SYMPTOMS BASED DISEASE PREDICTION USING MACHINE LEARNING

[1] MOHAMED SADIQ .B , [2] NAHUSH R JAIN

[1] *Professor, Department of Master of Computer Application, BIET, Davangere*
[2]Student, Department of MCA, BIET, Davangere

Abstract--Computer Aided Diagnosis (CAD) is quickly evolving, diverse field of study in medical analysis. Significant efforts have been made in recent years to develop computer-aided diagnostic applications, as failures in medical diagnosing processes can result in medical therapies that are severely deceptive. Machine learning (ML) is important in Computer Aided Diagnostic test. Object such as body-organs cannot be identified correctly after using an easy equation. Therefore, pattern recognition essentially requires training from instances. In the bio medical area, pattern detection and ML promises to improve the reliability of disease approach and detection. They also respect the dispassion of the method of decisions making. ML provides a respectable approach to make superior and automated algorithm for the study of high dimension and multi - modal bio medicals data. The relative study of various ML algorithm for the detection of various disease such as heart disease, diabetes disease is given in this survey paper. It calls focus on the collection of algorithms and techniques for ML used for disease detection and decision making processes.

Keywords: Machine Learning (ML), Artificial Intelligences, ML Technique.

I. INTRODUCTION

The machine can think through Artificial Intelligence. AI makes machines even more intelligent. The subfield of AI Research is ML. Different researchers think that knowledge cannot be produced without learning's. The objective of ML is on designing computer algorithms that can read and use data to know for themselves.

In order to search for trends in data and make informed choices in the future based on the examples we have, the learning process starts with observation or data, such as references, direct experience, or Guidance. The primary objective is to allow systems to learn automatically and change behavior according without human involvement or assistance.

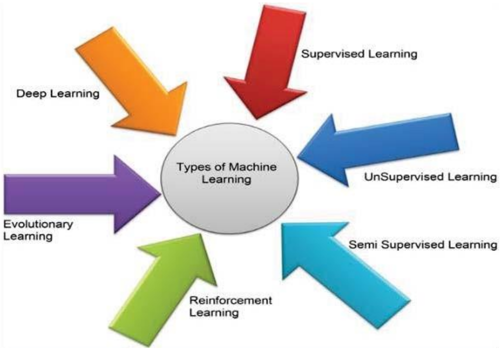


Fig. 1. Categories of ML technique.

There are several categories of techniques used for ML that are shown in Figure 1. The styles of machine learning methods are supervised, unsupervised, semi supervised, reinforcement, evolutional learning and deep learning. To identify the data collection, these techniques are used.

1.1. Supervised learnings

A training set of examples with acceptable target has been given and algorithms responded correctly to all possible inputs on the basis of this test data. Another name for Guided Learning is learning from examples. The forms of supervising learning, are classification and regress.

1.2. Unsupervised learning

Right answers or goals are not given. The unsupervised learning methodology seeks to figure out the similarity between the inputs data, and the un-supervised learning techniques classifies the data based on these similarities. It is also known as estimating densities [1].

1.3. Semi-supervised learning

A community of supervised learning methods is a semi-supervised learning technique. Unlabeled material for educational purposes was also used for this learning (Typically a small amount of label data with a significant amount of un label information). Unattended-learning (un label data) and supervised learning are semi-supervised learning (label data).

1.4. Reinforcement learning

Behavioral psychology encourages this learning. When the answer is incorrect, the algorithm is informed, but does not advise how to correct it. Before it finds the correct answer, it has numerous possibilities to exploring and examining. It's often referred to as studying with a critic. It does not suggest upgrades.

1.5. Evolutionary Learning

The biologic growing analysis can be viewed as a phase of learning: the biologic organism is evolved to advance its survival rate and to have the possibility to springing off. By using the theory of fitness to verify how effective the solution is, we could use this model on a computer [1].

1.6. Deep learning

This ML division is focused on a series of algorithm. In the data, these learning algorithm models large level abstraction. It used a deeper graph made up of several linear and nonlinear transformations of different processing layers.

II. MACHINE LEARNING TECHNIQUES

2.1 Naive Bayes Classifier Algorithm

A supervised learning algorithms based on the Bayes theorem and used to solve categorization issues is the Naive Bayes algorithm. It is used primarily in text classification, which requires a data-set of high-dimensional training. One of the easy and most powerful classification algorithms is the Naive Bayes Classifier, which helps to create fast ML models that can make quick predictions. It is a selective classifier, which implies, on the basis of an object likelihood, it predicts. Spam filtration, sentimental evaluation, and classifying papers are some typical examples of the Naïve Bayes Algorithm.

2.2. Random-Forest Algorithm

Random-Forest is a common algorithm for ML that belonging to the system of supervised learnings. It can be used for both problems with differentiation and regression in ML. It is focused on the idea of group modeling, which is a technique by which many classifiers are merged to address a specific problem and increase the effectiveness of the model.

Random-Forest is a classification algorithm which contains a set of decision trees and takes the averages to improve the predictive efficiency of that dataset on different subsets of the given dataset. The Random-forest takes the forecast from every tree and is based on the maximum voting of forecasts instead of relying on one decision tree, and it forecasts the final productions.

2.3. Decision Tree Algorithm

In real life, a tree has many analogs, and it turns out to have affected a large field of ML, includes both differentiation and regression. A decisions tree can be used in decisions evaluation to represent decisions and decision making visually and clearly. It uses a tree like models of decisions, as the name goes. While a commonly used method in data-mining to extract some strategies to achievedsome specific objectives, it is also widely used in ML, which will be the primary focus of this article.

2.4. Support Vector Algorithm

One of the most successful Supervised Learning algorithms, that is used for categorization and regression issues, is SVM. It is however, mainly used in Machine Learning for classification challenges. The goal of the SVM algorithm is to construct the best line or set point that can isolate n-dimensional spaces into categories so that we can conveniently position the latest data points in the correct category in the future. These boundaries of the right idea are called a hyper-plane. The extreme points/vectors which help to construct the hyper-plane are chosen by SVM. Such extreme cases are referred to as support vectors, and the algorithm is thus referred to as the Support Vector Machine.

2.5. Hybrid Technique in ML

A method for improving the predicted semi-structured sequential data performance. With simplistic music notation derived from midi files, we train and test the work on data that is beyond either an artificial neural system or the usual capability of a state-machines.

It is usually based on integrating two separate techniques of machine learning. For example, one unsupervised learners (or cluster) may be comprised of a hybrid classification model to pre-process the training data and one supervised learners (or classifier) to learn the outcome of the clustering or vice-versa.

2.6. Fuzzy Logic in ML

A process of reasoning that represents human reasoning is Fuzzy Logic (FL). This strategy is close to how decision making is carried out by humans. And all the intermediate considerations between YES and NO are involved. The conventional logic framework that a machine understands takes specific input and generates a TRUE or FALSE definitive output that is similar to the YES or NO of a human being. LotfiZadeh developed the Fuzzy logic, which found that humans have a wide set of possibility between YES and NO, unlike systems.

III. LITERATURE SURVEY

This segment discusses how many researchers have worked on various MLalgorithms for disease diagnostic. It has been acknowledged by researchers that machine-learning algorithms perform well for the diagnosis of various diseases. Diseases identified by MLT in this survey paper are heart and diabetes.

3.1. Heart Disease

Otoom [2] introduced a framework for research and tracking purposes. This proposed device detects and tracks coronary artery disease. UCI is extracted from the Cleveland heart data collection. This data set is made up of 304 cases and 77 features/attributes. Out of 76 features, 14 characteristics are used. For detection purposes, two experiments are performed with three algorithms: Bayes-Net, SVM, and FT. For identification, the WEKA tools is used. 88.3 percent accuracy is reached by using the SVM techniques after practicing with the Holdout test. The precision of 83.8 percent is given by both SVM and Bayes-Net in the Cross Validation test. After using FT, 81.5 percent accuracy is achieved. Using the Best First selected algorithm, FT.7 best characteristics are collected by and Cross-validation Checks are used for evaluation. Bayes Net achieved 84.5 percent accuracies by apply the test to 7 best selected feature, SVM offers 85.1 percent accuracies and FT properly classifies 84.6 percent.

Vembandasamy [3] proposed a research was conducted using the Naïve-Bayes algorithm to identify heart diseases. In Naïve-Bayes, Baye's theorem is included. Therefore, the Naïve-Bayes have a strong presumption of freedom. The data collection used was collected from one of Chennai's leading diabetics research institutes. 500 patients are part of the data collection. By using 70 per cent of Percentage Split, Weka is used as a method and executes classifying. Naive Bayes provides 86.419% precision.

Tan [4] proposed in which two ML algorithms called Genetics Algorithm (GA) and SVM are effectively joins by using the wrapper method, the proposed hybrid strategy. In this study, LIBSVM and the WEKA data mine tool are used. For this analysis, two data sets (Diabetes disease, Heart disease) will be obtained from the UC Irvine ML repository. An 84.07 percent precisions for heart disease is achieved after using the GA and SVM hybrid strategy. 78.26 percent accuracies are reached for a diabetes data collection. And some of the benefits are that it is a binary classifier to create right classifier and less over-fitting, resilient to noise and the drawbacks are. It may use pair wise identification for the classification of multi-classes. The cost of computation is high, so it works slowly.

3.2. Diabetes Disease

Iyer [5] is using decisions trees and Naïve-Bayes, they conducted a job to predict diabetes disease. Diseases arise when there is inadequate insulin production or there is excessive use of insulin. The Pima India diabetes data set is the data set used in this work. Various experiments were carried out using the data mining tool WEKA. The percentage division (71:31) predicts better than cross-verification in this data collection. By using Cross-verification and Percent Splitting Respectively, J48 indicates 74.8698 percent and 76.9565 percent precision. By using PS, Naïve-Bayes provides 79.5653 percent precisions. By using percent split checks, algorithms demonstrate the highest precision.

Sarwar and Sharma [6] proposed a study to predict diabetes type 2 on Naïve-Bayes has been suggested. Diabetes has 3 forms. A diabetes Type-1 will be the first type. The diabetes type 2 second type, and the third type is gestational diabetes. A diabetes Type 2 results from the production of opposition to insulin. The data collection consists of 416 cases for the reason of varieties; data are collected in India from different societies of sector. For model creation, MATLAB with aservers SQL is used. Naive Bayes achieves 95% accurate forecasting.

A diabetes diagnosis method has been developed by Ephzibah [7]. The GA and fuzzy logic are joined by the proposed model. This is used to select the finest subsets of features and also to improve classified accuracy. MLUCI Lab, which has 7 attributes and 768 instances, collects a dataset for study. For deployment, MATLAB is used. Only three good characteristics are picked by using the genetics algorithms. The fuzzy logical classifiersuse these three characteristics and provides 88 percent precision. Approximately 50 percent is less than the original expense.

3.3. Liver Disease

Vijayarani [8] Utilizing SVM and Naive Bayes Identification techniques to forecast liver diseases. ILPD is derived from the UCI data collection. 561 instance and 11 characteristics are part of the data collection. Comparative analysis is performed on the basis of precision and implementation of time. In 1770.00 ms, Naive Bayes demonstrates 61.29 percent correctness. In 3410.00 ms, SVM reaches 79.67 percent precision. MATLAB is used for deployment. Similar to the Naive bayes for liver disease forecasting, SVM displays the best precision. In terms of time implementation, relative to the SVM, Naives Bayes requires less effort.

The research of smart methods to identify liver patients is undertaken by the Gulia et al. [9]. The data collection used will be gained from the UCI. This analysis uses WEKA data mining platform and five intellectual classification methods for J48, MLP, Random Forest, SVM and Bayesian Network approaches. In the first step, both algorithms are added to the initial data set and the consistency percent is obtained. In the second stage, the process of feature selection is extended to the entire data set to get the significant sub-set of liver patients, and all these methods are used to evaluate the entire data set sub-set. In the third phase, before and after selecting features, they compare the results. Algorithms have the best accuracy after FS, as J48 has 70.668 percent accuracy, 70.8406 percent accuracy is obtained by the MLP algorithm, SVM gives 71.3552 percent accuracy, Random Forest produces 71.8686 percent reliability, and Bayes Net shows 69.1262 percent precision.

Rajeswari and Reena [10] The Naive Bayes, K star and FT tree data mining techniques were used to analyze liver diseases. UCI, which consists of 366 instance and 8 parameters, is taken from the set of data. Using the WEKA tool, 11 cross validation checks are added. In 0.5 sec., 96.53 percent accuracy is given by Naïve Bayes. 97.11 percent precision is obtained in 0.3 sec by using FT graph. The K star algorithm correctly classifies the instances in 0 seconds by around 83.48 percent. The FT tree on liver diseases dataset provides the best identification performance on the base of findings relative to other data mining techniques.

3.4. Dengue Disease

Tarmizi [11] Using Data Mining Techniques, the study was conducted for Malaysia Dengue Epidemic Prediction. Dengue is now a dangerous illness that is viral. In certain places where the environment is hot, such as Thailand, Indonesia and Malaysia, this causes problems. The classification methods used to forecast dengue fever in this research are DT, ANN and RS. The data collection is from the Department of Health of the State of Selangor. Two checks are used in the WEKA data mining tool. DT provides 99.96 percent accuracy by using 11-Cross fold authentication, ANN provides 99.97 percent consistency and RS displays 101 percent accuracy. Both the DT and the ANN have 99.93 percent of accuracy after using PS. 99.73 percent precision is obtained by RS.

Fathima [12] Study on Arbovirus-Dengue diseases prediction. The SVM is the data mining method used by these authors. The King Institute of Preventive Medicine and surveys of several Chennai and Tirunelveli hospitals and labs in India are the data collection for study. 28 characteristics and 5006 samples are used. Data is analyzed in version 2.12.3 of the R project. The precision obtained by the SVM is 0.9043.

IV. OVERVIEW OF DISEASE DIAGNOSTIC TECHNIQUES IN MACHINE LEARNING

SVM had the highest accuracy in current literature of 94.60 percent in 2015, as in Table 1. SVM demonstrates good output results in many application fields. Using the SVM version called SMO, Otoom et al. The FS approach used to detect the good characteristics. SVM respond to these characteristics and provides 84.5 percent accuracies, but as in 2015, it is comparatively poor. There are different learning and research strategies for all data sets, and also some several data types.

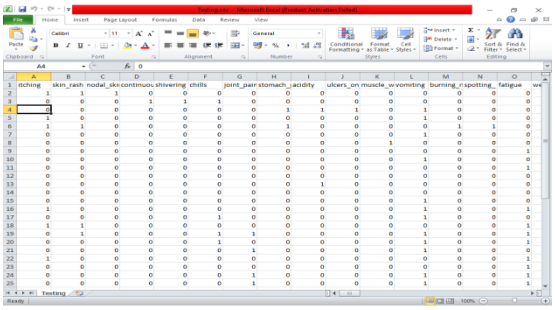
YEAR	AUTHOR	TOOL	DATA SET RESOURCE	ML TECHNIQUES	DISEASE	ACCURACY
2015	Vembandasamy et al.,	WEKA	Diabetic Research Institute, Chennai	Naive Bayes	Heart Disease	86.41%
2015	Otoom et al.,	Bayes Net SVM FT	UCI	WEKA	CAD	81.5% 94.60% 84.5%
2012	Sarwar , Sharma	MatLab (SQL)	Different Sectors of society, India	Naive Bayes	Diabetes Type-2	95%
2011	Ephzibah	MatLab	UCI	Fuzzy Logic + GA	Diabetes	87%
2009	Tan et al.,	LIBSVM, WEKA	UCI	Hybrid Techniques	Heart Disease	84.09%
2015	Vijayarani and Dhayanand	MATLAB	ILPD from UCI	SVM	Liver Disease	79.67%
2014	Gulia	WEKA	UCI	Random Forest	Liver Disease	71.87%
2019	Rajeswari and Reena	WEKA	UCI	K Star	Liver Disease	83.47%
2018	Tarmizi et al.	WEKA	Public Health Department of Selangor State	DT ANN RS	Dengue Disease	99.96%
2016	Fathima And Manimegalai	R project Version 2.12.2	King Institute of Preventive Medicine and surveys of many hospitals and laboratories of Chennai and Tirunelveli from India	SVM	Arbovirus-Dengue disease	90.43%

Table.1. Comprehensive view of machine learning techniques for disease diagnosis.

By excluding unrelated features, Naive Bayes improves the efficiency of the classification. Its productivity is good because there is less computing time. The only downside is that whole training examples must be stored because the algorithm requires a large amount of data to produce successful performance. The WAKE tool is often used for differentiation, but its efficiency is relatively low compare to Naive Bayes, since it provides only 70% split accuracy, whereas Naive Bayes provided 86.41% accuracy. The GA and fuzzy logic increase the accuracy of the classification and 87 percent accuracy is given and the cost is much lower. 84.07% and 76.26% precision of heart disease and diabetes is obtained by the hybrid approach.

The Naive Bayes-based approach is useful for diagnosing diabetes. In 2012, Naïve Bayes gave the highest accuracy of 96%. The findings show’s that with minimum error, this method can make efficient predictor and this method is also useful for diagnosing diabetes disease. But the precision provided by Naive Bayes in 2015 is poor. It offers 79.58% precision. This suggested framework for the detection of diabetes diseases will require the creation and evaluation of further data on training.

V IMPLEMENTATION



A4 f_x 0

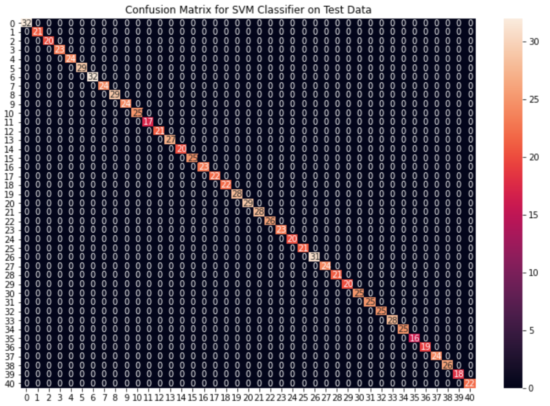
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	itching	skin_rash	nodal_ski	continuou	shivering	chills	joint_pair	stomach_	acidity	ulcers_on	muscle_w	vomiting	burning_n	spotting_	fatigue	we
2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	
5	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
6	1	1	0	0	0	0	0	1	0	0	0	0	1	1	0	
7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
16	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	
17	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	
19	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	
20	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	
21	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	
22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
24	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	
25	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	

Fig. 2. Datasets being used

Above figure shows complete Dataset consists of 2 CSV files . One of them is training and other is for testing your model.

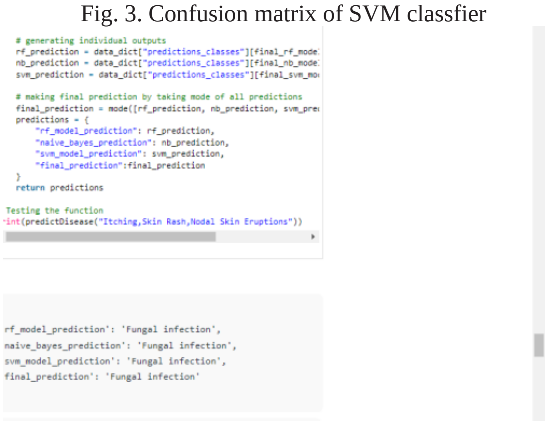
Each CSV file has 133 columns. 132 of these columns are symptoms that a person experiences and last column is the prognosis.

These symptoms are mapped to 42 diseases you can classify these set of symptoms to.


Confusion Matrix for SVM Classifier on Test Data

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22

Fig. 3. Confusion matrix of SVM classfier



```
# generating individual outputs
rf_prediction = data_dict["predictions_classes"][final_rf_mode
nb_prediction = data_dict["predictions_classes"][final_nb_mode
svm_prediction = data_dict["predictions_classes"][final_svm_mo

# making final prediction by taking mode of all predictions
final_prediction = mode([rf_prediction, nb_prediction, svm_pre
predictions = {
    "rf_model_prediction": rf_prediction,
    "naive_bayes_prediction": nb_prediction,
    "svm_model_prediction": svm_prediction,
    "final_prediction":final_prediction
}
return predictions

Testing the function
int(predictDisease("Itching,Skin Rash,Nodal Skin Eruptions"))
```

```
rf_model_prediction': 'Fungal infection',
naive_bayes_prediction': 'Fungal infection',
svm_model_prediction': 'Fungal infection',
final_prediction': 'Fungal infection'
```

Fig. 4. Figure showing results predicted

VI . CONCLUSION

The evaluation field has been flooded by statistical prediction models that are incapable of generating good quality outcomes. In maintaining generalized knowledge, statistical models are not efficient, coping with missing values and broad data points. The value of MLT stems from all of these causes. In many applications, ML plays a vital role, such as image recognition, data mining, processing of natural languages and diagnosis of diseases. ML provides potential solutions in all these fields. This paper discusses various techniques of ML for the diagnosis of various diseases such as heart, diabetes diseases. Most models have shown excellent results because they specifically describe the characteristic.It is noted from previous studies that SVM provides 94.60 percent improved performance for heart disease identification. Naive Bayes is a correctly diagnosed diabetes condition. It provides 95 percent of the highest classification precision. The survey shows the benefits and drawbacks of such algorithms. A suite of tools built in the AI community is also presented in this survey paper. These approaches are very useful for the analysis of certain problems and also provide opportunities for an improved decision making process.

REFERENCES

[1] Marshland, S. (2009) Machine Learnings an Algorithmic Perspectives. CRC Press, New Zealand, 6-7.

[2] Otoom et al., (2015) Effective Diagnosis and Monitoring of Heart Diseases. International Journal of Software Engineering and Its Application. 9, 143-156.

[3] Vembandasamy et al., (2015) Heart Disease Detection Using Naive Bayes Algorithms. IJISET-International Journal of Innovative Science, Engineering & Technology, 2, 441-444.

[4] Tan et al., (2009) A Hybrid Evolutionary Algorithm for Attribute Selections in Data Mining. Journal of Expert Systems with Application, 36, 8616-8630.https://doi.org/10.1016/j.eswa.2008.10.013.

[5] Iyer, A., Jeyalatha, S. and Sumbaly, R. (2015) Diagnosis of Diabetes Using Classification Mining

Technique. International Journal of Data Mining & Knowledge Management Process (IJDKP), 5, 1-14.

[6] Sarwar, A. and Sharma, V. (2012) Intelligent Naive Bayes Approaches to Diagnose Diabetes Type-2. Special Issues of International Journal of Computer Application (0975-8887) on Issues and Challenges in Networking, Intelligences and Computing Technologies-ICNICT 2012, 3, 14-16.

[7] Ephzibah, E.P. (2011) Cost Effective Approach on Feature Selection using Genetic Algorithm and Fuzzy Logics for Diabetes Diagnosis. International Journal on Soft Computing (IJSC), 2,

1-10. https://doi.org/10.5121/ijsc.2011.2101.

[8] Vijayarani, S. and Dhayanand, S. (2015) Liver Diseases Prediction using SVM and Naive Bayes Algorithm. International Journals of Science, Engineering and Technology Researches (IJSETR), 4, 816-820.

[9] Gulia, A et al., (2014) Liver Patients Classification Using Intelligent Technique. (IJCSIT) International Journal of Computer Science and Information Technology, 5, 5110-5115.

[10] Rajeswari, P. and Reena,G.S. (2019) Analysis of Liver Disorders Using Data Mining Algorithms. Global Journal of Computer Science and Technology, 10, 48-52.

[11] Tarmizi, N.D.A et al., (2018) Malaysia Dengue Outbreaks Detection Using Data Mining Model. Journals of Next Generation Information Technology (JNIT), 4, 96-107.

[12] Fathima, A.S. and Manimeglai, D. (2016) Predictive Analysis for the Arbovirus Dengue using SVM Classifications. International Journals of Engineeringand Technology, 2, 521-527.

[13] Rambhajani, M et al., (2015) A Survey on Implementations of Machine Learning Technique for Dermatology Disease Classifications. International Journal of Advance in Engineering &amp; Technology, 8, 194-195.