IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-11, Nov 2018 E-ISSN: 2347-2693

Performance Analysis of Classifier Models to Predict Thyroid Disease

M. Saktheeswari[1*], T. Balasubramanian[2]

[1,2]Dept. of Computer Science, Periyar University, Salem-636011, Tamilnadu, India

[*]*Corresponding Author: saktheeswarisvmc5@gmail.com , Tel: +91 9486231497*

Available online at: www.ijcseonline.org

Accepted: 25/Nov/2018, Published: 30/Nov/2018

Abstract- Machine Learning Algorithm aims at providing computational method for accumulating, changing and updating knowledge in health care systems. In particular learning mechanism will assist us to procure knowledge from the data set. The classification of machine learning algorithm is used not only to detect diseases, but also measure better fidelity. This article emphasizes on codification of disease symptoms on thyroid disease among the public. Thyroid disease is rampant worldwide. There are feasibility of thyroid disease and disorder including thyroiditis and thyroid cancer. We used 7200 sample thyroid dataset from the University of California Irvine Machine Learning Repository, a large and highly imbalanced dataset that comprises both discrete and continuous attributes. In this work, we collate machine learning classifiers such as Logistic Regression, Linear Discriminant Analysis, Naive Bayes, k-Nearest Neighbours, Classification and Regression Tree, Support Vector Machine using python to classify the disease symptoms. This work is carried out using different classifiers to achieve more verisimilitude. The selected algorithms are evaluated using five performance metrics namely accuracy, sensitivity, specificity, F1-score and kappa, and also estimated from the confusion matrix produced by the selected classifier.

Keywords- CART Decision Tree; KNN algorithm; Support Vector Machine; Thyroid Disease Diagnosis; Linear Regression; Linear Discriminant Analysis.

I. INTRODUCTION

Through worldwide about 200 million people or nearly 15% of the entire adult population, are affected by thyroid disease. In the United States, thyroid disease affects about 27 million people, half of whom remain undiagnosed. Thyroid disease is one of the most common endocrine disorders found in worldwide. Thyroid disease often leads to thyroid dysfunction involving either hypo or hyperthyroidism, which are both relatively prevalent among the general population [1].

Data mining based applications are very valuable, it consists of many types of machine learning algorithms which all playing an important role to predict diseases in healthcare and medical science. In health care, there are large amount of data, and these data have no organizational value until converted into information and knowledge, which can help control costs, increase profits, and maintain high quality of patient care.

Thyroid disease is a study of Endocrinology which at most seems to be an element in science of Medicine. Thyroid is one of the most wide spread diseases that is frequently misunderstood and misdiagnosed [30].

The thyroid or thyroid gland is one of the most important and commonly called an endocrine gland present in the human body and is located in the human neck below the Adam's apple. The main purpose of thyroid is to produce thyroid hormones thyroxin (T4) and triiodothyronine (T3) into the blood stream as the principal hormones to control the body's metabolic rate and growth. The failure of thyroid hormone will lead to thyroid disorder or disease. When this hormone is secreted very little it may lead to hypo-thyroidism and when this hormone is secreted too much it may lead to hyper-thyroidism. Both excess and less thyroid hormone secretion cause health problems and sometimes may lead to death [8].

Thyroid Disease Diagnosis is one of the very difficult and tedious tasks because it needs lot of experience and deep knowledge about it. However, thyroid disorder can be identified using a TSH test, even before symptom onset [2]. One of the main tasks is to diagnosis the disease at the early stage with better accuracy. It offers various classification techniques to predict the disease accuracy through consequence.

Table 1. Symptoms of hypo Vs hyperthyroidism

Symptoms of hypothyroid	Symptoms of hyperthyroid
Dry hair	Hair loss
Slow heart beat	Rapid heart beat
Weight gain	Weight loss
Constipation	Sleeping difficulties
Memory loss	Irritability
Heavy menstrual periods	Scant menstrual periods
Muscle Aches	Muscles weakness
Cold intolerance	Heat intolerance

Diagnosing thyroid disease is a three-class classification problem and for this work numerous supervised methods is successfully applied to classify the different types of thyroid dysfunction [1-20]. Also a comparative thyroid disease diagnosis is realized by using different machine learning classifiers. To perform this work, the UCI machine learning database is used [22].

For accuracy measuring, 10-fold cross validation techniques have been applied among the classification models [31,32].

II. RELATED WORK

Li.LN.outang et. al. (2012) used an effective Computer aided diagnosis (CAD) system based on principle component analysis (PCA) and extreme learning machine (ELM) to assist the task of thyroid disease diagnosis. ELM classifier is explored to train an optimal predictive model [1].

Chen, Hui-Ling, et al (2012) proposed Expert system, Fisher Score Particle Swarm Optimization Support Vector Machines (FS-PSO-SVM) to evaluate the thyroid disease dataset. It is commonly used among researchers who use machine learning methods for thyroid disease diagnosis [2].

Dogantekin et. al. (2010) introduced Least Square Support Vector Machine (LSSVM) classifier to classify the disease [4].

Feyzul lah Temurtas (2009) used multilayer, probabilistic, and learning vector quantization neural networks to diagnosis thyroid disease [6].

Ozyilmaz et. al. (2002) proposed a system that includes Generalized Disciminant Analysis and Wavelet Support Vector Machine System (GDA_WSVM) for diagnosing thyroid disease through the classification accuracy and confusion matrix methods [8].

III. DATA PREPARATION

The thyroid dataset for this work has been retrieved from UCI machine learning repository. It consists of 7200 instances with 21 attributes. Among 21 attributes, 15 attributes are binary type and 6 attributes are continuous values. From the total instances, 3772 are training instances and 3428 are testing instances and it is shown in Table 2. The Class Labels has been classified into three classes: hyper-thyroidism, hypo- thyroidism and normal function of the thyroid gland and its shown in Table 3.

In the class distribution, the class value1 is interpreted as "Tested Negative for Thyroid", class value2 for "Tested Positive for Hypothyroid" and class value3 for "Tested Positive for Hyperthyroid".

Table 2. General information

Thyroid Disease (ann) Data Set			
Type	Classification	Origin	Real World
Features	21	(Continuous/Categorical)	(6/15)
Instances	7200	Classes	3
Missing Values		No	

Table 3. Attribute Description

Attribute	**Domain**	**Attribute**	**Domain**
Age	[0.01,0.97]	Lithium	[0,1]
Sex	[0,1]	Goiter	[0,1]
On_thyroxine	[0,1]	Tumor	[0,1]
Query_ On_thyroxine	[0,1]	Hypopituitary	[0,1]
On_antithyroid_medication	[0,1]	Psych	[0,1]
Sick	[0,1]	TSH	[0.0,0.53]
Pregnant	[0,1]	T3	[0.0005,0.18]
Thyroid surgery	[0,1]	TT4	[0.002,0.6]
T131-treatment	[0,1]	T4U	[0.017,0.233]
Query_hypothyroid	[0,1]	FT1	[0.002,0.642]
Query_hyperthyroid	[0,1]	Class	{1,2,3}

IV. MATERIALS AND METHODS

a. Data pre-processing and Feature selection

The pre-processing step is necessary to resolve several types of problems including noisy data, redundant data, missing data values, etc. In data mining, the high quality data leads to high quality results and reduces cost. Using required algorithms, missing data should be pre-processed so as to allow the whole dataset to be processed. Most of the existing algorithms are able to extract knowledge from dataset that stores discrete features. If the features are continuous, then the algorithms can be integrated to convert as discrete attributes [45].

All the input attributes are first binarized by using binarizer method through Python, so that they lie in a suitable range. It is done by normalizing the attributes into the interval of [0,1]. In this work, a wrapper method named Recursive Feature Elimination (RFE) is used. The wrappers can achieve better feature selection result in most cases. For the stopping criterion, when the result starts to get worse or the number of features reaches a predefined threshold, the procedure stops [46].

Among 21 attributes, 10 high potential attributes have selected to predict the thyroid disease. To evaluate the classification accuracy, 10-fold cross validation method is used to guarantee the valid result. Using five performance metrics, *accuracy*, *sensitivity*, *specificity*, *F1-score* and *kappa statistics*, the selected algorithms are evaluated and estimated from the produced confusion matrix.

b. Basic Classifiers

Classification can be done through various classifiers. The classifier selection is based on its performance. For this work, the six different classifiers: Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naive bayes, K-Nearest Neighbours (KNN), Classification and Regression Tree (CART), Support Vector Machine (SVM) are chosen.

(i) Logistic Regression

Logistic regression is sometimes called as logistic model or logit model. It analyzes the relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to a logistic curve. There are two models of logistic regression: binary logistic regression and multinomial logistic regression. Binary logistic regression is typically used when the dependent variable is dichotomous and the independent variables are either continuous or categorical. When the dependent variable is not dichotomous and is comprised of more than two categories, a multinomial logistic regression can be employed [24-25]. With logistic regression we model the natural log odds as a linear function of the explanatory variable:

$$\text{logit}(y) = \ln(\text{odds}) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta x \quad \text{(Equation 1)}$$

where p is the probability of interested outcome and x is the explanatory variable. The parameters of the logistic regression are α and β. This is the simple logistic model.

Taking the antilog of equation (1) on both sides, one can derive an equation for the prediction of the probability of the occurrence of interested outcome as

p= P (Y=interested outcome/X= χ, a specific value)

$$= \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{1}{1+e^{-(\alpha+\beta x)}} \quad \text{(Equation 2)}$$

Extending the logic of the simple logistic regression to multiple predictors, one may construct a complex logistic regression as

$$\text{logit}(y) = \ln(\text{odds}) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 \ldots + \beta_k x_k \quad \text{(Equation 3)}$$

Therefore,

p=P (Y=interested outcome/X1= χ1, ... Xk=χk)

$$= \frac{e^{\alpha\alpha+\beta_1 x_1 \ldots+\beta_k x_k+\beta x}}{1+e^{\alpha\alpha+\beta_1 x_1 \ldots+\beta_k x_k+\beta x}} = \frac{1}{1+e^{-(\alpha\alpha+\beta_1 x_1 \ldots+\beta_k x_k+\beta x)}} \quad \text{(Equation 4)}$$

(ii) Linear Discriminate Analysis

LDA was first proposed by Fisher (1936) as a classification technique. So far, it has been reported as the most commonly used technique in handling classification problems (Lee, Sung, & Chang, 1999). In the simplest type of LDA, two-group LDA, a linear discriminate function (LDF) that passes through the centroids (geometric centres) of the two groups can be used to discriminate between the two groups. The LDF is represented by

$$LDF = a + b_1x_1 + b_2x_2 + \ldots + b_px_p \quad \text{(Equation 5)}$$

where a is a constant, and b1 to bp are the regression coefficients for p variables. LDA has been widely applied in a considerable wide range of application areas, such as business investment, bankruptcy prediction, and market segment (Kim, Kim, Kim, Ye, & Lee, 2000; Lee, Jo, & Han, 1997; Trevino & Daniels, 1995)

(iii) Naive Bayes

It uses bayes rule to find the probability of a given instance from the test sample belonging to the specific class. The learning of this classifier is done by computing the mean and the variance of each dimension in each class.

Naïve Bayes make use of Bayesian theorem as the underlying technique. It applies conditional independence assumption between the features. Let (x1, x2, x3,….., xn) represent the vector to be classified. It is represented as 'n' dimensional space, where n represents the number of features considered [9]. For each class c, it assigns

probabilities as p(Ck| (x1, x2, x3,….., xn), which is represented in (2).

$$p(c_k/X) = \frac{p(c_k)p(x/c_k)}{p(x)} \quad \text{(Equation 6)}$$

where 'k' represents the number of classes considered for classification. Equation (6) represents the posterior probability for each class [10]. The main advantage of Naïve Bayes algorithm is its simplicity and it works well when the data are of higher dimensional space.

(iv) K-Nearest Neighbours

The working of k-nn is based on subspace method. It begins with plotting the data points into the provided space. Plotting is done by considering the class labels. K-nn classifier classifies each unlabeled observation by taking majority voting among 'k' nearest neighbors. Most commonly used distance metric is Euclidean distance [34]. Equation shows how the computation is done using Euclidean distance.

$$D(x, y) = \sum\sqrt{(x_i^2 - y_i^2)} \quad \text{(Equation 7)}$$

D(x,y) represents the distance between two selected input vectors. xi and yi represents the data point contained in the vector space and new data point to be classified respectively.

(v) Classification and Regression Tree

CART, a statistical procedure [35], is primarily used as a classification tool with the objective to classify an object into two or more categories. A CART analysis generally consists of three steps. In a first step an overgrown tree is built, which closely describes the training set. This tree is called the maximal tree and is grown using a binary split procedure. In a next step the overgrown tree, which shows overfitting is pruned. During this procedure a series of less complex trees is derived from the maximal tree. In the final step, the tree with the optimal tree size is selected using a cross-validation (CV) procedure.

(vi) Support Vector Machine

Support Vector Machine is one of the most popular supervised algorithms used for binary classification. SVM tries to discriminate objects belonging to different classes [44]. Different kernel functions linear, polynomial, radial Basis and quadratic are used.

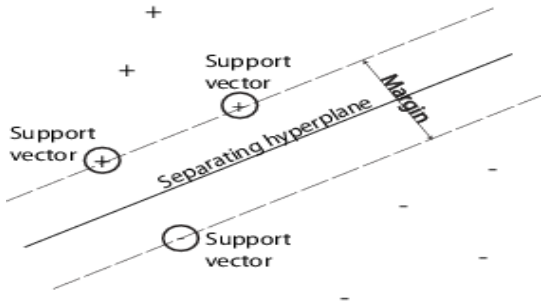

Fig. 1. Support Vector Machine

SVM starts working by mapping the input data for training into high dimensional space and a separating hyper plane is found out that can discriminate between two classes and then the margin for the hyper plane is maximized. A simplified representation of SVM is shown in Fig 1. Whenever a new data is given for testing, it will be mapped to the previous high dimensional space and determines in which class area the plotted points belong to. That area represents the class for the new data.

V. RESULT ANALYSIS AND DISCUSSION

The criteria taken for evaluating the performance our classifiers comparisons are accuracy, specificity, sensitivity, precision, kappa and area under ROC curve.

a. Evaluation Criteria

The above mentioned criteria are used to evaluate the performance of the implemented system. The actual output and predicted results are compared to create a confusion matrix. From the generated values accuracy, specificity and sensitivity can be calculated.

(i) Accuracy

Accuracy is calculated as the ratio of true cases to total cases in the test dataset i.e. the ratio of correctly classified samples (True positive (TP), True Negative (TN)) with the total samples available (TP, TN, False Negative (FN), False Positive (FP)), given in the equation(1):

Accuracy = (TP+TN) / (TP+FP+TN+FN) (Equation 8)

Where True Positive, False Negative, False Positive, True Negative values are obtained from confusion matrix [39].

(ii) Sensitivity

Sensitivity is also termed as true positive rate or recall, which gives the ratio of true positive to the sum of true positive and false negative. The values for TP and FN are derived from the confusion matrix [39].

Sensitivity = TP/ (TP+FN) (Equation 9)

(iii) Specificity

Specificity is also termed as true negative rate which gives the ratio of true negative to the total negative cases i.e. proportion of TN to (TN+FP)[39].

Specificity = TN / (TN+FP) (Equation 10)

(iv) Precision

Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances [39].

Precision = TP/(TP+FP) (Equation 11)

(v) Kappa

Cohen's kappa coefficient (κ) is a statistic which measures inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure

than simple percent agreement calculation, as κ takes into account the possibility of the agreement occurring by chance. There is controversy surrounding Cohen's Kappa due to the difficulty in interpreting indices of agreement. Some researchers have suggested that it is conceptually simpler to evaluate disagreement between items [40].

Cohen's kappa measures the agreement between two raters. The first mention of a kappa-like statistic is attributed to Galton (1892);[41] see Smeeton (1985) [42].

The definition of k is

$$K = \frac{p_o - p_e}{1 - p_e} \quad \text{(Equation 12)}$$

where p_o is the relative observed agreement among raters (identical to accuracy), and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category. Kappa is also used to compare performance in machine learning but the directional version known as Informedness or Youden's J statistic is argued to be more appropriate for supervised learning[43].

(vi) F1-Score

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = \frac{2}{1/precision + 1/recall} \quad \text{(Equation 13)}$$

(vii) Confusion Matrix

Confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing [39].

True Positive	False Positive
False Negative	True Negative

b. Performance Analysis

(i) Base Learners

From the below table, it is clear that the maximum accuracy obtained is only 99%, which is for CART. Algorithms comparison is given below as shown in fig. (2). CART achieves better accuracy of 99% than others.

Table 4. Overall performance measure of classification algorithms on datasets

Model	Kappa	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1 – Score
Logistic Regression	0.44	0.78	0.10	0.94	0.94	0.93
Linear Discriminant Analysis	0.27	0.50	0.93	0.93	0.93	0.91
Naive Bayes	0.70	0.98	0.25	0.97	0.95	0.95
k-Nearest Neighbors	0.68	0.10	0.10	0.96	0.96	0.96
Classification and Regression Tree	0.9	0.10	0.10	0.99	100	100
Support Vector Machines	0.06	0.11	0.81	0.87	0.89	0.88

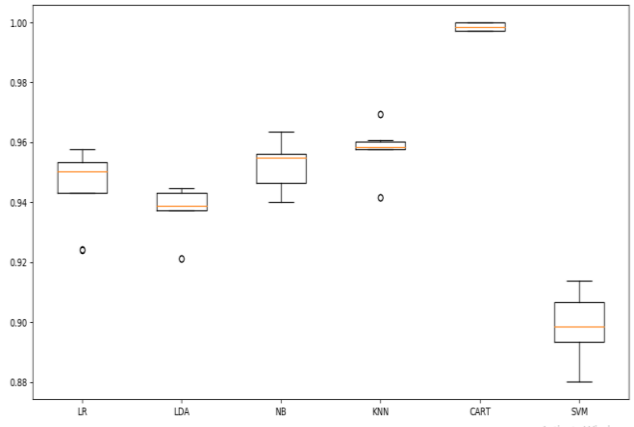

Fig. 2. Box plot representation of comparing six classifiers

(ii) Receiver operating characteristics (ROC) curves

Receiver operating characteristics graph is a technique for visualizing, organizing and selecting classifiers based on their performance. ROC graphs have been used in signal detection theory to depict the trade-off between hit rates and false alarm rates of classifiers (Egan, 1975; Swets et al., 2000). ROC analysis has been extending for use in visualizing and analyzing the behaviour of diagnostic system (Swets, 1998). The medical data mining community has an extensive literature on the use of ROC graphs for diagnostic testing (Zou, 2002).

The area under the receiver operating characteristics curves (AUC-ROC) in logistic regression model for the class 0 is 0.93, class 1 is 0.54, class 2 is 0.66 and total AUC is 0.71 as shown in Fig. 3. The area under the receiver operating characteristics curves (AUC-ROC) in linear discriminate

analysis model for the class 0 is 0.78, class 1 is 0.50, class 2 is 0.59 and total AUC is 0.62 as shown in Fig.4.

The area under the receiver operating characteristics curves (AUC-ROC) in k-nearest neighbors model for the class 0 is 0.90, class 1 is 0.74, class 2 is 0.78 and total AUC is 0.81 as shown in Fig. 5. The area under the receiver operating characteristics curves (AUC-ROC) in naive bayes model for the class 0 is 0.61, class 1 is 0.97, class 2 is 0.96 and total AUC is 0.85 as shown in Fig. 6.

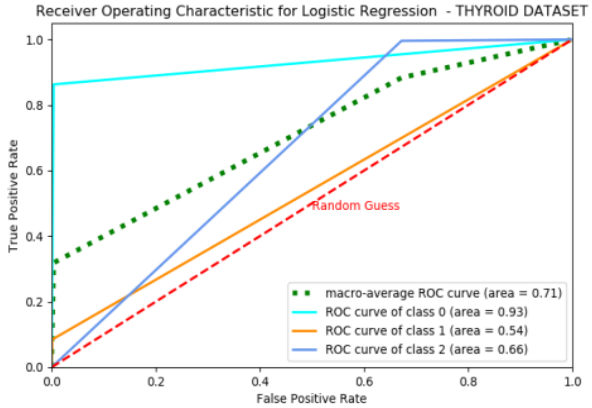

Fig. 3. AUC-ROC for Logistic Regression models

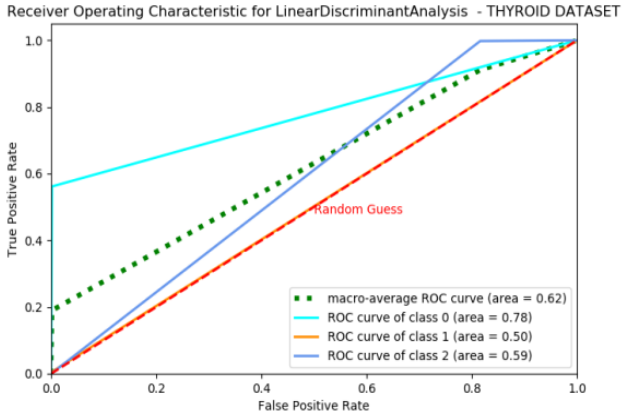

Fig. 4. AUC-ROC for Linear Discriminant Analysis models

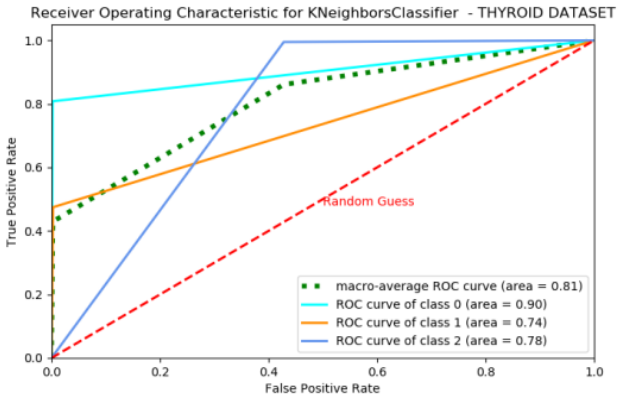

Fig 5. AUC-ROC for K-nearest neighbours models

The area under the receiver operating characteristics curves (AUC-ROC) in CART model for the class 0 is 0.93, class 1 is 0.54, class 2 is 0.66 and total AUC is 0.71 as shown in Fig. 7.

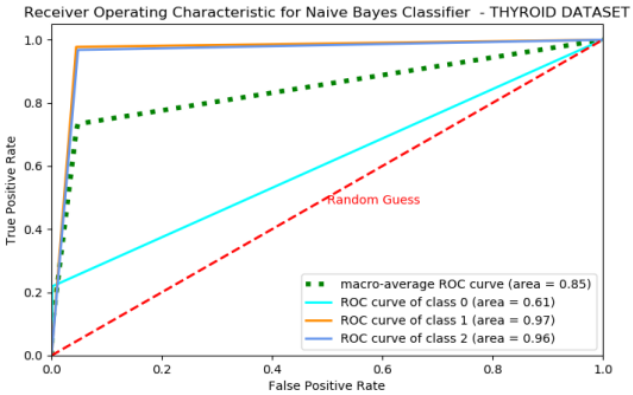

Fig. 6. AUC-ROC for Naive Bayes models

The area under the receiver operating characteristics curves (AUC-ROC) in support vector machine model for the class is 0 is 0.58, class 1 is 0.50, class 2 is 0.53 and total AUC is 0.53 as shown in Fig. 8.

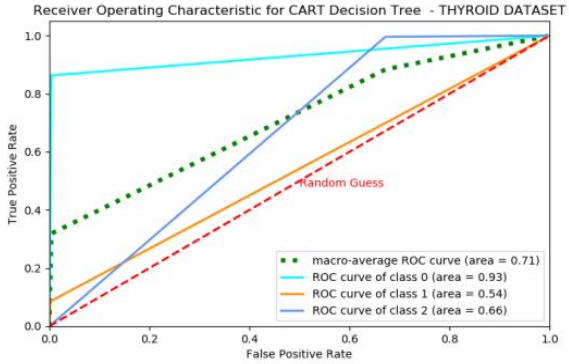

Fig. 7. AUC-ROC for Classification and Regression models

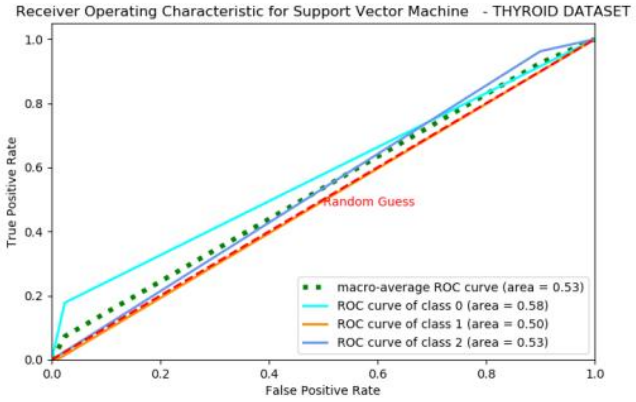

Fig. 8. AUC-ROC for Support Vector Machine models

VI. CONCLUSION

The objective of this work is to provide a comparative study of different classifiers in machine learning that can be

employed in diagnosis of thyroid disease dataset. In recent years, for the efficient diagnosis of various diseases under health care domain the defined classifier models are used in the existing researches. In this proposed work, it is identified that the Classification and Regression Tree (CART) is the best in diagnosis of thyroid disease that shows better performance than other classifier models used. Here, the evaluation is done based on criteria such as accuracy, sensitivity and specificity, kappa statistics, precision and f1-score. The future work will be to propose a new classifier machine learning algorithm for diagnosing other health care disease such as thyroid.

ACKNOWLEDGEMENT

The authors would like to highly appreciate and grateful knowledge to UCI Machine Learning Repository (http: //archive.ics.uci.edu/ ml/machine-learning-databases/ thyroid-disease) for obtaining the thyroid disease dataset.

REFERENCES

[1] Li.LN.ouyang JH,Chen HL,Liu DY. A computer aided diagnosis system for thyroid disease using extreme learning machine. J Med Syst 2012;36:3327-37.

[2] Chen HL, Yang B, Wang G, Liu J Chen YD, Liu DY. A three stage expert system based on support vector machines for thyroid disease diagnosis. J Med Syst 2012;36;1953-63.

[3] Dogantekin E.Dogantekin A,Avci D. An expert system based on Generalized Discriminant analysis and Wavelet Support Vector Machine for diagnosis of thyroid disease. Expert Syst Appl 2011;38:146-50.

[4] Dogantekin E.Dogantekin A,Avci D. An Automatic diagnosis system based on thyroid gland; ADSTG. Expert Syst. Appl 2010;37;6368-72.

[5] Pasi L Similarity classifier applied to medical data sets. 2004. 10 sivua. Fuzziness in Finland '04. International conference on soft computing, Helsinki. Estonia: Finland & Gulf of Finland & Tallinn; 2004.

[6] Feyzul lah Temurtas" A comparative study on thyroid disease diagnosis using neural networks", Elsevier, Expert Systems with Applications 36 (2009) 944–949.

[7] Serpen G. Jiang H, Allred L. Performance analysis of probabilistic potential function neural network classifier. In Proceedings of artificial neural networks in engineering conference viol. 7 : 1997. p.471-6.

[8] Ozyilmaz, L., Yildirim,T.(2002). Diagnosis of thyroid disease using artificial neural network methods. In proceedings of ICONIP'02 9th international conference on neural information processing (pp.2033-2036). Singapore: Orchid Country Club.

[9] Liu DY. Chen HL Yang B.XE L, Li LN, Liu J. Design of an enhanced fuzzy k-nearest neighbour classifier based computer aided diagnostic system for thyroid disease. J Med Syst 2012;36:3243-4354.

[10] Ng SK. McLachlan GJ. Extension of mixture-of-experts networks for binary classification of hierarchical data. Artif Intell Med 2007;41:57-67.

[11] Chang WW.Yeh WC, Huang PC, A hybrid immune-estimation distribution of algorithm for mining thyroid gland data. Artif Intell Med 2007;41:57-67.

[12] Kodaz H. Ozsen S,Arslan A,Gunes S. Arslan A, Gunes S, Medical application of information gain based artificial immune recognition system(AIRS): Diagnosis of thyroid disease. Expert syst. Appl 2009;36:3086-92.

[13] Keles A.Keles A.ESTDD: expert system for thyroid disease diagnosis.Expert Syst Appl. 2008;34:242-6.

[14] Duch W.Adamezak R. Grabezewski K. A new methodology of extraction,optimization and application of crisp and fuzzy logic rules. IEEE Trams Neural Netw 2001;12:277-306.

[15] Bologna G. A model for single and multiple knowledge based networks. Artif Intell Med 2003;28:141-63.

[16] Abe S.Thawonmas R. A fuzzy classifier with ellipsoidal regions. IEEE Trans Fuzzy Syst 1997;5:358-63.

[17] Abe S. Thawonmas R, Kayama M, A fuzzy classifier with ellipsoidal regions for diagnosis problems. IEEE Trans Syst Man Cybern C Appl Rev 1999;29:140-9.

[18] Sharaf-EI-Deen DA, Moawad IF,Khalifa ME. A new hybrid case-based reasoning approach for medical diagnosis systems. J Med Syst 2014;38;9-19.

[19] Falco ID. Differential evolution for automatic rule extraction from medical databases. Appl soft Comput 2013;13;1265-83.

[20] Zhu P.Hu Q. Rule extraction from support vector machine based on consistent region covering reduction. Knowl Based Syst 2013;42;1-8.

[21] Napierala K. Stefanowski J. BRACID: a comprehensive approach to learning rules from imbalanced data. J Intell Inf Syst 2012;39;335-73.s

[22] Yongchuan Tang, Wuming Pan, Haiming Li and Yang Xu. Fuzzy Naïve Bayes Classifier based on fuzzy clustering. IEEE International Conference on System, Man and Cybernetics, Vol.5, 2002.

[23] Yaguang Ji, Songnian Yu and Yafeng Zang. A Novel Naïve Bayes Model: Packaged Hidden Naïve Bayes. Sixth IEEE Joint International Conference on Information Technology and Artificial Intelligence. pp.484-487, 2011.

[24] Park, Hyeoun-Ae, An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain, J Korean Acad Nurs Vol.43 No.2 April 2013

[25] Le,Cessie S, Van Houwelingen JC.Ridge estimators in logistic regression. Applied statistics.1992;p.191-201.doi:10.2307/2347628

[26] Jiawei Han , Micheline Kamber, Data Mining Concepts and Techniques. Published by Elsevier 2006.

[27]University of California, Irvine Learning Repository. (http: //archive.ics.uci.edu/ ml/machine-learning-databases/ thyroid-disease/)

[28] Anupam Shukla, Prabhdeep Kaur, Ritu Tiwari and R.R. Janghel, Diagnosis of Thyroid disease using Artificial Neural Network. In Proceedings of IEEE IACC 2009.

[29] Umar Sidiq, "Comparative Study of Existing Techniques for Diagnosing Various Thyroid Ailments", Global Journal of Computer Science and Technology: E Network, Web & Security Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online ISSN: 0975-4172 & Print ISSN: 0975-4350

[30] Azar AT,Hassanien AE, Kim T(2012) Expert system based on neural fuzzy rules for thyroid disease diagnosis springer berlin.

[31] Liu B, Fang L, Liu F, Wang X, Chen J, Chou KC. Identification of real microRNA precursors with a pseudo structure status composition approach. PloS one. 2015;10(3):e0121501 doi: 10.1371/journal.pone.0121501 [PMC free article] [PubMed [32]

[32] Liu B, Fang L, Liu F, Wang X, Chou KC. iMiRNA-PseDPC: microRNA precursor identification with a pseudo distance-pair composition approach. Journal of Biomolecular Structure and Dynamics. 2016;34(1):223–235. doi: 10.1080/07391102.2015.1014422 [PubMed]

[34] Shiliang Sun, Rongquing Huang. An adaptive k-nearest neighbor. IEEE Seventh International Conference on Fuzzy System and Knowledge Discovery. pp.91-94, 2010.

[35] Thyroid disease diagnosis via hybrid architecture composing rough data sets theory & machine learning algorithms, V. Prasad, T. Srinivasa Rao, M.Surendra Prasad Babu, Springer, soft comput(2016) 20:1179-1189, DOI 10.1007/s00500-014-1581-5
[36] L.Breiman, J.Friedman, R. Olshen,et al; Classification & Regrssion trees, chapman & Hall, London(1984)
[37] L.Rokach, O.Maimon, classification trees, Data Mining & Knowledge Discovery, Springer New York Dordrecht Heidelberg London,(2010) 149-174
[38] Derrac, J., Chiclana, F., Garcia, S., & Herrera, F.(2016). Evloutionary fuzzy k-nearest neighbours algorithm using interval-valued fuzzy sets. Information sciences, 329, 144-163.
[39] Tom Fawcett, An introduction to ROC analysis, Institute for the study of learning and expertise,2164 staunton court, palo alto, CA 94306, USA,Elsevier
[40] Pontius, Robert; Millones, Marco (2011). "Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment". International Journal of Remote Sensing. 32: 4407–4429.
[41] Galton, F. (1892). Finger Prints Macmillan, London.
[42] Smeeton, N.C. (1985). "Early History of the Kappa tatistic". Biometrics. 41: 795. JSTOR 2531300.
[43] Powers, David M. W. (2012). "The Problem with Kappa" (PDF). Conference of the European Chapter of the Association for Computational Linguistics (EACL2012) Joint ROBUS-UNSUP Workshop.
[44] Nitin Aji Bhaskar. Performance Analysis of Neural Network and Support Vector Machine in Detection of Myocardial Infarction. International Conference on Information and Communication Technologies, pp.20-30, 2015.
[45] Kotsiantis, S. B., D. Kanellopoulos, and P.E. Pintelas. "Data preprocessing for supervised learning" International Journal of Computer Science 1.2(2006):111-117.
[46] Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. Gene selection for cancer classification using support vector machines, Machine Learning, Vol.46, Issue 1-3, Pp.389-422,2002

AUTHORS' PROFILES

M.Saktheeswari (Corresponding author) received her M.Sc Degree from V.V.V. Arts and Science College for Women , Virudhunagar affiliated with Madurai Kamaraj University and M. Phil Degree from Manonmaniam Sundranar University, Tirunelvelli. She has qualified in NET & SET eligibility examinations. Now, she is pursuing her Ph. D (Part time) in Periyar University, Salem. Currently she is working as a Assistant Professor in Department of Computer Science & Applications at Sri Vidya Mandir Arts and Science College, Uthangarai, Krishnagiri(Dt), Tamilnadu. Her research area is Machine Learning Techniques.

T. Balasubramanian received his Master Degree in Computer Science from Bharathidasan University, Trichy in 1996. He has received his Master of Philosophy degree from Periyar University Salem in year 2007. In the specialization of Data Mining, he received his Doctorate from Bharathiar University in 2015. He published more than 32 research papers in various International and National Journals. He also presented 8 research articles in various International Conferences and 12 papers in National and State level conferences & seminars. He has written 5 books under various domains of computer science. Currently he is working as a Associate Professor in the PG and Research Department of of Computer Science in Sri Vidya Mandir Arts and Science College, Uthangarai, Krishnagiri (Dt). He has more than 17 Years of teaching experience. His domain interest is Data mining, Big data, Network Security.

