

Exploring the Efficiency of Various Supervised Machine Learning Techniques to Predict the Heart Disease using Risk Factors

Jaishri Wankhede, Palaniappan Sambandam, S. Magesh Kumar

Abstract: Data Science in healthcare is a innovative and capable for industry implementing the data science applications. Data analytics is recent science in to discover the medical data set to explore and discover the disease. It's a beginning attempt to identify the disease with the help of large amount of medical dataset. Using this data science methodology, it makes the user to find their disease without the help of health care centres. Healthcare and data science are often linked through finances as the industry attempts to reduce its expenses with the help of large amounts of data. Data science and medicine are rapidly developing, and it is important that they advance together. Health care information is very effective in the society. In a human life day to day heart disease had increased. Based on the heart disease to monitor different factors in human body to analyse and prevent the heart disease. To classify the factors using the machine learning algorithms and to predict the disease is major part. Major part of involves machine level based supervised learning algorithm such as SVM, Naviebayes, Decision Trees and Random forest.

Keywords: Heart disease prediction, Machine Learning, LSVC, SVM.

I. INTRODUCTION

Healthcare research in 1990's researching the electronic healthcare data. It focusses on information management, operational management, research details and organizational management is a major challenges of healthcare systems. IT services to reduce the manual data work and reduce the cost. It can increase the productivity and efficient outcomes of the patient. Security services in healthcare sensitive data uploaded to the cloud [1] [2]. 20th century, Healthcare focuses and shifted from treating heart-based disease and to preventing it in those at risk. In 1998, to identify the heart disease risk factors based on family environment, hypertension, stress, age, smoking,

Revised Manuscript Received on November 22, 2019.

* Correspondence Author

Jaishri wankhede, Research Scholar, Department of CSE, Saveetha School of Engineering Simats , wankhede.jaishri@yahoo.com

Dr. Palaniappan Sambandam, Associate Professor, department of CSE, KCG College of Technology, s.palani.in@gmail.com

Dr. S. Magesh kumar , Associate professor, department of CSE, Saveetha school of Engineering simats , mnce6450@ gmail.com

cholesterol, diabetes and life style is the major risk factor in the patient. Now a day, risk factors can be improved in the public patient health.

II. DIFFERENT TYPES OF ATTACKS IN HEART

Heart attack is a form of ACS is known as a myocardial infarction. ACS (Acute coronary syndrome) is the arteries that carry blood, nutrients and get blocked. It occurs blood supply doesn't go properly in the heart.

Acute Coronary Syndrome refers to any group of clinical symptoms that occur when there isn't enough blood flow to heart. Three different types of ACS are ST segment elevation myocardial infarction (STEMI), Non-ST segment elevation myocardial infarction (NSTEMI) and coronary spasm, or unstable angina. The ST segment represents the isoelectric period when the ventricles are in between depolarization and repolarization.

STEMI: A STEMI is a full-blown heart attack caused by the complete blockage of a heart artery. A STEMI heart attack is taken very seriously and is a medical emergency that needs immediate attention.

SIGNS AND SYMPTOMS OF STEMI:

1. Chest pain or discomfort
2. Shortness of breath
3. Dizziness or light-headedness
4. Nausea or vomiting
5. Diaphoresis

NSTEMI: An NSTEMI differs from a STEMI, which is the most common type of heart attack, by causing less damage to a person's heart. While there may be less heart damage, an NSTEMI is still a serious condition. Treatment for an NSTEMI depends on how blocked the coronary artery is as well as the severity of the heart attack itself.

SIGNS AND SYMPTOMS OF NSTEMI:

- Feeling Short of breath
- Pain or tightness in the jaw
- Excessive sweating

CORONARY SPASM: Coronary artery disease causes narrowing in one or more of the coronary arteries. Chest pain is a common symptom of coronary artery disease. Coronary artery spasm is a brief and temporary.[4]

SIGNS AND SYMPTOMS OF CORONARY SPASM:

- Chest tightness
- A feeling of constriction

III. PROPOSED ALGORITHM:

The algorithm is diagrammatically depicted as shown in the Fig I.

- Step 1: The first step is to collect the input image and the image processing system is applied to the data.
- Step 2: The image processing system applies the grey scale conversion to the input image
- Step 3: The converted image is then classified

- Step 4: The features of the image is extracted from the classified image. The feature extraction is to segregate the data based on the categories
- Step 5: In this step the Disease is predicted based on the threshold level
- Step 6: Disease is categorized based predicted value
- Step 7: Then we apply the decision making system to identify the newly inserted data
- Step 8: Finally we apply the Feature reduction to reduce the size of the data

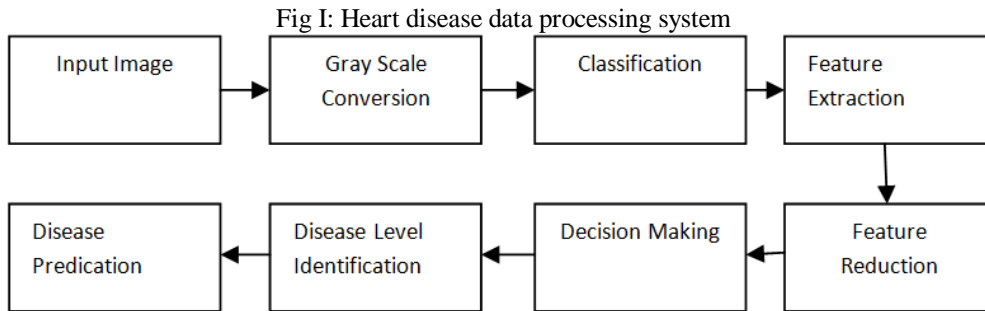


Fig I: Heart disease data processing system

Parameters used in the system

1. Age age in years
2. sex(1 = male; 0 = female)
3. cpchest pain type
4. trestbpsresting blood pressure (in mm Hg on admission to the hospital)
5. cholserumcholesterol in mg/dl
6. fbs(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecgresting electrocardiographic results
8. thalachmaximum heart rate achieved
9. exangexercise induced angina (1 = yes; 0 = no)
10. oldpeakST depression induced by exercise relative to rest
11. slopethe slope of the peak exercise ST segment
12. canumber of major vessels (0-3) colored by flourosopy
13. thal3 = normal; 6 = fixed defect; 7 = reversable defect target1 or 0

During the attacks there are different factors are measured to identify the risk factors. The factors are, age, sex, chest pain type, blood pressure, cholestorol, blood sugar, electrocardiographic, heartrate, exang.

The main factor is blood sugar can be blood sugar > 120 mg/dl. If blood sugar decreases, the risk factor of the patient increased in the heart disease.

IV CLASSIFIERS

4.1 Naïve Baye’s:

Naïve bayes (as shown in Fig.II) is a very popular classification algorithm. It identifies the base accuracy of the given dataset. It can be correlate non correlated data [5].

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Fig. II. Naïve Bayes Theorem

4.2 KNN (K Nearest Neighbor):

K- Nearest Neighbors is a supervised machine learning algorithm,it knows the target variable. It also uses classification and regression.

It uses similarity values and to predict the new values from the cluster. It is not an assumption of non-parametric distribution value pattern.

At the time of prediction, all the data points are mainly used.No training step,prediction step so costlier [6].

In our experimentation, we have age and testbps in a dataset along with the different categories of heart disease. We want to predict the heart disease of different ages, whose age and testbps affected in the heart disease [3].

- Step 1: Choose a value for K. K should be an odd number (age or testbps).
- Step2: Find the testbps values of the new point >=120 to each of the training data.
- Step 3: Find the K nearest neighbors to the new data point in testbps and different age group.
- Step 4: For classification, count the number of data points in test bps and age each category among the k neighbors. New data point will belong to class that has the most neighbors.



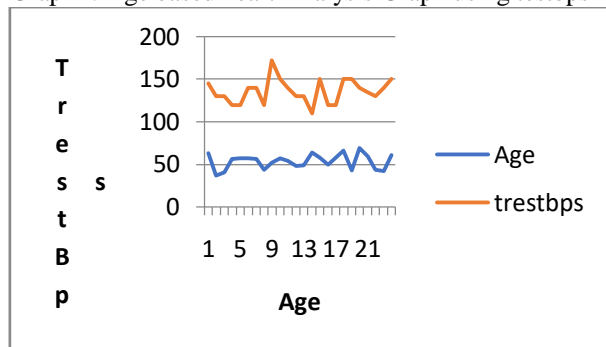
4.3 LSVC (Longitudinal support vector classifier): LSVC is based on support vectoemachine, it separates data from the datasets based on different classes. It also detects the cost of data based on the regression techniques [7].

4.4 SVM (Support Vector Machine): Support vector machine is to build a multi-dimensionally separating a data from different classes. It also increase and maximizing the distance from multidimensional data to the near set data of the class. It can assign the nearest data point .These data points are called support vectors [8][9].

Table I: Age based heart Analysis Data

S.No	Age	Sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Target
1	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
2	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
3	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
4	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
5	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
6	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
7	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
8	44	1	1	120	263	0	1	173	0	0	2	0	3	1
9	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
10	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
11	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
12	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
13	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
14	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
15	58	0	3	150	283	1	0	162	0	1	2	0	2	1
16	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
17	58	0	2	120	340	0	1	172	0	0	2	0	2	1
18	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
19	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
20	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
21	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
22	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
23	42	1	0	140	226	0	1	178	0	0	2	0	2	1
24	61	1	2	150	243	1	1	137	1	1	1	0	2	1

Graph I: Age based heart Analysis Graph using trestbps



As shown in Graph I, we predicted the heart disease of different ages, in that ages and trestbps affected in the heart disease.

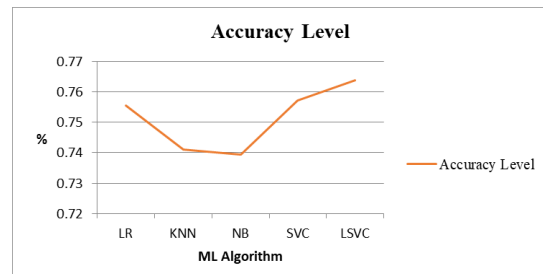
V. RESULT

Table. II. Accuracy Level of Different Classifiers

S.No	Classifier Name	Accuracy Level
1	LR	0.75563
2	KNN	0.74098
3	NB	0.73945
4	SVC	0.75727
5	LSVC	0.7638

Table II shows the Accuracy level of prediction [10] of heart disease in the different classification techniques 75, 74, 73, 75, 76. These values help us to predict the heart disease based on the accuracy and threshold values. Compared to the classification techniques LSVC is better than the other classification techniques based on the evidence observation.

Graph II: Accuracy level of heart disease data using different classifier



As shown in Graph II, we compared different classification techniques. In all those techniques LSVC is better than others.

VI. CONCLUSION

In this paper, mainly discussed on different effective techniques that can be used for predict the heart diseases using classification techniques is evaluated based on the different parameters using classification algorithm. A main challenge in machine learning technique is to predict the heart disease using different algorithms in the healthcare applications. It is mainly discussed to predict different ages of patient and their blood sugar level, cholesterol, heart rate and chest pain. These techniques very performance of Naive Bayes shows high level compare with other classifiers. This automated system to detect efficient trust based intelligent system which predicts the heart attack accurately based on the symptoms according to gender/age and knowledge of ANN in the field at the lowest cost.

REFERENCES

1. F. Miao, P. Zhang, L. Jin and H. Wu, "Chinese News Text Classification Based on Machine Learning Algorithm," 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2018, pp. 48-51.
2. L. N. Pondhu and G. Kummari, "Performance Analysis of Machine Learning Algorithms for Gender Classification," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICT), Coimbatore, 2018, pp. 16261628.
3. <https://medium.com/datadriveninvestor/k-nearest-neighbors-knn-7b4bd0128da7>
4. S. Turgut, M. Dağtekin and T. Ensari, "Microarray breast cancer data classification using machine learning methods," 2018 Electric Electronics, Computer Science, Biomedical Engineering's Meeting (EBBT), Istanbul, 2018, pp. 1-3.
5. S. R. Jadhav, R. Das, S. D. Thepade and S. De, "Applications of Hybrid Machine Learning for Improved Content Based Image Classification," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6.
6. R. B. G. Luta, R. G. Baldovino and N. T. Bugtai, "Multi-label Classification of pH Levels using Support Vector Machines," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 2018, pp. 1-4.
7. S. Wang, Z. Tang, S. Li, "Design and implementation of an audio classification system based on SVM", *Procedia Engineering*, 2011
8. R. B. G. Luta, R. G. Baldovino, N. T. Bugtai, "Image pre-processing using quick color averaging approach for color machine vision (CMV) systems", *HNICEM 2017 - 9th International Conference on Humanoid Nanotechnology Information Technology Communication and Control Environment and Management*, 2018.
9. M. Sewak, S. K. Sahay and H. Rathore, "Comparison of Deep Learning and the Classical Machine Learning Algorithm for the Malware Detection," 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence Networking and Parallel/Distributed Computing (SNPD), Busan, 2018, pp. 293-296.
10. Ashu Sharma, Sanjay kumar Sahay, "An effective approach for classification of advanced malware with high accuracy", *International Journal of Security and Its Applications*, vol. 10, no. 4, 2016.

AUTHORS PROFILE



Mrs. Jaishripandhar Wankhede, Saveetha University, Chennai is doing full time research scholar in Saveetha University and her research area is Image Processing. She has completed her M.TECH in Computer Science and Engineering from JNTU Hyderabad and BE in Computer Science and Engg. from Amravati university Maharashtra. She has also completed diploma course in. Net Tech. from NIELIT Calicut. She has published & presented a paper on Denial of service attack. She has got one year experience as lecturer in Muchhala institute of technology, Mumbai



Dr. S Palaniappan, PhD, working as an Associate Professor in the Department of Computer Science and Engineering at KCG College of Technology, Chennai. Graduated as Engineer in the year 2000 and done my masters in the year 2002 and completed my doctorate from JNTU Hyderabad in the year 2017. Has a total of 17 years of teaching experience and 1 year of research experience at NUS, Singapore. Authored or co-authored over 21 International Journal publications in the field of data science.



Dr. S. Magesh Kumar, graduated B.Sc., Computer Science at Sri Sankara Arts& Science College from University of Madras, Chennai, in the year 1997, Post graduate in Master of Computer Application .from Bharathidasan University, Thiruchirappalli in the year 2000, Post graduate in M.E. Computer Science and Engineering from Sathyabama University in the year 2006. Started the research carrier from the year 2013 at Bharath Institute of Higher Education and Research (BIHER), in the field of "An Integrated Framework For Cloud Storage System To Handle Data And Resource Management" in the domain of cloud computing and published around 12 research works in various journals, in which 10 are Scopus indexed and other papers are under non Scopus journals. Awarded Doctorate Degree in the year of December 2018 from BIHER. Having 16+ years of Teaching experience in Engineering College, worked as Lecturer/Assistant Professor/Associate Professor cum HOD in Thirumalai Engineering College, Lord Venkateswara Engineering College and Current Employer Saveetha School of Engineering. Areas of Interests are Computer Networks, Storage Optimization and Resource Management in Cloud Computing.