ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

Thyroid Prediction System using Auto Associative Neural Network

Komal Agrawal[1], Mradul Dhakar[2]

M. Tech Student, Dept. of Computer Science, ITM UNICERSITY, Gwalior, India[1]

Assistant Professor, Dept. of Computer Science, ITM UNIVERSITY, Gwalior, India[2]

ABSTRACT: Thyroid is across the board around in the world. The thyroid gland is an organ situated at the lower section of the human neck that produces the hormones that help regulate many body processes, including growth, energy balance, body temperature, and heart rate which is in indicts of delivering wide-ranging thyroid hormones. This work focused on developed of a predictive system for thyroid detection such as hypothyroidism, hyperthyroidism, sick people and normal people. In this work, we have applied auto associative neural network (AANN) for thyroid patient dataset. We have collected thyroid patient dataset from UCI Machine Learning Repository. Then overall prediction system accuracy for training state is found to be 100% and for testing over unknown testing data is found to be 95.1%.

KEYWORDS: Thyroid, Hyper Thyroid, Hypo Thyroid, Data Mining (DM), Data Pre-processing (DP), Auto-associative neural network (AANN).

I.INTRODUCTION

According to an analysis "DAILY TIME OF INDIA" while one in ten adults in India's people is suffering from hypothyroidism. This estimation is found on the premise of an analysis conduct by India thyroid society. The study also alert for thyroid and thyroid is 9th ranked in comparison to other type common disease like asthma, cholesterol, depression, diabetes etc. medical practitioner say that thyroid are same as other disorders, however, the investigation population are alert to thyroid disorders, know that there are diagnostic tests for finding of this disease[3]. Thyroid gland is broken into two section (1) Normal category of thyroid gland (2) this category the gland create abnormal type of thyroid hormones like as hypothyroid and hyperthyroid. Hypothyroidism (underactive thyroid or low thyroid) that is called the thyroid hormones are not generating as much as necessary of certain important hormones [22]. Hypothyroidism can justification various health problems such as: heaviness, joint pain, unfruitfulness and heart disease. Hyperthyroidism (overactive thyroid) belongs to a position is the thyroid gland delivers a lack of the hormone thyroxin. For this situation, the body's digestion system is quickening essentially, bringing about sudden weight reduction, a fast or irregular heartbeat. Predictive system is a group of accurate systems from machine learning, & information mining that consider present and reliable reality to make opportunity concerning future or generally difficult to understand occasions [22]. Disease determination is not easy task it requires lots of experience and knowledge. This is single of the traditional ways for diagnosis is doctor's examination various blood tests. The unique task is to give problem ending at before time stages with higher correctness. Data mining is a unique part in medicinal field for disease finding. Show predictive detection system of thyroid or illness finding systems important part of data mining.

II. OVERVIEW OF THYROID

THYROID:

Thyroid is worldwide. In India also, the thyroid gland is positioned at the lower section of the human neck that produce the hormones that help regulates body processes, including growth, energy balance, and heart rate, which is produce two kind of thyroid hormones, levothyroxine (T4) and triidothyronine(T3) [14], [7] that hormones involve some function of the body such as stabilizing body temperature, blood pressure, changeable the heart rate[3] Thyroid gland disease are different from other disease in conditions of their ease of diagnosis, accessibility, of health conduct, and the virtual visibility, In thyroid hormones is categorized into two common types Hyperthyroidism or hypothyroidism[4] .

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

1. HYPERTHYROIDISM: hormone is increase then this is generating reason of "hyperthyroidism". In healthcare field, "hyper" indicates too much. Hyperthyroidism collects when the gland produces overload hormones. This is called as an overactive thyroid; the hormone overload can cause an extensive range of physical changes. Many symptoms have similarities during Hypothyroidism as well as lessening hair, arid skin and high temperature feeling. The symptoms that show the occurrence of hyperthyroidism such as weight failure in bad feeling of a good food eating, an increment of heart rate, high blood pressure, anxiety, large sweating, addition in your neck and shaking hands[3]

1. Heart intolerance
2. Fatigue
3. Diarrhoea
4. Increased perspiration
5. Muscle weakness
6. Nervousness
7. Sleep disturbances
8. Goiter
9. Weight loss

Table I**:** List of symptoms or sign of Hyperthyroidism

2. HYPOTHYROIDISM: hormone is decrease then this is generating reason of hypothyroidism. In healthcare field, term hypo means short or not sufficient. For example, hypoglycaemia is a term for low blood sugar. Hypothyroidism is a form that the thyroids do not create required hormones. Swelling and break to the gland reasons hypothyroidism. Likes as: weight gain, to go down weight despite an accurate weight failure system. Table II shows the list of symptoms of hypothyroid. An increased risk of thyroid happens if there is a family unit history of thyroid like a type I diabetic, over 50 years of age and a stressful life [3].

1. Brittle nails
2. Cold hands and feet
3. Cold intolerance
4. Constipation
5. Depression
6. Difficulty swallowing
7. Dry skin
8. Elevated cholesterol
9. Essential hypertension
10. Hair loss
11. Poor memory
12. Slower heartbeat

Table II**:** List of symptoms or sign of Hypothyroidism

Both are hypothyroid and hyperthyroid be able to be diagnosed with thyroid function tests, which measures the levels of Thyroid-Stimulating Hormones (TSH) in bloodstream of human body.
In medical health care data mining (DM) is a set of large quantity of data involving with different technique, methods and algorithms which exist at the collection of fields includes artificial intelligence, machine learning and database

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

systems [24]. Hospitals, clinics and medical analysis laboratories save a huge quantity of patient data over the years. These data provide a basis for the analysis of danger features for many diseases (various types of disease like as cancer, heart diseases, diabetes, hepatitis etc [24].

DM is a key part in disease detection. AANN is able to be available for the disease finding, in classify to improve the quality of nature of analysis. The thyroid hormones dataset collected from UCI machine learning repository. These dataset 29 attributes for hyper thyroids, hypo thyroids, thyroid sick. Thyroid domains (records provided by the Garavan Institute of Sydney, Australia) in dataset includes several missing attribute values (signified by "?") in each data set 2800 instances [23].

III. METHODOLOGY

The diagram of the proposed system is shown in figure 1 which consists of three main parts:

i. Data Preprocessing.
ii. Strong Feature selection.
iii. Predictive model development

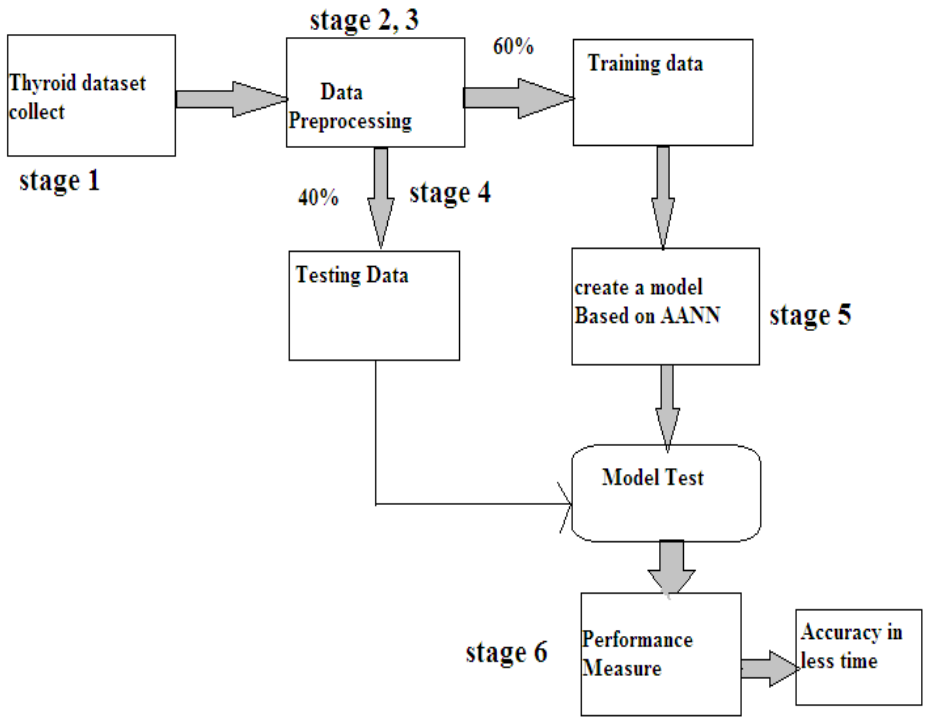

Figure 1: Block diagram of the proposed system

A. DATA PREPROCESSING:

Data pre-processing (DP) is an often neglected but it is main process in the DM. The phrase is “Garbage in, Garbage Out”, It is mostly applicable to information mining and machine learning, impossible data combinations (e.g. Gender is male or female, fever is yes or no) to the help recover the feature of the data, as to recover the efficiency and easy to the mining process.

DP includes, cleaning, normalization, transformation, feature extraction and selection etc.

Raw data has lower signal to noise ratio, missing values, and inconsistency that affects the DM results, In classify to help improve the feature of the data, as to improve the efficiency and easy to the mining process. DM pre-processing has been following categories:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

- Data Cleaning
- Data Integration
- Data Transformation
- Data Reduction

In the used dataset attributes are in redundant format, to decrease the redundancy and for making dataset convenient for processing, we have transform some attributes into digital form, which are given here:

1. M stands for male is convert into 1
2. F stands for female is convert into 0
3. t stands for true is convert into 1
4. f stands for false is convert into 0
5. Sick convert into 3
6. Hyper convert into 1
7. Hypo convert into 2
8. Normal convert into 0

B. FEATURE SELECTION / ATTRIBUTE REDUCTION:

Feature selection (FS) is also well-known as variable selection or attributes selection. It is the usable selection of attributes in dataset that are most important to the predictive modeling [42]

FS is the process to select a separation of important features for use in model construction. FS methods include reject attributes present in the dataset without changing them. In this work, we have applied Correlation Based feature selection (CFS) algorithm for attribute selection. CFS algorithm show highest valuation or ability of a separation of features. This heuristic takes useful of features for predicting the class label along with the level of inter correlation among them. The hypotheses on which the heuristic is based can be stated.

$$Rzc = \frac{K\overline{rca}}{\sqrt{k + k(k-1)\overline{raa}}}$$

Where,

Rzc = correlation between the summed attribute and the outside variable

$\overline{rca}$ = average attribute- class correlation

k = number of total attribute.

$\overline{raa}$= average attribute-attribute correlation

C. MODEL EVALUATION:

Model evaluation is a basic step of the prediction system process. It will help to get the best result that provides our data and how to work well the chosen prediction system will work in the future case. Model evolution has two type of evaluating method in data mining1) Cross- validation 2) Hold-out. In this work, Hold out method is using for model development. In this method, the mostly large amount dataset is unplanned separated into two parts.

Step-I: divide data into train set and test set

This step thyroid dataset is divided into two part 1) Training part 2) Testing part

In training part contain 60% thyroid dataset or in testing part contain 40 % data dataset shown in figure 2.

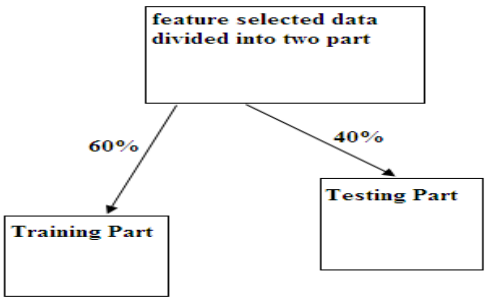

Figure 2: Data division into training and testing set.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

Step-II: Training data model building using Auto associative Neural Network.
This step create a Model Builder of AANN is Trained by using training dataset shown in figure 3

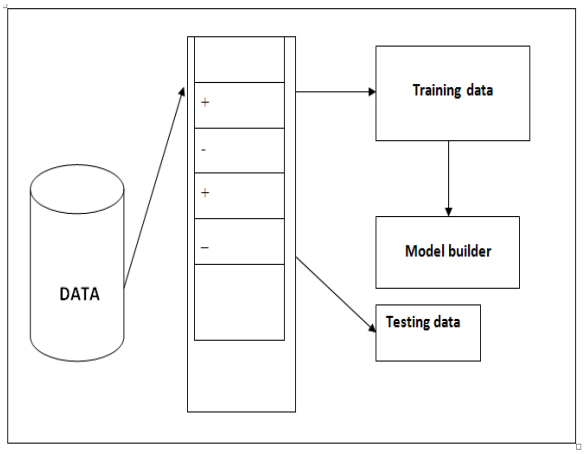

Figure 3: diagram of Training data model building process

Step-III: Evaluate an test set result
This step Trained AANN is tested by using Testing dataset shown in figure 4

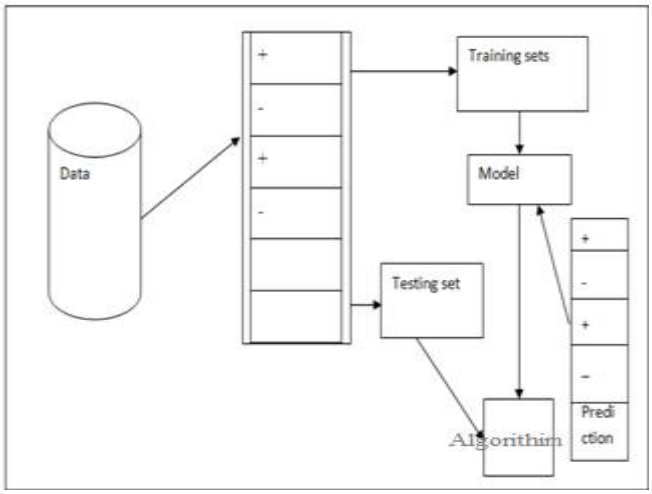

Figure 4: Block diagram of Evaluation and Testing Process

D. TRAINING OF THE NETWORK:

An AANN has a very straight forward architecture. It has an input layer and an output layer. The input layer is allied to the output layer passing through hidden layer with associated weights. In this process the input layer and the output layer is the similar that is, the same data set is used for input and output. The network is trained with as train dataset and associated weight is calculated by equation (ii) i.e. when the output consider by [O], the input consider by [I] or the weight consider by [w] then:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

$$\begin{bmatrix} w11 & w12 & w13 & \dots & w1m \\ w21 & w22 & w23 & \dots & w2m \\ \vdots & \vdots & \vdots & & \vdots \\ w31 & w32 & w33 & \dots & w3m \end{bmatrix} = \begin{bmatrix} I1 \\ I2 \\ I3 \\ \vdots \\ Ikn \end{bmatrix} * \begin{bmatrix} O1 \\ O2 \\ O3 \\ \vdots \\ Okm \end{bmatrix} \qquad \text{.....(i)}$$

Where the input consider by the train dataset as [I], the output consider by the same as train dataset as matrix [O] and the weigh [w] consider by calculated by using (ii). As the input and output is the equal matrix of the equal dataset. Then [O] = [I], so we can rewrite (iii) as:

$$\begin{bmatrix} w11 & w12 & w13 & \dots & w1m \\ w21 & w22 & w23 & \dots & w2m \\ \vdots & \vdots & \vdots & & \vdots \\ w31 & w32 & w33 & \dots & w3m \end{bmatrix} = \begin{bmatrix} I1 \\ I2 \\ I3 \\ \vdots \\ Ikn \end{bmatrix} * \begin{bmatrix} I1 \\ I2 \\ I3 \\ \vdots \\ Ikn \end{bmatrix}^{r} \qquad \text{.....(ii)}$$

E. TESTING THE NETWORK:

Case 1: Testing with same training dataset

For testing, first the same training dataset is given to the network as input. The training dataset is read as matrix. The weight matrix is similar as calculated during training the network time by equation (iii).The output [O] was calculated with i.e.

$$\begin{bmatrix} O1 \\ O2 \\ O3 \\ \vdots \\ Okm \end{bmatrix} = \begin{bmatrix} w11 & w12 & w13 & \dots & w1m \\ w21 & w22 & w23 & \dots & w2m \\ \vdots & \vdots & \vdots & & \vdots \\ w31 & w32 & w33 & \dots & w3m \end{bmatrix} * \begin{bmatrix} I1 \\ I2 \\ I3 \\ \vdots \\ Ikn \end{bmatrix}$$

Now the output matrix [O] is passed through activation function and then compared with input matrix [I]. As the input and output matrix is same the training dataset is declared to be "**same**".

Case 2: Testing with different/ testing dataset

A different dataset is given as input. The weight matrix is same as calculated during training the network by equation (iii).Now the output is calculated as in

$$\begin{bmatrix} O1 \\ O2 \\ O3 \\ \vdots \\ Okm \end{bmatrix} = \begin{bmatrix} w11 & w12 & w13 & \dots & w1m \\ w21 & w22 & w23 & \dots & w2m \\ \vdots & \vdots & \vdots & & \vdots \\ w31 & w32 & w33 & \dots & w3m \end{bmatrix} * \begin{bmatrix} I1 \\ I2 \\ I3 \\ \vdots \\ Ikn \end{bmatrix}$$

Now the output matrix [O] is passed through threshold function and the compared to the input matrix [I]. this time input matrix and output matrix is not "**same**" the different/testing dataset is confirmed to be" **not same**".

IV. RESULT AND DISCUSSION

In Original data, total number of attributes is 29. Upon these attributes, we have applied CFS algorithm and select 10 attribute. In implementation, from total thyroid dataset we have applied only 360 patient data. Further this dataset is divided into two parts by using holdout method. These two parts are: 1) training part 2) testing part. In training part, we have applied 216 patient data in which hyperthyroid training dataset contains 36 patient data, hypothyroid training data contains 60 patient data, normal training data contains 60 patient data, and sick training dataset contains 60 patient data.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

For testing part we have used 144 patient data in which hyperthyroid training dataset contains 24 patient data, hypothyroid training data contains 40 patient data, normal training data contains 40 patient data, and sick training dataset contains 40 patient data.

A. NETWORK FORMATION & ITS TRAINING:

We create an AANN using the command line functionality of Mat Lab toolbox, and then modify them to customize the network training. After that we have applied CFS algorithm to convert the originally provided thyroid data into ten most discriminant attributed data. After that we have applied hold out method to divide these pre-processed data into training and testing data. With the help of these training data we are going to create and train an AANN for different type of thyroid dataset.

	EPOCH	TIME	PERFORMANCE
Hyper	9 Epoch	9 Sec	7.11 *10^-25
Hypo	9 Epoch	29 Sec	1.21 * 10^-5
Normal	9 Epoch	27 Sec	4.88 *10^-6
Sick	18 Epoch	1 Min 8 Sec	3.08 * 10^-6

Table III: Performance Comparison of training states of different Thyroid data

In Table III consider about Trained AANN by using different training dataset like as hyperthyroid dataset, hypothyroid dataset, Normal Patient, Sick Patient. hyperthyroid dataset trained AANN window show 9 epoch in 9 sec with performance 7.11*10^25, hypothyroid dataset trained AANN window show 9 epoch in 27 sec with performance 1.38*10^25, Normal Patient dataset trained AANN window show 10 epoch in 30 sec with performance 2.16*10^-09, sick patient dataset trained AANN window show 9 epoch in 27 sec with performance 5.85*10^-25.

B. NETWORK TESTING RESULT

After, creation and training of the AANN we have applied the same training data to test upon all four trained network. After testing we have found that all instances of hyper thyroid, Hypo thyroid, sick patient and Normal Person data has been classifying into their respective target classes. And thus accuracy of this network is 100 % for all respective classes. The confusion matrix for overall training data is given in figure 6.

C. CLASSIFICATION RESULT OF TESTING DATA ON THE NETWORK:

After, creation and training of AANN, we have used this network to test upon all four class of testing data. After testing we have found that all instances of hyper thyroid has been classified into hyper thyroid class whereas some of the testing data of hypo, sick and normal classes are miss classified into wrong target classes. The confusion matrix for overall testing data is given in figure 7. And thus overall accuracy of this network is 95.1 % for all respective classes.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

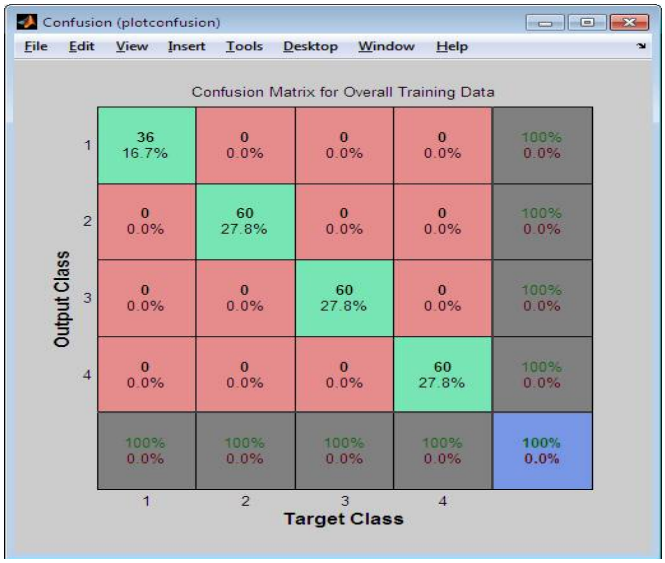

Figure 6: Confusion matrix for classification of all four classes.

Figure 7: Confusion matrix for classification of all four classes.

V. CONCLUSION

In thyroid have been two common type of the thyroid hormones and one are added sick otherwise patients is normal. Proper explanation of the thyroid disorder dataset, also clinical test and matching survey, is an essential problem in the diagnosis of thyroid. The thyroid prediction system will develop will reduce the attributes used in classifying thyroid The classifier models for thyroid analysis using AANN are successfully developed. The thyroid data set is a selected data set containing 360 dataset with 10 attributes and 4 classes. AANN could model a classifier for such a selected data

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 4, April 2017

set with good accuracy of higher than 95.1%. The results indicate that the models developed AANN have better accuracy and suitable for systems like thyroid..

REFERENCES

[1] JAMIL AHMED, M. ABDUL REHMAN SOOMRANI, ATTAULLAH SAHITO, SIRAJ AHMED ABBASI, "TDV: Intelligent System for Thyroid Disease Visualization", IEEE, PP: 1-7, 2016.
[2] IRINA IONITA, LIVIU IONITA, "Prediction of Thyroid Disease Using Data Mining Techniques", BRAIN-broad research in artificial intelligence and Neuron science, volume 7,Issue 3, PP: 115-124, 2016
[3] K.GEETHA & Capt. S. SANTHOSH BABOO, "An Empirical Model for Thyroid Disease Classification using Evolutionary Multivariate Bayesian Prediction Method", Global Journal of Computer Science and Technology: E Network, Web & Security Volume 16 Issue 1, PP: 1-9, 2016.
[4] EBRU TURANOGLU-BEKAR, GOZDE ULUTAGAY, SUZAN KANTARC-SAVAS, "Classification of Thyroid Disease by Using Data Mining Models: A Comparison of Decision Tree Algorithms", Oxford Journal of Intelligent Decision and Data Science, PP: 13-28, 2016.
[5] N MOHANA SURDARAM, VRENUPRIYA, "Artificial Neural Network Classifiers for Diagnosis of Thyroid Abnormalities", International Conference on Systems, Science, Control, Communication, Engineering and Technology, PP: 206-2011, 2016.
[6] VIKRAM V HEGDE, DEEPAMALA N, "Automated Prediction of Thyroid Disease using ANN", International Journal of Innovative Research in Science, Engineering and Technology ,Volume 5, Special Issue, PP: 268, 2016.
[7] SHAIK RAZIA and M. R. NARASINGA RAO, "Machine Learning Techniques for Thyroid Disease Diagnosis - A Review", Indian Journal of Science and Technology, Volume 9, PP: 1- 9, 2016.
[8] FEYZULLAH TEMURTAS, "A Comparative Study On Thyroid Disease Diagnosis Using Neural Networks", Expert Systems with Applications PP: 944–949, 2016.
[9] SERAVANA KUMAR, Dr. ADIMULAM YESUBABU, S ANUPKANT, "Autoimmune diseases - a wider perspective", International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 08, PP: 317-323, 2015.
[10] VADAMODULA PRASAD, THAMADA SRINIVASA RAO, ANKIT KUMAR SURANA, "Standard Cog Exploration on Medicinal Data", International Journal of Computer Applications, Volume 119 – No.10, PP: 34- 38, 2015.
[11] NILESH BORISAGAR, PROF.AMIT MARU, "ANN in Medical Field Survey Paper", International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 9, PP: 8210- 8216, 2015.
[12] PRERANA, PARVEEN SEHGA, KHUSHBOO TANEJA, "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network", International Journal of Research in Management, Science & Technology, Volume 3, No. 2, ISSN: 2321-3264, PP: 75:80, 2015
[13] I Md. DENDI MAYASANJAYA, HANUNG ADI NUGROHO, NOOR AKHMAD SETIAWAN, "a comparison of classification methods on diagnosis of thyroid disease", PP: 89-93, 2015.
[14] K.SARAVANA KUMAR and DR. R. MANICKA CHEZIAN, "Support Vector Machine and K- Nearest Neighbor Based Analysis for The Prediction Of Hypothyroid", International Journal of Parma and Bio Sciences, ISSN 0975-6299, PP: 447-453, 2014.
[15] K. RAMA LAKSHMI and S.PREM KUMAR," utilization of data mining techniques for prediction and diagnosis of major life threatening Diseases survivability", International Journal of Scientific & Engineering Research, Volume 4, Issue 6, ISSN 2229-5518, PP: 923-932, 2013.
[16] SHIVANEE PANDEY, ROHIT MIRI, S. R.TANDAN, "Diagnosis and Classification of Hypothyroid Disease Using Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), Volume 2 Issue 6, ISSN: 2278-0181, PP: 3188- 3193 , 2013.
[17] JASDEEP SINGH BHALLA, ANMOL AGGARWAL, "A Novel Method for Medical Disease Diagnosis Using Artificial Neural Networks Based on Back propagation Algorithm", IEEE, PP: 55-61, 2013.
[18] M. R. NAZARI KOUSARRIZI, 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-IJENS, Volume 12, PP: 13-19, 2012.
[19] K. SARAVANA KUMAR, Dr. R. MANICKA CHEZIAN, "Analysis on Suspicious Thyroid Reorganization Using Association Rule Mining", Journal of Global Research in Computer Science, 2012.
[20] FATEMEH SAITI , AFSANEH ALAVI NAINI, MAHDI ALIYARI SHOOREHDELI, MOHAMMAD TESHNEHLAB, "Thyroid Disease Diagnosis Based on Genetic Algorithms using PNN and SVM", IEEE, PP:1-4, 2009
[21] LALE OZYILMZ, TULAY ULDIRIM, "Diagnosis of Thyroid Disease Using Artificial Neural Network Methods" IEEE In the proceedings of the 9th International Conference on Neural Information Processing (ICONIP'OZ), Volume 4, pp: 2033-2036, 2002.
[22] https://en.wikipedia.org/wiki/Predictive_analytics.
[23] http://archive.ics.uci.edu/ml/support/thyroid+Disease.