International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

NNSVM: A Novel Approach for Early Prediction of Human Lifestyle Related Diseases

[1]Ms. Apeksha V. Sakhare, [2]Dr. Prasad Lokulwar

Submitted: 27/11/2023 **Revised**: 07/01/2024 **Accepted**: 17/01/2024

Abstract: Due to the tremendous advancements in public health over the past few decades, life expectancy has increased by roughly 15-20 years on average. Improvements in public policy, economic growth, medical diagnosis, and treatment protocols have accompanied the industrialization and urbanization of the 21st century. Deviations in lifestyle and eating habits dramatically raise the risk of noncommunicable diseases (NCDs) such as obesity, type 2 diabetes, hypertension, dyslipidemia, hypertension, osteoarthritis, sleep apnea, and various forms of cancer. Nearly 36 million deaths worldwide (or 63%) were caused by NCDs. In addition to microbial infection and an unhealthful diet, sedentary work habits and insufficient physical activity were other newly discovered risk factors that mostly caused metabolic diseases. Environmental factors and human lifestyle choices bring on numerous diseases that account for a significant portion of global mortality, and diagnosing these illnesses can occasionally be difficult. We need a trustworthy, practical, accurate, and robust method to diagnose noncommunicable diseases (NCDs) based on lifestyle context to treat them early and effectively. NCDs include cardiovascular diseases (CVD), stroke, diabetes, and some types of cancer, usually called lifestyle related diseases (LRDs), since lifestyle choices significantly impact these ailments. This paper discusses the model concept, Prediction methods. Here, NNSVM model is proposed and implemented to predict the diseases like Diabetes, Depression and Hypertension. The purpose of this study is to suggest a framework for early prediction of disorders related to lifestyle that consists of three essential modules: a feature selection module, a disease prediction module, and to suggest remedies.

Keywords*: Feature selection, lifestyle Related diseases (LRDs), noncommunicable diseases, prediction, NNSVM,*

1. Introduction

The immune system protects the human body, but occasionally it cannot shield it against infections on its own. According to McKinsey's analysis, percent of people suffer from chronic conditions. Numerous chronic diseases have long been linked to lifestyle choices. People's lifestyle choices contribute to an increase in the likelihood of developing chronic diseases. [1] In India, disease prevalence has climbed along with people's healthy environment. About 60% of deaths have been caused by non-communicable conditions like heart, cancer, and diabetes. Such illnesses are frequently brought on by people's lifestyle choices and the environment [2].

The work was completed to extract the necessary data for disease prediction. Historical patient health information, research, and big data could identify various ailments. These illness prediction models are crucial for determining whether a disease is present [3]. supervised, semi-supervised, unsupervised learning, etc., are necessary machine learning techniques to detect diseases. Machine learning algorithms can utilize the raw data to learn from it, and then they can forecast the sickness in the future based on what they have learned [4]. The performance of the classification model may need to be improved by the noisy, irrelevant, redundant, and missing information in most medical datasets. The accuracy of the medical data and the models applied throughout the classification process determine how well the classifier performs (disease prediction). Therefore, employing classifiers to correctly and accurately interpret sensitive medical data to forecast and detect diseases is crucial. The problem is to predict the risk of lifestyle-related diseases based on individuals lifestyle and health related Factors.

The expected outcome of this research work is that if the disease is detected and identified at its early stage it can be cured very effectively and easily with more accuracy and thus reducing the risk factor of life of a patient and the cost of treatment required to cure the disease [5].

2. Literature Review

A wide range of machine-learning methods and algorithms that can be applied in medical settings have been used in several research to predict diseases. After reviewing the papers on disease prediction, it is found that the prediction of diseases is already made based on the patient's symptoms. Various Data Mining and Machine Learning algorithms are used and compared with each other to give the best prediction model. Here in this paper, only Lifestyle based habits are taken, and based on those habits, Noncommunicable Diseases will be predicted at the early stages. For lifestyle Habits and noncommunicable diseases, only the study part is done by many researchers. However,

[1]*Research Scholar, School of Engineering & Technology, GHRU Amravati University, Amravati.*
apeksha.sakhare@raisoni.net
[2]*Professor, Department of Computer Science and Engineering, G H Raisoni College of Engineering , Nagpur.*
prasad.lokulwar@raisoni.net

very little work is still found for predicting those diseases, which will reduce the rate of deaths globally caused by Lifestyles [6]. We will discuss some papers that will give results based on Machine Learning Techniques, Deep Learning where patients' symptoms are taken in the datasets.

A. Reviews of Diseases Prediction

Machine learning has been researched by Y. Deepthi et al. to predict disease based on symptoms. On the given dataset, the disease is predicted by machine learning techniques such as the Naive Bayes, Decision Tree, and Random Forest. They focused on three diseases Jaundice, Heart Attack, and Diabetes. Performance performances are checked and studied using Naïve Based, Decision Tree algorithms, and Random Forest. This system's accuracy is 82% for naïve based, 98% for Random Forest, and 79% for Decision Tree Algorithm [7].

The prediction of particular diseases utilizing machine learning methods that comprised the Decision Tree, the Naive Bayes classification, and the Support Vector Machine methods was considered by Dhomse Kanchan B. et al. [8] using Principal Component Analysis(PCA). This system's accuracy for diabetes is 34.89 percent, and its accuracy for heart illness is 53%.

Pahulpreet Singh Kohli et al. [9] provided a prediction of diseases using the Decision Tree, Logistic Regression, Support Vector Machine, Adaptive Boosting, and Random Forest using machine learning applications and techniques. This essay focuses on the prognosis of heart disease, diabetes, and breast cancer. The Logistic regression yields the maximum precision rates of 96.71 percent for breast cancer, 85.42 percent for diabetes, and 86.12 percent for heart disease.

Reddy et al. [10] studied the uses of data mining for the prediction of diabetic disease using Nave Bayes and KNN algorithms. This system can predict diabetes, and KNN's accuracy is higher than Nave Bayes'.

Utilizing distributed machine learning classifiers, L. Jena and R. Swain et al. [11] concentrated on risk prediction for chronic diseases using methods like the Naive Bayes and the Multilayer Perceptron. The accuracy of the Naive Bayes and the Multilayer Perceptron in this paper's attempt to forecast Chronic Kidney Disease is 95% and 99.7%, respectively.

The Random Forest Algorithm was used by Rashmi G. Saboji et al. [12] to discover a scalable approach to predict cardiac disease via classification mining. By using this approach, the Nave-Bayes classifier is compared. However, Random Forest yields 98% accurate findings.

The Decision Tree, a Support Vector Machine, a Random Forest, the Naive Bayes Neural Network, and the KNN algorithm, were some of the techniques used by Senthil Kumar Mohan et al. [13]. They focused on hybrid machine learning methods that can reliably forecast heart illness. The accuracy rating of this system is 88.47%.

Using machine learning approaches, including the Decision Tree, the Support Vector Machine, the Random Forest, the Naive Bayes, the Neural Network, and the KNN, Rati Shukla et al. [14] proposed prediction and diagnosis for breast cancer. The Support Vector Machine provides more accurate outcomes in this system than any other technique.

Aakash Chauhan et al. used evolutionary rule learning.[15] to propose cardiac disease prediction model. Here suggested approach takes advantage of the Association Rule. Due to its accuracy of 53%, this technique could be more effective.

Ankita Dewan et al. suggested a disease prediction system that leverages a data mining classification hybrid technique [16] for predicting cardiac disease. The Neural networks, Decision trees, and the naive Bayes are some of the strategies this system employs. This system is 87 per cent accurate.

By N. M. Lutimath et al. [17], machine learning-based heart disease prediction was proposed. A multilayer perceptron model is used in this system. This method predicts cardiac illness based on typical symptoms including age, sex, pulse rate, and other characteristics. The suggested system has a 91 per cent accuracy rate.

B. A comparative study using various algorithms in the literature review

TABLE I. COMPARATIVE ANALYSIS OF THE LITERATURE USING DIFFERENT ALGORITHMS

Year	Name of Author	Objective	Algorithm	Result
2020	Y. Deepthi [7]	to predict disease based on symptoms	• Naive Bayes, • Decision Tree, • Random Forest	82% for naïve based, 98% for Random Forest and 79% for Decision Tree Algorithm.
2017	B. Dhomse Kanchan and M.	to study the prediction of specific diseases using	• Naive Bayes classification	The accuracy of diabetes is 34.89%,

	Mahale Kishor[8]	machine-learning algorithms	• Decision Tree, • Support Vector Machine approaches.	and heart disease is 53%.
2020	P. S. Kohli and A. [9]	Using machine learning applications and techniques	• The Logistic Regression, • The Decision Tree, Support	breast cancer 96.71% diabetes 84.42%, heart disease 87.12
2019	I. S. V. Reddy, S. Magesh Kumar, and S. P. Chokkalingam [10]	investigated the applications of data mining for diabetic illness prediction	• Using Nave Bayes • KNN algorithms	This system can predict diabetes, and KNN's accuracy is higher than Nave Bayes'.
2018	L. Jena and R. Swain[11]	Utilizing distributed machine learning classifiers	• Naive Bayes • Multilayer Perceptron	Chronic Kidney Disease is 95% and 99.7%, respectively.
2018	R. G. Saboji and P. K. Ramesh[12]	to find a scalable system that can predict heart disease via classification mining	• Random Forest	Random Forest provides results that are 98% accurate.
2019	S. Mohan et al.[13]	They concentrated on hybrid machine-learning techniques	• The Decision Tree • The Support Vector Machine, • The Random Forest, • The Naive Bayes, • The Neural Network, • The KNN	This system has an accuracy rate of 88.47 per cent
2019	R. Shukla et.al. [14]	proposed prediction and diagnosis for breast cancer	• The Decision Tree • The Support Vector Machine, • The Random Forest, • The Naive Bayes, • The Neural Network, • The KNN,	The Support Vector Machine provides more accurate outcomes in this system than any other technique.
2018	A. Chauhan et. al. [15]	to propose a disease prediction model for cardiac disease	• approach takes advantage of the Association Rule	Accuracy is 53%. Not so effective.

2015	A. Dewan and M. Sharma [16]	a data mining classification hybrid technique	• Neural networks, • decision trees, • naive Bayes	87% Accuracy
2019	N. M. Lutimath [17]	Machine learning-based cardiac disease prediction	• a multilayer perceptron model	91% Accuracy

C. Limitation of Existing Work

Numerous illness prediction strategies have been examined in the literature review using data mining, machine learning, and deep learning techniques. Previous prediction methods focused on the patient symptoms only. Many of the researchers focused on machine learning techniques. Here we replace these patient symptoms data with the patients' bad habits or lifestyles. Lifestyle habits can help diagnose diseases at their early stages, where death rates can be decreased. To overcome the limitation of the existing work, we can use random Forest Feature Extraction with Neural Network with support Vector Machine. We can use performance evaluation metrics to measure the performance of the model. Here the main challenge is choosing different parameters or attributes for predicting lifestyle diseases.

3. Proposed Methodology

This work presents a methodology to predict models of non-communicable diseases. Here in this work, only noncommunicable diseases [18] [19]based on lifestyle factors [20] are focused. An overview of the phases for the recommended approach is shown in Fig.1.

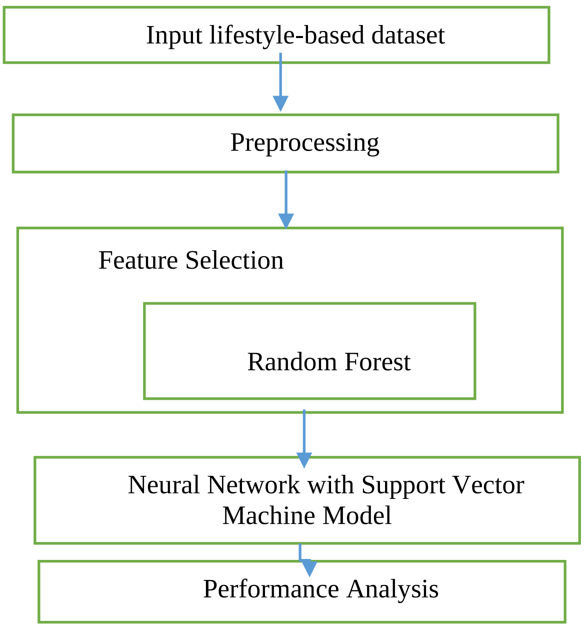


Fig. 1. Overview of Proposed Methodology

The dataset for lifestyle diseases is used as the input for the proposed work. Following preprocessing, Random Forest Feature Selection will be used to choose features. Based on the selected features, using Neural Network and machine Learning approach prediction can be done for human lifestyle disease. The effectiveness of the suggested methodology will next be evaluated in terms of several characteristics. Fig. 1 depicts the proposed system's overall layout.

D. Datasets:

Datasets for predicting lifestyle diseases are composed of daily lifestyle parameters such as smoking, alcohol intake, eating junk food, Body Mass Index, daily exercise, Waist circumference etc. The dataset is collected from a Kaggle source. It’s a labelled dataset.

E. Features selection for predictinng Noncommunicable Diseases

Variable collection, also identified as feature selection [21], is a popular data preparation technique for data mining that mainly minimizes data by deleting pointless and supplementary elements from any dataset [22]. Additionally, by using this technique, learning algorithms may be trained faster, predictions can be made with greater accuracy, and the data is simpler to comprehend and visualize. Numerous healthcare-related uses of relevant feature identification algorithms exist. Some frequently employed variable selection approaches include filters, wrappers, ensembles, and embedded methods. Removing noisy and inconsistent data is usually preferable to achieve more accurate answers faster before applying any model to the data. It's essential to reduce a dataset's dimensionality for use in practical applications. Moreover, the complexity falls off dramatically if the most crucial traits are chosen.

In recent years, a number of feature selection methods have been used to datasets of healthcare to extract more beneficial information.. On clinical datasets, feature selection techniques are used to predict various different chronic diseases, including the diabetes, a heart disease, strokes, the hypertension, the thalassemia, etc. Different algorithms operate more effectively and if the data has more relevant and non-redundant features, the results will be more accurate. An effective parameter assortment strategy is required to sort out intriguing features pertinent to the condition because medical datasets contain many redundant and irrelevant features.

According to literature reviews, the majority of existing prediction studies (91%) concentrate on single disease prediction, with only two studies focusing on multiple diseases, including 6 papers on diabetes prediction, 9 papers on hypertension prediction, 8 papers on overweight or obesity prediction, and 5 papers on cardiovascular disease prediction.

Following additional examination, the analyzed publications' quantifiable relationships between models and illnesses were as follows.

• One to one: Almost every publication that was examined exclusively addressed the prognosis of a single condition.

• One-to-many: To predict numerous diseases, three examined articles employed the same model and various datasets.

• Many-to-many: Only one study paper used various models to forecast various diseases in various datasets.

However, the flaw and the secret to pressuring the algorithm for higher performance is each article's limitation, which is described above. Thus, in our framework for the early diagnosis of the diseases for Hypertension, diabetes and Depression based on the lifestyle context is proposed and implemented where the performance evaluation metrics were applied to a model efficiency analysis.

According to the aforementioned study, current studies are unable to accurately identify essential characteristics of diseases when developing prediction models with various architectures and levels of robustness for various LRDs(Lifestyle Related Diseases) . Our goal is to develop an intelligent framework for LRD risk prediction that can reliably estimate the risk of LRDs, accurately identify essential characteristics of various LRDs from unclean real medical data, and visually display prediction findings.

F. Materials and methods:

Based on the analysis and studies data characteristics three research issues need to be considered in the proposed disease prediction framework for LRDs:

Because medical data has been obtained is unclean and has many missing values, it is difficult for individuals to utilize most popular prediction approaches in accordance with normal processes.

Accurate and thorough risk factor identification is necessary because reducing duplicate variables can reduce model complexity and make model predictions simpler to understand and analyze.

It is critical to investigate resilient models since data noise may reduce the model's accuracy and convergence rate. Increasing model resilience can improve model accuracy, minimize sensitivity to noisy data, and provide more dependable auxiliary services.

Last but not least, the suggested disease prediction framework for LRDs must consider the aforementioned three factors: 1) accurate illness prediction; 2) identification of essential traits; and 3) analysis and remedies for the diseases.

The proposed framework should particularly address these three problems by incorporating three key methods.

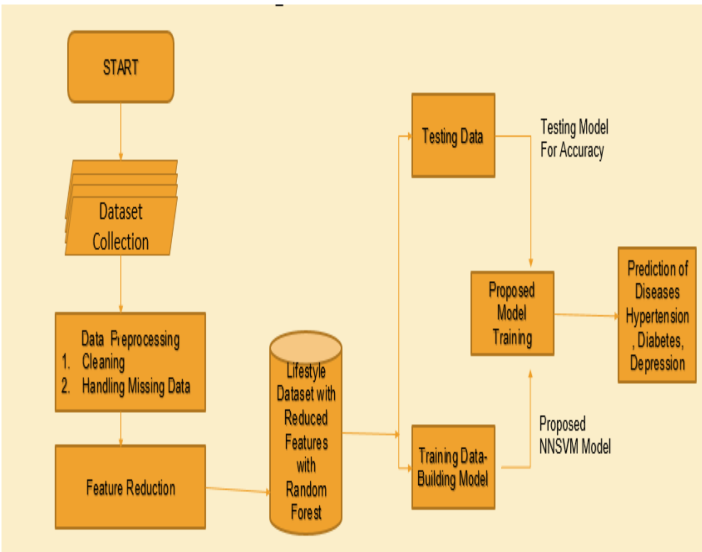


Fig. 2. Overview of the NNSVM Model

G. Proposed Algorithm NNSVM with Random Forest Feature Selection:

The dataset for lifestyle diseases is used as the input for the proposed work. Following preprocessing, Random Forest Feature Selection will be used to choose features. Based on the selected features, using machine Learning approach prediction can be done for human lifestyle disease. The effectiveness of the suggested methodology will next be evaluated in terms of several characteristics. Fig. 2 depicts the proposed system's overall layout.

Datasets for predicting lifestyle diseases are composed of daily lifestyle parameters such as smoking, alcohol intake, eating junk food, Body Mass Index, daily exercise, Waist circumference etc. The dataset is collected from a Kaggle source. It's a labelled dataset.

Features selection for predicting Noncommunicable Diseases is again very challenging task as here three diseases are focused and discussed.

Variable collection, also identified as feature selection, is a popular data preparation technique for data mining that mainly minimizes data by deleting pointless and supplementary elements from any dataset. Additionally, by using this technique, learning algorithms may be trained faster, predictions can be made with greater accuracy, and the data is simpler to comprehend and visualize. Numerous healthcare-related uses of relevant feature identification algorithms exist. Some frequently employed variable selection approaches include filters, wrappers, ensembles, and embedded methods. Removing noisy and inconsistent data is usually preferable to achieve more accurate answers faster before applying any model to the data. It's essential to reduce a dataset's dimensionality for use in practical applications. Moreover, the complexity falls off dramatically if the most crucial traits are chosen.

In recent years, a number of feature selection methods have been used to datasets of healthcare to extract more beneficial information. On clinical datasets, feature selection techniques are used to predict various different chronic diseases, including the diabetes, a heart disease, strokes, the hypertension, the thalassemia, etc. Different algorithms operate more effectively and if the data has more relevant and non-redundant features, the results will be more accurate. An effective parameter assortment strategy is required to sort out intriguing features pertinent to the condition because medical datasets contain many redundant and irrelevant features.

Classifying medical disorders using machine learning algorithms requires a lot of work with the nature of the dataset, which can comprise inadequate, unclear, and imprecise information. The performance of the classification model is impacted by the presence of such data in the dataset.

After Implementation of the Random Forest Feature selection Algorithm, following features are extracted and taken for further implementation.

BMI

Age

Smoking

Alcoholism

Stress and Anxiety

Lack of Physical Activity.

Performance Analysis:

For testing and evaluating the performance of any detection system, dataset assessment is essential. A high-quality dataset is essential for generating effective and worthwhile outcomes.

a. Metrics for Evaluation

To enhance our prediction model's effectiveness, we calculate the precision, recall, support, accuracy, and F-1 measure. The metrics that will be utilized for evaluation are described below. Every technique can perform differently regarding correctly classified instances, as demonstrated in Equation (4), which measures accuracy as the number of correctly predicted examples divided by the total number of forecasts produced by the model.

$$\text{Accuracy} = (Tp + Tn) / (Tp + Fp + Fn + Tn) \quad (4)$$

Generally, precision prediction expected successful outcomes. The formula shown in Equation (5) can be used to calculate the number of people who were classed as depressed and who are anticipated to be depressed.

$$\text{Precision} = Tp/(Tp+Fp) \quad (5)$$

Recall indicates how many positive instances out of all the positive cases were correctly predicted by the model. Calculating the recall percentage involves multiplying the TP by the sum of the TP and FN, given by Equation (6).

$$\text{Recall} = Tp/(Tp+Fn) \quad (6)$$

Because they test different properties represented in Equation (7), the precision and recall values of the F1-Measure evaluate the harmonization of two factors.

$$\text{F1 Measure} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall}) \quad (7)$$

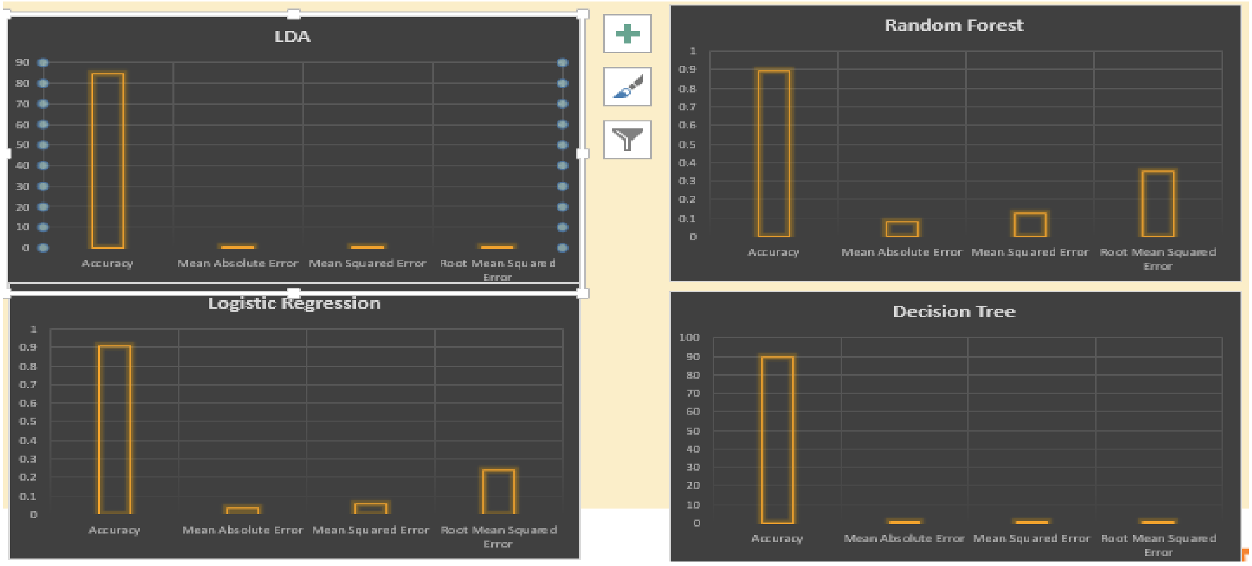
These equations will check the result, and then it will be compared with the other models which already exist.

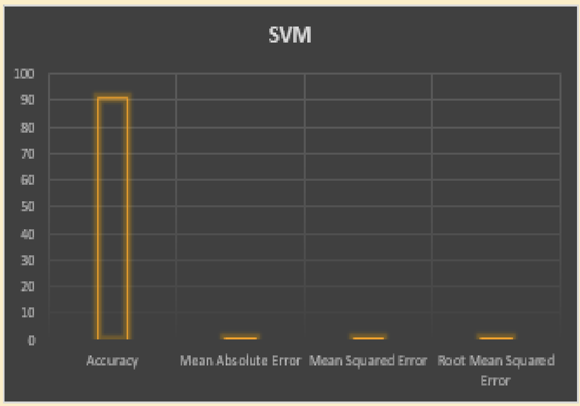
H. Result and Discussion:

In this paper, Random Forest Feature Selection Model with NNSVM technique is implemented and discussed. Here Feature selection model is implemented. This section compares and discusses some results without and some with feature selection. Here some diseases are focused such as Depression, Diabetes and Hypertension.

Some Machine Learning Models are initially implemented without any feature selection Algorithm to see the accuracy. The accuracy of Linear Discriminant Analysis, Random Forest, Logistic Regression and Decision Tree and Support Vector Machine.

The results of the above algorithms are as follows:





Sr. No.	ML Algorithms	Accuracy
1	LDA	84.94983278
2	SVM	91.30435
3	Decision Tree	89.63210702
4	Random Forest	89.6454849
5	Logistic Regression	91.2244147

Fig. 3 Machine Learning Algorithms and Accuracy Comparison

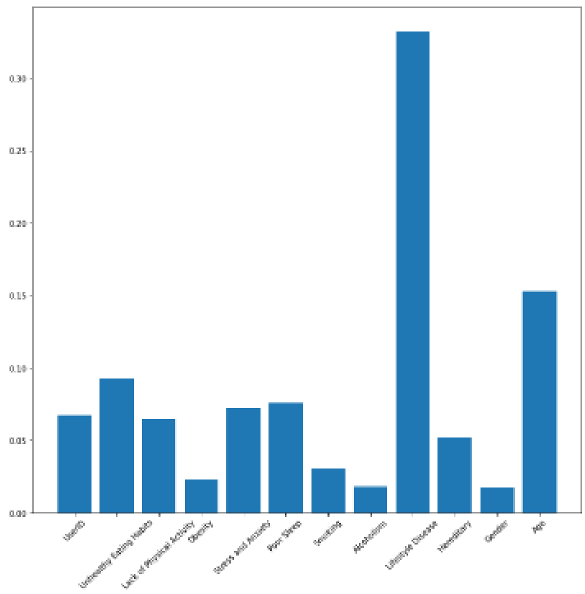


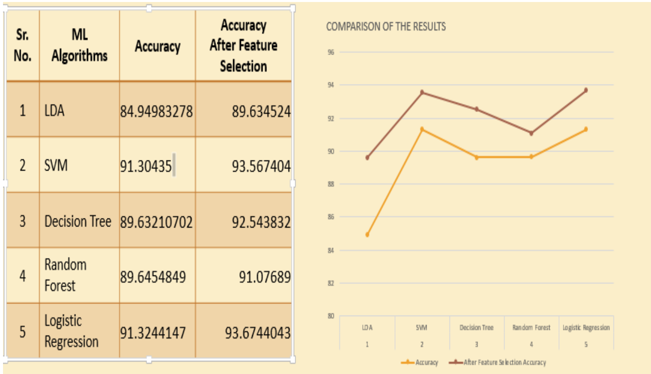
Fig.4. Best Features after RF FE Algorithm

I. Analysis of Feature Selection

A key factor in raising the predictive model's overall performance is feature selection. To accomplish this, a hybrid NNSVM technique is used in the proposed work. To evaluate various feature subsets and choose the highest-scoring feature set, cross validation is used. Here we are working for three diseases.

1. BMI
2. Age
3. Smoking
4. Alcoholism
5. Stress and Anxiety
6. lack of Physical Activity

Machine Learning Algorithms are again applied with the best features in the RF FE model, and results are compared in Fig.5. It gives the comparison of the machine learning with and without Feature Selection.



Sr. No.	ML Algorithms	Accuracy	Accuracy After Feature Selection
1	LDA	84.94983278	89.634524
2	SVM	91.30435	93.567404
3	Decision Tree	89.63210702	92.543832
4	Random Forest	89.6454849	91.07689
5	Logistic Regression	91.3244147	93.6744043

Fig.5. Comparison of the results with and without Feature selection

Ensemble Method:

Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs), two different machine learning models, combine their predictions to form the ensemble.

Support Vector Machine Model:

For regression and classification, supervised learning models like SVM are employed.

Using a subset of the dataset, train an SVM model with the goal variable being the labels of health disorders. SVMs work well with high-dimensional data and are especially helpful with non-linear decision boundaries.

ANN Framework:

A family of machine learning models called artificial neural networks is modelled after the composition and operations of the human brain. Utilizing the same dataset used for SVM, train an ANN model. ANNs are capable of feature learning and are able to record intricate correlations between features.

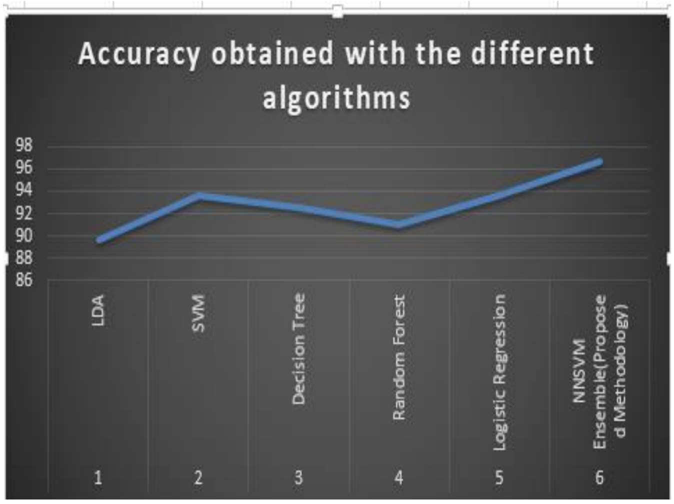
Merging Forecasts:

Once the SVM and ANN models (NNSVM) have been trained, use a different validation or test dataset to generate predictions. Add the two models' projections together. Combining these forecasts can be done in a number of ways, such as: the output probabilities are averaged. by use of a majority vote process. Assembling a weighted ensemble in which the forecast weight of each model is determined by its past performance.

Performance Evaluation:

Utilizing pertinent assessment criteria including accuracy, precision, recall, F1-score, and ROC-AUC, assess the ensemble's prediction ability. After implementation all results are compared with the proposed method. Results are discussed as follows:

Sr. No.	Algorithms	Accuracy obtained with the different algorithms
1	LDA	89.634524
2	SVM	93.567404
3	Decision Tree	92.543832
4	Random Forest	91.07689
5	Logistic Regression	93.6744043
6	NNSVM Ensemble(Proposed Methodology)	96.65543



J. Classification Performance Analysis:

A key factor in raising the predictive model's overall performance is feature selection. The suggested work approaches classification performance analysis using a hybrid NNSVM methodology. A comparative analysis is conducted between a numbers of cutting-edge algorithms that are frequently utilized for prediction in order to assess the effectiveness of the suggested NNSVM technique.

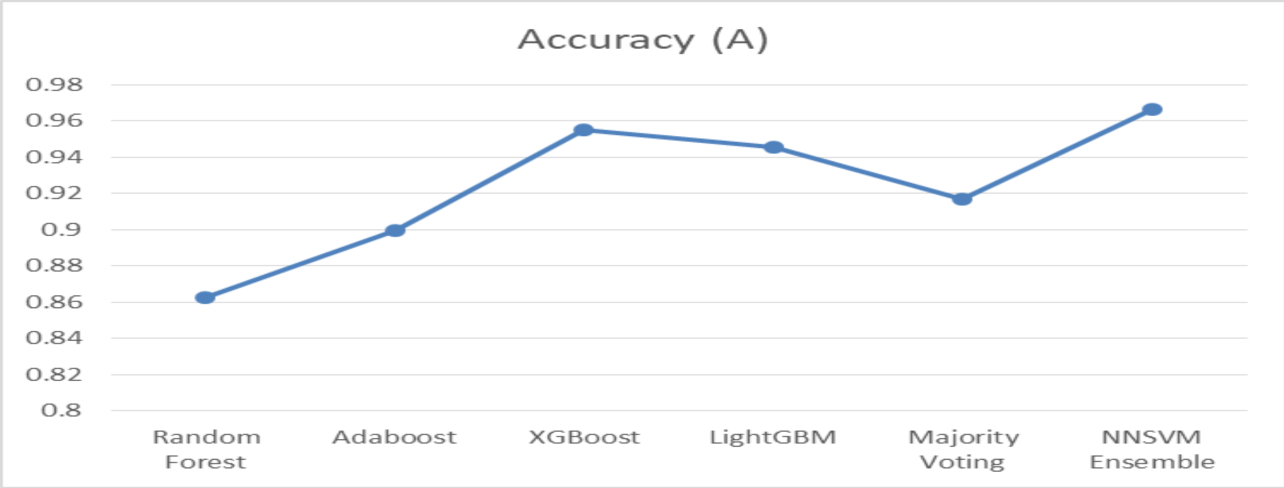


Fig 6: Comparison of the results with other model

In order to research and develop a better classification machine learning algorithm, several classification techniques are first employed. Both of the findings are compared and analyzed following the application of the Random Forest feature selection approach.

Based on the comparison above, SVM outperforms the other categorization techniques. Here, artificial neural network and support vector machine are integrated to increase the accuracy of the outcomes shown in fig.6

Predicting human health diseases can be effectively achieved by combining Artificial Neural Networks (ANNs) with an ensemble of Support Vector Machines (SVMs). The ANN and SVM strengths are combined in this method to increase prediction accuracy.

Analysis and Remedies for Diseases:

The NNSVM Method outperforms the other approaches after a thorough analysis of the data obtained from the various methods. We can then recommend treatments for the illnesses brought on by certain lifestyle choices shown in fig.7 and fig. 8

K. Conclusion:

This paper discusses various disease prediction algorithms, and the accuracy is studied on different parameters. Based on the literature survey, all the predictions are made on the patient's symptoms. Here we analyzed that no prediction is made on lifestyle-based factors. It is found difficult to search such datasets. Here NNSVM is implemented with the Random forest Feature Selection Method. We recommended some remedies for the disease like Depression, Diabetes, Hypertension etc.

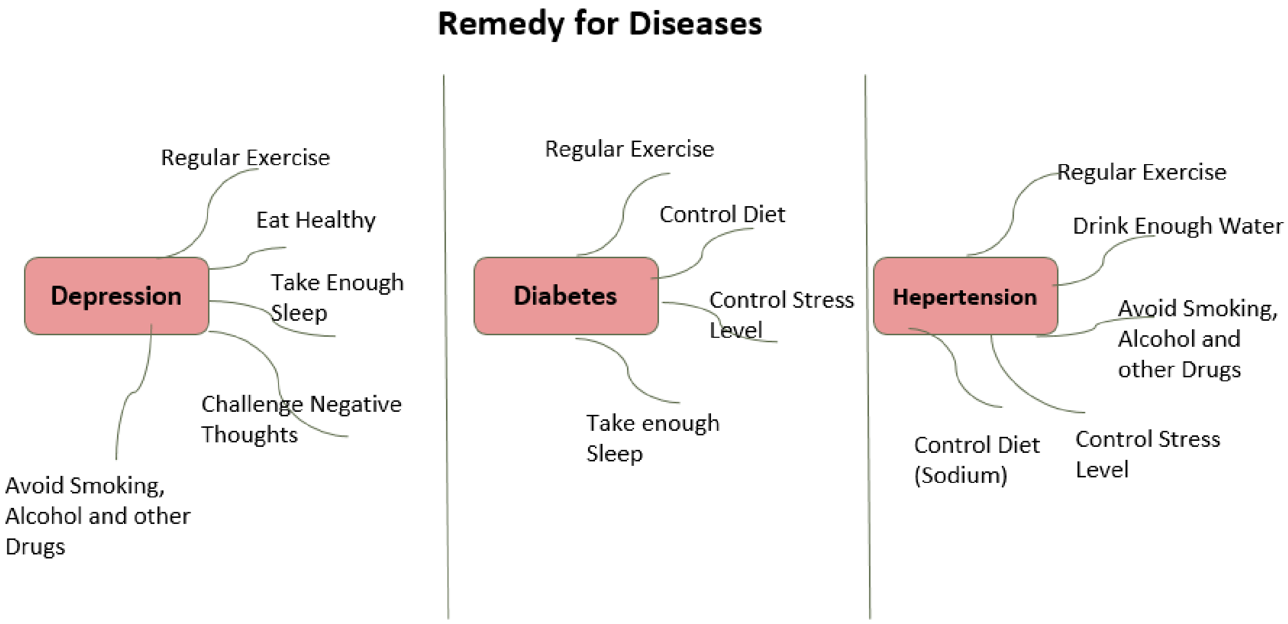


Fig 7: Remedies suggested for the Diseases

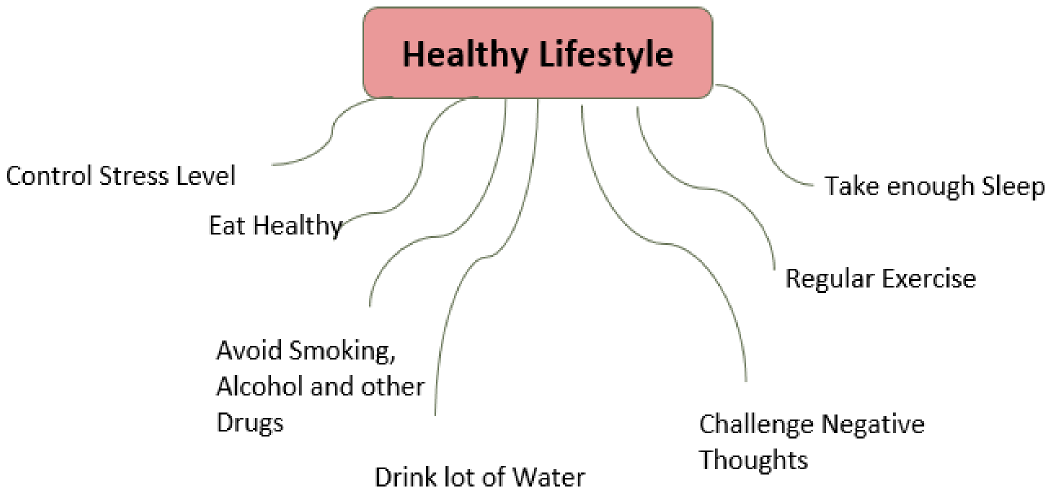


Fig 8: Healthy Lifestyle in reference with WHO Standards [1]

References

[1] Farooqui ME, Ahmad DJ. A Detailed Review on Disease Prediction Models That Uses Machine Learning. Int J Innov Res Comput Sci Technol. 2020;8(4):326–330. doi:10.21276/ijircst.2020.8.4.14.

[2] Ambekar S, Phalnikar R. Disease Risk Prediction by Using Convolutional Neural Network. Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018. 2018;1-5. doi:10.1109/ICCUBEA.2018.8697423.

[3] Farooqui ME, Ahmad DJ. Disease Prediction System using Support Vector Machine and Multilinear Regression. SSRN Electron J. 2020;331–336. doi:10.2139/ssrn.3673232.

[4] Dahiwade GP, Dahiwade EM. Designing Disease Prediction Model Using Machine Learning Approach. In: 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). 2019;1211–1215.

[5] Das H, Naik B, Behera HS. Medical disease analysis using Neuro-Fuzzy with Feature Extraction Model for classification. Informatics Med Unlocked. 2020;18. doi:10.1016/j.imu.2019.100288.

[6] Sakhare AV, Sharma AK. Analysis and modelling of contextual activity pattern for identifying lifestyle disorders. In: 12th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2021. 2021;203–207.

[7] Deepthi Y, Kalyan KP, Vyas M, Radhika K, Babu DK, Krishna Rao NV. Disease prediction based on symptoms using machine learning. Lecture Notes in Electrical Engineering. 2020;664:561–569. doi:10.1007/978-981-15-5089-8_55.

[8] Kanchan BD, Kishor MM. Study of machine learning algorithms for special disease prediction using principal of component analysis. In: Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016. 2017;5–10. doi:10.1109/ICGTSPICC.2016.7955260.

[9] Kohli PS. Application of Machine Learning in Disease Prediction. In: 2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020. 2020;1–4.

[10] Reddy ISV, Kumar SM, Chokkalingam SP. Disease prediction using data mining. Test Eng. Manag. 2019;81(11–12):5490–5493.

[11] Jena L, Swain R. Work-in-Progress: Chronic Disease Risk Prediction Using Distributed Machine Learning Classifiers. In: Proceedings - 2017 International Conference on Information Technology, ICIT 2017. 2018;170–173. doi:10.1109/ICIT.2017.46.

[12] Saboji RG, Ramesh PK. A scalable solution for heart disease prediction using classification mining technique. 2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput. ICECDS 2017. 2018;1780–1785. doi:10.1109/ICECDS.2017.8389755.

[13] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542–81554. doi:10.1109/ACCESS.2019.2923707.

[14] Shukla R, Yadav V, Pal PR, Pathak P. Machine learning techniques for detecting and predicting breast cancer. International Journal of Innovative Technology and Exploring Engineering. 2019;8(7):2658–2662.

[15] Chauhan A, Jain A, Sharma P, Deep V. Heart Disease Prediction using Evolutionary Rule Learning. In: Int. Conf. "Computational Intell. Commun. Technol. CICT 2018. 2018;doi:10.1109/CIACT.2018.8480271.

[16] Dewan A, Sharma M. Prediction of heart disease using a hybrid technique in data mining classification. In: 2015 Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2015. 2015;704–706.

[17] Lutimath NM, Chethan C, Pol BS. Prediction of heart disease using machine learning. International Journal of Recent Technology and Engineering. 2019;8(2 Special Issue 10):474–477. doi:10.35940/jute.B1081.0982S1019.

[18] Senthilkumar B, et al. Ensemble Modelling for Early Breast Cancer Prediction from Diet and Lifestyle. IFAC-PapersOnLine. 2022;55(1):429–435. doi:10.1016/J.IFACOL.2022.04.071.

[19] Tabish SA. Lifestyle Diseases : Consequences, Characteristics, C and Control. 2020;2–5.

[20] Naheeda P, Sharifullah K, Ullah SS, Azeem AM, Shahzad Y, Kinza W. Development of a cost-effective CVD prediction model using lifestyle factors. A cohort study in Pakistan. Afri Health Sci. 2020;20(2):849. doi:10.4314/ahs.v20i2.39.

[21] Tang L, Alelyani S. Feature selection for classification: a review. 2014.

[22] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3:1157–1182.