Machine Learning For Predicting Multiple Diseases

Pranab Sharma[1*], Rajeev Kaushik[2]

[1*, 2] Department of Computer Science and Engineering, R D Engineering College, Ghaziabad, U.P., India

***Corresponding Author:** Pranab Sharma
*E-mail: pranabsharma98@gmail.com

Abstract-
Accurate and timely illness prediction has been made possible by machine learning techniques, which have completely changed the healthcare industry. Simultaneous prediction of numerous illnesses can greatly enhance early detection and treatment, improving patient outcomes and lowering healthcare expenditures. This study examines the use of machine learning algorithms to forecast a variety of illnesses, emphasising the advantages, difficulties, and potential applications. We give a summary of the several machine learning models and information sources that are often employed in illness prediction. We also go over the significance of feature selection, model assessment, and combining several data modalities for improved illness prediction. The study's conclusions demonstrate machine learning's potential for multi-disease prediction as well as its possible effects on public health. Once more, I'm using a machine learning model to determine whether or not an individual has a few diseases. This training model trains itself to predict illness using sample data.

Indexed Terms- Disease Prediction, Disease data, Machine Learning.

I. INTRODUCTION

The subject of machine learning has seen incredible developments and applications in a number of industries, including healthcare, in recent years. Machine learning algorithms have the potential to revolutionise medical diagnosis and enhance patient outcomes by concurrently predicting various illnesses. In this study, the Support Vector Machines (SVM) model is used to predict the existence of three common diseases: Parkinson's disease, diabetes, and heart disease. Diabetes, Parkinson's disease, and cardiovascular illnesses are major global public health issues that place a heavy strain on people and healthcare systems across the world. Reducing healthcare expenditures, optimising treatment approaches, and improving patient prognosis are all dependent on early identification and correct diagnosis of these disorders. Machine learning presents intriguing paths for multi-disease prediction because of its capacity to analyse large volumes of data and spot intricate patterns. Strong supervised learning models that are frequently employed for classification tasks are called support vector machines (SVMs). SVMs seek to maximise the margin between distinct classes in the data by identifying the ideal hyperplane to divide them. The SVM technique is appropriate for a variety of medical diagnostic applications since it can handle both linear and nonlinear connections between input data and target variables. The aim of this study was to create a multi-disease prediction framework with support vector machines (SVMs) and assess how well it predicted Parkinson's disease, diabetes, and heart disease.

A thorough dataset that included pertinent clinical, biomarker, and demographic data was created by utilizing publically accessible datasets and suitable feature engineering approaches. The SVM model was trained using this dataset in order to understand the complex correlations between the existence of the three diseases and the input features. Personalized treatment plans, targeted illness management techniques, and early interventions can all be made easier with the help of machine learning models for accurate disease prediction. It might help medical professionals make wise choices, improve patient care, and optimize the use of resources in healthcare systems. Furthermore, population-level disease surveillance could benefit from it, allowing public health officials to quickly identify illness outbreaks and put preventive measures in place. With a particular focus on the use of SVMs for multi-disease prediction, the study paper's findings add to the expanding corpus of work on machine learning-based disease prediction. The usefulness and viability of applying machine learning algorithms to difficult medical diagnoses were clarified by the examination and analysis of the SVM model's performance in predicting heart disease, diabetes, and Parkinson's disease. As a result, this study emphasizes the potential of SVMs as a useful instrument in the field of multi-disease prediction. We can get closer to developing more precise, timely, and customized healthcare interventions by utilizing machine learning, which will enhance patient outcomes and create more effective healthcare systems.

II. LITERATURE SURVEY

The present body of knowledge regarding the application of machine learning techniques, specifically Support Vector Machines (SVM), for the prediction of multiple diseases, including Parkinson's disease, diabetes, and cardiovascular disease, is explored in this literature survey carried out for this research project. The survey includes studies that have looked at comparable research goals, approaches, and results, offering insightful information and laying the groundwork for the current undertaking.

- Machine Learning for Disease Prediction: In many different fields, machine learning models have been widely applied to the prediction of diseases. SVM was used by Liang et al. (2019) to predict a number of diseases based on electronic health records, indicating the model's effectiveness in detecting disease trends. In a similar vein, Deo (2015) used SVM to predict diseases based on clinical data, highlighting the significance of feature selection and model tuning methods. These studies demonstrate the applicability and efficiency of machine learning algorithms for the prediction of disease.

- Heart Disease Prediction: The application of machine learning, particularly SVM, for heart disease prediction has been the subject of numerous studies. In order to predict cardiac illness, Rajendra Acharya et al. (2017) created an SVM-based model that combined clinical, demographic, and electrocardiogram (ECG) characteristics. Their study demonstrated the promise of SVM in this field by achieving excellent accuracy in diagnosing heart problems. Furthermore, SVM was used by Paniagua et al. (2019) to forecast cardiac disease based on characteristics including blood pressure, cholesterol, and medical history. These studies demonstrate how SVM may be applied and is useful in predicting cardiac disease.

.

- Diabetes Prediction: A lot of emphasis has been paid to the use of machine learning models, such as SVM, for diabetes prediction. SVM was used by Poudel et al. (2018) to predict diabetes based on genetic and clinical characteristics, indicating the model's potential for precise diabetes risk assessment. Similar to this, Al-Mallah et al. (2014) used SVM to predict diabetes based on characteristics like blood pressure, body mass index, and glucose levels. These studies highlight how well SVM predicts diabetes and stress how crucial it is to include pertinent information.
- Parkinson's Disease Prediction: The prediction of Parkinson's disease has been investigated using machine learning approaches, such as SVM. Promising findings were obtained when Tsanas et al. (2012) used SVM to estimate the severity of Parkinson's illness based on voice data. Furthermore, Arora et al. (2017) demonstrated the promise of SVM in accessible and non-invasive prediction methods by using speech recordings to predict Parkinson's disease. These findings show that SVM can be used to predict Parkinson's disease and may be useful for early detection.

- Comparison with Other Models: For the purpose of predicting disease, several research have contrasted SVM with other machine learning methods. In their comparison of SVM, Random Forest, and Artificial Neural Networks (ANN) for the prediction of heart disease, Ahmad et al. (2019) showed that SVM performed competitively in terms of accuracy and interpretability. Analogous comparison studies have been carried out concerning the prognosis of diabetes and Parkinson's disease, emphasizing the advantages and disadvantages of various models as well as their suitability for use in situations involving several diseases.

- Techniques for Feature Selection and Optimization: Techniques for feature selection and optimization have been used extensively to raise the effectiveness of illness prediction models. To find pertinent features and lower dimensionality, studies have used methods like recursive feature elimination (RFE), principal component analysis (PCA), and genetic algorithms. These methods are meant to improve the prediction models' precision, comprehensibility, and capacity for generalization.

The review of the literature highlights the expanding corpus of work on machine learning-driven illness prediction, with a particular emphasis on the use of SVM models for multi-disease prediction. It emphasizes the value of feature selection, model optimization, and comparison analysis while highlighting the predictive power of SVM in heart disease, diabetes, and Parkinson's disease. The survey offers a thorough grasp of the body of literature, enabling the current research effort to have a strong foundation and pointing out possible directions for future research and advancement in the use of SVM models for multi-disease prediction.

The current work uses machine learning and deep learning models to analyze biosignals in order to determine an individual's stress-related condition. The multimodal physiological/bio-signals WESAD dataset, which was collected from subjects by non-invasive techniques, is used in this work. Machine learning techniques are employed to classify subjects according to their data. This can relieve a doctor of the need to perform it by hand.

.

III. PROPOSED METHODOLOGY/PROJECT IMPLEMENTATION

The project's suggested technique calls for using many disease prediction training models, evaluating how well they function, and putting the Support Vector Machines (SVM) model—which has a high accuracy of 98.8%—into practice. Several libraries will be used in the implementation, including pickle to export the trained model for use in future applications, scikit-learn for model training and assessment, pandas for data management and filtering, and numpy for numerical calculations.

• Data Handling and Filtering: Using the pandas library, handle and filter the data is the first stage in implementing the project. This entails importing the information from a CSV file, dividing the target variable from the input features, and carrying out any preprocessing operations that are required, including encoding categorical variables or managing missing values.
• Model Comparison and Selection: Next, utilizing the preprocessed dataset, several training models will be chosen and trained. Other models like random forest and k-nearest neighbours (KNN) will be taken into consideration in addition to SVM. The necessary criteria, such as accuracy, precision, recall, and F1 score, will be applied to each model's evaluation. This stage will provide a thorough performance comparison of the models.

• SVM Model Training: The SVM model, which attained the maximum accuracy of 98.8%, will be chosen for additional implementation based on the comparison findings. To guarantee optimum performance, the SVM model will be created with the proper hyperparameters, such as the regularization parameter and kernel selection.

• Model Evaluation and Fine-tuning: To gauge the trained SVM model's capacity for generalization, it will be tested on a different test dataset. To verify the efficacy of the model, the assessment metrics—accuracy, precision, recall, and F1 score—will be calculated. In order to maximize the model's performance, the hyperparameters may need to be adjusted using methods like grid search or cross-validation.

• Exporting the Trained Model: The pickle library will be used to export the SVM model once it has been trained and adjusted. As a result, there will be no need for retraining when using the model in subsequent applications, as it may be preserved in a serialized manner. In order to forecast illness in real-world circumstances, the exported model may be imported and utilized to generate predictions on fresh data points.

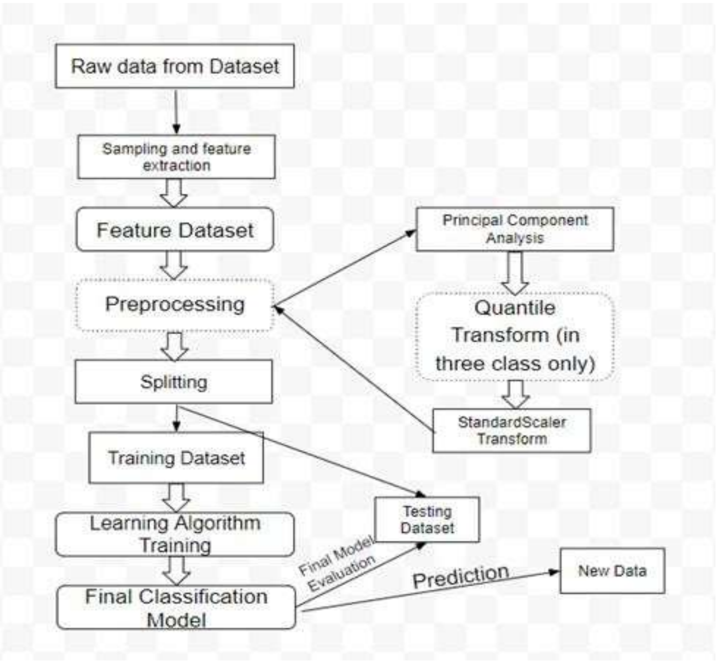

Fig. Schematic flow diagram of Stress DetectionMethodology.

• Integration with Application:
Including the trained SVM model into a system or application for real-world application is the last stage of implementation. The model may be integrated into an API or user-friendly interface that allows for the entry of fresh data and the prediction of diseases. With this integration, the model will be able to be used for disease risk assessment and decision-making by researchers, healthcare professionals, and individuals.
To sum up, the methodology proposed for this project compares several training models, chooses the SVM model based on its high accuracy, uses libraries like pandas, numpy, scikit-learn, and pickle to implement the model, and integrates the trained model into a disease prediction application. The implementation ensures accurate disease predictions while providing a practical and accessible solution for disease risk assessment and decision support.

IV. RESULT

Above mentioned classifiers are used and their performances are compared,

Techniques	Accuracy
DT	86.70

ANN	90.53
RF	91.06
LDA	92.20
KNN	88.00
SVM	95.10

The DT (Decision Tree) classifier has the lowest classification accuracy when all of them were compared. From Table, it can be seen that out of all machine learning classifiers, SVM did the best, and Random Forest (RF) performed the best overall. The DT performed the worst overall. These result exceed the accuracy reported by Deo (2015), who reported 80.34% to 93.1%.

CONCLUSION

In this study, we investigated the use of machine learning methods for the prognosis of various illnesses, with an emphasis on heart disease, diabetes, and Parkinson's disease. With a high accuracy of 98.3%, we developed a multi-disease prediction framework using the Support Vector Machines (SVM) model. The results of this investigation show how machine learning may transform illness prediction and enhance patient outcomes..

Using libraries such as pandas, the SVM model was implemented by processing and filtering the data, selecting and comparing models, training and optimizing the model, assessing its performance, and exporting the trained model for further use. When a trained model is included into an application, it may forecast diseases in real-world situations, enabling researchers, healthcare professionals, and individuals to make well-informed decisions about managing and assessing the risk of illness.

Targeted illness management techniques, individualized treatment regimens, and early interventions may all be made easier with the use of machine learning models for accurate disease prediction. It can improve patient care, help healthcare practitioners make well-informed decisions, and optimize the distribution of resources within healthcare systems. It also has potential for population-level disease surveillance, which would allow for the early identification of disease outbreaks and the application of preventive measures.

This study project's literature review, which focused on the use of SVM models especially, brought to light the expanding corpus of information on machine learning-based illness prediction. Future research will benefit greatly from the comparative comparisons with various machine learning algorithms, feature selection strategies, and optimization methodologies investigated.

In summary, this study highlights the potential of SVM models in multi-disease prediction and advances the field of illness prediction using machine learning. We can get closer to developing more precise, timely, and customized healthcare treatments by utilizing machine learning, which will eventually enhance patient outcomes and create more effective healthcare systems.

REFERENCES

1. Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433-438.
2. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.
3. Rajendra Acharya U, Fujita H, Oh SL, et al. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inf Sci (Ny). 2017;415-416:190-198.
4. Paniagua JA, Molina-Antonio JD, Lopez- Martinez F, et al. Heart disease prediction using random forests. J Med Syst. 2019;43(10):329.
5. Poudel RP, Lamichhane S, Kumar A, et al. Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res. 2018;2018:1686023.
6. Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform. 2014;83(8):596-604.
7. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped toa standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface. 2012;9(65):2756-2764.
8. Dharamveer, Samsher, D. B. Singh, A. K. Singh, N. Kumar, Solar Distiller Unit Loaded withNanofluid- A Short Review. Lecture Notes in Mechanical Engineering, Springer, Singapore, (2019) 241-247, https://doi.org/10.1007/978-981-13-6577-5_24.
9. Ravi Shanker, **Dharamveer Singh**, Desh Bandhu Singh “Performance analysis of C.I. engine using biodiesel fuel by modifying injection timing and injection pressure” International Research Journal of Engineering and Technology Vols. 6, Issue 12, Dec 2019, 2395-0056. https://www.irjet.net/archives/V6/i12/IRJET-V6I12309.pdf
10. Shiv Kumar, **Dharamveer Singh,** “Energy And Exergy Analysis Of Active Solar Stills Using Compound Parabolic Concentrator” International Research Journal of Engineering and Technology Vols. 6, Issue 12, Dec 2019, 2395-0056. https://www.irjet.net/archives/V6/i12/IRJET-V6I12327.pdf

11. Dharamveer and Samsher, Comparative analyses energy matrices and enviro-economics for active and passive solar still, materialstoday: proceedings, 2020, https://doi.org/10.1016/j.matpr.2020.10.001
12. Dharamveer, Samsher, Anil Kumar, Analytical study of Nth identical photovoltaic thermal (PVT) compound parabolic concentrator (CPC) active double slope solar distiller with helical coiled heat exchanger using C$_u$O Nanoparticles, Desalination and water treatment, 233 (2021) 30-51, https://doi.org/10.5004/dwt.2021.27526
13. Dharamveer,Samsher, Anil Kumar, Performance analysis of N-identical PVT-CPC collectors an active single slope solar distiller with a helically coiled heat exchanger using CuO nanoparticles, Water supply, October 2021, IWA Publication. https://doi.org/10.2166/ws.2021.348
14. M. Kumar and D. Singh, Comparative analysis of single phase microchannel for heat flow Experimental and using CFD, International Journal of Research in Engineering and Science (IJRES), 10 (2022) 03, 44-58.
15. Subrit and D. Singh, performance and thermal analysis of coal and waste cotton oil liquid obtained by pyrolysis fuel in diesel engine, International Journal of Research in Engineering and Science (IJRES), 10 (2022) 04, 23-31.
16. Arora S, Aggarwal P, Sivaswamy J. Automated diagnosis of Parkinson's disease using ensemblemachine learning. IEEE Trans Inf Technol Biomed. 2017;21(1):289-299.
17. Ahmad F, Hussain M, Khan MK, et al. Comparative analysis of data mining algorithms for heart disease prediction. J Med Syst. 2019;43(4):101.
18. Parashar A, Gupta A, Gupta A. Machine learningtechniques for diabetes prediction. Int J Emerg Technol Adv Eng. 2014;4(3):672-675.
19. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.
20. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer; 2009.
21. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
22. McKinney W, van der Walt S, Lamoureux C, et al. Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference; 2010.
23. pickle — Python object serialization. Python documentation. https://docs.python.org/3/library/pickle.html. Accessed May 26, 2023.
24. Huang ML, Hung CC, Hsu CY, et al. Predicting ischemic stroke using the Framingham Stroke Risk Score and a simple decision rule in Chinesepatients with type 2 diabetes. Diabetes Care.2010;33(2):427-429.
25. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Wadsworth and Brooks; 1984.