Heart Attack Analysis and Prediction using SVM

Madhu H.K.
Assistant Professor
Department of MCA
BIT, Bengaluru, India

D. Ramesh, PhD
Professor& Head
Department of MCA
SSIT, Tumakuru, India

ABSTRACT

Smart gadgets from tiny oximeter to wrist watches collect data from human body to analyse and predict future occurrences. The most wanted model for this high active environment is the prediction model. Many algorithms have been developed by various researchers and today tools are available in software like MATLAB, Phyton and Tenser flow. In this paper SVM a supervised model is implemented to predict heart attack. The 13 features are considered which include personal details like chest pain type, blood pressure, collestral level and heart rate. The implemented model is tested on UCI health care heart disease data set. The efficacy of the model proposed is justified using performance and confusion matrix. The accuracy obtained is 83%.

Keywords

Machine Learning (ML), Support Vector Machines (SVM), UCI Health Care Dataset.

1. INTRODUCTION

Health care data collected from various electronic devices of a patient reside in a format specified in electronic health record (EHR). The data available today arein abundant which could be temporal, specific, contextual and voluminous. Analysing these data for identification, recognition and predictions are the areas of research interest using AI and ML technology. Considering the supervised model to learn the occurrences and allow the model to predict the occurrences from the new data collected in EHR is most wanted model in health care. Support Vector Machine (SVM) is a supervised model which is basically implemented as classifier which classifies the sample on a hyper plane using quadratic kernal. A two-class classifier can also be built to train and predict occurrences to classify the samples as normal and abnormal. The data set consists of 76 attributes in the subset of Cleveland database where 13 attributes are considered for analysis. As most of the healthcare data is not clean, a pre-processing model is included to remove unwanted data and missing sections are filled up by averaging the particular attribute values. A detailed implementation has been covered in section IV.

The proposed paper is segregated as follows: A comprehensive review of the literature on SVM and the various methods used to predict heart attacksare provided in section II ,sectionIII on the data set considered to test the efficacy of the proposed model, section IV describes the implementation and results achieved, conclusions are discussed in section V.

2. LITERATURE SURVEY

Health care data analytics poses challenges to research committee as it is voluminous and scanty in nature. Today many predictive models are implemented in smart devices like smart watch by apple/android/fit bit and many more. The challenges which is still live-in research committee is accuracy.

The survey conducted here focuses on understanding SVM model and also various predictive models implemented to achieve high accuracy and low computational complexity.

Three supervised model along with SVM Classifier is implemented by Padmavathi Janardhananet al.[1]. The prediction model's performance in predicting diseases is measured and evaluated using various medical datasets. The datasets were binary in nature, with a varying number of attributes in each. Heart datasets, cancer datasets, and diabetes datasets are considered for the research work. The SVM classifier is found to produce a higher percentage of accuracy.Aditianilghive and DR. Patil [2] proposed employing SVM and NB classification algorithms to identify diseases and therapies in brief text. Experimental data reveal that SVM outperforms nave bayes in classification. Accuracy, Precision, Recall, and F-measure are the four criteria used to evaluate these algorithms.To forecast heart disorders, Dr. S. Anitha and Dr. N. Sridevi [3] suggested supervised machine learning techniques such as SVM, KNN, and Naive Bayes. The R programming language is used to implement the machine learning algorithms. The algorithms' accuracy is used to assess their performance. KNN, Naive Bayes, and SVM are three supervised machine learning algorithms that are compared in terms of accuracy. The testing results show that the SVM algorithm correctly predicts heart illness 86.6 percent of the time.

Perumal, P.T. and Priyanka [4] proposes that the technology be used to increase the accuracy of test results by using its wings. Age, blood pressure, artery thickness, and other factors are considered in the SVM and Principal Component Analysis(PCA), which are used to estimate a person's risk of heart attack. The best algorithm, according to experimental data, is the SVM algorithm with PCA component, which has a maximum accuracy value of 87 percent. This study presents a comprehensive overview of machine learning strategies for identifying heart disease that improve accuracy.Research paper by E. Laxmi Lydia et al. [5] proposesmethods for categorising diseases based on disease factors and using several developed intense classification methods such as SVM, Decision Tree, Logistic Regression, K-nearest neighbour, and Naive Bayes. The suggested classification algorithms use disease datasets to measure diseases and provide reliable predictions. The experimental studies were performed on three disease datasets: namely thyroid, diabetes, and cancer.

The proposed approach by Madhura Patil et al. [6] uses the SVM algorithm to detect cardiac disease. By taking into account dangerous factors connected with heart disease, the approach aids in the prediction of heart disease. The system uses a SVM algorithm on the patient's historical data and gives features such as age, sex, smoking, obesity, alcohol intake, bad cholesterol, blood pressure, and heart rate to make a more accurate prediction of coronary heart disease.A. Jamuna et al.

[7] uses a dataset to discuss various machine learning techniques for predicting diabetes disease. Machine learning algorithms such as Decission Tree, SVM, Naive Bayes, K-mean clustering, and Logical Regression algorithms are used to detect diabetes at an early stage in their experiments. Experiments are conducted using the PIDDfrom the UCI machine learning repository. In comparison to other algorithms, SVM outperforms with a high accuracy of 82 percent.R.R.Ade et al. [8] provides a classifier strategy for heart disease identification and demonstrate how SVM and Naive Bayes can be employed for classification. When an unknown sample is given as an input, this algorithm classifies the data into several categories and also forecasts the risk of heart disease.

Purusothaman G. et al. [9], helps in discovering the optimum model for subsequent work by providing a quick and easy knowledge of several prediction models in data mining. This is a unique approach since different strategies are enumerated and expressed in a bar chart to comprehend the level of accuracy of each. These methods were chosen for their efficacy in the literature.

The suggested effort by Ahmed Ismail et al. [10], intends to construct a real-time prediction system for health issues based on cloud-based big medical data processing. The goal of this study is to see how well machine learning algorithms work with EHR. In compared to previous research, the proposed system of prediction could provide great accuracy, with a predictability of 90.6 for heart disease. The methodology used in this study is parameterization for parameters on SVM in order to increase the likelihood of prediction utilising the most effective features.

Amit Kisan Pagare and Vijay Kumar Verma [11], compare several data mining classification strategies such as Neural Network, DT, Bayesian Classification, Back Propagation, SVM, Knearestneighbor, and Genetic Algorithms (GA).

Himanshu Sharma and M A Rizvi [12], present a wealth of material on cutting-edge machine learning and deep learning methodologies. It also illustrates the potential of machine learning algorithms in the treatment of cardiac disease in the future. This research also examines the use of deep learning in the realm of heart disease prediction.According to Mythili T et al. [13],In order to propose an accurate model of predicting heart disease, this study presents a rule-based model that compares the accuracies of applying rules to individual SVM, decision tree, and logistic regression findings on the CHD Database.SVM had the greatest accuracy of 90.5 percent, followed by neural networks at 88.9%, Bayesian at 82.2 percent, decision tree at 77.9%, and logistic regression at 73.9 percent, based on 102 data sets.

M. Marimuthu et al. [14], discuss existing data mining and machine learning algorithms for predicting cardiac illness, including Artificial Neural Networks (ANN), Decision Trees, Fuzzy Logic, K nearest neighbour , Nave Bayes, and SVM.S.Vinothini et al. [15] recommended that decision trees be integrated using the k-means algorithm. They also used learning techniques such as SVM and Logistics regression. The outcomes of SVMand Logistics regression for heart disease prediction were evaluated.

V.V. Ramalingam et al. [16], evaluate and analyse the performance of numerous models such as Artificial Neural Networks (ANN), Decision Trees, Fuzzy Logic, K nearest neighbour , Nave Bayes, and SVMbased on such methods and methodologies. Researchers favour models based on supervised learning techniques including SVM, KNN, Nave Bayes, DT, Random Forest (RF), and ensemble models.

Nahla Farid et al. [17],proposes the usage of SVMclassifiers with kernel functions for identification of normal and autoaggressive actions from EMG data.

From survey it is evident that the models considered are SVM [1,2,6,9,12], naive base [2,3,4,7,13], decision tree [13,14,16] and the accuracy of the model depends on the data set considered to evaluate the proposed model. The need of the hour is to check the accuracy of the proposed model on real-time data generated from smart devices. This builds a collaborated model which amalgamates IOT, data storage and client/edge/server computing. Motivated on this backdrop SVM model is implemented as an hybrid of classification and prediction as a supervised model. Model is tested on UCI data set.

3. DATASET COLLECTION

UCI health care dataset consists 76 attributes, but most of the researchers use 13 of them. Till data Machine Learning researchers use only clevland data base. Sample UCI data set is shown in figure 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
	44	1	1	120	263	0	1	173	0	0	2	0	3	1
	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
	58	0	3	150	283	1	0	162	0	1	2	0	2	1
	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
	58	0	2	120	340	0	1	172	0	0	2	0	2	1
	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1

Figure 1: Sample UCI data set

Attribute Information:

Age which is of integer.

Sex which is of Male/Female.

Four different types of chest pain

Blood pressure of type resting

Serum cholestoral in mg/dl

FBS> 120 mg/dl

Values of resting electrocardiographic results

Heart beatwhich is of maximum

Exercise induced angina

Oldpeakis equal to ST depression induced by exercise relative to rest

The slope of the peak exercise ST segment0-3 are the major vessels colored by flourosopy

thal: Three = normal; Six = fixed defect; Seven= reversable defect.

4. METHODOLOGY

In this research work, SVM binary classification model is implemented using phyton. The optimal hyper plane is set through quadratic kernel function and set using the width of the margin using maximization. The unique global minimum value is identified.Its primary goal is to identify the optimal hyper plane for separating data into its two classes. Multiple binary

SVMs were recently combined to accomplish multiclass classification. Figure2 depicts the SVM [17] architecture.
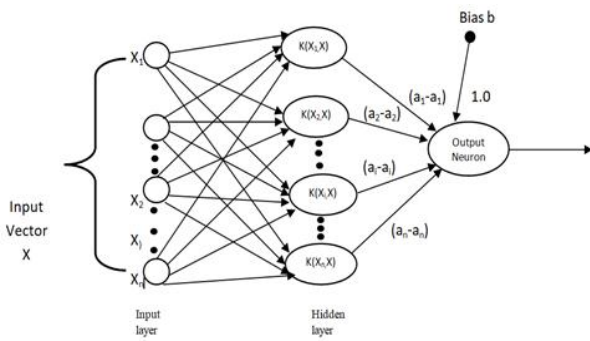

Figure 2: SVM Architecture

Let l be the number of training examples { xi, yi }, i = 1,..., l, with each instance containing an input xi and a class label yi ∈ {-1,1}. Each hyper plane has a weight vector (w) and a bias (b) that may be described using the equation below.

$$w. x + b = 0 \quad -------------- \quad (1)$$

Given the hyper plane, the following function can be used to classify training and testing data:

$$d(x) = \sin(w.x + b) \quad ---------------- \quad (2)$$

The previous function can be stated as follows when working with kernel functions.

$$d(x) = \sin \sum_{i=1}^{N} \boldsymbol{\alpha i y i} k(xi , x) + b) \quad ------ \quad (3)$$

Herenumber of samples used for training is represented as N, xi is the training samples input and yi is the label for matching class, bias is represented by b, and the kernel function K(xi,x) that translates into an enlarged feature space from input vectors.

From the below two constraints given in (4) and (5), the αi coefficients are obtained:

$$0 \leq \alpha i, i=1,.....,N \quad ------------ \quad (4)$$

$$\sum_{i=1}^{N} \alpha i \, yi = 0 \quad ------------ \quad (5$$

The SVM algorithm is the most used kernel learning algorithm. Using well-established notions from optimization theory, it delivers relatively robust pattern recognition performance.

Quadratic Kernel Function

The Rational Quadratic kernel requires less computing needs than the Gaussian kernel and can be used as a fallback if the Gaussian becomes too expensive.

$$K(xa,xb) = 1 - ||xa,xb||2/(||xa,xb||2+c)$$

On several groups of actions, SVMis used along with various kernel functions. The accuracy of each group was determined using polynomial, quadratic, and radial basis functions. On each group, the classification accuracy of kernel function is compared.

The proposed model consists of data pre-processing, attribute selection, visualizing the corelation among columns present in the data set, classification of normal and abnormal data sample is done by implementing SVM with classification using quadratic kernel function.

Data Pre-processing

Data pre-processing is to overcome redundant and missing data as the data is value attributed with 13 features, redundancy is

identified by clustering the samples using frequency histogram and missing data is regenerated based on the average of the attribute values of the corresponding samples with in cluster. The proposed model is hybridized with concept combination of clustering and classification. The pre-processed phase generates voluminous data in clean and suitable for analysis and feature selection through visualization and corelation.

The feature selection is based on the variance and co-variance of all the attributes which features the range of occurrences and supports the consideration of features. The selected features are considered to train and test the SVM implemented. The input variable matching is computed and also feature selection process is evaluated.The prediction model is built to predict the occurrences of the abnormalities based on the disturbances observed from normal attribute values.

Visualizing the correlation among columns present in the dataset

Visualizing the count of each target classes in order to check whether the dataset is imbalanced or not. Figure 3 depicts corelation comparisons.
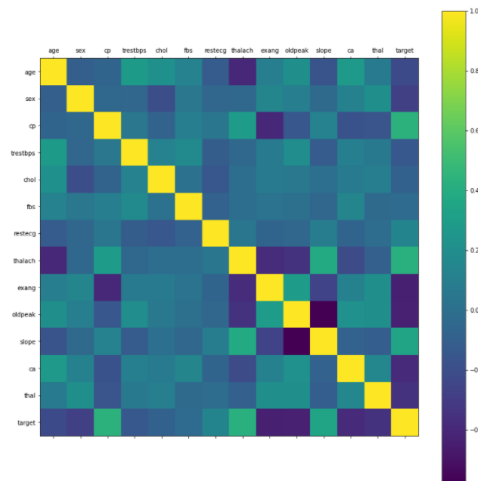

Figure 3: Corelation Comparisons.

Input variable matching

The number of input attributes included in every ML model has a big impact on how well it works. To ensure superior execution, selecting the relevant qualities rather than selecting multiples is critical.

Prediction

At the end, the classification funcion has achieved and analyzed.The function value calculated as the percentage of correctly predicted ouput divided by the total integer of samples.

$$Performance = \frac{True\ predicted\ outcomes}{Total\ number\ of\ samples}\ X\ 100$$
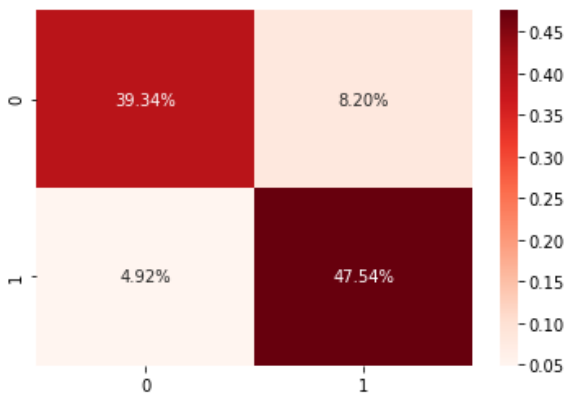
Confusion Maxtrix

Figure 4: Confusion Matrix

- ✓ Here, Type-1 Error is 8.20% which is also known as False Positive.

- ✓ Type-2 Error is 4.92% which is also known as False Negative.

- ✓ while other % value in the confusion matrix represents that they are correctly predicted in their specific categories.

- ❖ When mapping has done by input attributes with targeted output group column, the machine will run the SVM algorithm consequently.

```python
model_svc = SVC(kernel = main_kernel)
model_svc.fit(X_train,y_train)
y_pred = model_svc.predict(X_test)
target_names = ['class-0','class-1']
print(classification_report(y_test, y_pred, target_names=target_names))
```

```
              precision  recall  f1-score  support

     class-0     0.83     0.81     0.82       48
     class-1     0.83     0.85     0.84       52

    accuracy                       0.83      100
   macro avg     0.83     0.83     0.83      100
weighted avg     0.83     0.83     0.83      100
```

Figure 5: Accuracy output

User can enter the values of different attribuiteson the basis of which risk factor will be determined utilizing machine learning algorithms.

5. CONCLUSION

Heart care requires technology innovation to hinder human as a resource in hospitals/clinicals. This requirement has triggered many technology researchers to implement intelligent models to identify/classify/predict occurances on health care data. In this paper quadratic kernel function has been implemented in SVM to generate a hyper plane for a two class classifier. Expected accuracy has not been met. There is still scope for improving the proposed model. SVM – machine learning algorithm has low computational complexity makes it more suitable compared to other techniques.

6. REFERENCES

[1] Padmavathi Janardhanan, Heena L., and Fathima Sabika "Effectiveness of Support Vector Machines in Medical Data mining". JOURNAL OF COMMUNICATIONS SOFTWARE AND SYSTEMS, VOL. 11, NO. 1, MARCH 2015.

[2] ADITI ANIL GHIVE, D. R. PATIL "IMPLEMENTATION OF SVM AND NB ALGORITHMS FOR CLASSIFICATION OF DISEASES AND THEIR TREATMENTS". International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009, Spl. Issue-4 Oct.-2015.

[3] Dr. S. Anitha, Dr. N. Sridevi "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES". Journal of Analysis and Computation (JAC) (An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XIII, Issue II, February 2019.

[4] P. Perumal, P.T. Priyanka "SUPERVISED HEART ATTACK PREDICTION USING SVM WITH PCA". Journal of critical reviews. ISSN- 2394-5125 VOL 7, ISSUE 19, 2020.

[5] E. Laxmi Lydia, N. Sharmil, K. Shankar and AndinoMaseleno "Analysing the Performance of Classification Algorithms on Diseases Datasets". International Journal on Emerging Technologies 10(3): 224-230(2019) ISSN No. (Print) : 0975-8364 ISSN No. (Online) : 2249-3255.

[6] Madhura Patil, Rima Jadhav, VishakhaPatil , Aditi Bhawar, Mrs. Geeta Chillarge "Prediction and Analysis of Heart Disease Using SVM Algorithm International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 7 Issue I, Jan 2019- Available at www.ijraset.com.

[7] A.Jamuna, Dr. R. Jemima Priyadarsini, Dr.S.Titus "Survey on Predictive Analysis of Diabetes Disease Using Machine Learning Algorithms". A.Jamuna et al, International Journal of Computer Science and Mobile Computing, Vol.9 Issue.10, October- 2020, pg. 19-27.

[8] Ms. R.R.Ade, Dhanashree S. Medhekar, Mayur P. Bote "Heart Disease Prediction System Using SVM and Naive Bayes". INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY [Ade, 2(5): May, 2013] ISSN: 2277-9655.

[9] G. Purusothaman, P. Krishnakumari "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease". Indian Journal of Science and Technology, Vol 8(12), DOI: 10.17485/ijst/2015/v8i12/58385, June 2015 ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645.

[10] AHMED ISMAIL, SAMIR ABDLERAZEK, I. M. EL-HENAWY "BIG DATA ANALYTICS IN HEART DISEASES PREDICTION". Journal of Theoretical and Applied Information Technology 15th June 2020. Vol.98. No 11. ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195.

[11] Amit Kisan Pagare, Vijay Kumar Verma "Heart Attack

Prediction Using Data Mining Classification Techniques:A Study". International Journal of Technology Research and Management ISSN (Online): 2348-9006 Vol 2 Issue 6 June 2015.

[12] Himanshu Sharma, M A Rizvi "Prediction of Heart Disease using Machine Learning Algorithms: A Survey". International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 5 Issue: 8.

[13] Mythili T., Dev Mukherji, Nikita Padalia, and Abhiram Naidu "A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)". International Journal of Computer Applications (0975 – 8887) Volume 68– No.16, April 2013.

[14] M. Marimuthu, M. Abinaya, K. S. Hariesh, K. Madhankumar and V. Pavithra "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach". International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 18, September 2018.

[15] S.Vinothini, Ishaan Singh, Sujaya Pradhan and Vipul Sharma "Heart Disease Prediction". International Journal of Engineering &Technology, 7 (3.12) (2018) 750 -753.

[16] V.V. Ramalingam, AyantanDandapath, M Karthik Raja "Heart disease prediction using machine learning techniques : a survey". International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687.

[17] NAHLA FARID, BASSANT MOHAMED ELBAGOURY, "A Comparative Analysis for Support Vector Machines For Stroke Patients". ISBN: 978-960-474-304-9.

7. AUTHOR'S PROFILE

Mr. Madhu H.K., is a research scholarat SSIT,Tumkur(SSAHE),having 21 years of Teaching experieance at Department of M C A, in BIT, His research interest includes Data Miningand Big Data Analytics.

Dr. D. Ramesh, Professor and HoD from SSIT,Tumkur,India. His vision is to make SSIT as a centre ofexcellence for imparting technical knowledge in thefield of computer applications, nurturing technicalcompetency and social responsibility amongbudding software professionals.