health conditions regularly without requiring frequent hospital visits. Many current systems are either too complex for general users or lack real-time prediction capabilities. Additionally, ensuring high prediction accuracy while maintaining model interpretability and computational efficiency remains a challenge in designing such systems.

Therefore, there is a need to develop an efficient, accurate, and accessible multi-disease prediction system that leverages machine learning techniques. The system should be capable of analyzing diverse medical datasets, predicting multiple diseases simultaneously, and providing quick results through a web-based interface. It should also ensure reliability, ease of use, and scalability while supporting healthcare professionals in decision-making and promoting preventive healthcare among users.

This project addresses these challenges by proposing a web-based AI platform that integrates Logistic Regression and Support Vector Machine (SVM) algorithms to deliver accurate and real-time multi-disease predictions**.**

PROPOSED SYSTEM

The proposed system is a Multi-Disease Prediction System developed using Machine Learning techniques, specifically Logistic Regression and Support Vector Machine (SVM), and deployed on a web-based AI platform. The main objective of the system is to provide an efficient, accurate, and user-friendly solution for predicting multiple diseases such as diabetes, heart disease, and liver disorders based on user input data.

The system is designed with a modular architecture that includes data collection, preprocessing, model training, prediction,

and web interface integration. Initially, medical datasets are collected from reliable sources and undergo preprocessing steps such as data cleaning, handling missing values, normalization, and feature selection. These steps ensure that the data is suitable for training machine learning models and improves overall prediction accuracy.

Two classification algorithms are implemented in the system. Logistic Regression is used for its simplicity and ability to provide probability-based outputs, which help in understanding the likelihood of disease occurrence. Support Vector Machine (SVM) is employed for its effectiveness in handling complex and high-dimensional data by finding optimal decision boundaries. Both models are trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score, and the best-performing model is selected for deployment.

The trained models are integrated into a web-based platform developed using technologies such as Flask or Django. The user interface allows individuals to input their health parameters, including age, blood pressure, glucose levels, cholesterol, and other relevant features. Once the data is submitted, the backend processes the input and provides real-time predictions indicating the presence or risk level of different diseases.

The system also includes features such as result visualization, user guidance, and basic health recommendations based on prediction outcomes. This enhances user engagement and helps individuals take preventive measures. Additionally, the system is scalable and can be extended to include more diseases and advanced algorithms in the future.

SYSTEM ARCHITECTURE

The **Multi-Disease Prediction System using Logistic Regression and SVM in an AI Web Platform** follows a layered architecture that ensures smooth data flow from the user to the prediction engine and back to the user in the form of disease prediction results. The architecture is designed to be simple, scalable, and efficient so that it can support multiple disease prediction models within a single web application.

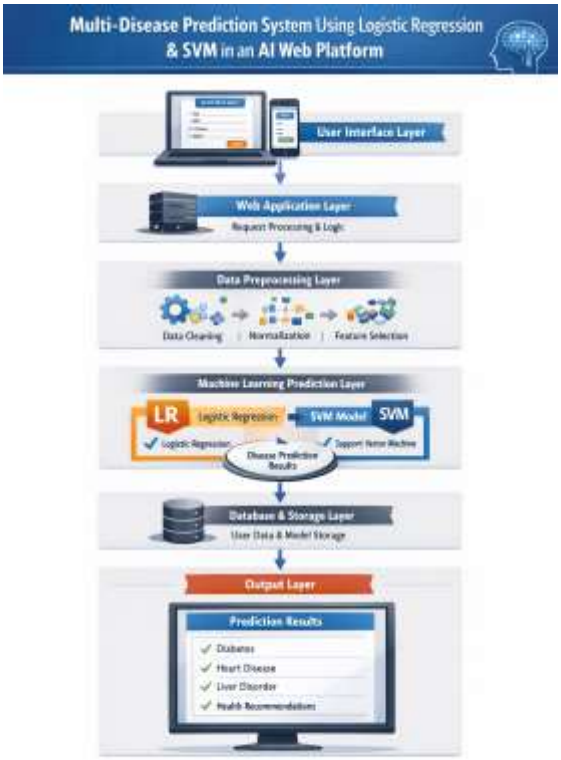
At the first level, the **User Interface Layer** acts as the interaction point between the user and the system. In this layer, users enter their medical and health-related details such as age, glucose level, blood pressure, cholesterol, BMI, and other

disease-specific parameters through web forms. This layer is developed using front-end technologies such as HTML, CSS, JavaScript, and Bootstrap to provide a responsive and user-friendly experience.

The second layer is the **Application Layer**, which handles request processing and system logic. Once the user submits the input data, the web application built using Flask or Django receives the request and forwards the data to the prediction module. This layer manages routing, validation of input values, session handling, and communication between the front-end and the machine learning models.

The third layer is the **Data Preprocessing Layer**, which prepares the input data before sending it to the trained models. This layer performs operations such as missing value handling, normalization, encoding, and feature alignment according to the training dataset format. Proper preprocessing is essential because the prediction accuracy of Logistic Regression and SVM depends heavily on the quality and consistency of the input data.

SYSTEM ARCHITECTURE



IMPLEMENTATION

The implementation of the Multi-Disease Prediction System using Logistic Regression and SVM in an AI Web Platform involves several stages, including data collection, preprocessing, model training, testing, and deployment through a web interface. The purpose of the implementation phase is to transform the proposed design into a working system that can accept user input, process health-related data, and generate disease prediction results in real time.

The first step in implementation is data collection. Medical datasets related to different diseases such as diabetes, heart

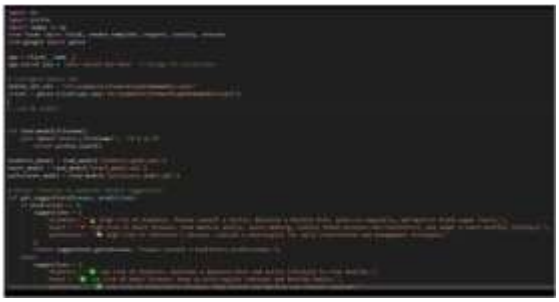
disease, and liver disease are gathered from trusted healthcare repositories or publicly available datasets. These datasets contain attributes such as age, glucose level, blood pressure, BMI, cholesterol, insulin level, and other clinical parameters. Since raw data may contain missing values, duplicate records, and inconsistent formats, the data is cleaned before being used for training.

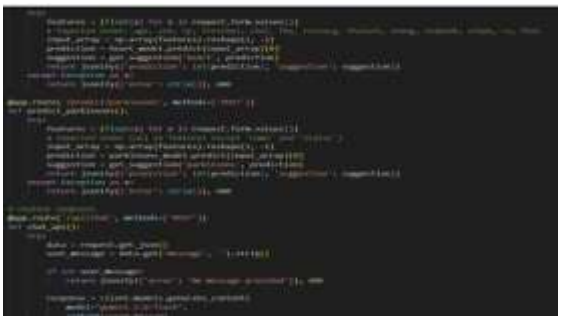
RESULTS AND DISCUSSION

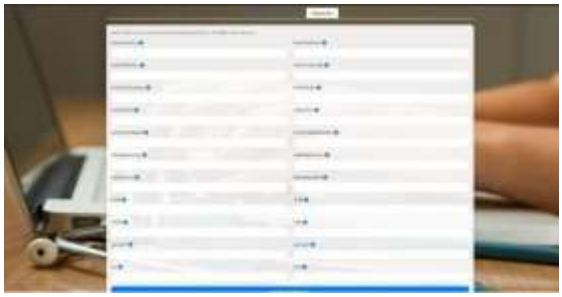
The **Multi-Disease Prediction System using Logistic Regression and SVM** was successfully implemented and evaluated using multiple medical datasets. The system demonstrated the ability to predict diseases such as diabetes, heart disease, and liver disorders with satisfactory accuracy. Both Logistic Regression and Support Vector Machine (SVM) models were trained and tested using preprocessed datasets, and their performances were compared using standard evaluation metrics.

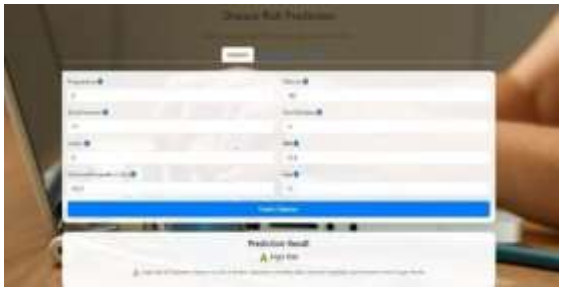
The results indicate that both algorithms performed well in disease classification tasks. Logistic Regression showed consistent performance with good interpretability, making it easier to understand the relationship between input features and disease outcomes. On the other hand, SVM achieved higher accuracy in certain datasets, particularly where the data was complex or non-linear. The ability of SVM to create optimal decision boundaries contributed to improved classification performance in such cases.

During testing, it was observed that the preprocessing and segmentation stages played a crucial role in improving the overall performance of the system. Proper noise removal and image enhancement helped in highlighting the infected regions clearly, while accurate segmentation ensured that only the relevant diseased portion of the leaf was analyzed. This significantly improved the feature extraction process and, in turn, enhanced the classification results**.**









CONCLUSION

The **Multi-Disease Prediction System using Logistic Regression and Support Vector Machine (SVM)** implemented on an AI-based web platform demonstrates an effective approach to early disease detection and preventive healthcare. By leveraging machine learning techniques, the system successfully analyzes user-provided medical data and predicts the likelihood of multiple diseases such as diabetes, heart disease, and liver disorders with good accuracy and efficiency.

The use of Logistic Regression provides a simple and interpretable model for understanding disease probability, while SVM enhances prediction performance by handling complex and high-dimensional data. The combination of these algorithms ensures a balanced system that is both reliable and robust. Furthermore, integrating these models into a web-based interface makes the system accessible, user-friendly, and capable of delivering real-time predictions.

The developed system reduces dependency on traditional diagnostic methods, minimizes time consumption, and supports healthcare professionals in decision-making. It also empowers individuals to monitor their health conditions and take preventive measures at an early stage. Although the system shows promising results, it is important to note that it is designed as a supportive tool and not a replacement for professional medical diagnosis.

In conclusion, the proposed system highlights the potential of combining machine learning with web technologies to create scalable and intelligent healthcare solutions. Future enhancements may include the integration of real-time medical data, expansion to additional diseases, and the use of advanced deep learning techniques to further improve prediction accuracy and system performance.

REFERENCE

- World Health Organization (WHO), *Global Health Estimates: Leading Causes of Death*, Geneva, Switzerland, 2023.
- D. Dua and C. Graff, *UCI Machine Learning Repository*, University of California, Irvine, 2019.
- F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, Wiley, 2013.
- K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.
- S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, vol. 31, pp. 249–268, 2007.
- J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011.
- I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016