International Journal on Emerging Technologies **10**(3): 224-230(2019)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Analysing the Performance of Classification Algorithms on Diseases Datasets

E. Laxmi Lydia[1], N. Sharmil[2], K. Shankar[3] and Andino Maseleno[4]
[1]Professor, Department of Computer Science and Engineering,
Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India.
[2]Associate Professor in CSE Department,
Gayatri Vidya Parishad College of Engineering for Women, Viskahpatnam, Andhra Pradesh, India.
[3]Department of Computer Applications, Alagappa University, Karaikudi, India.
[4]Institute of Informatics and Computing Energy, University Tenaga Nasional, Malaysia.

(Corresponding author: E. Laxmi Lydia)
(Received 16 June 2019, Revised 29 August 2019 Accepted 23 September 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Change in regular food habits and physical activities of the human body, some of the genetic diseases were inherited from generation to generation. The most common hereditary diseases that stay lifetime are thyroid, diabetics, cancer. Predicting cancer-like diseases consumes time; cure for such hereditary diseases can be identified at an early stage. Medical technology has been improved for the prognosis of healthcare. Healthcare using prediction analysis enhances medical technology. Researchers have advanced Prediction modelling under three phases. In the first state, they define the issue, collection of data and progress the data. In the second state, they choose a model and perform training and testing and in the third state, they apply the model in real-world. This has become a crucial task in the medical field for immediate disease diagnosis. To advance such automatic healthcare prediction system, modern Artificial Intelligent technology has been developed an easy way to identify the existence of the diseases. The proposed research papers examine the diseases through the disease parameters and classify them using various developed intense classification algorithms such as Support Vector Machine, Decision tree, Logistic Regression, K-nearest neighbor, Naive Bayes. The proposed classification algorithms measure the diseases using the disease datasets which estimates the accurate prediction. The experimental analyses have been carried out over three disease datasets namely Thyroid dataset, diabetics data set, cancer dataset.

Keywords: Classification techniques, Disease Datasets, Healthcare, Support Vector Machine, Prediction accuracy.

I. INTRODUCTION

Healthcare using prediction analysis enhances medical technology. Researchers have developed Prediction modelling under three phases. In the first phase, they define the problem, collection of data and process the data. In the second phase, they choose a model and perform training and testing. And In the third Phase, they apply the model in the real-world.

Diabetes is a disease that is deep-rooted (continual) into the human body. It mainly occurs due to the changes obtained inside the blood such as a change in insulin levels and an increase of the sugar levels in the blood. Nowadays diabetes has very advanced that young people without any perceptive knowledge getting diabetes [19]. In general, diabetes is classified as type1 diabetes, type 2 diabetes, and gestation diabetes. Diabetic patients under these type1 are known as Insulin Diabetes Dependent Patients, types 2 are known as Non-Insulin Dependent patients [21], gestation diabetes occurs during pregnancy period. Diabetes instances are fatigue, hungry, excess thirst, and urinary, weight gain or loss, blurry vision, change in BP high pressure or Low pressure, smoking and Body Mass Index. Based on the different parameters of the body selects optimal features like BMI, sugar levels, blood pressure is used to predict thyroid disease.

Thyroid disease is the most commonly affected worldwide disease in humans. It disturbs the major functioning parts of the body and led to many other disorders like diabetes, heart problems, depressions, hormonal imbalance so on. Thyroid disease is classified into two categories such as Hypothyroidism and Hyperthyroidism. This occurs mostly due to the thyroid gland that exists over our neck part in a butterfly shape. This gland generates hormones to all parts of the body. Hormones that are released by the thyroid gland are T3 and T4. The pituitary gland releases TSH hormones. A person who has hypothyroidism shows that the thyroid gland produces fewer hormones required for the human body, which leads to muscle weakness, infertility, puffy eyes, etc. Simply it is described as the deficiency of hormones. A person who has hyperthyroidism indicates that the thyroid gland is generating an excessive number of hormones into the body, which causes weight loss, increase in heart rate, nerves weakness, etc. Based on the different parameters of the body selects optimal features like T3, T4, TSH, blood pressure are used to predict thyroid disease [17].

Breast cancer disease is caused by cancer tissues/cells inside the body. Cancer cells gradually increase inside body causing damage to the organ. This can be diagnosed in many different stages. Once effected in body, hard to be cured completely. People with cancer need to undergo a biopsy to get clarification of any tumors. The early stage of predicting cancer is very advantageous for a person's life [20, 23].

Algorithms like Support vector machine classification is used to operate on continuous and categorical values. Any classification process divides the input dataset into two sets, i.e., training data and testing data, Decision tree find out regression as well as classification problems. It follows a tree structure, Logistic Regression identifies the relation between the dependent and independent variables, K-nearest neighbour algorithm provides fast training phase and slow testing phase, Naive Bayes is a statistical approach. More detail

procedures of algorithms will be discussed in methodology

II. LITERATURE SURVEY

Researchers started their experiments on health care disease predictions using machine learning classification algorithms [24]. They have identified some of the strong classification methodologies for classifying diabetics disease, whether the person got effected with diabetes or not. Support Vector Machine classifier performs functioning upon the linear plane method, which aims to provide a propitious tool for diabetic patient predictions [26].It identifies the patterns using training dataset and predicts the class [3]. Later in the authors proposed research work on classification techniques comparative analysis [4]. Being on the parameters accuracy, application, type of datasets and time for execution. Algorithms like M5P, K start, M5Rul, and MLP are implemented. They have considered the workflow of the selection step, pre-processing step, transformation step, data mining step, and Interpretation step for Knowledge Discovery Process. Every classification technique has identified its limitations and advantages. They used Java programming language for implementation with respect to the graphical user interface. KNN with Euclidean and Manhattan distance calculation for thyroid dataset discussed by Chalekar *et al.,* [16].

Furthermore, in the advancement of technology, this comparative study moved on to various sectors. In [5], they implemented the analysis of efficient Disease Prediction over different classification techniques [25]. Based on the classification, classifier develops a model of samples. Algorithms like Decision trees and Bayesian networks work on datasets like breast cancer and heart disease [18]. They have identified the differences in a large number of attributes varying in size. The obtained training accuracy for cancer dataset for the C5.0 algorithm is 98% and testing accuracy for cancer dataset for the C5.0 algorithm is 95%. The obtained training accuracy for heart disease dataset for the C5.0 algorithm is 92% and testing accuracy for heart disease dataset for the C5.0 algorithm is 76%. The obtained training accuracy for cancer dataset for Bayesian algorithm is 99% and testing accuracy for cancer dataset for Bayesian algorithm is 85%. The obtained training accuracy for heart disease dataset for Bayesian algorithm is 91% and testing accuracy for heart disease dataset for Bayesian algorithm is 70%. The use of R programming for predicting healthcare datasets were implemented [6]. They have considered breast cancer datasets, which is usually occurring in most women at present days. R package has implemented Decision tree algorithm by considering three classifiers like rpath, ctree and random forest. Parameter measures like recall, precision, sensitivity, accuracy, and specificity are achieved. Workflow of this paper is a collection of data, train data, breast cancer dataset, pre-process data, Anomaly the classification algorithms, performance of the algorithms. The main observation of this classification is the work flow performance at data analysis.

Use of Matlab for predicting thyroid datasets using Neural Network, the authors experimented the thyroid disease by 244 subjects [11]. These are classified as healthy thyroid subjects, subclinical hyperthyroidism subjects, suffering from subclinical hypothyroidism subjects for inspecting. Neural network architectures like MLP, PNN, GRNN, and FTDNN are estimated through three validations in MATLAB. Regression graphs were drawn using these neural networks. Later a thyroid dataset from UCI is considered to analyze the set of records in the dataset with more attributes [12]. They recommended thyroid dataset with 215 samples for three class types Normal, hyper, hypo. The proposed work implemented Linear SVM, Quadratic SVM, Cubial SVM, Fine KNN, Medium KNN, Weighted KNN, Cubic and Decision trees. Classification algorithms are estimated the True positive, True negative, False positive and False-negative values of the confusion matrix. Above all ASVM has achieved 96% accuracy. Previously, Linear Discriminant Analysis is used for the prediction of thyroid dataset [15]. Later on the for the prediction hypothyroid disorder dataset, Random forest approach is implemented. The overall observation was analyzed using weka tool.

Work on sentiment analysis for the identification and classification of social media text environment in [1]. They have used three machine-learning algorithms on movie review dataset like Naive Bayes, SVM, and Decision trees. They have measured the efficiency of the classifiers using general parameters like anger, anticipation, disgust, fear, joy, so on. All the sentiments were generated through the R language syuzhet package. The authors in this paper [2] suggested the comparative analysis study of different classification techniques in equivalence to the consistency of accuracy. They have considered three healthcare datasets like carcinoma dataset, breast cancer dataset, and cardiovascular disease dataset. All the threes datasets were experimentally analyzed through weka tool. The classification techniques dealt in this proposed work are Simple Cart, FT, Random forest and LMT. They experimented result analysis on training and testing datasets using weka tool. Classification techniques (DT, KNN, SVM, NB, and NN) for numerical and categorical attributes with newly assigned class labels [9].

Datasets like Breast cancer Wisconsin and Hepatitis are analyzed and performed Single classification and associating ensemble methods [7]. Single Classification applied is Naive Bayes and ensemble methods are boosting, bagging, and stacking. They have specified 10-fold cross-validation for data mining tool. They introduced and implemented various multi-classifier techniques like MLP with the combination of stacking. Navie Bayes with the combination of bagging. The obtained accuracies are 97.51% for MLP+ stacking and 86.25% for NB + bagging. In 2018, [8] diagnosis of heart disease is diagnosed accurately by the use of machine learning classification techniques. This model has progressed based on one specific process. Initially, a dataset is maintained, pre-processing of the dataset, a model is applied to the pre-processed dataset, and finally, the classifier is used to predict the condition of the data. Attributes that are specified in this heart disease dataset are age, gender, cp (chest pain type), trestbps (blood pressure), chol (serum cholesterol), FBS (fasting blood sugar), ca (colored by fluoroscopy), thal (Thallium heart scan), num(angiographic disease status), slope (ST segment).

To the more specific areas of the medical field, thyroid disease classification based on the types hyperthyroidism and hypothyroidism is selected [10]. Medical field involving with technological field emerges every attribute for training the system and acquire accurate prediction values. Irrespective of the Laboratory results, technology has come forward with its

amazing classification techniques using RT3 and basal metabolic temperature for diagnosis.
A Hybrid classification, clustering, and ensemble models are used for predictions in Healthcare [13]. To predict thyroid disease authors have suggested random hybrid implementations for automatic thyroid disease aided computer systems. The proposed model has designed a model in which the feature input, feature selection, training feature set, testing feature set are processed and given to K-M Clustering, EM Clustering, KNN classifier, and SVM Classifier, which are processed and further implementation the data is given to Ensemble Classifier and Ensemble Clustering. This will provide the prediction Result.
In 2019, the classification techniques are implemented using Logistic Regression, SVM, KNN for Breast Cancer [22] were implemented using python code [14]. The results were analyzed and studied that SVM has got 92% of accuracy. This has got an automatic diagnosis system.

A. Methodology
The architectural process to perform classification techniques for prediction of diseases:
To perform the classification process, the Input data is loaded using online databases or offline databases. Input Data can be with missing values, noisy values, in consistent data. Such data need to be pre-processed in a pre-processing step. For better classification, optimal selection of features plays a major step. Best features are selected and for further process of classification, the dataset is divided as training and testing data. Training data is loaded to the classification algorithm and validated using testing data. Overall performance of the data is evaluated through accuracy or recall or precision or speed etc. Finally, results show the prediction values. Fig. 1 describes the flow diagram of the Prediction Process.

1. Experimental procedure for Support Vector Machine:
Support vector machine classification is used to operate on continuous and categorical values. Any classification process divides the input dataset into two sets, i.e., training data and testing data. SVM has a special feature known as marginal hyperplane that classifies the data. To minimize the error in classification optimal hyperplane is activated. Data points that are close to the hyperplane are identified as the Support vectors. These data points determine the margin line separation. While working with SVM, determination of hyperplane plays a major role. Depending upon the data points and the hyper plane, the good margin is selected. SVM algorithm is implemented based on the kernel selection. There are three types of kernels, Linear, Polynomial, Radial basis function.
Following is the python code procedure using Jupiter notebook for SVM classification:
Step1: Import the basic libraries like scikit-learn, pandas, numpy, seaborn, matplotlib, cross validation/ model selection for the entire process.
Step2: Load dataset (diabetes, thyroid, breast cancer).
Step3: Explore the data by replacing the missing values and select the features.
Step4: Split the dataset using train test split ().
Step5: IImport SVM model by choosing Kernel.
Step6: Use fit () function for training data into model and predict () for predictions.
Step7: Estimate the accuracy of the model and draw confusion or ROC curves if needed.

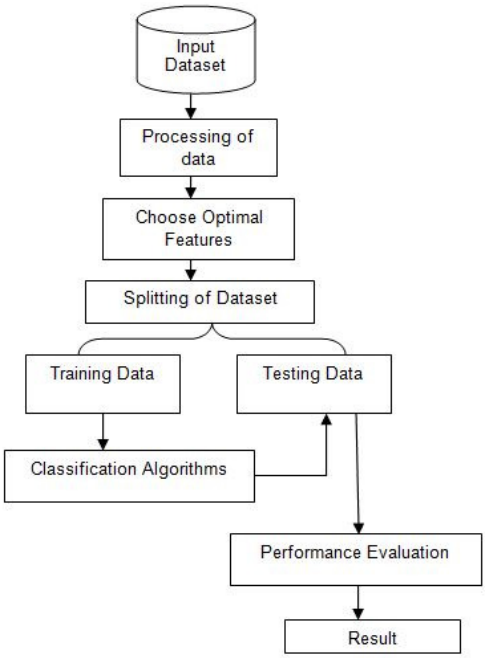

Fig. 1. Procedure for Prediction of Disease through Classification techniques.

2. Experimental procedure for Logistic Regression:
Logistic regression is mostly used for two-class classification. It identifies the relation between the dependent and independent variables. It mostly to predict data using statistical methods. Here the resultant variable is also a categorical value. The basic functional equation for logistic regression using sigmoid function is

$$a=1/1+e^{-b} \quad (1)$$

Where b is the dependent variable. logistic regression follows Bernoulli Distribution. There are three types of logistic regression, namely, Binary Logistic Regression, Multinomial Logistic Regression, and Ordinal Logistic Regression.
Following is the python code procedure using Jupiter notebook for Logistic Regression classification:
Step1: Import the basic libraries like scikit-learn, pandas, numpy, seaborn, matplotlib, cross_validation/ model_selection for the entire process.
Step2: Load dataset (diabetes, thyroid, breast cancer).
Step3: Explore the data by replacing the missing valuesand select the features.
Step4: Split the dataset using train test split().
Step5: Import Logistic Regression model by choosing a linear model.
Step6: Use fit() function for training data into model and predict() for predictions.
Step7: Estimate the accuracy of the model and draw confusion or ROC curves if needed.

3. Experimental procedure for Naive Bayes:
Naive Bayes classification algorithm is a straightforward classification technique for statistical approach that relies on Bayes theorem. This algorithm primary role is to assume the specified feature in a class which is independent of other features. Major steps of the Navie Bayes algorithm after selecting a single feature are
Step1: Converting data into a frequency table by calculating the prior probabilities of the class labels.
Step2: Finding probabilities by generating Likelihood table for every attribute of every class.
Step3: Assign these values into the Bayesian equation

$$P(h/D) = \frac{P\left(\frac{D}{h}\right)P(h)}{P(D)} \quad (2)$$

and evaluate the posterior probability.
Step4: Check for the higher probability of a class
Step5: Higher probability class is the outcome of the prediction.
Following is the python code procedure using Jupiter notebook for Naive Bayesclassification:
Step1: Import the basic libraries like scikit-learn, pandas, numpy, seaborn, matplotlib, cross validation/ model selection for the entire process.
Step2: Load dataset (diabetes, thyroid, breast cancer).
Step3: Explore the data by replacing the missing values and select the features.
Step4: Split the dataset using train test split().
Step5: Import Naive Bayes classifier.
Step6: Use a fit () function for training data into the model and carry out predictions.

4. Experimental procedure for Decision Tree:

Decision tree algorithm is used to find out regression as well as classification problems. It follows a tree structure by selecting the first node as the root node based on the attributes. Initially, to know the root node, all the attributes are evaluated to find the best attribute among all using Attribute Selection measure (i.e, Information gain, gain ratio, Gini ratio). Later split the dataset depending on the attributes. This starts constructing a tree, until all the attributes complete. This process checks for the same attribute value, no attributes, and no instances.
Following is the python code procedure using Jupiter notebook for Decision Tree classification:
Step1: Import the basic libraries like scikit-learn, pandas, numpy, seaborn, matplotlib, cross validation/ model selection for the entire process.
Step2: Load dataset (diabetes, thyroid, breast cancer).
Step3: Explore the data by replacing the missing values and select the features.
Step4: Split the dataset using train test split ().
Step5: Build a Decision Tree model.
Step6: Use fit () function for training data into model and predict () for predictions.

5. Experimental procedure for K Nearest Neighbor (KNN):

KNN classification algorithm is a process of a fast training phase and slow testing phase. K is defined by the user as an odd number, k describes the nearest neighbors. To perform KNN, it initially calculates the distance between the data points from k. The distance can be calculated using Euclidean distance or Hamming distance or Manhattan distance or Minkowski distance depending on the data. After estimating the distance measure among the data point, find the closest neighbors and finally vote for the labels.
Following is the python code procedure using Jupiter notebook for K Nearest Neighbor classification:
Step1: Import the basic libraries like scikit-learn, pandas, numpy, seaborn, matplotlib, cross validation/ model selection for the entire process.
Step2: Load dataset (diabetes, thyroid, breast cancer).
Step3: Explore the data by replacing the missing values and select the features.
Step4: Split the dataset using train test split ().
Step5: Import KNN by defining K
Step6: Use fit() function for training data into model and predict () for predictions.
Step7: Estimate the accuracy of the model and draw confusion or ROC curves if needed.

III. RESULT AND DISCUSSION

In this paper, three datasets are collected from Kaggle and performed accuracy using five classifiers. Such as Support Vector Machine, Logistic Regression,Naive Bayes, Decision Tree, and K nearest Neighbor.
Diabetes Dataset contains patient health features like pregnant, glucose, BP, skin, insulin, BMI, pedigree, age, label. Among them, select some features to perform analysis.769 patient records were taken for performance with 9 selected features. Following Table 1 describes the accuracy of each classification technique for a diabetic dataset for selected features. Out of which SVM and Navie Bayes classifier perform best results among all other classifiers i.e, 75% accuracy.
Thyroid Data set contains patient health features like age, sex, on thyroxine, quer on thyroxine, on anti-thyroid medication, thyroid surgery, query hypothyroid, query hyperthyroid, pregnant, sick, tumor, lithium, goitre, TSH measured, TSH, T3 measured, T3, T4 measured,T4,TT4 measured,TT4, T4U measured, T4U, FT1measured, FT1, TBG measured, TBG.1030 patient records were taken for performance with 9 selected features.
Following Table 2 describes the accuracy of each classification technique for Thyroid dataset for selected features. Out of which SVM and Logistic Regression classifier perform best results among all other classifiers i.e, 96% accuracy.
Breast Cancer Dataset contains patient health features like mean radius, perimeter, area, smoothness, compactness, concavity so on. 570 patient records were taken for performance with 9 selected features. Following Table 3 describes the accuracy of each classification technique for Breast Cancer dataset for selected features. Out of which SVM classifier performs best results among all other classifiers i.e, 91% accuracy.

Table 1: Accuracy for Diabetes Dataset.

Classifier	Accuracy
Support Vector Machine	0.750017
Logistic Regression	0.742191
Naive Bayes	0.751299
Decision Tree	0.669258
K Nearest Neighbor	0.703178

Table 2: Accuracy for Thyroid Dataset.

Classifier	Accuracy
Support Vector Machine	0.962098
Logistic Regression	0.962098
Naive Bayes	0.934913
Decision Tree	0.927089
K Nearest Neighbor	0.908681

Table 3: Accuracy for Breast Cancer Dataset.

Classifier	Accuracy
Support Vector Machine	0.912281
Logistic Regression	0.907895
Naive Bayes	0.903509
Decision Tree	0.907895
K Nearest Neighbor	0.868421

Fig. 2 describes the graph plot for the Diabetes dataset using five classification algorithms. The accuracy for each classifier is plotted in a Python environment.
Fig. 3 describes the graph plot for the Thyroid dataset using five classification algorithms. The accuracy for each classifier is plotted in a Python environment.

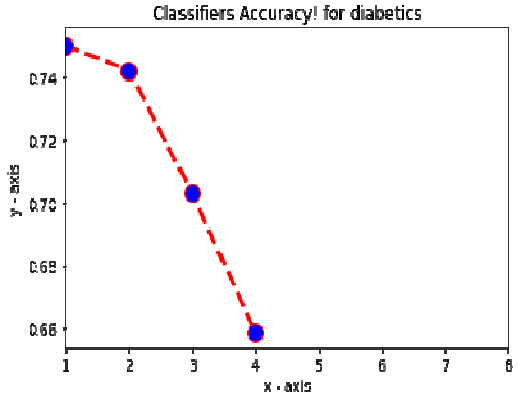

Fig. 2. Classifiers accuracy for diabetics dataset.

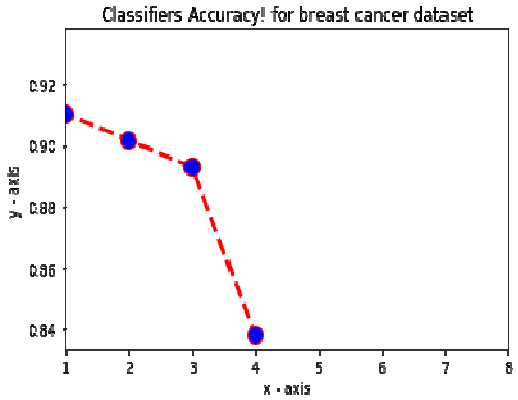

Fig. 3. Classifiers accuracy for thyroid dataset.

Fig. 4 describes the graph plot for the Breast Cancer dataset using five classification algorithms. The accuracy for each classifier is plotted in Python environment.

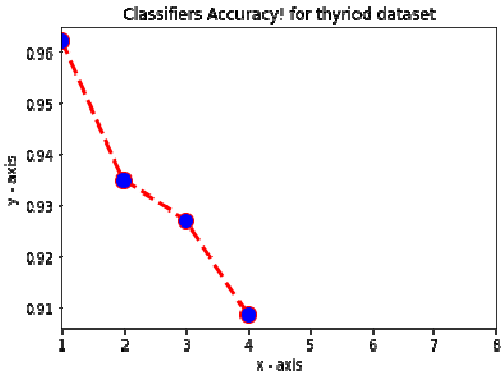

Fig. 4. Classifiers accuracy for breast cancer dataset.

Fig. 5 describes the outcome of all the patient feature values. 1 represents the positive value i.e., the patient is with diabetic disease and 0 represents negative value i.e., the patient is not with the diabetic disease. This graph is plotted among a total number of patients to count the number of diseased patients and not diseased patients. It is drawn using a sea born package in Python.

Fig. 6 describes the outcome of all the patient feature values. 1 represents the positive value i.e., the patient is with Thyroid disease and 0 represents negative value i.e., the patient is not with Thyroid disease.

This graph is plotted among a total number of patients to count the number of diseased patients and not diseased patients.

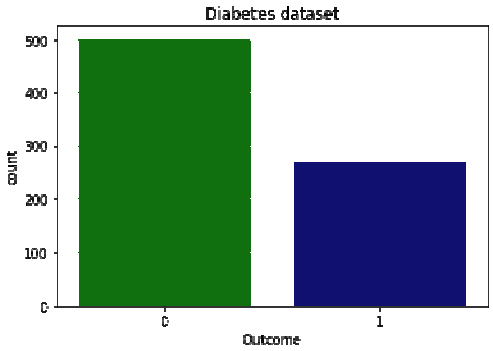

Fig. 5. Outcome plot for Diabetes.

Fig. 7 describes the outcome of all the patient feature values. 1 represents the positive value i.e., the patient is with Breast cancer and 0 represents negative value i.e., the patient is not with breast cancer. This graph is plotted among a total number of patients to count the number of diseased patients and not diseased patients.

Fig. 8 describes the bar graph for all classification techniques in each dataset. Five classifiers, it is observed that SVM results best and Navie Bayes and Logistic regression also provided good results for the considered data sets.

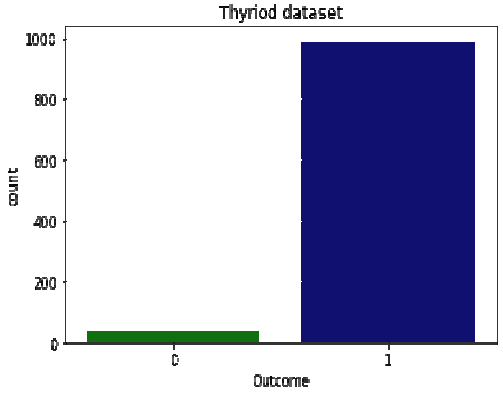

Fig. 6. Outcome plot for Thyroid.

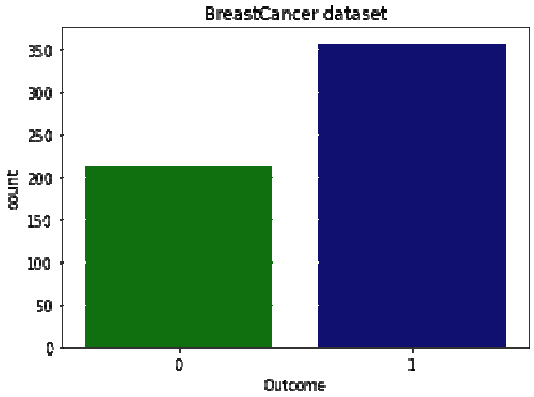

Fig. 7. Outcome plot for Breast Cancer.

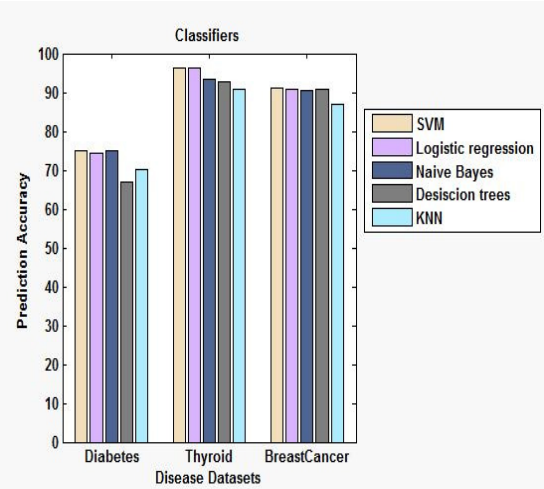

Fig. 8. Bar graph for three diseases data set with five classification techniques.

IV. CONCLUSION

Healthcare using classification techniques brought new ways to advance the technology with automatic predictions and quick treatment of patients. In this paper, five classification techniques were implemented for the prediction of diseases. The classification model is generated based on the training data, and the data is tested though predictions in the prediction stage. Three datasets namely Diabetics, Thyroid, Breast cancer were tested using classification algorithm using Python environment. Classification and regressions problems were solved mostly by these supervised classification algorithms. It is clear that Support Vector Machine classification Algorithm has given the best accuracy results when compared to other techniques. Thus by this, it is very clear that Healthcare organization can adopt predicting technologies using classing algorithms for healthy living.

FUTURE SCOPE

Areas like Machine Learning, Artificial Intelligence, and Deep Learning construct future world for better Prediction analysis of diseases.

Conflict of Interest: Nil

REFERENCES

[1]. Ragupathy, R., Maguluri, L. P. (2018). Comparative analysis of machine learning algorithms on social media test. *International Journal of Engineering & Technology,* Vol. *7*(2.8): 284-290.

[2]. Rajeswara, R. D., Vidyullata, P., Sathish, T., & Ramya, H. T. (2015). Performance Analysis of Classification Algorithms using Health care Data set. *International Journal of Computer Science and Information Technologies,* Vol. *6*(2): 1103-1106.

[3]. Aishwarya, R., Gayathri, P., & Jaisankar, N. (2013). A method for classification using machine-learning technique for Diabetes. *International Journal of Engineering and Technology (IJET),* Vol. *5*(3): 2903-2908.

[4]. Sharma, R., Kumar, S., Maheshwari, R. (2015). Comparative Analysis of Classification Techniques in Data Mining using different datasets. *International Journal of Computer Science and Mobile Computing (IJCSMC),* Vol. *4*(12): 125-134.

[5]. Sandhya, N., Sharanya, M. M. (2016). Analysis of Classification techniques for efficient Disease Prediction. *International Journal of Computer Applications*, *155*(8): 20-24.

[6]. Sudhamanthy, G., Thilagu, M., Padmavathi, G. (2016). Comparative Analysis of R Package Classifiers using Breast Cancer Data set. *International Journal of Engineering and Technology (IJET),* Vol. *8*(5): 2127-2136.

[7]. Rosly, R., Makhtar, M., Awang, M. K., Awang, M. I., & Rahman, M. N. A. (2018). Analyzing the performance of classifiers for medical data sets. *International Journal of Engineering and Technology (IJET),* Vol. *7*(2.15): 136-138.

[8]. Maryam, I., Janabi, A., Mahmoud, H. Q., & Hijjawi, M. (2018). Machine Learning classification techniques for heart disease prediction: a review. *Internatioal Journal of Engineering and Technology (IJET),* Vol. *7*(4): 5373-5379.

[9]. Gorade, S. M., Deo, A., & Purohit, P. (2017). A Study of some data mining classification techniques. *International Research Journal of Engineering and Technology (IRJET),* Vol. *4*(4): 3112-3115.

[10]. Sumathi, A., Nithya, G., & Meganathan, S. Classification of thyroid disease using data mining techniques. *International Journal of Pure and Applied Mathematics,* Vol. *119*(12): 13881-13890.

[11]. Obeidavi, M. R., Rafiee, A., & Mahdiyar, O. (2017). Diagnosing thyroid disease by neural networks. *Biomedical and Pharmacology Journal,* Vol. *10*(2): 509-524.

[12]. Gopinath, M. P. (2017). Comparative study on classification algorithm for Thyroid dataset. *International Journal of Pure and applied mathematics,* Vol. *117*(7): 53-63.

[13]. Pavya, K., Srinivasam, B. (2018). Hybrid thyroid stage prediction models combining classification, clustering and ensemble systems. *International Journal of Engineering and Technology (IJET),* Vol. *7*(4.7): 297-302.

[14]. Shravya, Ch, Pravalika, K., & Subhani, S. (2019). Prediction of breast cancer using supervised machine learning techniques. *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* Vol. *8*(6): 1106-1110.

[15]. Ammulu, K., & Venugopal, T. (2017). Thyroid data Prediction using data classification algorithm. *International Journal for Innovative Research in science and Technology, International Journal of Engineering and Technology (IJET),* Vol. *4*(2): 208-212.

[16]. Chalekar, P., Shroff, S., Pise, S., & Panicker, S. (2014). Use of K-nearest neighbor in thyroid disease classification. *International Journal of current engineering and Scientific Research (IJCESR), International Journal of Engineering and Technology (IJET),* Vol. *1*(2): 36-41.

[17]. Geeta, K., & Baboo, S. S. (2016). An Empirical model for thyroid disease classification using evolutionary multivariate Bayesian prediction model. *Global Journal of Computer science and technology; E Network, Web & security,* Vol. *16*(1): 1-10.

[18]. Kavitha, M., Lavanya, G., Janani, J., & Balaji, J. (2018). Enhanced SVM Classifier for Breast Cancer Diagnosis. *International Journal of Engineering Technologies and Management research.* Vol *5*(3): 67-74.

[19]. Sneha, N., & Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big Data,* Vol. *6(13)*: 1-19.

[20]. Sultana, J., & Jilani, A. B. (2018). Predicting Breast Cancer using Logistic Regression and Multi-Class Classifiers. *International Journal of Engineering and Technology (IJET),* Vol. *7*(4.20): 22-26.
[21]. Juliet, L. P., & Bhavadharam, T. (2019). An improved prediction model for type 2 diabetes mellitus disease using clustering and classification algorithms. *International Research Journal of Engineering and Technology,* Vol. *6*(2), 1179-1186.
[22]. Jafarpisheh, N., Nafisi, N., & Teshnehlab, M. Breast Cancer Relapse prognosis by classic and modern structures of Machine Learning algorithms. *Iranian Joint Congress on fuzzy and Intelligent Systems (CFIS),* pp.120-122, 978-1-5386-2836-2/18.
[23]. Shailaja, K., Seetharamulu, B., Jabbar, M. A. (2018). Prediction of Breast cancer using big data analytics. *International Journal of Engineering and Technology (IJET),* Vol. *7*(4.6): 223-226.
[24]. Joshi, S., & Nair, K. M., (2018). Survey of classification based prediction techniques in Healthcare. *International Journal of science and technology,* Vol. *11*(15): 1-19.
[25]. Shirsath, S. S., & Patil, S. (2018). Disease prediction using machine learning over big data. *International Journal of innovative research in Science Engineering and Technology,* Vol. *7*(6): 6752-6757.
[26] Deepti, S., & Dilip, S. S. (2018). Prediction of diabetics using Classification algorithms. *International Conference on Computational Intelligence and Data science(ICCIDS) Science Direct,*Vol.132(2018):1578-1585.

How to cite this article: Lydia, E.L., Sharmil, N., Shankar, K. and Maseleno, A. (2019). Analysing the Performance of Classification Algorithms on Diseases Datasets. *International Journal on Emerging Technologies*, 10(3): 224– 230.