Multi-Disease Prediction Method Based on Machine Learning and Computational Approaches

Tarun Kumar Agarwal[1], Baldev Singh[2], Nilam Choudhary[3]

[1]Research Scholar, Department of Computer Science & Engineering VGU, Jaipur, Rajasthan, India
[2]Professor, Department of Computer Science & Engineering VGU, Jaipur, Rajasthan, India
[3]Associate Professor, Department of Computer Science & Engineering SKIT, Jaipur, Rajasthan, India

Abstract— In today's era, various life-threatening illnesses such as those related to the heart, diabetes, hypothyroid, breast cancer, and others are rapidly increasing the rate of sudden deaths. Early detection is crucial for saving lives and preventing severe illnesses. While existing research has shown promising forecasting rates ranging from 75% to 92%, there remains an accuracy gap of 8% to 25% to be addressed. This paper introduces a novel approach to multi-disease forecasting, aiming to significantly improve prediction accuracy through the collaborative application of machine learning (ML) techniques. Our proposed method analyses the predictive capabilities of five different ML algorithms: Hidden Markov Model (HMM), Support Vector Machine (SVM), DTNB (an ensemble algorithm combining Naïve Bayes and decision trees), Artificial Neural Network (ANN), and Radial Basis Function Network (RBNF). By leveraging the complementary strengths of these algorithms, our approach mechanically combines the forecasting abilities of top two well matched algorithms into a layered structure, enabling data analysis at multiple stages. To evaluate our method, we utilize diverse disease datasets sourced from the well-established UCI benchmarked open access research repository. The comparative experimental results demonstrate the effectiveness of our proposed approach, indicating significant improvements in disease forecasting accuracy.

***Keywords*—**Machine Learning (ML), Multi Diseases, Classification, AI, SVM,DTNB,RBNF,HMM,ANN

1. INTRODUCTION

Foreseeing illnesses is an intricate undertaking that frequently includes the investigation of different factors like clinical history, hereditary qualities, way of life decisions, ecological elements, from there only sky is the limit. However, in recent era various ML procedures can be utilized to assemble prescient models for specific illnesses. Nonetheless, it's critical to take note of that sickness expectation isn't generally precise, and forecasts ought to be deciphered circumspectly and related to clinical experts. The fundamental elements of this research are represented in this chapter. It commences with an introduction that demonstrates the needs and benefits of diseases forecasting scheme, follows by a list of current issues, and ends with the arrangement of the chapter organization of this thesis.

1.1 OVERVIEW

In the present period of innovation and the Web, individuals don't focus on their wellbeing or prosperity. Because of work and unanticipated conditions, ordinary visits to the clinic for check-ups are phenomenal, this has prompted an expansion in ailment among individuals. These days, most of individuals experience the ill effects of various ongoing medical issue like heart, diabetes, kidney infection. The World Wellbeing Association (WHO) considers such ongoing sicknesses are the main sources of unexpected passing, answerable for half of all passings in many nations. Early distinguishing proof of patients can keep the sickness from advancing. Figure 1 represent the usual approach of diseases diagnosis and for the treatment.

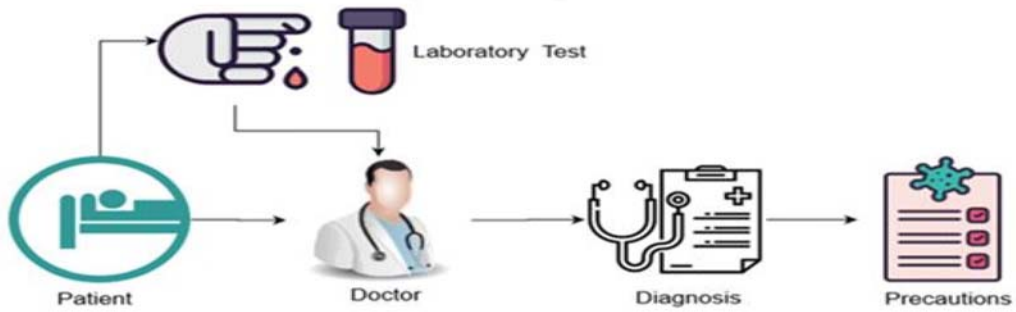

Figure 1. General Diseases Diagnosis & Treatment Suggestion Approach

1.2 Aids & Needs of ML Strategies in the Medical services Area

Machine learning techniques have completely changed the way that medical services are provided. They provide a number of advantages that fundamentally enhance care cycles and patient consideration. The ability to facilitate early sickness recognition and finding is one of the top benefits. Massive and complex patient records can be efficiently broken-down using ML algorithms, which can also spot subtle patterns and anomalies that can elude human awareness.

2. Literature Review

2.1 Settled Practices for Diseases Detection

Since ML was first introduced into the healthcare industry, numerous ML models have attempted to enhance disease detection procedures. To expand a framework's power, every technique for ML puts forth particular attempts. However, the application's nature, type, and associated characteristics greatly influence each method's suitability; A technique that is effective in one context may not be effective in another. The few approaches to creating an effective prediction model from the various methods is of particular interest to researchers. Figure 2.6 presents a discussion of the most well-known ML schemes that have been extensively utilized to construct diseases prediction models both in the past and in the present.

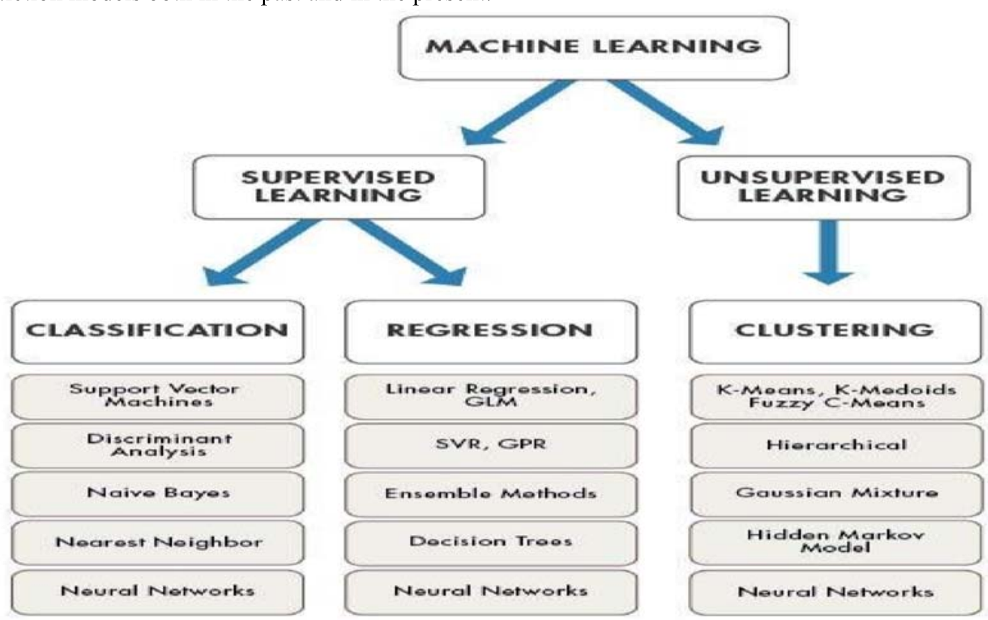

Figure 2. Most Common Practices for Diseases Detection

As shown in Figure 2.6, numerous researchers have developed a wide array of algorithms aimed at assisting medical professionals in achieving highly precise disease diagnoses. Each of these approaches has strived to enhance diagnostic accuracy. However, it is not feasible to illustrate every single algorithm in this chapter. Therefore, the literature review in this research work focuses on a select few methods chosen for analysis and the development of the proposed approach. For this investigative work we have consider a disease forecasting act of 5 different ML algorithms recognized as Hidden-Markov model (HMM, an unsupervised ML algorithm), SupportVector-Machine (SVM, Supervised Algorithm), DTNB (a form of ensemble algorithm of Naïve Bayes and the decision tree), Artificial-Nural-Network (ANN) and the deep learning algorithm Radial Basis Function Network (RBNF). Additionally, some researchers have found ways to increase the effectiveness of the construction process by including the Gain-Ratio (GR) feature selection scheme and the Correlation based Feature Selection (CFS) method, both of techniques are studied in this work. For each of the chosen methods, the relevant literature has been included to demonstrate the method's suitability, advantages, and associated challenges.

TABLE I. Recent Forecasting Efforts of Multi Diseases by utilizing of ML Techniques

Investigators	ML Technique	Achieved Success	Diseases
V.A. and R. Chitra [40]	SVM	78%	Diabetes
V. Sharma [41]	NB	95%	
K. Sivakami,[42]	Hybrid	91%	Breast Cancer
Shrivastav [43]	Decision Tree	98.14%	
M. R.	Support Vector	98.62%	Thyroid
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	

3. Methodology

3.1 Process Flow Framework & Employed Methodology Of Intended Approach

The operational framework of the proposed algorithm put primarily focused on mitigating the limitations of existing algorithms. In essence, the proposed approach places significant importance on minimizing false alarms by selecting and integrating optimal forecasting features and a feature selection process into a unified method, thereby enhancing forecasting accuracy for informed and more efficient decision-making. The procedural steps taken in this research are illustrated in the subsequent figures.

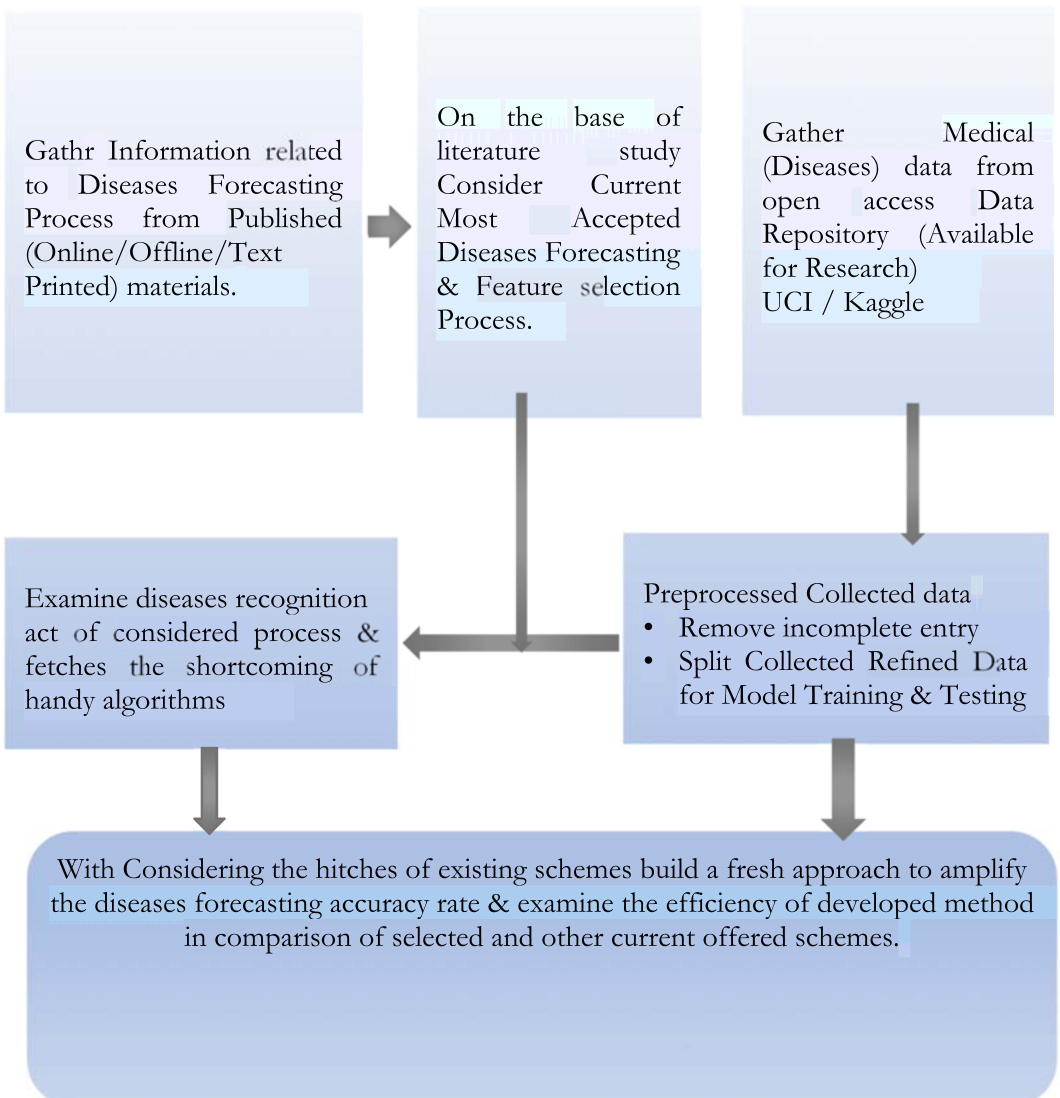

Figure 3. Process flow diagram

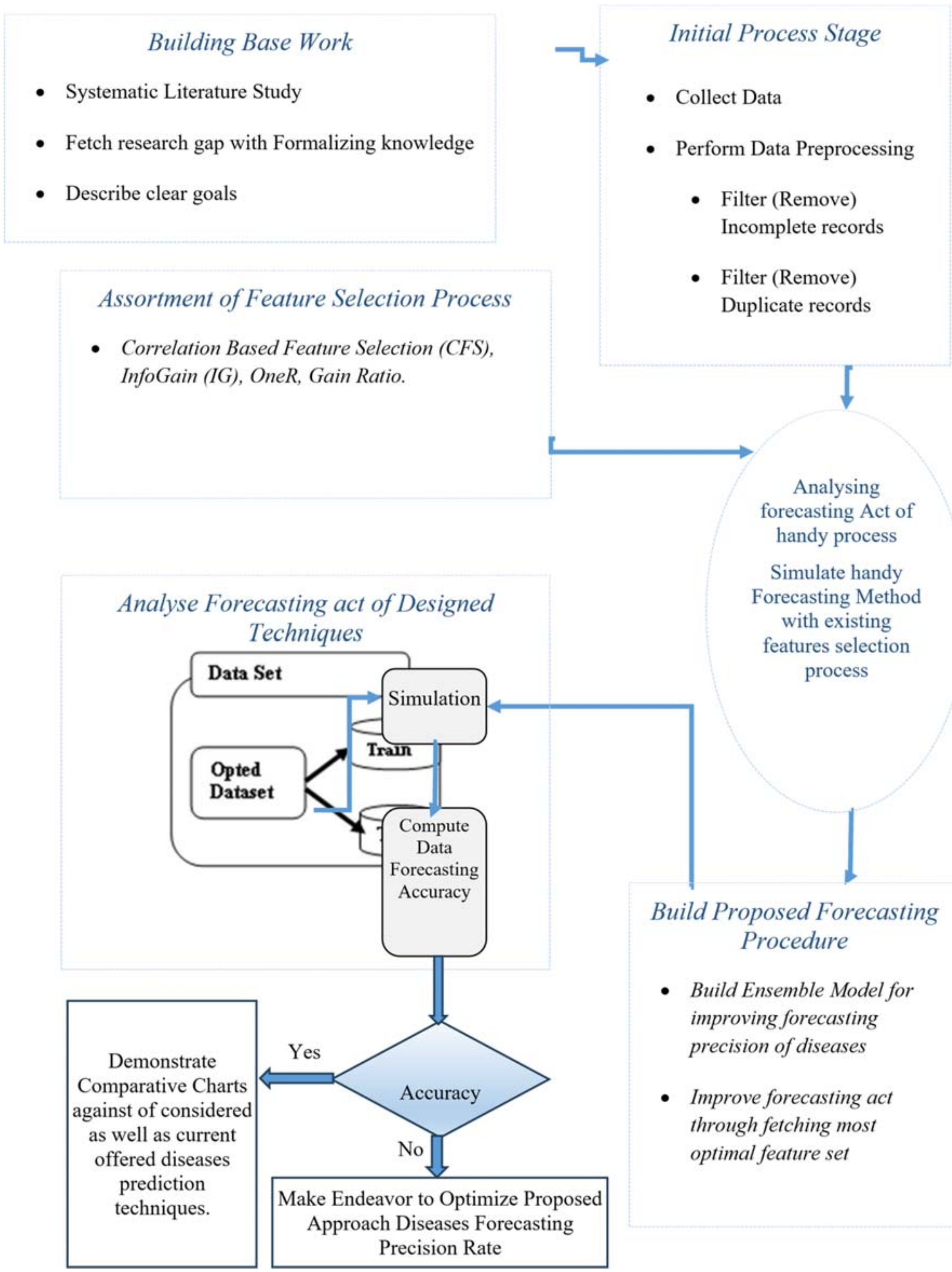

Figure 4. Process Flow Framework

We Can say a hybrid method that executes in a layer format, executes one after the other in a way that outcome of first layer is use as the input of next layer, depicted in figure 3.2.

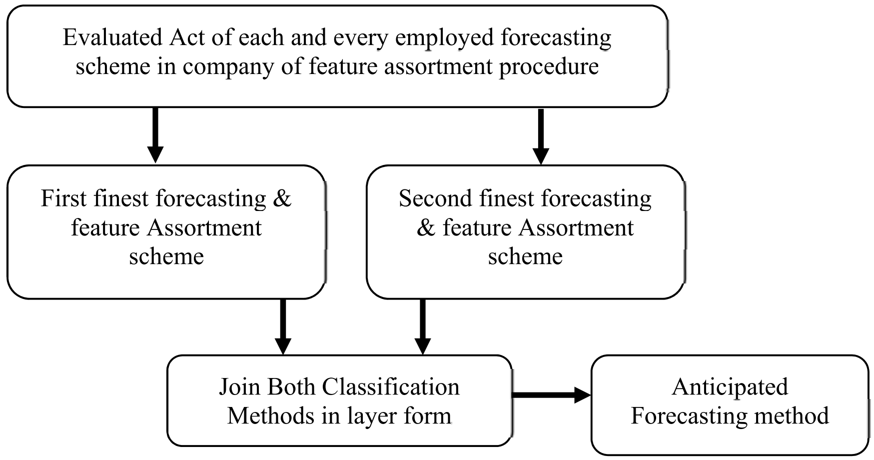

Figure 5. Anticipated Actions for Building Hybrid Forecasting Scheme

As depicted in the above figure 3.2, the suggested strategy uses distinct forecasting techniques for each layer, which may or may not follow the same feature assortment process.

3.2 Collection of Medical Statistics

Initially, we collect vital information on the disease; medical datasets used in this research are sourced from openaccess sites like UCI and Kaggle.

TABLE 2. Exploited Dataset for Testing an Act of Proposed Approach

S. No.	Name of Diseases Datasets	No. of Associated Attributes	Medical Statistics		
			Normal	Diseases Affected	Total
1.	Heart Diseases Data	14	171	132	303
2.	Pima Indians Diabetes Database	9	500	268	768
3.	Dermatology Diseases	35	112	254	366
4.	Breast-cancer Dataset	10	201	85	286
5.	Hypothyroid Diseases	30	3481	291	3772

4. EXPERIMENT SETUP & RESULTS

For recognizing the fair compete act of proposed approach it has evaluated along with 5 other classical forecasting techniques of heart diseases over considered dataset. Whole the considered techniques i.e. HMM, SVM, DTNB, ANN and the RBFN has evaluated with the suggested feature set of each accountable feature selection schemes i.e. CFS, GR, IG and OneR. Such process has done to attain the finest precision amount of each evaluated algorithm.

TABLE 3. Comparative Heart Diseases Forecasting Accuracy of Assessed Methods

Evaluating Approaches	Accuracy (%) With Suggested Features set			
	CFS	GR	IG	OneR
HMM	***85.149***	***83.169***	82.189	82.839
SVM	81.189	79.868	81.849	71.288
DTNB	84.159	83.169	84.489	***84.839***
ANN	83.499	79.538	***82.509***	80.859
RBFN	81.849	82.509	82.509	83.169

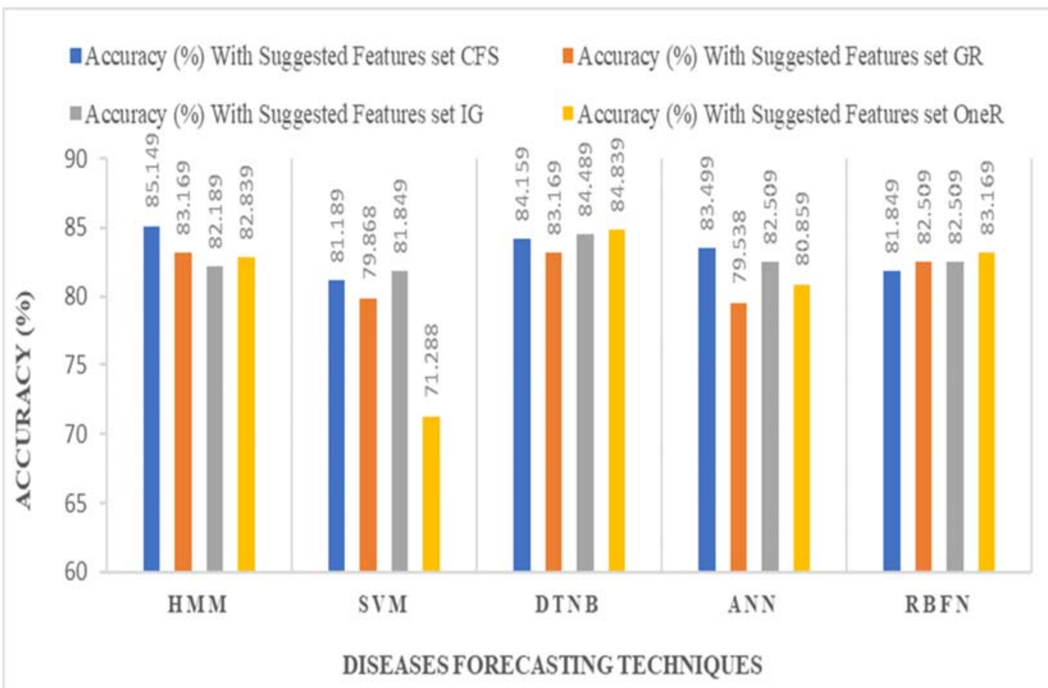

Figure 6. Heart Diseases Forecasting Accuracy of Assessed Methods

Table 4. Heart Diseases Forecasting Accuracy: Proposed Vs Classical Techniques

S.No.	Assessed Diseases Forecasting Technique	Accuracy (%)
1	HMM	85.149
2	SVM	81.849
3	DTNB	84.489
4	ANN	83.499
5	RBFN	83.169
6	**Proposed Technique**	**97.00**

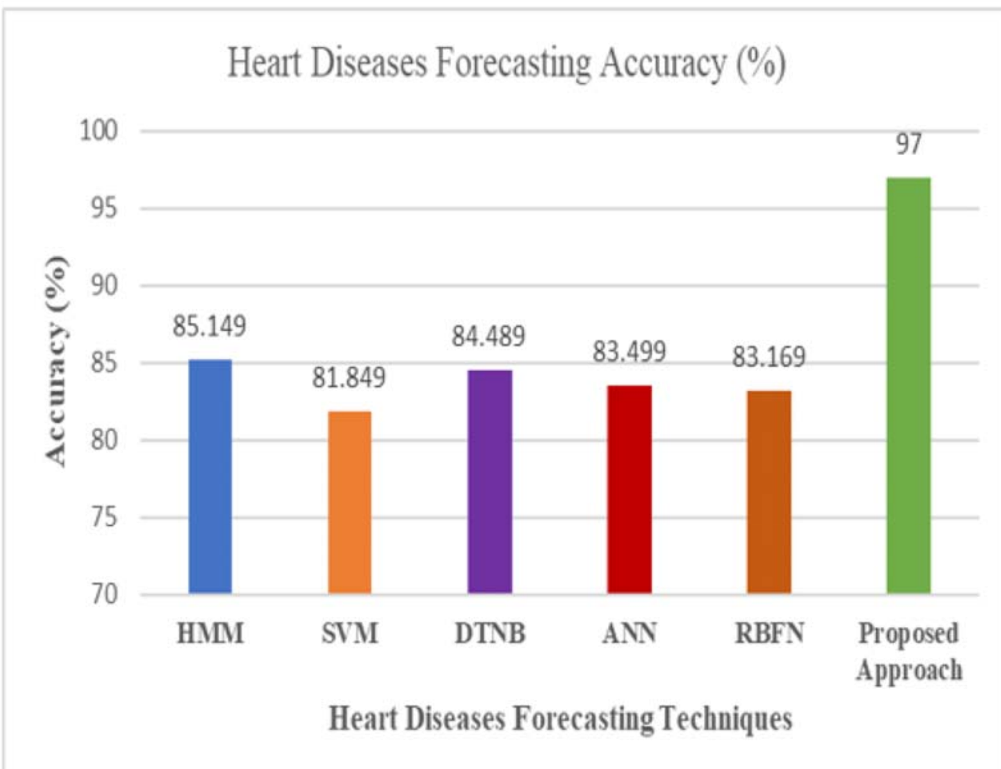

Figure 7. Heart Diseases Forecasting Accuracy: Proposed Vs Classical Techniques

TABLE 5. Dermatology Diseases Forecasting Accuracy of Assessed Methods

Evaluating Approaches	Accuracy (%) With Suggested Features set			
	CFS	GR	IG	OneR
HMM	97.541	83.334	77.323	97.268
SVM	98.088	84.7	78.416	97.268
DTNB	97.541	83.607	77.596	96.995
ANN	96.449	84.154	76.23	95.629
RBFN	96.995	84.427	78.143	98.088

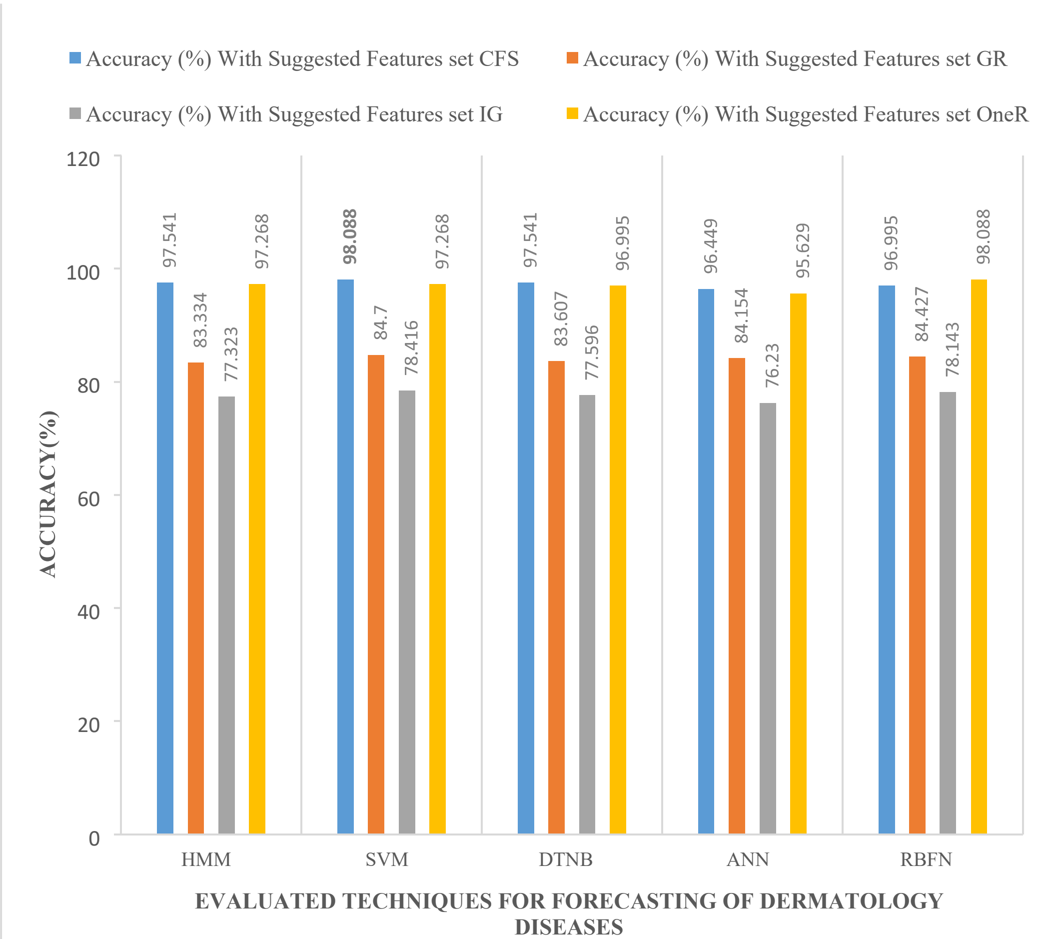

Figure 8. Dermatology Diseases Forecasting Accuracy of Assessed Methods

Table 6. Forecasting Accuracy of Dermatology: Proposed Vs Classical Methods

S. No.	Diseases Forecasting Technique	Accuracy (%)
1	HMM	97.541
2	SVM	98.088
3	DTNB	97.541
4	ANN	96.449
5	RBFN	98.088
6	**Proposed**	**100**

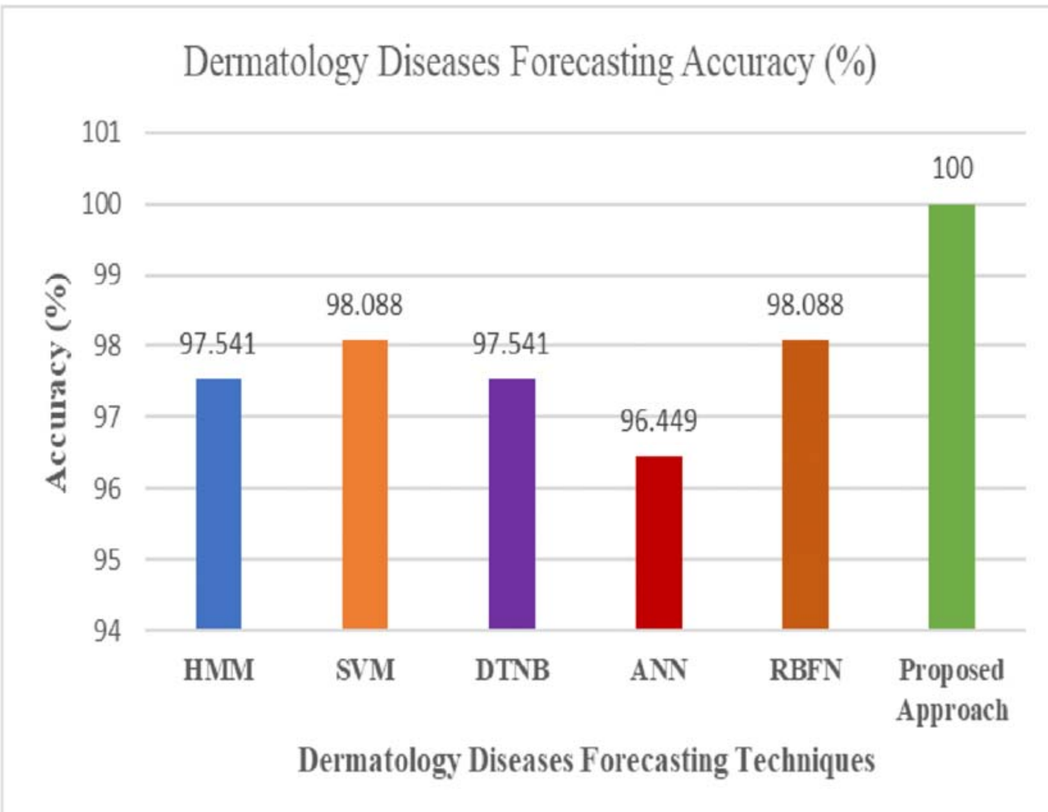

Figure 9. Forecasting Accuracy of Dermatology: Proposed Vs Classical Methods

TABLE 7. Hypothyroid Diseases Forecasting Accuracy of Evaluated Approaches

Evaluating Approaches	Accuracy (%) With Suggested Features set			
	CFS	GR	IG	OneR
HMM	98.039	97.906	97.906	**99.55**
SVM	95.175	97.668	97.668	94.672
DTNB	95.122	94.99	94.99	**98.039**
ANN	96.899	92.286	*92.286*	95.282
RBFN	95.069	95.149	95.149	95.043

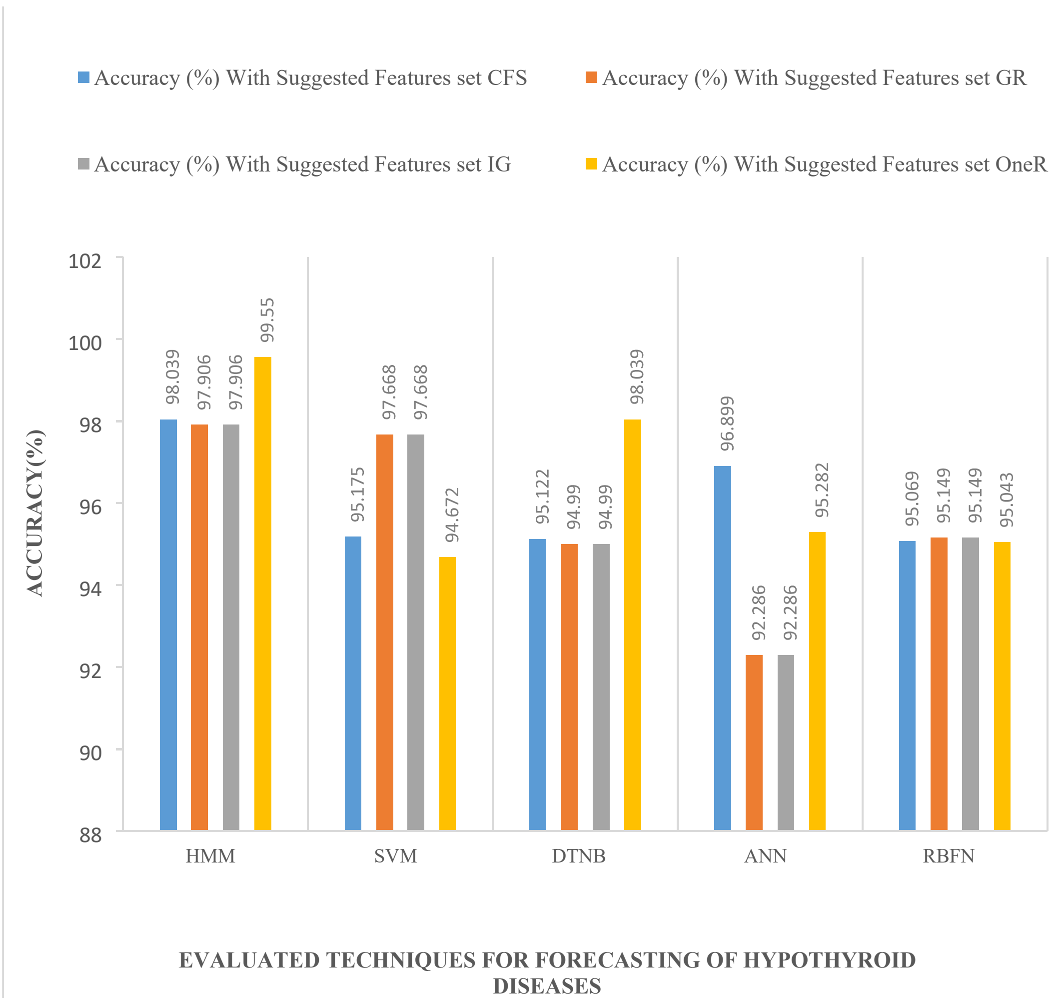

Figure 10. Hypothyroid Diseases Forecasting Accuracy of Evaluated Approaches

Table 8. Forecasting Accuracy Of Hypothyroid: Proposed Vs Standard Methods

S. No.	Diseases Forecasting Technique	Accuracy (%)
1	HMM	99.55
2	SVM	97.688
3	DTNB	*98.039*
4	ANN	96.899
5	RBFN	95.149
6	Proposed	100

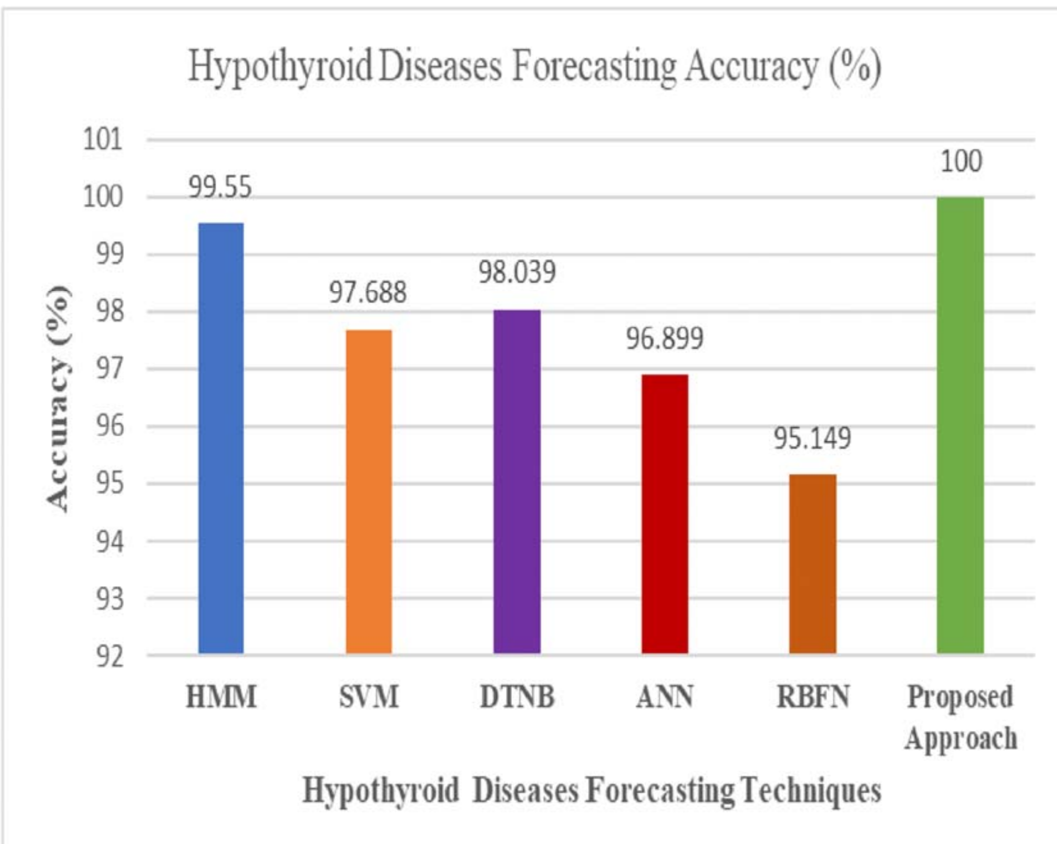

Figure 11. Forecasting Accuracy of Hypothyroid: Proposed Vs Standard Methods

TABLE 9. Diabetes Diseases Forecasting Accuracy of Evaluated Approaches

Evaluating Approaches	Accuracy (%) With Suggested Features set			
	CFS	GR	IG	OneR
HMM	**77.214**	75.521	74.87	76.433
SVM	62.37	63.282	71.485	65.105
DTNB	76.563	76.042	73.568	74.48
ANN	75.521	76.433	74.089	75.912
RBFN	76.563	76.433	73.698	75.652

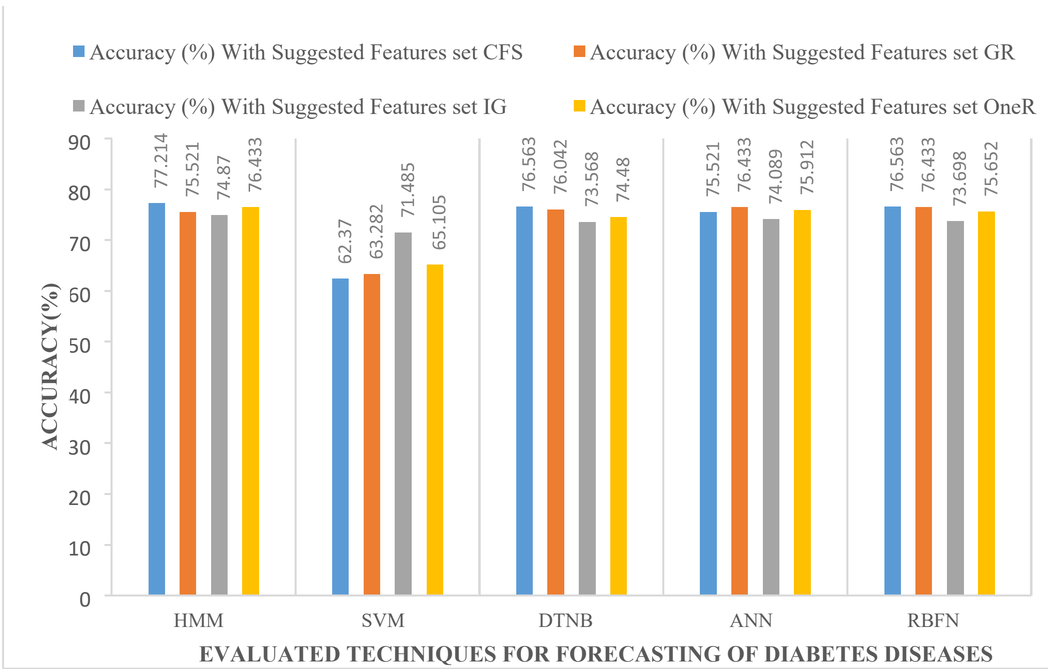

Figure 12. Diabetes Diseases Forecasting Accuracy of Evaluated Approaches

Table 10. Forecasting Accuracy of Diabetes: Proposed Vs Standard Methods

S. No.	Diseases Forecasting Technique	Accuracy (%)
1	HMM	77.214
2	SVM	71.485
3	DTNB	76.563
4	ANN	76.433
5	RBFN	76.563
6	Proposed	95

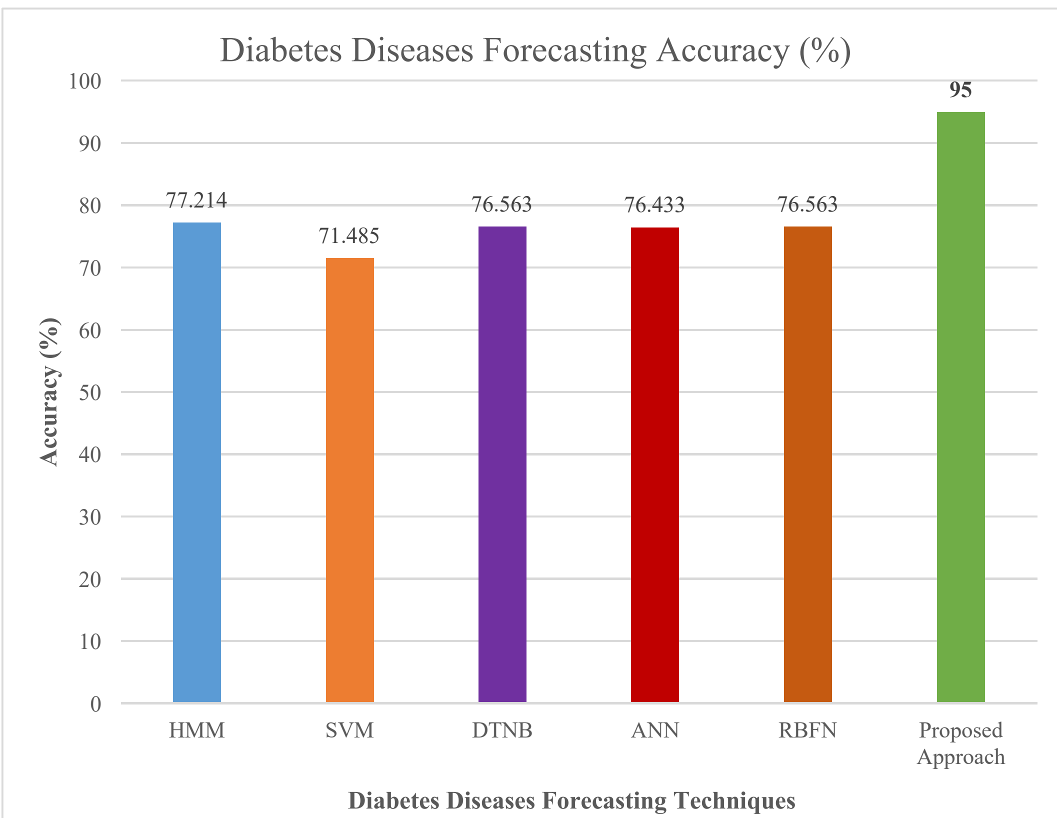

Figure 13. Forecasting Accuracy of Diabetes: Proposed Vs Standard Methods

TABLE 11. Breast Cancer Forecasting Accuracy of Evaluated Approaches

Evaluating Approaches	Accuracy (%) With Suggested Features set			
	CFS	GR	IG	OneR
HMM	75.525	69.931	**76.224**	75.525
SVM	69.231	72.378	70.63	70.28
DTNB	73.777	70.98	**76.224**	73.777
ANN	71.679	69.581	73.427	68.182
RBFN	71.329	72.378	75.875	70.98

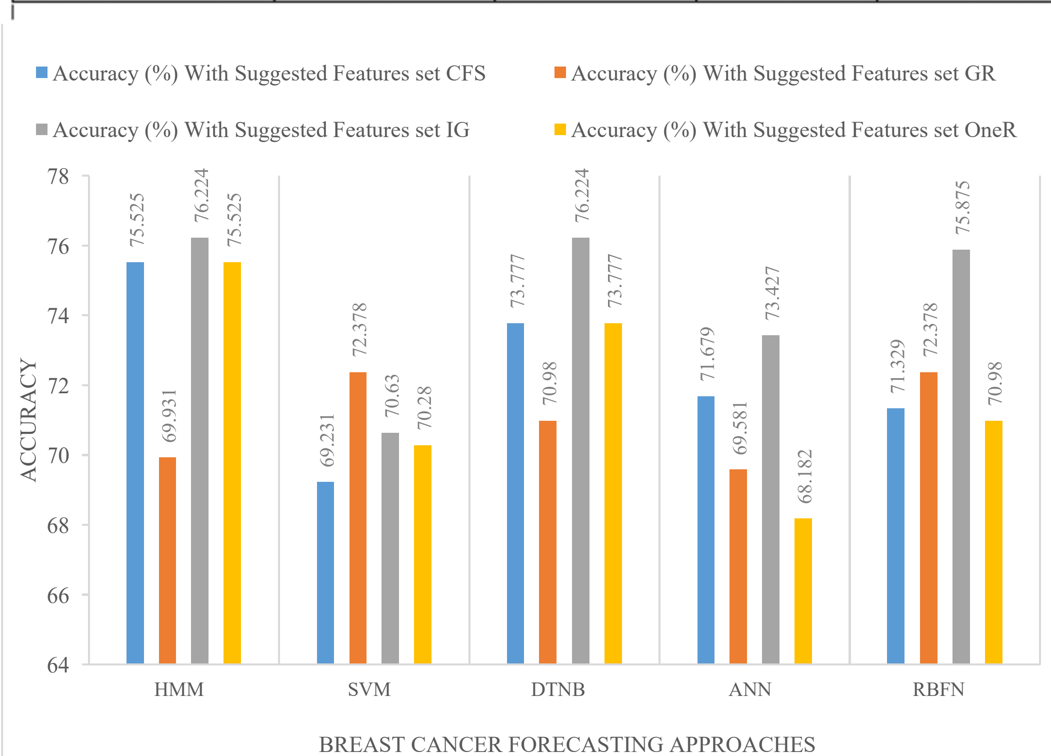

Figure 14. Breast Cancer Forecasting Accuracy of Evaluated Approaches

Table 12. Forecasting Accuracy of Breast Cancer: Proposed Vs Standard Methods

S. No.	Diseases Forecasting Technique	Accuracy (%)
1	HMM	76.224
2	SVM	72.378
3	DTNB	76.224
4	ANN	73.427
5	RBFN	75.875
6	**Proposed**	**98**

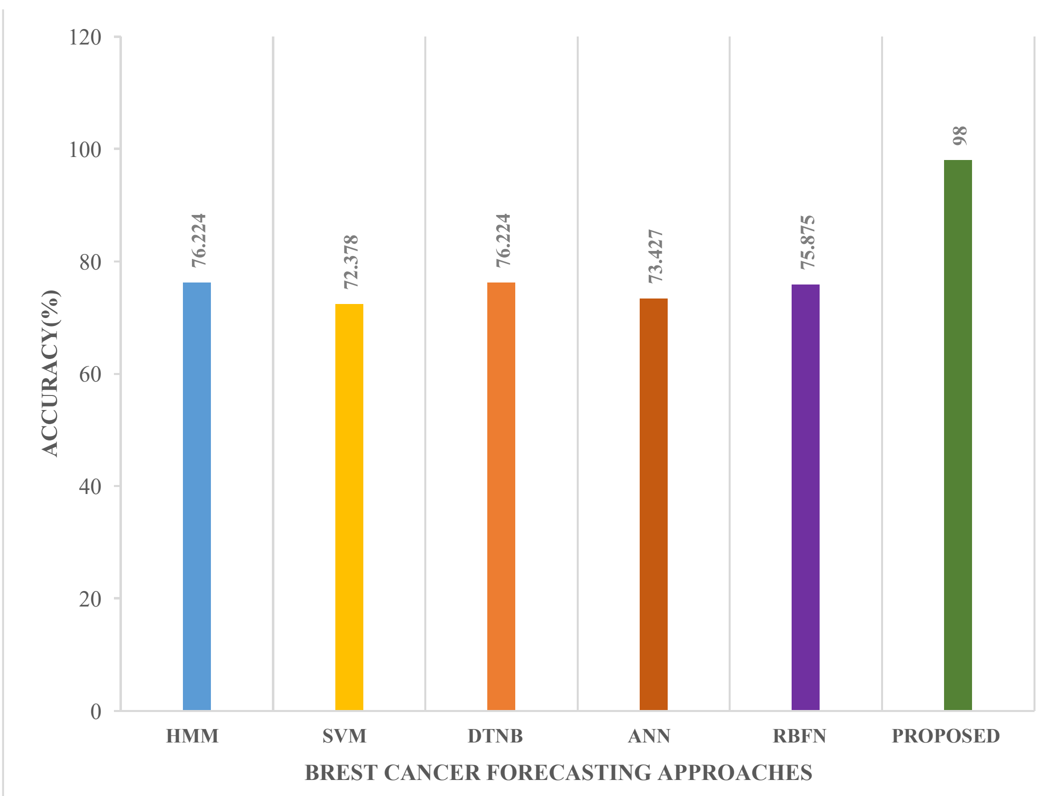

Figure 15. Forecasting Accuracy of Breast Cancer: Proposed Vs Standard Method

6. Conclusion

Accurately forecasting of diseases in their early phases remains a problem for medical practitioners across the globe. But in light of recent technological developments, the field of medical research has been working to enhance patient outcomes. Despite these efforts, the absence of a trustworthy disease forecasting technique that can precisely forecast the major signs of various illnesses is a significant factor contributing to high death rates. This research introduces an innovative approach to overcome these limitations by presenting a method for forecasting of various diseases including those related to heart, dermatology, hypothyroidism, diabetes, and breast cancer with enhanced accuracy.

6.1 Future Scope

Future endeavors could involve developing disease prediction models using different techniques and minimizing prediction errors with even more effective models to achieve superior performance.

7. 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 nearesthyperrectangle 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] A. Oza, A. Bokhare, Diabetes prediction using logistic regression and K-nearest neighbor, in: Congress on Intelligent Systems, Springer, Singapore, 2022, pp. 407–418.

[12] P. Sharma, S. Saxena and Y. Mohan Sharma, "An Efficient Decision Support Model Based on Ensemble Framework of Data Mining Features Assortment & Classification Process," *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2018, pp. 487-491, .

[13] Rashi Rastogi, Mamta Bansal, Diabetes prediction model using data mining techniques, Measurement: Sensors, Volume 25, 2023, 100605.

[14] Botlagunta, M., Botlagunta, M.D., Myneni, M.B. et al. Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Sci Rep 13, 485 (2023)..

[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, Multimedia, 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.