2022 International Conference on Business Analytics for Technology and Security (ICBATS) | 978-1-6654-0920-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICBATS54253.2022.9758996

Novel Intelligent Model for Heart Disease Prediction using Dynamic KNN (DKNN) with improved accuracy over SVM

Kummita Sravan Kumar Reddy
Department of Computer Science and Engineering,
Saveetha School of Engineering ,
Saveetha University
Chennai, India
kummitasravankumarreddy17@saveetha.com

K.V.Kanimozhi
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha University,
Chennai, India
kanimozhikv.sse@saveetha.com

***Abstract*— Aim: In comparison to Support Vector Machine, the major goal is to forecast the Novel Intelligent model for Heart Disease prediction using Dynamic KNN (SVM). Materials and Procedures: Two machine learning methods, Dynamic KNN (N=92) and Support Vector Machine (N=92), are used to predict heart disease. Dynamic KNN is asimple algorithm used for disease prediction. Heart disease dataset is used for disease prediction. For each group 20 samples are taken and it is divided into training and testing dataset. Result and Discussion: Accuracy of Dynamic KNN is 84.44% and Support Vector Machine is 67.21%. There exists an analytical significant difference between Dynamic KNN and SVM. Conclusion: Dynamic KNN appears to perform significantly better than Support Vector Machine for Novel Heart Disease Prediction.**

Keywords*— *Novel Intelligent Model, Machine Learning, Dynamic KNN, Support Vector Machine, Accuracy, Heart Disease Prediction.

I. INTRODUCTION

Heart Disease Prediction is the process of predicting heart disease at an early stage. [1]. Early heart disease prediction helps to control heart disease of patients in the modern world and there is a need in the clinical scenario [2]. This type of predicting heart disease is mainly used in the medical image processing environment [3]. Machine learning algorithms can be used to forecast heart disease in the healthcare system [21]. For the past five years, there have been around 59 IEEE Xplore articles and 37 Google Scholar publications. [4] has proposed a research on heart disease predictionusing Dynamic KNN and SVM. Cleveland dataset with 72 attributes. Sensitivity, specificity and accuracy values are displayed and concatenated. Comparative analysis is done for KNN, SVM and Naive-Bayes. [5] has proposed research on supervised machine learning algorithms. Using the histogram, the accuracy, precision, f1 score is shown. Chi square statistics is used to test the data. [6] has proposed a research on comparative study for SVM, KNN, Random forest for heart disease prediction using a Cleveland dataset. For the purpose of disease prediction, we use classifiers and feature extraction techniques. This research investigates accuracy for predicting the disease and features that will help in increasing the speed of prediction. [7] has proposed a research on heart disease prediction. Data classification, extraction features used to separate the data. Accuracy is calculated to get the predicted results for disease prediction.

Previously, our team had a wealth of expertise working on a variety of research projects across a wide range of fields. From [8 - 11]. This initiative was prompted by the growing popularity of this topic in recent years.

The drawbacks of Support Vector Machines are the accuracy is less for larger datasets. It takes more time to load the dataset. It takes more time to train the data. It overlaps the data when a large amount of data takes place. It is expensive and slower for large datasets. Based on the literature survey, Dynamic KNN has more features than Support Vector Machine. [12] has proposed research on heart and disease prediction using SVM. Cleveland dataset is used which has approximately 303 rows and 74 attributes used for disease prediction. It shows different accuracy rates and speed for different datasets. The aim is to improve accuracy by implementing a novel intelligent model in the Support Vector Machine in comparison with dynamic KNN for Heart Disease Prediction.

II. MATERIALS AND METHODS

At Saveetha School of Engineering, the suggested work will be carried out. The study has two groups of participants. SVM and Dynamic KNN are the first two groups.

Heart dataset is used for disease prediction (Heart Disease UCI | Kaggle). This dataset comprises more than 303 rows of information about the heart, as well as 14 key features. In order to compute the sample size, Clinical.com is utilised, and the alpha value is 0.05, the power is 0.8, and the beta value is 0.2. We take a total of 20 samples from each group for two different groups. The dataset is divided into two sections: the training dataset and the testing dataset. For training data, we take 20 samples and for testing data 20 samples taken for both groups. After splitting the dataset, we fit the algorithm into training and testing to predict the accuracy values.

A. Dynamic KNN

In this research, Dynamic KNN is the proposed algorithm. Dynamic KNN is a supervised machine learning algorithm. The distance is calculated between the query and current example in the data. Order collection in the data is sorted and preprocesses the data and implements the data with k-entries from the sorted collection. It gives more accuracy. The performance is also high compared to the existing one because the algorithm fits for larger datasets. It takes less time and is less difficult to put the algorithm into action. The technique can also be applied to classification and regression analysis.

978-1-6654-0920-9/22/$31.00 ©2022 IEEE

B. Pseudocode

Input - Heart dataset
Output - Improved accuracy
Begin
Read the dataset
Preprocess the data
Train the model
Store the predicted values
Calculate the accuracy
Dknn=round(accuracy_score(Y-pred_knn,Y_test)*100,2)
End for
Return accuracy
End

C. Support Vector Machine

Specifically, the Support Vector Machine (SVM) is the technique that was employed in this investigation. The support vector machine is a machine learning method that operates under the supervision of a trained operator. SVM is simple to use and performs better with less datasets than other methods. When it comes to classification and regression analysis, it's commonly used. The interpretation of a support vector machine is straightforward.

Pseudocode

Input- Heart Dataset
Output- Accuracy
Begin
Read the dataset
Pre-process the data
Train the algorithm
Calculate accuracy
svm=round(accuracy_score(Y_pred_svm,Y_test)*100,2)
Return accuracy
End

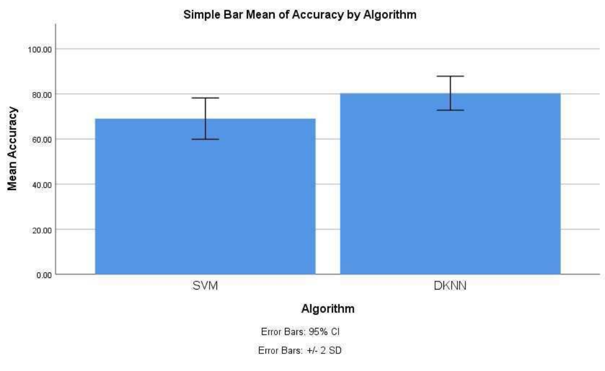

Figure 1 Bar graph representing the comparison of Mean Accuracy of Support Vector Machine and Dynamic KNN. Mean Accuracy of Dynamic KNN is better than Support Vector Machine and standard deviation is slightly better for Dynamic KNN than SVM. X-axis: SVM vs Dynam

A system with Windows operating system and Hard disk capacity of 50GB is used. 8GB RAM and language used is Python, Jupyter Notebook (Anaconda). Processor used is Intel i5. Independent variables for heart disease prediction are trestbps, chol, oldpeak, exang,slope, thal. Dependent variables are improved accuracy values.
Software used for our study is IBM SPSS version 26. Using SPSS we calculate T-Test,group statistics for the accuracy and loss. Independent Sample test and Independent Sample Effect sizes for accuracy and loss are calculated. According to the analysis done between the Dynamic KNN and Support vector machine, Dynamic KNN appears to have better accuracy and performance than the Support vector machine.

III. RESULTS

There is an explanation of the dataset used for the heart disease prediction in Table 1. Table 2 shows group data for the accuracy and loss of Dynamic KNN and SVM. Mean, Median, and Standard Error There were two groups of people. As compared to SVM, Dynamic KNN has an average accuracy of 80.29.
Table 3 displays the results of accuracy and loss T-tests on the dataset. Using the two groups, the mean difference, standard error, and 95% confidence interval are all calculated. p 0.05 is regarded as significant. As a result, the 95 percent confidence interval for the difference is -15.19000, with the mean difference being -11.24900 and the SD difference being 1.87584. Sign (two-tail) should be less than 0. For Dynamic KNN and SVM, 10 samples are collected for each algorithm.

Table 1. Attributes Description in Heart Dataset

S.No	ATTRIBUTE	DESCRIPTION	TYPE
1	AGE	PATIENT'S AGE(29 TO 77)	NUMERIC
2	SEX	GENDER OF PATIENT(MALE-0, FEMALE-1)	NOMINAL
3	CP	CHEST PAIN TYPE	NOMINAL
4	TRESTBPS	BLOOD PRESSURE RESISTANCE (IN MM HG ON ADMISSION TO HOSPITAL, VALUES FROM 94 TO 200)	NUMERICAL
5	CHOL	SERUM CHOLESTEROL(IN MG/DL.VALUES FROM 126 TO 564)	NUMERICAL
6	FBS	FASTING BLOOD SUGAR>120 MG/DL,TRUE-1 FALSE-0	NOMINAL
7	RESTING	STOP ELECTROCARDIOGRAPHS ARE THE OUTCOME (0 TO 1)	NOMINAL
8	THALACH	MAXIMUM HEART RATE ACHIEVED(71 TO 202)	NUMERICAL
9	EXANG	EXERCISE INCLUDED AGINAL(1-YES 0-NO)	NOMINAL
10	OLDPEAK	EXERCISE RELATED TO REST INTRODUCES STATIC DEPRESSION (0 TO 2)	NUMERICAL
11	SLOPE	THE SLOT OF THE PEAK EXERCISEST SEGMENT (0 TO 1)	NOMINAL
12	CA	NUMBER OF MAJOR VESSELS (0-3)	NUMERICAL
13	THAL	3-NORMAL	NOMINAL
14	TARGETS	1 OR 0	NOMINAL

Table 2. Group statistics for Mean,Standard Deviation and Standard Error Mean for two groups accuracy and loss. Mean accuracy for SVM is 69.0480% and Mean accuracy for Dynamic KNN is 80.2970%

Group Statistics					
	Algorithm	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	SVM	10	69.0480	4.57982	1.44827
	Dynamic KNN	10	80.2970	3.77003	1.19219
Loss	SVM	10	31.2000	4.51664	1.42829
	Dynamic KNN	10	19.4660	2.76855	.87549

The similar findings of the related work found in the previous study are discussed. [13] has proposed a research on comparative study of disease prediction in comparison with KNN and SVM. To predict heart disease prediction accuracy values are calculated. Accuracy for KNN is 78% and SVM is 73%. This methodology shows KNN is more efficient than SVM. [14] has proposed a research on comparison between data mining techniques for disease prediction. Here knn classifiers are used for training the dataset. Accuracy for combining all the algorithms have increased to 80%. Accuracy for KNN has more accuracy than the remaining data techniques. [15] research on analysis of heart disease prediction using KNN, SVM, Decision tree and Naive-Bayes. Adaboost Classifier is used for implementing algorithms. According to the analysis results show more efficiency for Dynamic KNN when compared to the SVM and Decision tree. Dissimilar findings for related papers [16] has proposed a research on comparative study for Support Vector Machine and K-nearest Neighbour, Convolutional neural network, Decision Tree for disease prediction. This research finds the speed of predicting the disease for algorithms and comparison between the algorithms. In this research accuracy for SVM is 78%, KNN is 67%, CNN is 58%. According to the analysis, SVM has more accuracy than the remaining algorithms. [17] proposed a research on comparative study of classification techniques for predicting heart disease. Using the histogram graphs sensitivity, specificity and accuracy is compared between the algorithms. The accuracy for SVM is 92% and is a higher accuracy rate than the other algorithms.

Table 3. Independent Sample t-test for accuracy and loss for dataset, 95% confidence interval of the difference of two groups.

Independent Samples Test										
Leven's Test for Equality of Variances				T-test for Equality of Means						
		F	Sig	t	df	Sig(2-tailed)	Mean difference	Std.Error Difference	95% Confidence Interval of the difference Lower	Upper
Accuracy	Equal variances assumed	.646	.432	-5.997	18	<.001	-11.24900	1.87584	-15.19000	-7.30800
	Equal variances not assumed			-5.997	17.359	<.001	-11.24900	1.87584	-15.20046	-7.29754
Loss	Equal variances assumed	2.340	.144	7.004	18	<.001	11.73400	1.67526	8.21441	15.25359
	Equal variances not assumed			7.004	14.926	<.001	11.73400	1.67526	8.16174	15.30626

The bar graph shows the difference between Dynamic KNN and the support vector machine from Fig. 1 shows a comparison graph of Dynamic KNN and SVM. Mean accuracy for Dynamic KNN is better than SVM and standard deviation is slightly better for Dynamic KNN than SVM. Dynamic KNN appears to have higher accuracy than the Support vector machine.

IV. DISCUSSION

The p-value of this study indicates that the dynamic KNN is more accurate at predicting heart disease than the Support Vector Machine. The sample dataset is tested statistically using SPSS Tool.

Our institution is passionate about high-quality evidence-based research and has excelled in various fields [18 - 20].

This study inspiration and methodology is based on [22-24]. We hope this study adds to this rich legacy.

In our study, 5 works, 3 works are similar findings and 2 works are dissimilar findings. Based on the above discussion we can conclude that there is an increase in success rate and accuracy. Execution time is differing for algorithms. It is concluded that Dynamic KNN appears to have higher accuracy and performance in all conditions than Support Vector Machine. Limitations of Dynamic KNN is that the prediction is slow for large datasets. It requires more memory for training the dataset. Memory allocation for training is more expensive. In the future we will overcome limitations of Dynamic KNN like algorithm, training speed. Also we can try to implement these algorithms for live datasets.

V. CONCLUSION

Using two machine learning algorithms, Dynamic KNN and Support Vector Machine, we were able to develop a Heart Disease Prediction system in this study. Support Vector Machines appear to have lesser accuracy than Dynamic KNN algorithms. Dynamic KNN appears to perform significantly better than Support Vector Machine for Heart Disease Prediction.

REFERENCES

[1]. Jiang, Shu. 2020. Heart Disease Prediction Using Machine Learning Algorithms.

[2]. Kasabe, Riddhi, and Trinity college of engineering and research. 2020. “Heart Disease Prediction Using Machine Learning.” International Journal of Engineering Research and. https://doi.org/10.17577/ijertv9is080128.

[3]. Dokare, Indu, Anjali Prithiani, Hanish Ochani, Sachin Kanjan, and Dinesh Tarachandani. n.d. “Prediction of Having a Heart Disease Using Machine Learning.” SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3697289.

[4]. Mirza, Imran, Arnav Mahapatra, Daryl Rego, and Kenneth Mascarenhas. 2019. “Human Heart Disease Prediction Using Data Mining Techniques.” 2019 International Conference on Advances in Computing, Communication and Control (ICAC3). https://doi.org/10.1109/icac347590.2019.9036836.

[5]. Princy, R. Jane Preetha, R. Jane Preetha Princy, Saravanan Parthasarathy, P. Subha Hency Jose, Arun Raj Lakshminarayanan, and Selvaprabu Jeganathan. 2020. “Prediction of Cardiac Disease Using Supervised Machine Learning Algorithms.” 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). https://doi.org/10.1109/iciccs48265.2020.9121169.

[6]. Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. 2020. “Heart Disease Prediction Using Machine Learning Techniques.” SN Computer Science. https://doi.org/10.1007/s42979-020-00365-y.

[7]. Hamdaoui, Halima El, Halima El Hamdaoui, Said Boujraf, Nour El Houda Chaoui, and Mustapha Maaroufi. 2020. “A Clinical Support System for Prediction of Heart Disease Using Machine Learning Techniques.” 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). https://doi.org/10.1109/atsip49331.2020.9231760.

[8]. Kumar, Muthusamy Senthil, Gelli Vamsi, Ramasamy Sripriya, and Praveen Kumar Sehgal. 2006. “Expression of Matrix Metalloproteinases (MMP-8 and -9) in Chronic Periodontitis Patients with and without Diabetes Mellitus.” Journal of Periodontology 77 (11): 1803–8.

[9]. Mehta, Meenu, Deeksha, Devesh Tewari, Gaurav Gupta, Rajendra Awasthi, Harjeet Singh, Parijat Pandey, et al. 2019. “Oligonucleotide Therapy: An Emerging Focus Area for Drug Delivery in Chronic Inflammatory Respiratory Diseases.” Chemico-Biological Interactions 308 (August): 206–15.

[10]. Danda, Anil Kumar, M. R. Muthusekhar, Vinod Narayanan, Mirza F. Baig, and Avinash Siddareddi. 2010. “Open versus Closed Treatment of Unilateral Subcondylar and Condylar Neck Fractures: A Prospective, Randomized Clinical Study.” Journal of Oral and Maxillofacial Surgery: Official Journal of the American Association of Oral and Maxillofacial Surgeons 68 (6): 1238–41.

[11]. Gopalakannan, S., T. Senthilvelan, and S. Ranganathan. 2012. “Modeling and Optimization of EDM Process Parameters on Machining of Al 7075-B4C MMC Using RSM.” Procedia Engineering 38: 685–90.

[12]. Nahiduzzaman, Md, Md Julker Nayeem, Md Toukir Ahmed, and Md Shahid Uz Zaman. 2019. “Prediction of Heart Disease Using Multi-Layer Perceptron Neural Network and Support Vector Machine.” 2019 4th International Conference on Electrical Information and Communication Technology (EICT). https://doi.org/10.1109/eict48899.2019.9068755.

[13]. Sinha, Parul, Poonam Sinha, and BUIT. 2015. “Comparative Study of Chronic Kidney Disease Prediction Using KNN and SVM.” International Journal of Engineering Research and. https://doi.org/10.17577/ijertv4is120622.

[14]. Thomas, J., and R. Theresa Princy. 2016. “Human Heart Disease Prediction System Using Data Mining Techniques.” 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT). https://doi.org/10.1109/iccpct.2016.7530265.

[15]. Pushpalatha, Mrs K. 2019. “Analysis of Heart Disease Prediction System Using Data Mining Techniques.” International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2019.4002.

[16]. Farooqui, Md Ehtisham, and Jameel Ahmad. n.d. “Disease Prediction System Using Support Vector Machine and Multilinear Regression.” SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3673232.

[17]. Khanna, Divyansh, Rohan Sahu, Veeky Baths, and Bharat Deshpande. 2015. “Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease.” International Journal of Machine Learning and Computing. https://doi.org/10.7763/ijmlc.2015.v5.544.

[18]. Sridharan, Gokul, Pratibha Ramani, Sangeeta Patankar, and Rajagopalan Vijayaraghavan. 2019. “Evaluation of Salivary Metabolomics in Oral Leukoplakia and Oral Squamous Cell Carcinoma.” Journal of Oral Pathology & Medicine: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology 48 (4): 299–306.

[19]. Pc, J., T. Marimuthu, and P. Devadoss. 2018. “Prevalence and Measurement of Anterior Loop of the Mandibular Canal Using CBCT: A Cross Sectional Study.” Clinical Implant Dentistry and Related Research. https://europepmc.org/article/med/29624863.

[20]. Ramadurai, Neeraja, Deepa Gurunathan, A. Victor Samuel, Emg Subramanian, and Steven J. L. Rodrigues. 2019. “Effectiveness of 2% Articaine as an Anesthetic Agent in Children: Randomized Controlled Trial.” Clinical Oral Investigations 23 (9): 3543–50.

[21]. Erdogan, Alperen, and Selda Guney. 2020. “Heart Disease Prediction by Using Machine Learning Algorithms.” 2020 28th Signal Processing and Communications Applications Conference (SIU). https://doi.org/10.1109/siu49456.2020.9302468.

[22] Suseendran, G., Zaman, N., Thyagaraj, M., & Bathla, R. K. (2019, December). Heart Disease Prediction and Analysis using

PCO, LBP and Neural Networks. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 457-460). IEEE.

[23] Anandan, R., Suseendran, G., Zaman, N., & Brohi, S. N. (2020). Echinacea purpurea to treat Novel Coronavirus (2019-nCoV).

[24] Khan, N. A., Brohi, S. N., & Zaman, N. (2020). Ten deadly cyber security threats amid COVID-19 pandemic.