Studies on Thyroid Disease Classification

Oleti Saisravan

Dept of Computer Science, Sri Venkateswara University, Tirupati

Abstract— *Recently, thyroid diseases are more and more spread worldwide. The main objective of thyroid is to produce thyroid hormones. Factors that affect the thyroid function are: stress, infection, trauma, toxins, low-calorie diet, certain medication etc. It is very important to prevent such diseases rather than cure them, because the majority of treatments consist in long term medication or in chirurgical intervention. The current study refers to thyroid disease classification in two of the most common thyroid dysfunctions (hyperthyroidism and hypothyroidism) among the population. The produced hormones go through the bloodstream to all the other organs which help to control metabolism and growth development in both in adults and in children.*

The authors analyzed and compared three SVM classification kernel methods: Linear, Polynomial and RBF kernels. The results indicate a significant accuracy for all the three kernels mentioned above, the best is RBF kernel, which is hot highest accuracy rate 98.5% being that of the SVM classification model.

I. INTRODUCTION

The thyroid gland secretes hormones which controls a lot of things in the human body system like metabolize the food, use energy, and sleep patterns, temperature preferences, body weight balance and a lot more. In this research work to comparative thyroid disease diagnosis were performed by using Machine learning techniques that is Support Vector Machine (SVM). Factors that affect the thyroid function are: stress, infection, trauma, toxins, low-calorie diet, certain medication etc. It is very important to prevent such diseases rather than cure them, because the majority of treatments consist in long term medication or in chirurgical intervention. The current study refers to thyroid disease classification in two of the most common thyroid dysfunctions (hyperthyroidism and hypothyroidism) among the population.

Nowadays, thyroid disorders destruction the normal functioning of the thyroid gland which causes anomalous production of hormones leading to hyperthyroidism [1]. The occurrence of hypothyroidism in the developed world is estimated to be about 4- 5%. Hypothyroidism may cause high cholesterol levels, an increase in blood pressure, cardiovascular complications, decreased fertility, and depression if not properly treated. Thyroid is a butterfly-shaped gland, which is located at the bottom of the throat responsible for producing two active thyroid hormones, levothyroxine (T4) and triiodothyronine (T3) that affect some functions of the body such as: stabilizing body temperature, blood pressure, regulating the heart rate etc. Reverse T3 (RT3) is manufactured from thyroxine (T4), and its role is to block the action of T3.

The technology and information in medical sciences, the computer science professionals are capable of providing expert advisory system [5] .To diagnose different kinds of diseases with high accuracy. The medical professionals are made to use these systems due to some developed errors during general diagnosis process [6]. Disease diagnosis operations using EAS are performed based on sets of disease symptoms. These systems are based on machine learning technique which helps the physician to minimize the costs and time in effective diagnoses.

An abnormal function of the thyroid implies the occurrence of hyperthyroidism and hypothyroidism, two of the common thyroid affections. Hypothyroidism (underactive thyroid or low thyroid) means that the thyroid gland doesn't produce enough of certain important hormones. Without an adequate treatment, hypothyroidism can cause various health problems such as: obesity, joint pain, infertility and heart disease. Hyperthyroidism (overactive thyroid) refers to a condition in which the thyroid gland produces too much of the hormone thyroxin.

II. SUPPORT VECTOR MACHINES (SVM)

SVM is a managed AI procedure. For the most part, Support Vector Machines is viewed as a characterization approach, it however can be utilized in the two sorts of arrangement and relapse issues [2][3]. It can undoubtedly deal with various persistent and unmitigated factors. SVM develops a hyperplane in multi-layered space to isolate various classes. SVM

produces ideal hyperplane in an iterative way, which is utilized to limit a mistake. The center thought of SVM is to find a most extreme minimal hyperplane that best partitions the dataset into classes [9].

Support vectors are the data of interest, which are nearest to the hyperplane. These focuses will characterize the isolating line better by working out edges. These focuses are more pertinent to the development of the classifier. A hyperplane is a choice plane what isolates between a bunch of items having different class participations [4][7][10]. An edge is a hole between the two lines on the nearest class focuses. This is determined as the opposite separation from the line to help vectors or nearest focuses. In the event that the edge is in the middle between the classes, it is viewed as a decent edge, a more modest edge is a terrible edge.

The principal objective is to isolate the given dataset in the most ideal manner. The distance between the either closest focuses is known as the edge. The goal is to choose a hyperplane with the most extreme conceivable edge between help vectors in the given dataset.

III. SVM KERNELS

The SVM calculation is executed by and by utilizing a piece. A bit changes an info information space into the necessary structure. SVM utilizes a strategy called the piece stunt. Here, the piece takes a low-layered input space and changes it into a higher layered space. At the end of the day, you can say that it switches nonseparable issue over completely to distinct issues by adding more aspect to it. It is most valuable in non-straight partition issue. Portion stunt assists you with building a more exact classifier.

3.1 Linear Kernel

A straight bit can be utilized as would be expected dab item any two offered viewpoints. The item between two vectors is the amount of the increase of each sets of information values.

K(x, xi) = sum(x * xi)

3.2 Polynomial Kernel

A polynomial piece is a more summed up type of the direct bit. The polynomial portion can recognize bended or nonlinear information space.

K(x,xi) = 1 + sum(x * xi)^d

Where d is the level of the polynomial. d=1 is like the direct change. The degree should be physically indicated in the learning calculation.

3.3 Radiasl Basis Function Kernel

The Radial premise work part is a well known piece work normally utilized in help vector machine order. RBF can plan an info space in boundless layered space.

K(x,xi) = exp(- gamma * sum((x - xi^2))

Here gamma is a boundary, which goes from 0 to 1. A higher worth of gamma will impeccably fit the preparation dataset, which causes over-fitting. Gamma=0.1 is viewed as a decent default esteem. The worth of gamma should be physically indicated in the learning calculation.

IV. EXPERIMENTAL RESULTS

This part gives results and related conversation on information driven analysis of thyroid dataset was gathered from UCI repositary [8]. WEKA is a cutting edge office for creating AI (ML) methods and their application to true information mining issues. The information record typically utilized by WEKA is in ARFF document design. ARFF represents Attribute Relation File Format, which comprises of extraordinary labels to demonstrate separating in the information document. WEKA implements algorithms for data pre-processing, classification. The dataset contains 3772 instances and 30 attributes. There are two distinct classes namely negative and sick. The negative class has 3541 instances and sick 231 instances. The analyses

were performed considering 70% of the complete examples were preparing information and 30% were trying information. The statistical summary of the dataset as shown in the figure-1.

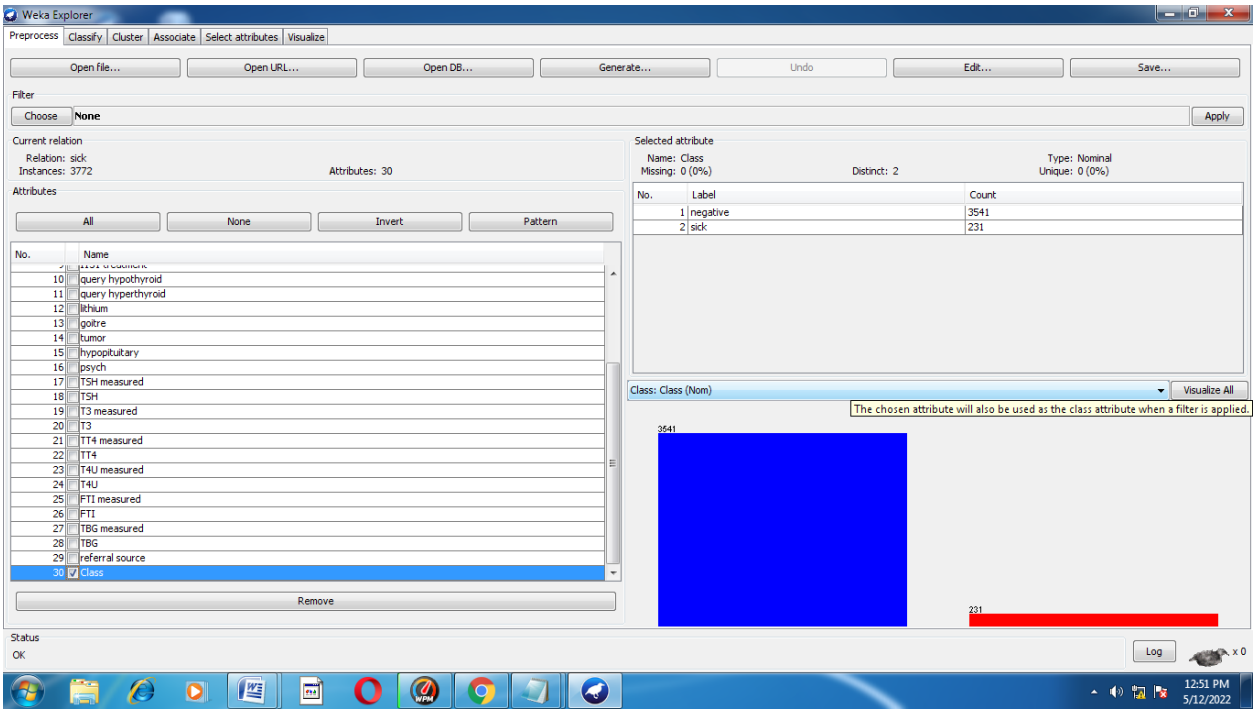

FIGURE 1: Statistical Summary of Dataset

The Experimental results of SVM classification with kernel selection compared the on basis of correctly classified instances is shown in the table-1 and same shown in the figure-2.

TABLE 1
PERFORMANCE OF CLASSIFIERS OF SVM WITH DIFFERENT KERNELS

Algorithm	Accuracy	precision	Recall
Polynomial kernel	93	88	93
Linear kernel	96.7	96.5	96.8
RBF kernel	**98.5**	98.5	98.5

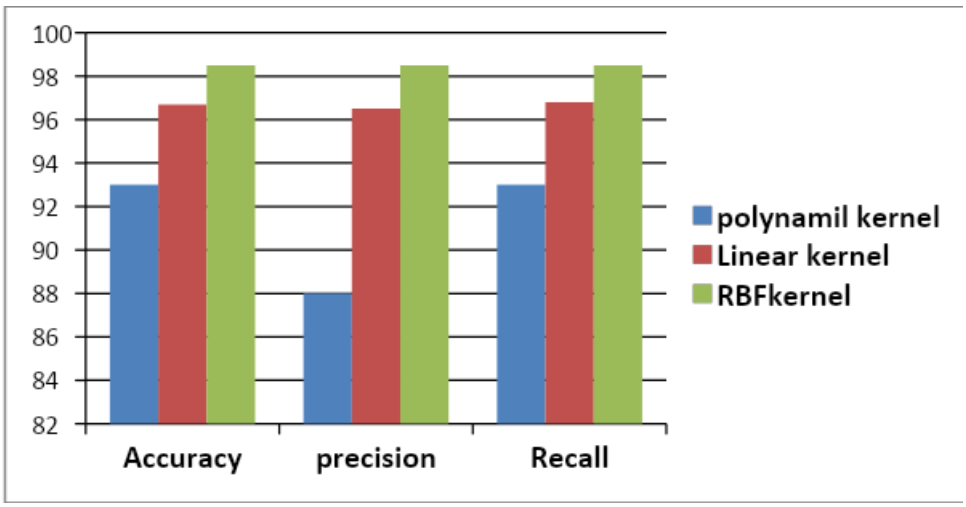

FIGURE 2: Experimental Results

From the figure-2, we notice the exhibition of SVM with the polynomial kernel has got 93%, linear kernel as achieved 96.7% accuracy and the RBF kernel got 98.5% of accuracy. So, in the SVM classification kernel selection is important. Accuracy

will depend on kernel selection. So in the thyroid diseases prediction RBF kernel has got highest accuracy (98.5%) when compared to linear and polynomial kernels. The screen shots of experimental results are shown in the figure-3 and figure-4.

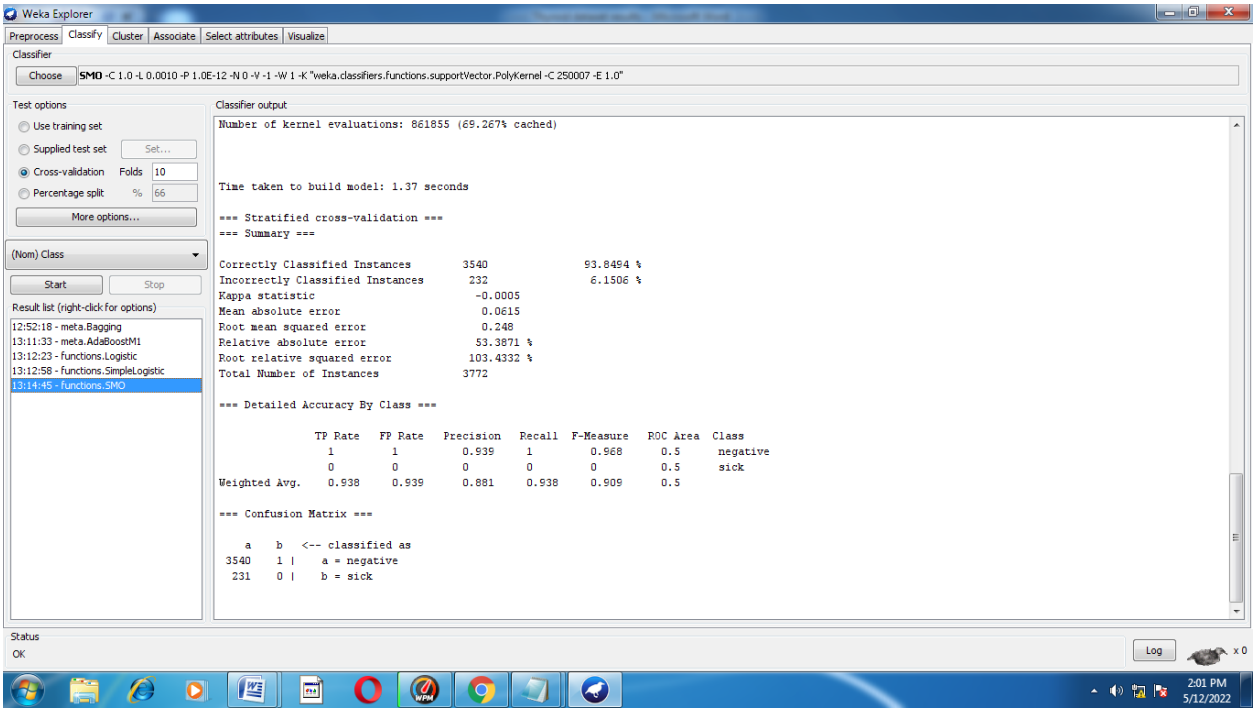

FIGURE 3: Experimental results Screen shot

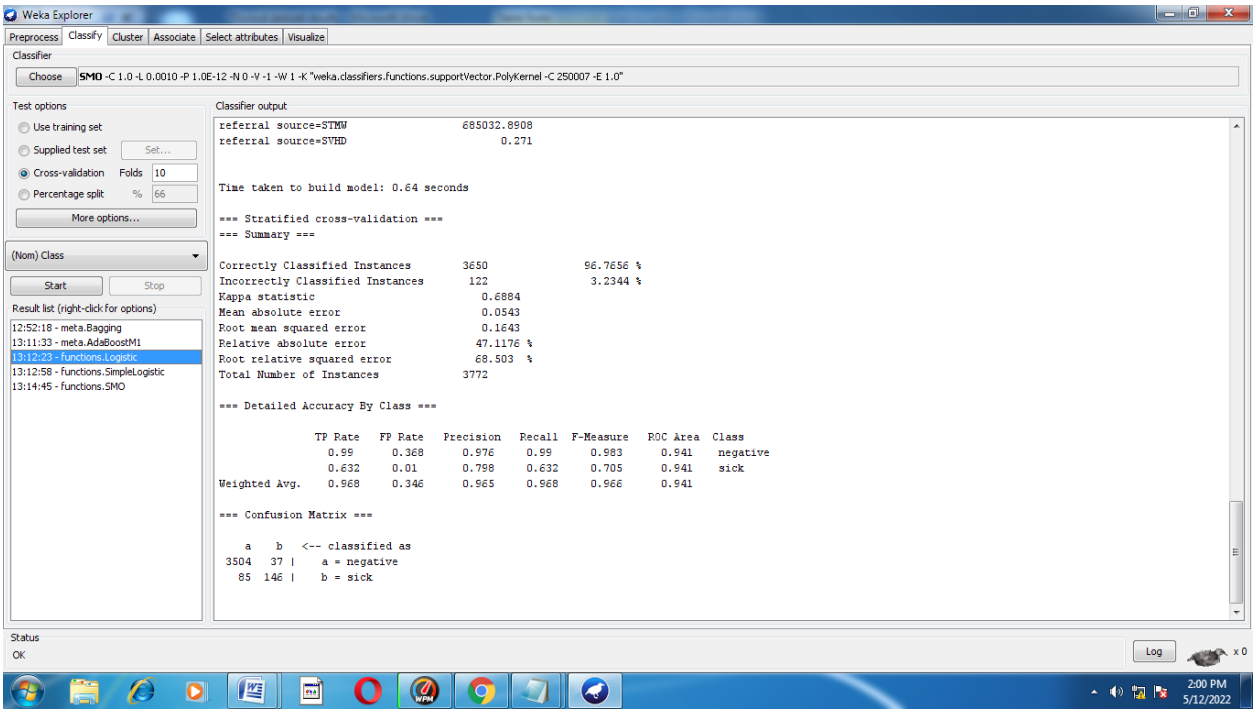

FIGURE 4: Experimental results Screen shot

V. CONCLUSION

The thyroid gland is the primary and biggest gland in the endocrine system. The objective of this work is aimed to show the classes of thyroid from the available raw medical dataset that helps to the accurate diagnosis. This paper examines thyroid disease diagnosis using SVM classification with three kernels computations. Our preliminary outcomes showed that the SVM with RBF kernel estimation gives better gathering accuracy achieved in distinctive hypothyroid dataset to determine the positive and the negative cases from the entire dataset. The enhancement can be made by using SVM with optimization

of kernels or rule extraction algorithms. It was compared and it can be seen that SVM successfully used to help the diagnosis of thyroid disease. It is observed that the SVM with RBF kernel performed in three kernels with respect to the accuracy of the diagnose the thyroid disease.

An acknowledgement section may be presented after the conclusion, if desired.

REFERENCES

[1] Geetha, K., and CAPT S. SANTHOSH BABOO. "EFFICIENT THYROID DISEASE CLASSIFICATION USING DIFFERENTIAL EVOLUTION WITH SVM." Journal of Theoretical & Applied Information Technology 88.3 (2016).

[2] G. Ravi Kumar, K.Nagamani and G.Anjan Babu, “A Framework of Dimensionality Reduction utilizing PCA for Neural Network Prediction”, Lecture Notes on Data Engineering and Communications Technologies, Volume-37, Pages:173 – 180, Springer Nature Singapore Pte Ltd, 2020

[3] H. Witten and E. Frank, “Data mining: practical machine learning tools and techniques with Java implementations”, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., (2000)

[4] J. Han and M. Kamber,” Data Mining concepts and Techniques”, the Morgan Kaufmann

[5] Karthikeyan, S., and T. Christopher. "A hybrid clustering approach using artificial bee colony (ABC) and particle swarm optimization." International Journal of Computer Applications100.15 (2014).

[6] Margret, J., B. Lakshmipathi, and S. Aswani Kumar. "Diagnosis of Thyroid Disorders using Decision Tree Splitting Rules." International Journal of Computer Applications 44.8 (2012): 43-46.

[7] S.Rahamat Basha and Surya Bhupal Rao G.Ravi Kumar, “A Summarization on Text Mining Techniques for Information Extracting from Applications and Issues”, Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, PP: 324-332, 2020, Institute of Mechanics of Continua and Mathematical Sciences

[8] UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)

[9] Vapnik V.N, “Statistical learning Theory”, John Wiley and Sons, New York, USA, 1998.

[10] Vapnik V.N,”The Natural of Statistical Learning Theory, Springer-Verlag, New York, USA, 1995.