A Predictive Study of Machine Learning and Deep Learning Procedures Over Chronic Disease Datasets

---

Nimay Seth[*]

[*]*JayShree periwal international school, Jaipur, Rajasthan, India.*

Corresponding Email: [*]*nseth@jpischool.com*

Abstract: People's health and well-being are not given priority in the technological and Internet-savvy world we live in. People are becoming worse because they don't regularly attend the hospital for checkups due to job and unanticipated events. Most people nowadays suffer from one or more chronic illnesses, such as diabetes, hypothyroidism, heart disease, breast cancer, and dermatology. According to the World Health Organization (WHO), these chronic illnesses account for half of all fatalities in most nations and are the main cause of premature mortality. Patients who are identified early on potentially have their condition stop progressing. Many dispersed studies clearly demonstrated that conventional approaches to diagnosing chronic illnesses are prone to prejudice and heterogeneity among physicians, making it difficult to promptly and precisely diagnose problems. Still, Despite the availability of up-to-date information and a variety of machine learning-based methods, there have been enormous published efforts demonstrating that machine learning (ML)/deep learning (DL) based approach can considerably enhance the timely estimation of various health conditions. However, precise diagnosis of such diseases remains a difficulty. There are many machine learning-based techniques and current knowledge available, however despite this, a great deal of published research has shown that machine learning/deep learning based approach can considerably enhance the timely estimation of various health conditions. However, precise diagnosis of such diseases remains a difficulty. In order to tackle this problem, this work uses the UCI/KAGGLE ML/DL disease dataset to evaluate various ML/DL procedures and explores how different machine learning algorithms forecast chronic diseases. Accuracy and confusion matrix are used to verify the results. In order to help inexperienced researchers comprehend the disease prediction function of ML/DL-based techniques and determine the direction of Upcoming research, this study also discusses the advantages and disadvantages of accessible disease prediction schemes.

Keywords: Machine Learning (ML), Deep Learning (DL), Chronic Diseases, Classification, AI, SVM, RBFN.

1. INTRODUCTION

It is now well acknowledged that machine learning techniques are important in the medical field, aiding in the diagnosis, classification, and prediction of health disorders and their outcomes. Using machine learning techniques, sophisticated genetic data analysis may be adopted quickly. It is possible to find better models for accurate prediction and to change and completely study medical records for accurate evaluation. Various researchers have employed different algorithms to predict different health concerns [1-4]. But high-dimensional data is a common problem in machine learning, needing a lot of memory to evaluate the data and possibly leading to over fitting and exponential growth [5-7]. The study indicates that weighted features can be used to lessen the dataset's redundancy, which can then lead to decrease processing time. Various strategies, such as component design and element selection, can be employed to reduce the dimensionality of the dataset by eliminating less significant information, as suggested by source [7-10]. To evaluate the efficacy of the machine learning and deep learning approach for identifying Chronic diseases, this study compares and contrasts a number of ML/DLalgorithms, including Support Vector Machine (SVM) , Logistic Model(LM) and DL algorithms Artificial Neural Network (ANN) , Radial Basis Function (RBF) Networks.

The remainder of the paper is coordinated as: An appraisal of the current ML/DL approaches has been presented in Section 2. The assessment methodology that used in this study is outlined in 3rd segment. The experiment's findings are addressed in Section 4, and in Section 5, the paper is concludes with an advice for further studies.

2. RELATED WORK

Advances in machine learning have been extremely beneficial to the healthcare sector in recent years. These methods have been extensively accepted and shown to be successful in a number of healthcare contexts, most notably in medical cardiology. Thanks to the rapid capture of clinical data, clinicians are now in a position to create and test novel algorithms in this field with unprecedented ease. Since coronary disease is still the largest cause of death in non-industrialized countries, it is critical to identify risk factors and early indications of the condition. Several published research [11–17] suggested that specialists in this field might be able to identify and prevent cardiac disease earlier than ever before through the appropriate application of machine learning techniques..

A strategy for predicting cardiovascular sickness was developed by researchers in [18] using machine learning techniques. The 303 cases and 17 characteristics of the Cleveland heart disease dataset from the UCI machine learning repository were used to create the data for the study. The authors used a variety of supervised classification methods, such as random forests, k-nearest neighbor (KKN), naive Bayes, and decision trees. The investigation's findings showed that, with an accuracy rate of 90.8%, the KKN model had the highest degree of precision. In order to achieve the greatest results, the study emphasizes how crucial it is to select the appropriate algorithms and strategies as well as how machine learning techniques may be helpful in forecasting cardiovascular sickness. Using a KNN classifier with constrained parameters, the authors of [19] were able to predict heart disease with an accuracy rate of 81.85%. However, as

the number of parameters and training data rose, the KNN classifier became computationally expensive. In order to train for the diagnosis of cardiac illness, a Chroniclayer Perceptron neural network (MLP-NN) with backpropagation is used, according to [20]. The authors examined recall, accuracy, precision, and F-measure in order to analyze performance. They processed the 303 cases and 76 attributes of the UCI Cleveland dataset by eliminating six entries that had missing values. The MLP-NN used the top 14 critical attributes for training and testing to reach an exceptional accuracy of 93.39% in just 3.86 seconds.

A study on the full prediction of cardiac disease using widely used machine learning classifiers was carried out in [21]. After cleaning the data to produce a dataset of 296 records, the authors used just 14 characteristics for training and testing out of the 303 records in the Cleveland (UCI) dataset. In the experiment, the SVM classifier had a greater accuracy of 90.00%.

With a precision of almost 85.55%, a different study [22] investigated the possibility of forecasting cardiac illness using learning vector quantization techniques. The UCI machine learning collection provided the 303 items and 76 attributes that made up the dataset that was used. After preprocessing the data to remove missing values, a sample of 302 records and 14 characteristics that could be indicative of heart disease were obtained. For testing and training, the data set was split into two halves. In a different study [23], a web-based tool for heart illness prediction was created using machine learning techniques. Models for classification were developed and tested using LR, NB, and SVM. With 25% of the data set aside for validation and 75% for learning, the authors utilized the Cleveland records from the UCI machine learning repository. With preprocessing to eliminate inconsistent and missing variables, SVM achieved a higher precision of 64.4%. The study's incapacity to detect risk variables in patients with coronary artery disease at an early stage was its main flaw. Researchers combined neural networks, artificial neural networks, and fuzzy neural networks to construct an algorithm to categorize heart disease.

To detect cardiac disease, the authors of [26] used a variety of cutting-edge deep learning techniques and algorithms, such as adjusting hyperparameters. Strategic relapse, SVM, and gathering methods such as irregular backwoods were among the models employed. Neural networks had an accuracy rate of 78.3%. Studies have shown that, even when traditional risk factors like diabetes and hypertension are taken into consideration, men are still more likely than women to get cardiac illness. Four major datasets—Cleveland, Switzerland, Long Beach V, and Hungary—with crucial characteristics for heart disease prediction have been studied by numerous researchers since 1998 [27]. By utilizing the Spark framework and the Random Forest technique, the authors of [28] were able to execute a 98% correct evaluation of the data. The authors of [29] concentrated on information collecting and the application of several algorithms. To improve the forecast accuracy of heart disease, they employ a limited set of variables. Optimization methods are applied to the prediction of cardiac disease in another work [30]. A study [31] examined the available technologies for data processing and classification. When SVM was used with seven distinct methodologies, it outperformed other tactics.

The authors of [32] created a diabetes estimation model with SVM and logistic regression. Creating visually pleasing results was the goal of the pre-processing of the data. They claim that the applied scheme SVM performed better, with an accuracy of 79%. The researchers in [33] used the Decision Tree, SVM, and Naive Bayes techniques. Ten-fold cross authentication was used as a way to increase performance. The maximum accuracy of 76.30 percent was attained

with the Naive Bayes approach. For the study, the Pima Indian Diabetes dataset was examined. The authors of [34] used strategic relapse on PIDD for diabetic expectation. They discovered that, out of all the PIDD features, the number of pregnancies, BMI, and glucose are the primary predictors for diabetes expectation. The Pima Indian Diabetes dataset's study aftereffects are processed and displayed using RStudio. With an accuracy record of 75.32%, their model produces reasonably accurate forecasts.

The authors of [24] reported that the success precision rate for predicting breast cancer by the use of ML/DL approach was 95.15 percent. Another approach identified this type of cancer with a 94% precision rate using the DT scheme. Several more associated efforts that are now being made to forecast different diseases by consuming large databases include illustrating in table 1.

Table 1. Current Chronic Disease Forecasting Efforts Using ML/DL Techniques

Investigators	ML/DL Technique	Achieved Success %	Diseases
Shorewall et. al. [35]	RF, SVM, KNN	75.1	Heart
Maiga [36]	RF, NB, LR, KNN	70	
Waigi, [37]	DT	72.77	
Khan [38]	NN, LR, SVM	72.22	
ElSeddawy [39]	RF	89.01	
V.A. and R. Chitra [40]	SVM	78%	Diabetes
V. Sharma [41]	NB	95%	
Ramalingaswamy Cheruku et. al., [45]	RBNF	70%	Diabetes
K. Sivakami,[42]	Hybrid	91%	Breast Cancer
Shrivastav[43]	Decision Tree	98.14%	
M. R. Kousarrizi [44]	Support Vector Machine	98.62%	Thyroid

3. METHODOLOGY

To assess the efficacy of machine learning techniques in predicting the risk of chronic illness, a useful tool for both patients and healthcare providers. To do this, we employed a variety of machine learning techniques using an open access dataset of diseases from UCI/KAGGLE. The results are reported in this study report. Furthermore, several strategies for selecting only the best features have received particular attention in order to assess the effect of feature reduction. After preprocessing the data, before assessing the approach, the dataset has been partitioned using the k-modes approach and missing or unnecessary values eliminated. The employed methodology for this paper work was depicted in Figure 1 presented the methodology that was used for this paper.
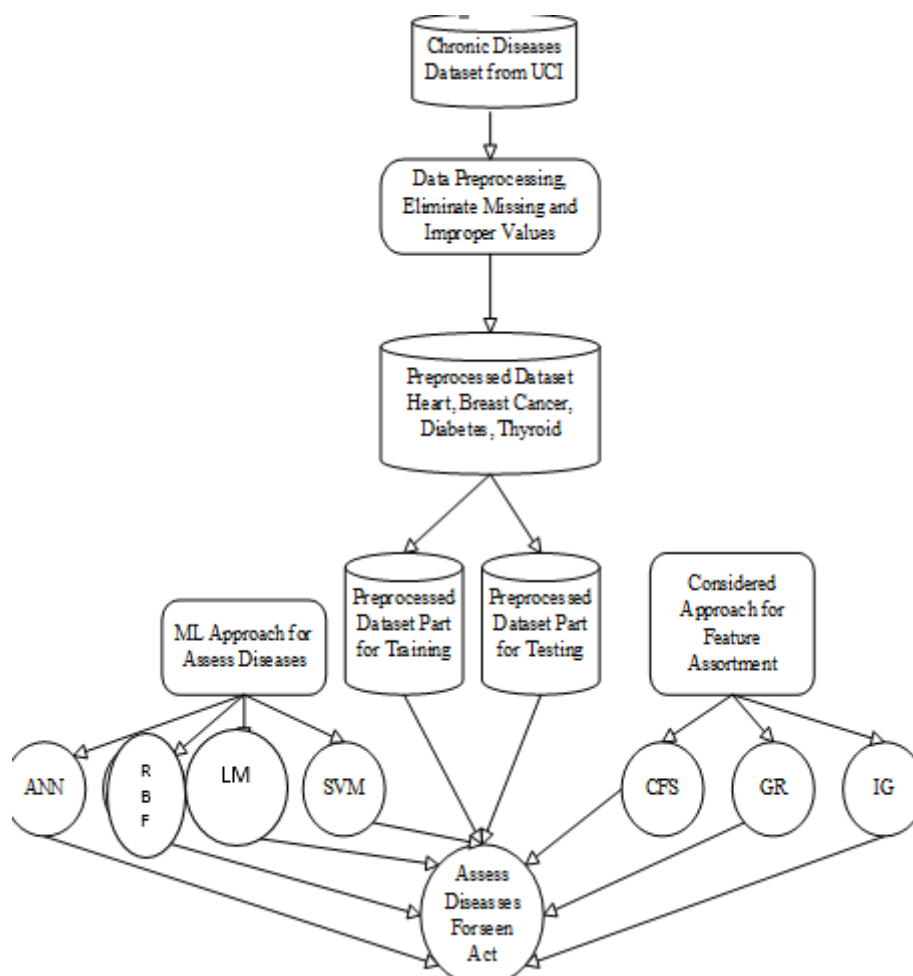
Fig. 1: Procedure for Analyzing ML/DL Method Predicting the onset of Chronic Diseases

The acquired dataset has been pre-processed to remove any missing or unsuitable attributes before the assessment cycle starts. Furthermore, the dataset is split into two parts: one for testing the process's real workings and the other for training the model. Two methods of implementation have been used: first, the conventional ML/DL methodology without the optimization of the feature selection process, and second, the features recommended by the CFS, GR, and IG feature selection schemes. Ultimately, a tabular and graphical presentation of all the results shows how the ML/DL technique affects disease prediction.

Accuracy, precision, and confusion matrix scores are used in the assessment procedure. The true attributes and expected values—referred to as the true positives and true negatives—are handled in a disorder lattice. There are four sections: First, there is true positive, or TP, where the values are determined to be authentic. The next section is bogus positive (FP), where the attributes that are highlighted as evident are actually deceptive. A false negative (FN) is the third component; in this case, the value was discovered to be false despite really being positive. The fourth component, true negative (TN), shows that the value was appropriately determined to be negative. The table is shown in its entirety in Figure 2.

		Predicted value	
		P	N
True value	P	TP	FN
	N	FP	TN

Fig. 2: Confusion Matrix

Method for verdict an accuracy is

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \qquad (1)$$

4. EXPERIMENTAL RESULTS AND DISCUSSION

Cleveland datasets, which are recognized as heart disease datasets, were taken into consideration for the first experimental process. Pre-processing has also been considered in order to remove related defects and only extract the best attributes. The entire and chosen attributes offered by the taken into consideration feature selection systems are shown in Table 1.

Table 2. Complete and Recommended Features for Cleveland Heart Disease Datasets as Determined by CFS, GR, and IG

		Full Attributes
CFS suggested Attributes	cp restecg thalach exang oldpeak ca thal	age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
GR suggested Attributes	sex chest_pain fbs exang oldpeak	
IG Suggested Attributes	cp oldpeak ca thal	

Cleveland Heart Diseases Data Set has been processed with all features at the initial assessment stage. The effectiveness of ANN, RBFN, LM, and SVM in predicting cardiac disorders is shown in Figure 3.
The accuracy of heart disease forecasting over Cleveland heart disease datasets without feature selection is presented in

Table 3.

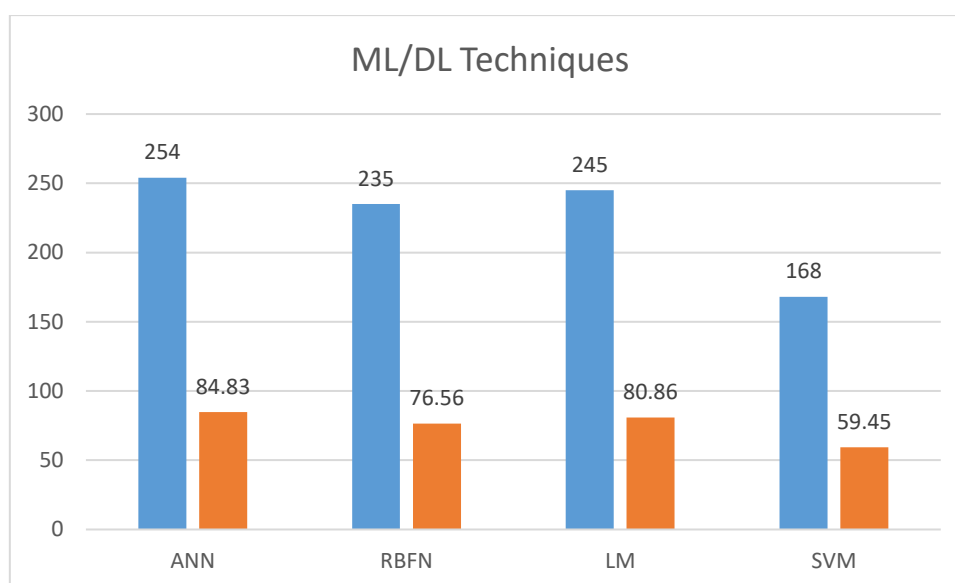|	(number of Instances 303)	Prediction Accuracy (%)
ANN	254	84.83
RBFN	235	76.56
LM	245	80.86
SVM	168	59.45

Fig. 3 shows the classification accuracy of the ML/DL technique over datasets related to Cleveland heart disease without the use of feature selection.

Figure 3's values, which show that the NB technique performs better than other practices under consideration, accurately identified 254 out of the 303 cases. The same experiment has been repeated using the recommended features set selection technique in order to confirm the effectiveness of ANN and determine the influence of include feature selection practices. Table IV and Figure 4 illustrate the experimental outcomes.

Table 4. Forecasting accuracy for diseases related to dermatology using datasets containing proposed features by the CFS, GR, and IG schemes

Classifiers	Prediction Accuracy (%)		
	With CFS Selection Feature set	With Gain Ratio (GR) Feature set	With IG Feature set
ANN	83.83	81.17	83.17
RBFN	82.19	78.57	76.57
LM	79.54	79.21	80.21
SVM	81.51	77.57	76.57

The achieved values of the algorithms under consideration for the forecasting of dermatology disorders unambiguously demonstrate how effective it is to include feature selection practices. SVM may improve fallout by attempting to retrieve the most appropriate feature selection practice as it presents a better act with the recommended feature set by each practice. The graphical comparison of the evaluated methodologies is shown in Figure 4.

Breast cancer datasets are a naïve illnesses data set that has been used in another evaluation practice to confirm the effectiveness of the methodologies under consideration. There are 366 instances of the 10 attributes in this data collection. Similar to the previous assessment scheme, the entire collection of procedures has been assessed here in both ways, starting with the entire Only the recommended feature set provided by taken into consideration practices should be included in the assessment forecasting process after the feature selection procedure is included. feature set and then evaluating the predicting act again.
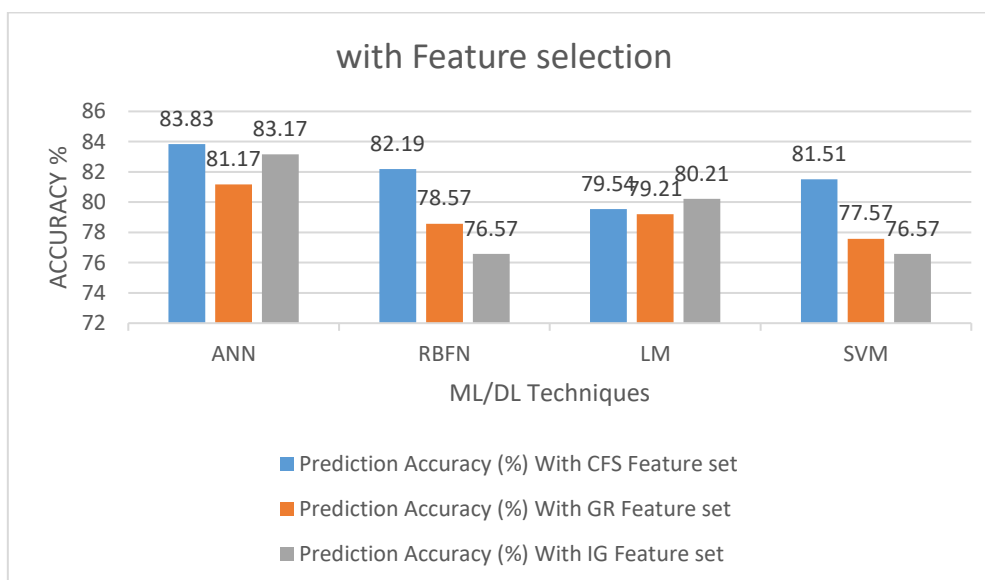

Figure 4: ML/DL approach classification accuracy over Cleveland heart disease datasets with feature selection practice included

Table 5 presented the evaluative values of the implemented technique. The table indicates the performance of the implemented algorithms to help identify the differences between each approach's feature selection classification act and its absence.

Table 5. Breast Cancer Forecasting Accuracy

Classifiers	Prediction Accuracy (%)			
	With CFS Feature set	With Gain Ratio (GR)Feature set	With Info Gain (IG)Feature set	With Full and all Feature Set
ANN	73.08	71.73	72.73	72.38
RBFN	73.78	73.08	73.08	73.08
LM	75.17	70.63	69.63	71.68
SVM	72.38	69.62	70.62	69.23

The differences between implementing algorithms without considering the characteristics of investigated techniques and using several feature selection methodologies are demonstrated by the numbers in Table 5. Figure 5 shows the compared graphical display of the assessed fallouts. Diabetes, which is recognized as one of the most chronic diseases in recent times, has been the subject of additional simulation work using a new dataset. There are 286 instances and 10 attributes in the collection. The entire set of categorization techniques that were taken into consideration for this paper's assessment was put into practice during the same phase as the previous evaluation. Table 6 showed the relative drawbacks of every assessed strategy. Nevertheless, no single strategy has yet been identified in the executed and assessed attempts that indicated his superiority over other simulation processes.
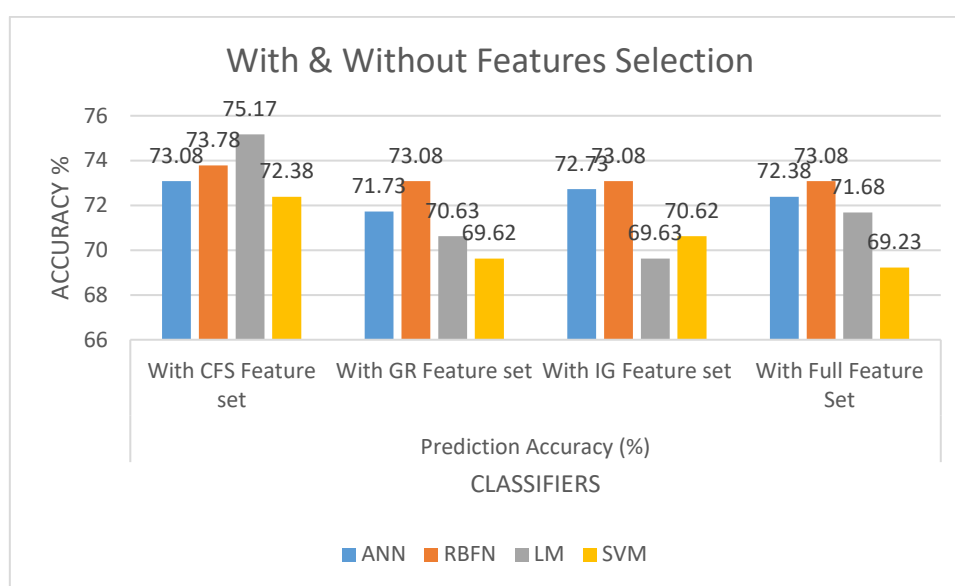

Fig. 5 Accurate classification of breast cancer datasets using ML/DL algorithm with and without feature assortment practice included

Classifiers	Prediction Accuracy (%)			
	With CFS Feature set	With Gain Ratio (GR) Feature set	With Info Gain (IG)Feature set	With Full Feature Set
ANN	75.52	74.87	75.39	74.35
RBFN	74.87	74.86	74.35	73.83
LM	75.52	76.30	76.69	75.39
SVM	62.37	64.58	64.58	65.10

Table 6. Diabetes Forecasting Accuracy

Table 6 illustrates the comparative fallouts, which highlight the advantages of incorporating a feature selection approach to improve illness recognition. But no single feature assortment strategy has been shown to be effective in providing the best selection of features for every dataset and enhancing the precision effectiveness of all methods. Fallouts indicated that although one technique performed better when using the recommended feature set of the methodology under consideration, it was unable to demonstrate its effectiveness because the

evaluative method or the dataset were altered. There are 403 instances and 9 attributes in the collection. The insufficiency of current approaches indicated a research void in this domain, drawing the attention of inexperienced researchers. The graphical depiction of the achieved accuracy by each was shown in figure 6.
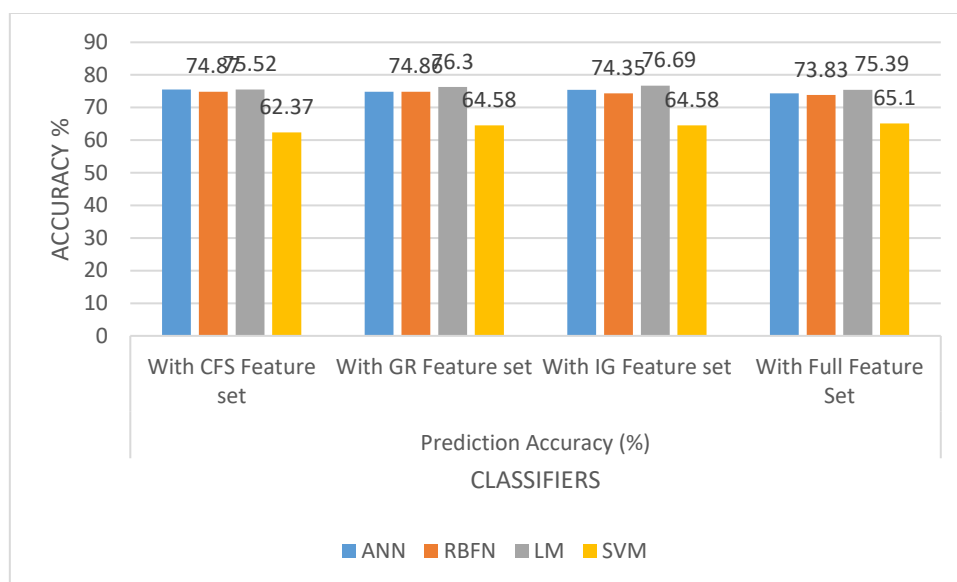

Fig. 6 Accurate classification of Diabetes datasets using ML/DL approach with and without feature assortment practice included

In the final stage of evaluating the effectiveness of the ML/DL methodology, a naïve illnesses dataset was considered. The silent killer of human life is the condition known as hypothyroidism. The 2800 records in the dataset each have a set of 27 features. The subsequent table

Table 7. Thyroid Diseases Forecasting Accuracy

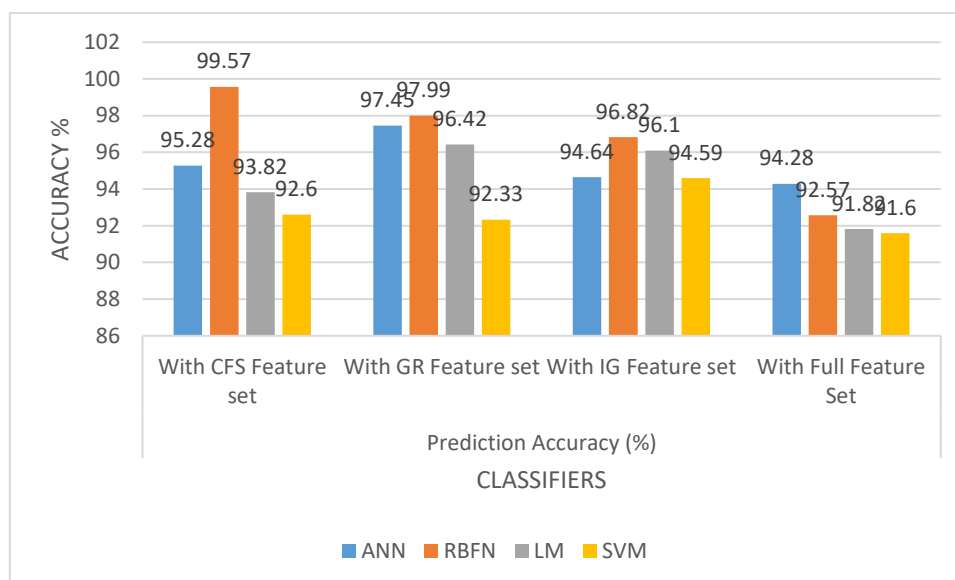| Classifiers	Prediction Accuracy (%)			
	With CFS Feature set	With GR Feature set	With IG Feature set	With Full Feature Set
ANN	95.28	97.45	94.64	94.28
RBFN	99.57	97.99	96.82	92.57
LM	93.82	96.42	96.10	91.82
SVM	92.60	92.33	94.59	91.60

Fig. 7 Accurate classification across hypothyroid datasets with and without feature selection practices included

5. CONCLUSION

Numerous readable studies have shown that ML/DL is a useful method for predicting health problems. A big community reported that when the quantities of medical datasets increased, the accessibility algorithms got worse. Efforts suggest that the methods provide improved prediction performance when combined with feature selection procedures. This work investigates the prediction of real illnesses using machine learning techniques such as ANN, RBFN, LM, and SVM. An investigation study using four distinct disease dataset sizes—heart, breast cancer, diabetes, and thyroid datasets—was carried out for this paper. To investigate the effectiveness of incorporating feature selection methods, an evaluation was conducted in two stages. The findings demonstrate that no approach can consistently outperform, and that performance varied when simulative parameters changed. These inadequacies highlight the need for additional study in this field.

6. REFERENCES

1. TEstes, C.; Anstee, Q.M.; Arias-Loste, M.T.; Bantel, H.; Bellentani, S.; Caballeria, J.; Colombo, M.; Craxi, A.; Crespo, J.; Day, C.P.; et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. J. Hepatol. 2018, 69, 896–904.
2. Droz˙dz˙ , K.; Nabrdalik, K.; Kwiendacz, H.; Hendel, M.; Olejarz, A.; Tomasik, A.; Bartman,W.; Nalepa, J.; Gumprecht, J.; Lip, G.Y.H. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. Cardiovasc. Diabetol. 2022, 21, 240.

3. S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," Journal of Intelligent Learning Systems and Applications, vol. 5, no. 3, 2013, pp. 176–183.

4. D. Al-Jumeily, S. Iram, F. B. Vialatte, P. Fergus, and A. Hussain, "A novel method of early diagnosis of Alzheimer's disease based on EEG signals," Scientific World Journal, vol. 2015, article 931387, 11 pages, 2015.

5. E. Keogh and A. Mueen, "Curse of dimensionality," in Encyclopedia of Machine Learning and Data Mining, C. Sammut and G. I. Webb, Eds., Springer, Cham, Switzerland, 2017.

6. D. Wettschereck and T. G. Dietterich, "An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms," Machine Learning, vol. 19, no. 1, pp. 5–27, 1995.

7. D. Wettschereck, D. W. Aha, and T. Mohri, "A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms," Lazy Learning, vol. 11, no. 1/5, pp. 273–314, 1997.

8. M.-S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," IEEE Transactions on Fuzzy Systems, vol. 26, no. 2, pp. 817–835, 2018.

9. R. Chen, N. Sun, X. Chen, M. Yang, and Q. Wu, "Supervised feature selection with a stratified feature weighting method," IEEE Access, vol. 6, pp. 15087–15098, 2018.

10. M. Imani and H. Ghassemian, "Feature extraction using weighted training samples," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 7, pp. 1387–1391, 2015.

11. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.

12. Gietzelt, M.;Wolf, K.-H.; Marschollek, M.; Haux, R. Performance comparison of accelerometer calibration algorithms based on 3D-ellipsoid fitting methods. Comput. Methods Programs Biomed. 2013, 111, 62–71.

13. K, V.; Singaraju, J. Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks. Int. J. Comput. Appl. 2011, 19, 6–12.

14. Waigi, R.; Choudhary, S.; Fulzele, P.; Mishra, G. Predicting the risk of heart disease using advanced machine learning approach. Eur. J. Mol. Clin. Med. 2020, 7, 1638–1645.

15. Shalini, Saini, P.K., Sharma, Y.M. (2021). An Intelligent Hybrid Model for Forecasting of Heart and Diabetes Diseases with SMO and ANN. In: Shorif Uddin, M., Sharma, A., Agarwal, K.L., Saraswat, M. (eds) Intelligent Energy Management Technologies. Algorithms for Intelligent Systems. Springer, Singapore.

16. Sharma, Y.M., Saini, P.K., Shalini, Sharma, N. (2021). Effective Decision Support Scheme Using Hybrid Supervised Machine Learning Procedure. In: Goyal, D., Gupta, A.K., Piuri, V., Ganzha, M., Paprzycki, M. (eds) Proceedings of the Second International Conference on Information Management and Machine Intelligence. Lecture Notes in Networks and Systems, vol 166. Springer, Singapore.

17. Jaiswal, O., Saini, P.K., Shalini, Sharma, Y.M. (2021). Analyze Classification Act of Data Mining Schemes. In: Goyal, D., Gupta, A.K., Piuri, V., Ganzha, M., Paprzycki, M. (eds) Proceedings of the Second International Conference on Information Management and Machine Intelligence. Lecture Notes in Networks and Systems, vol 166. Springer, Singapore.

18. Shah, D.; Patel, S.; Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN Comput. Sci. 2020, 1, 345.

19. I. Ketut Agung Enriko, M. Suryanegara, and D. Agnes Gunawan, Heart Disease Prediction System Using K-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters, Springer, Berlin, Germnay, 2016.

20. K. Subhadra and B. Vikas, "Neural network based intelligent system for predicting heart disease," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 5, pp. 484–487, 2019.

21. S. N. Khan, N. M. Nawi, A. Shahzad, A. Ullah, and M. F. Mushtaq, "Comparative analysis for heart disease prediction," International Journal on Informatics Visualization, vol. 1, no. 4-2, pp. 227–231, 2019.

22. S. Anitha and N. Sridevi, "Heart disease prediction using data mining techniques," Journal of Analysis and Computation, vol. 8, no. 2, pp. 48–55, 2019.

23. A. Jagtap, P. Malewadkar, O. Baswat, and H. Rambade, "Heart disease prediction using machine learning," International Journal of Research in Engineering, Science and Management, vol. 2, no. 2, pp. 352–355, 2019.

24. H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," Expert Systems with Applications, vol. 35, no. 1–2, pp. 82–89, 2008.

25. G. Parthiban and K. Srivatsa, "Applying machine learning methods in diagnosing heart disease for diabetic patients," International Journal of Applied Information Systems, vol. 3, no. 7, pp. 25–30, 2012.

26. B. Dun, E. Wang, and S. Majumder, Heart disease diagnosis on medical data using ensemble learning, Stanford Publisher, vol. 1, no. 1, pp. 1–5, 2016,

27. R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," Computational Intelligence and Neuroscience, vol. 2021, Article ID 8387680, 11 pages, 2021.

28. Rashmi G Saboji, "A scalable solution for heart disease prediction using classification mining technique", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).

29. Sarath Babu, E M Vivek, K P Famina, K Fida, P Aswathi, M Shanid, M Hena, "Heart disease diagnosis using data mining technique", 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA).

30. Chaitanya Suvarnam, Abhishek Sali, Sakina Salmani, "Efficient heart disease prediction system using optimization technique", 2017 International Conference on Computing Methodologies and Communication (ICCMC).

31. Seyedamin Pouriyeh,Sara Vahid,Giovanna Sannino,Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez,"A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease",2017 IEEE Symposium on Computers and Communications (ISCC).

32. D. Sisodia, D.S. Sisodia, Prediction of diabetes using classification algorithms, Procedia Comput. Sci. 132 (2018) pp. 1578–1585.
33. N.P. Tigga, S. Garg, Predicting type 2 Diabetes using Logistic Regression accepted to publish in: Lecture Notes of Electrical Engineering, Springer.
34. Salim Amour Diwani, Anael Sam, Diabetes forecasting using supervised learning techniques, Adv. Comput. Sci.: Int. J. [S.l.] (ISSN: 2322-5157) (2014) pp. 10–18.
35. Shorewala, V. Early detection of coronary heart disease using ensemble techniques. Inform. Med. Unlocked 2021, 26, 100655
36. Maiga, J.; Hungilo, G.G.; Pranowo. Comparison of Machine Learning Models in Prediction of Cardiovascular Disease Using Health Record Data. In Proceedings of the 2019 International Conference on Informatics, Chronicmedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 24–25 October 2019; pp. 45–48.
37. Waigi, R.; Choudhary, S.; Fulzele, P.; Mishra, G. Predicting the risk of heart disease using advanced machine learning approach. Eur. J. Mol. Clin. Med. 2020, 7, 1638–1645.
38. Khan, I.H.; Mondal, M.R.H. Data-Driven Diagnosis of Heart Disease. Int. J. Comput. Appl. 2020, 176, 46–54.
39. Ouf, S.; ElSeddawy, A.I.B. A proposed paradigm for intelligent heart disease prediction system using data mining techniques. J. Southwest Jiaotong Univ. 2021, 56, 220–240.
40. Kumari, V.A. and R. Chitra, "Classification of Diabetes Disease Using Support Vector Machine", International Journal of Engineering Research and Applications, vol.3, pp. 1797-1801, 2013.
41. Sarwar, A. and Sharma, V. , "Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2. Special Issue", International Journal of Computer Applications and Challenges in Networking, Intelligence and Computing Technologies, vol. 3, pp.14-16, 2012.
42. Shrivastava et al., "An Overview on Data Mining Approach on Breast Cancer data", International Journal of Advanced Computer Research, vol.3, pp. 256-262, 2013.
43. K. Sivakami, "Mining Big Data: Breast Cancer Prediction using DTSVM Hybrid Model", International Journal of Scientific Engineering and Applied Science vol.1, pp.418-429, 2015.
44. M. R. Nazari Kousarrizi et al., "An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification", International Journal of Electrical& Computer Sciences, vol.1,pp.13-19,2012.
45. Cheruku, R., Edla, D.R. and Kuppili, V. (2017) 'Diabetes classification using radial basis function network by combining cluster validity index and bat optimization with novel fitness function', International Journal of Computational Intelligence Systems, 10(1), p. 247. doi:10.2991/ijcis.2017.10.1.17.
46. Yadav, D.K. et al. (2022) 'Genetic algorithm and Gaussian Radial Basis Function Network (GAGRBFN) based diabetes mellitus prediction system', Lecture Notes in Networks and Systems, pp. 279–286. doi:10.1007/978-981-16-4807-6_27.