Int J Pharm Bio Sci 2014 Oct; 5(4): (B) 447 - 453

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

SUPPORT VECTOR MACHINE AND K- NEAREST NEIGHBOR BASED ANALYSIS FOR THE PREDICTION OF HYPOTHYROID

K.SARAVANA KUMAR*[1] AND DR. R. MANICKA CHEZIAN[2]

[1]Research Scholar, NGM College, Pollachi, India.
[2]Associate Professor, NGM College, Pollachi, India.

ABSTRACT

The thyroid gland produces two important hormones T3 (triiodothyronine) and T4 (thyroxine). These hormones contain in the blood and control many undertakings in human bodies. Calories burning, heartbeats are supervised by T3 and T4. This is called human body metabolism. The thyroid, which is in good condition, produces as required amounts of hormones for the function of good mechanism of metabolism. The thyroid illness mostly affects women than men. Hypothyroidism and Hyperthyroidism are the two levels of thyroid malfunction. In data mining, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are the two important modes applied to the prediction of hypothyroid. This paper discusses that predictions of Hypothyroid using K- Nearest Neighbor better than the Support Vector Machine.

KEYWORDS: Thyroid, Thyroid Gland, Data Mining, K-Nearest Neighbor, Support Vector Machine, Prediction.

K.SARAVANA KUMAR
Research Scholar, NGM College, Pollachi, India.

***Corresponding author**

DR. R. MANICKA CHEZIAN
Associate Professor, NGM College, Pollachi, India.

Co-author

INTRODUCTION

In the human body, the thyroid gland is very important organ. It produces thyroid hormones to maintain our body metabolism [1]. The thyroid gland is located in the front of the neck and below the Adam's apple. The thyroid produces two major hormones called T3 (triiodothyronine) and T4 (thyroxine). These T3 and T4 hormones travel in our blood to all parts of our body and affect almost every cell in the body, and helps to control our body's functions [2]. If the amount of thyroid hormone decreased in our blood, our body function gets slow down, this condition is called hypothyroidism[3]. The indications of hypothyroid are depression, exhaustion of body strength, tiredness, constipation, excess weight, cramps, dry skin, sexual disorders and infertility. If the increased amount of thyroid hormones seen in our blood, our body functions will speeds up. This condition is called hyperthyroidism[4]. The symptoms of hyperthyroid are nervousness, palpitation, exhausted body strength, tremors, loss of weight, diarrhea, menstrual disorder and ex-ophthalmic. The thyroid in a good condition will produce the right amounts of hormones needed to keep your body's metabolism working, i.e. not too fast or too slow [5]. Nowadays, there are a number of clinical tests are widely applied for thyroid diagnosis[6]. Data mining is to extract the required information from the large database[7]. Data mining have lots of techniques (Association Rule Mining, Clustering, Classification, Regression, and Summarization). It is very helpful for Medical domain. The data mining helps to reduce the cost and increase profit in each and every business. It is very useful in the medical domain. The paper mainly focused on two data mining techniques.

DATASET AND METHODOLOGY

The Dataset are taken from University of California Irvine (UCI) Repository. Hypothyroid dataset are used for experimental purposes. The dataset has 3090 data. 2941 is belongs to negative, 149 belongs to hypothyroid. In table 1, the last column is the class, so all six features are used to classify data and table 1 shows first column as parameters set and the next column expands corresponding values attained.

Table 1
Hypothyroid Parameter Details

Parameters	*Value*
Age	Continuous
Gender	Male (M), Female (F)
TSH	Continuous
T3	Continuous
T4	Continuous
Results	**Negative, Hypothyroid**

KNN (K-Nearest Neighbor)

K- Nearest Neighbors algorithm is a non-limitation, method used for classification. The input consists of the *k* closest instructing instances in the characteristic space. The output depends on whether KNN is used for classification or regression. In the classification point, k is a user defined constant. All the neighbors have equal Vote, and the maximum class voters choose among k neighbors. Ties are broken randomly or a weighted vote in the referendum [8]. Usually k takes Dice to reduce the number of odd. In KNN classification, the output is the property importance of the point. This point is mean of the points of its *k* nearest neighbors.

SVM (Support Vector Machine)

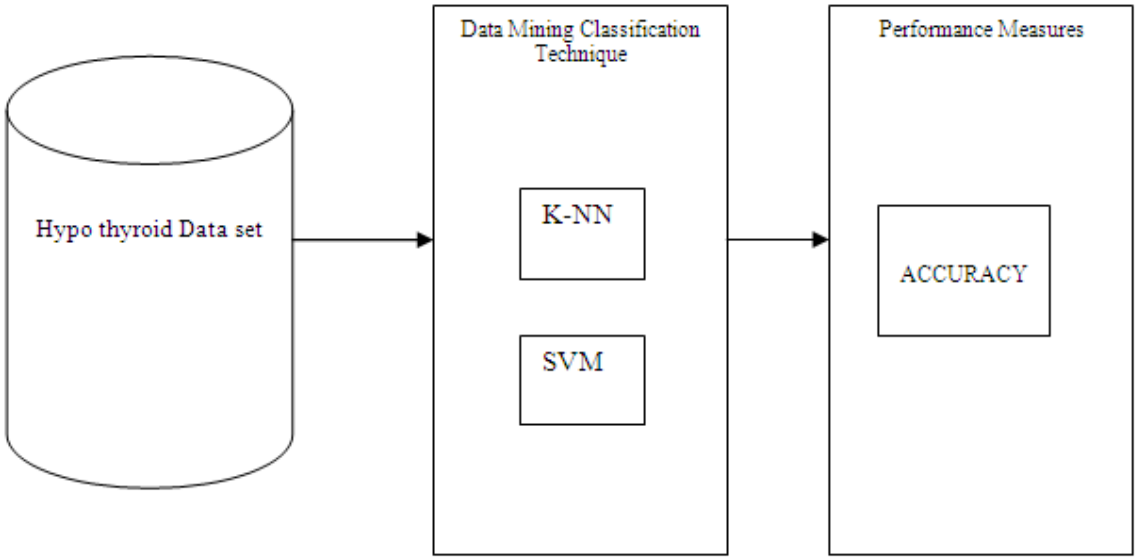
Support Vector Machine non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus

providing great classification or regression performance [9]. It is one of the most famous algorithms. It uses a technique called the kernel trick to transform our data. Then, based on these transformations; it finds an optimal boundary between the possible outputs. To put it in brief, SVM does some extremely complex data transformations and it separates our data, based on the labels or outputs that we have defined.

SYSTEM IMPLEMENTATION

The dataset has two categories. One is Hypothyroid, the next one is negative. The dataset is taken from theUCI repository. The hypothyroid dataset is loaded into Matlab software and got a predicted result. Data Mining in K- Nearest neighbor and Support Vector Machine are used in the dataset to predict the new results from the dataset.

Figure 1
Hypothyroid Prediction Process



RESULTS AND ANALYSIS

The dataset is containing 3090 data. 2941 is belongs to negative, 149 belongs to hypothyroid. 111 women are suffering from hypothyroid, 38 men are suffering from hypothyroid.

Table 2
Hypothyroid dataset based on Gender

Gender	*Hypothyroid*	*Negative*
Female	111	2071
Male	38	870

Figure 2 shows a diagram of hypothyroid precious people, which indicated by blue color, and the red color indicates negative results. The Experimental results shows, women are suffering by thyroid disease[10]. X axis represents sex and Y axis represents type.

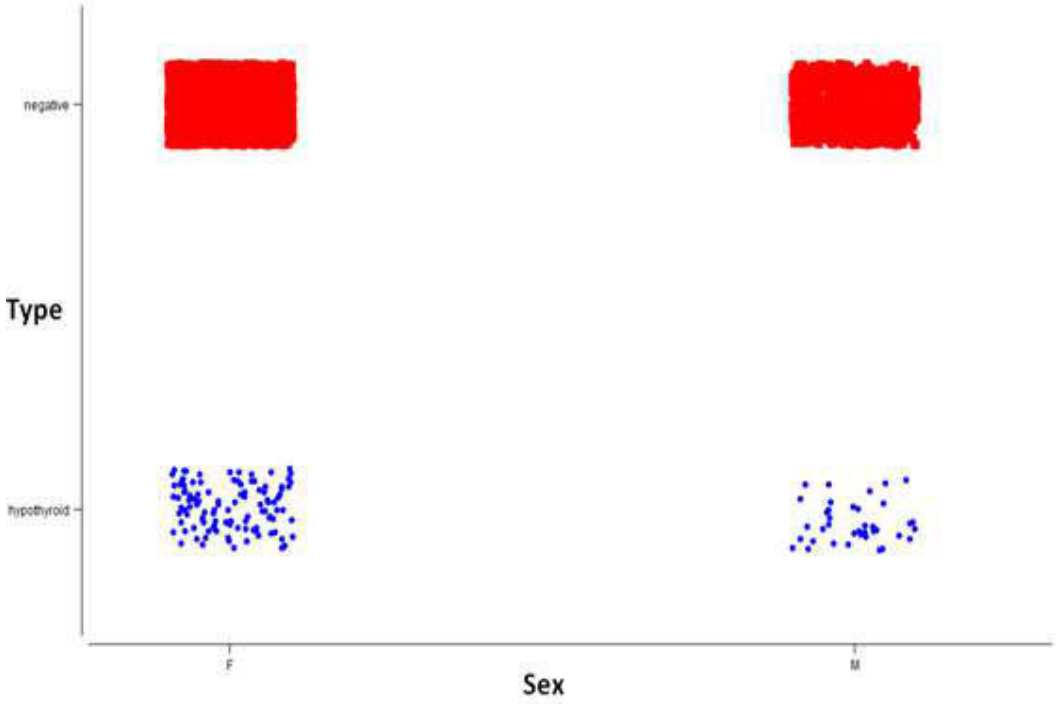


Figure 2
Hypothyroid data based on Gender

Where, x_i is the data points, where $i = 1,2,\ldots,n$. x_1 is number of accurate predictions that's occurrence is false, 'x_2' is the number of inaccurate predictions that occurrence is true, 'x_3' is the number of inaccurate of predictions that's occurrence is false, 'x_4' is the number of exact predictions that's occurrence is true.

Table 3
Confusion Matrix based Prediction Particulars entry for Hypothyroid

	PREDICTED	
	HYPOTHYROID	***NEGATIVE***
HYPOTHYROID	x_1	x_2
NEGATIVE	x_3	x_4

The table 3 shows the 2 – class matrix applied in the confusion matrix to extract the results

Table 4
SVM based hypothyroid Prediction

	PREDICTED	
	HYPOTHYROID	***NEGATIVE***
HYPOTHYROID	107	42
NEGATIVE	130	2811

The table 4 is used to recognize the prediction of hypothyroid by SVM.
Hypothyroid dataset is loaded into Support Vector Machine for Prediction Process. The dataset have 3090 data. 149 people were affected in hypothyroid disease. On x_1 cell hypothyroid categories is correctly predicted 107 data, On x_2 cell hypothyroid categories is incorrectly predicted 42 data. 2941 people are negative categories. On x_3 cell negative category is incorrectly predicted 130. On x_4 cell a hypothyroid category is correctly predicted 2811. Finally 2918 people categories are correctly predicted and 172 people are predicted incorrectly. When predicting positive hypothyroid categories the Support Vector Machine (SVM) gives better result than using K- Nearest Neighbor (K-NN).

Table 5
KNN based hypothyroid Prediction

	PREDICTED	
	HYPOTHYROID	***NEGATIVE***
HYPOTHYROID	58	91
NEGATIVE	22	2919

The table 5 is used to recognize the prediction of hypothyroid by KNN.
Hypothyroid dataset is loaded into K-Nearest Neighbor for Prediction Process. The dataset has 3090 data. 149 people are affected in hypothyroid disease. On x_1 cell hypothyroid categories is correctly predicted 58 data, On x_2 cell hypothyroid categories is wrongly predicted 91 data. 2941 people are negative categories. On x_3 cell negative category is wrongly predicted 22 data. On x_4 cell negative category is correctly predicted 2919. When predicting negative categories the K-Nearest Neighbor (KNN) gives better result than using Support Vector Machine (SVM). The Accuracy (AC) of the data points x_i is calculated using equation 1.

$$AC = \left(\sum_{i=1}^{n} x_i\right)^{-1} . (x_{1+}x_n) \ldots (1)$$

Accuracy (AC) is exact predictions of the entire number ratio. It is determined using Equation 1

Table 6
Hypothyroid Accuracy Details

Techniques	*Accuracy*
SVM	94.4336
KNN	96.3430

The Accuracy details in table 6 gives detail about prediction in K-NN and SVM.

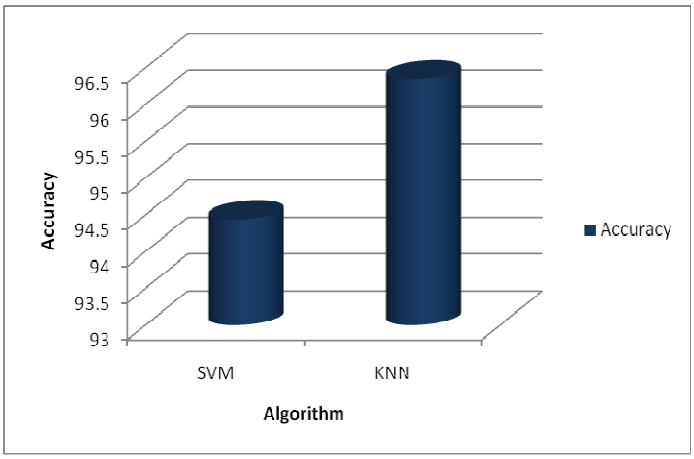


Figure 3
Hypothyroid data based on Accuracy

The figure 3 shows the comparison of Support Vector Machine and K – Nearest Neighbor accuracy. X axis represents algorithm, Y axis represents a prediction of hypothyroid data accuracy.

CONCLUSION

The thyroid gland is the one of the most important gland and the largest gland of the endocrine system. It consists of two connected lobes. It is found in the neck and below the Adam apple and shapes like a butterfly. The thyroid gland produces two thyroid hormone thyroxine (T3) and triiodothyonine (T4), these hormones are very helpful to control the body's metabolism. The trial of six parameters such as age, gender, TSH (Thyroid Stimulating Hormone), T3 (Triiodothyronine), T4 (Thyroxine), type (hypothyroid, negative) is used. SVM and KNN methods applied to the collected data to find hypothyroid. In SVM, the prediction accuracy is 94.4336. However, KNN accuracy is 96.3430. The difference / variance is 1.9094. Therefore, comparatively, KNN performs better than the SVM. These experimental techniques may be extended to other complaints such as breast cancer, heart attack, etc.

ACKNOWLEDGEMENT

The authors are thankful to Dr.Amit Kumar M.B.B.S, [DNB] General Surgery, Sri Ramakrishna Hospital, Coimbatore, Tamilnadu, India for helpful to carry out this work.

REFERENCES

1. D.Keran Hanirex, Dr.K.P..Kaliyamurthie, **"**Multi-Classification Approach for Detecting Thyroid Attacks", International Journal of Pharma and Bio Sciences (IJPBS,)**,** Volume 4 no 3, pp: 1246 -1251, July (2013).
2. Anurag Upadhayay, Suneet Shukla, Sudsanshu Kumar**,** "Empirical Comparison by data mining Classification algorithms (C 4.5 & C 5.0) for thyroid cancer data set, International Journal of Computer Science & Communication Networks, Vol: 3, No: 1, PP :64- 8, February – March – (2013).
3. M. R. NazariKousarrizi, F.Seiti, and M.Teshnehlab, "An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification," International Journal of Electrical & Computer Sciences IJECS, Vol: 12 No. 01, pp. 13-20, February (2012).
4. Satish N. Kulkarni and Dr. A. R. Karwankar, "Thyroid disease detection using modified fuzzy hyperline segment clustering neural network", International Journal of Computers & Technology, Vol: 3 No. 3,, pp. 466-469 , Nov-Dec, (2012).

5. Suneel.B,J.N.NaiduandAparna.R.R," Mineral Metabolism In Hyper Thyroidism",International Journal of Pharma and Bio Sciences (IJPBS Vol 3, Issue 2, pp. 602-606, April-June (2012).
6. Shivanee Pandey, Rohit Miri, "Diagnosis And Classification Of Hypothyroid Disease Using Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), Vol: 2, No: 6, PP: 3188 – 3193, June (2013).
7. J. Jacqulin Margret, B. Lakshmipathi, S.Aswani Kumar, " Diagnosis of Thyroid Disorders using Decision Tree Splitting Rules", International Journal of Computer Applications (IJCA), Vol: 44, No.8, pp 43-46, April (2012).
8. Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi, Mohammed K. Ali Shatnawi Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm", Vol: 3 No. 3, pp: 32 -43, March (2013).
9. S. Sathiya Keerthi, Olivier Chapelle, Dennis DeCoste "Building Support Vector Machines with Reduced Classifier Complexity" Journal of Machine Learning Research, Vol: 7, PP 1493– 515, January – (2006).
10. K.Saravana Kumar and Dr.R.Manicka Chezian "Analysis on Suspicious Thyroid Recognition Using Association Rule Mining"Journal of Global Research in Computer Science (JGRCS), Vol: 1, No-9, pp. 47-50, September (2013).