IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-1 E-ISSN: 2347-2693

Intelligent Thyroid prediction system using Big data

K. Vijayalakshmi[1*], S. Dheeraj[2], B.S.S. Deepthi[3]

[1*]Department of IT, SNIST, Hyderabad, India
[2]Department of CSE, SNIST, Hyderabad, India
[3]Medico, D.No.5-7-200/8/A[3]Mamatha Medical College, Pakabanda , Khammam, India

**Corresponding Author: vldms@yahoo.com,*

Available online at: www.ijcseonline.org

Received: 06/Dec/2017, Revised: 14/Dec/2017, Accepted: 08/Jan/2018, Published: 31/Jan/2018

Abstract – Thyroid hormones delivered by the thyroid organ help control of the body's digestion. The thyroid, a butterfly-formed organ situated in the human neck and ace organ of digestion. At the point when thyroid doesn't work, it can influence each part of human wellbeing, particularly heaviness, causative or directing to gloominess and uneasiness, liveliness levels, and cardiac issues. Assortments of strategies have been suggested for thyroid illness. Healing of thyroid infection is simple, but the treatment taken by the greater part of the patients ceaselessly like blood pressure and diabetic patients. The principle goal is to build up a prototype intelligent thyroid Prediction System utilizing Big data and information mining displaying strategies. This framework can find and concentrate concealed information (examples and relationships) related to the thyroid ailment from a chronicled thyroid database. It can answer complex inquiries for diagnosing thyroid and consequently help medicinal services specialists to settle on wise clinical choices which conventional choice emotionally supportive networks. By giving compelling medicines, it likewise diminishes treatment costs. The social insurance industry gathers tremendous measures of enormous information which, shockingly, are not mined. Medicinal determination is viewed as an essential undertaking that should be executed precisely and capably. The computerization of this framework would be to a great degree worthwhile. Accordingly, a medicinal diagnosis system like the thyroid prediction framework would probably be exceedingly useful.

Keywords: Hormones, clinical, Hypo Thyroid, Treatment, patients, Risk Prediction.

I. INTRODUCTION

Thyroid leads different infections like corpulence, coronary illness, barrenness, and furthermore prompts the tumor in some cases. It making a move to get a legitimate finding and treatment is basic. Female patients confront a more serious danger of thyroid illness than men. Pregnant and new conceived youngster's likewise confronting the thyroid sickness. Framework of endocrine and specific organs of endocrine, with the thyroid, endures correspondingly to further organ frameworks and significant utilitarian changes on maturing. Various acceptable and anatomical changes of the thyroid amid the way toward maturing are notable [1, 2, 3].

The present audit concentrates on the most up to date discoveries concerning the changes in thyroid capacity amid the way toward maturing. Thyroid often cause indications in the neck zone where thyroid exist and it will swallow when the patient brought about by thyroid illness. The most widely recognized thyroid classes are Goiter (thyroid organ augmented). Thyroid organ delivered [5] a greater number of hormones than the human body obliged prompts Hyperthyroidism, when thyroid organ does not make enough thyroid hormones prompt Hypothyroidism, Thyroid growth, Thyroid knobs are protuberances in the thyroid organ and Thyroiditis is only swelling of the thyroid organ.

The common symptoms of thyroid are low pulse rate, irregular blood pressure, abnormal body temperature, swelling of hands and legs, hair loss, unexpected weight gain or loss, depression, mood changes regularly, metabolism, loss of memory power, dry skin, itching skin, sensitivity in eye-sight etc., In female patients [4] who suffering with thyroid their menstrual changes either irregular or heavy menstrual. In pregnancy, the female thyroid patients failed to assist reproduction, Miscarriage, depression and breast feeding. Females who are in the range of 40 to 50 facing, Menopause is starting typical or having difficult menopausal symptoms.

These horrible symptoms are leads to cardiac effects. Thyroid disease can make unpredictable pulse and harm the heart muscle. The most noticeably awful of these indications are

cardiovascular impacts. An overactive thyroid can bring about unpredictable pulse and even harm to the heart muscle..

II. RELATED WORK

Most of the people are not willing to spend time and money to know the prediction for thyroid disease. Banu et al [6] system explains about people to know the prediction for thyroid disease and also to know the prediction details and level of disease anywhere in the world. They used classification and clustering method to find the prediction details. Iodine acquires part a noteworthy role of the thyroid organ. It empowers thyroid hormones, and is essential for their production. Iodine existed in food and water. Where there is an iodine deficiency, iodine must be added to the salt or bread.

Thyroid [10] is classified into two types one is Hyper Thyroidism and another one is Hypothyroidism. Hyperthyroidism occurs when hormones produces more by the gland which leads a broad range of physical changes. When hormones produces by gland very less then hyper Thyroidism occurs. The symptoms of both are

Hypothyroidism symptoms are

1. Fatigue syndrome
2. feeling tired
3. inability to exercise
4. lethargy, weakness
5. feeling cold
6. Snoring
7. Abnormal menstruation
8. Delayed puberty
9. Gastrointestinal: water retention
10. Muscular: flaccid muscles
11. Mood: mood swings
12. Hair dryness and loss

Also patient will have fragile nails, gloominess, dry skin, puffy thyroid, cholesterol more, irritation feeling, memory loss, swollen eyes, cold, sleepless, heart rate become slow, sluggishness, swelling, or weight gain

Hyperthyroidism symptoms are

1. Fatigue or heat intolerance
2. Hyperactivity, irritability
3. Abnormal heart rhythm,
4. Insomnia
5. Irregular menstruation
6. Apathy
7. Abnormal protrusion

Also common: diarrhea, hair loss, muscle weakness, nervousness, premature ejaculation, the tremor, warm skin, or weight loss.
Other Types of Thyroid are

Goiter: swelling and coughing.

Thyroiditis: patient has no symptoms. When symptoms occur, they may change contingent upon the phase of the irritation.

Thyroid cancer: No symptoms only lump in the neck.

Thyroid diagnosis is identified by clinical tests leading to the decision [7], including thyroxine and tri-iodothyronine percentage of hormones and thyroid stimulating hormone (TSH) percentage. Identifying thyroid functional data is an important issue of the diagnosis of disease. Categorization is a data mining technique utilized the machine learning algorithm to predict thyroid patients for data examples [6, 8]. Rasitha Banu [9] proposed LDA data mining categorization techniques to classifying the Hypo Thyroid disease by K-fold cross validation.

III. FRAMEWORK

Big Data is a broad term for datasets, it is extensive in size and complex to perform conventional information handling applications are deficient. Significant challenges of Big Data incorporate analysis, a limit, information set, search, sharing, storage, exchange, visualization, data protection and security. The usage of predictive analysis or other certain advanced methods to retrieve valuable information from data, and often refer to a particular size of dataset. Accurate analysis result in big data more confident decision making.

The framework contains thyroid data set, updated thyroid data set and the Hadoop framework which has HBase and mahout.

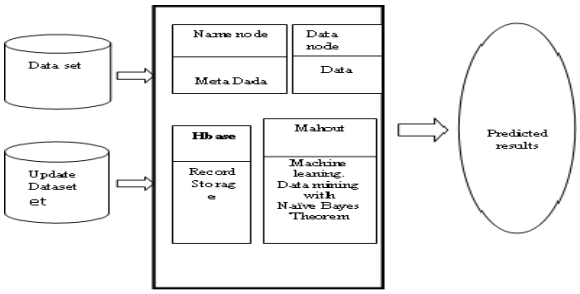

Fig 3. 1. Framework for the prediction system for hypothyroid disease

III. a. Data set:

For experimental purpose, we used Hypo Thyroid data set down loaded from the website (http://repository. seasr.org/Datasets/UCI/csv). The Hypo Thyroid data contains 3772 instances from which 3481 instances have negative,195 long-sufferings originate hypothyroid , 94 long-sufferings have primary-hypothyroid, while 2 long-sufferings have secondary-hypothyroid. The last characteristic is the class; thus there are 29 features taking all things together, which will be utilized to classify the data. The Hypo Thyroid data is representing in the following table.

TABLE 3.1
The attributes of Hypo-Thyroid Dataset including null Values.

S.No.	Field Name	Value of Field
1	Age	Continuous
2	Sex	Male(M), Female(F)
3	TSH measured	True (t), False(f)
4	TSH	Numeric value
5	T3 measured	True (t), False(f)
6	T3	Numeric value
7	Class	Stage of Thyroid disease

III b. Update Dataset:

After removing null values Dataset containing 2790 cases (out of 3772 cases). The data set after preprocessing is

Table 3.2
The attributes of Hypo-Thyroid Dataset without null value

S.No.	Field Name	Value of Field
1	Age	Continuous
2	Sex	Male(M), Female(F)
3	TSH measured	True (t), False (f)
4	TSH	Numeric value
5	T3 measured	True (t), False (f)
6	T3	Numeric value
7	Class	Stage of Thyroid disease

III c. Hadoop Framework

It is reliable and scalable data storage system. Hadoop is an open source, free and Java based programming system offers a capable appropriated stage to the store and oversee Big Data. There are two noteworthy parts of the Hadoop system. 1. MapReduce: to part a bigger information issue into littler lump and circulate it to various ware servers. 2. Hadoop Distributed File System (HDFS) is a virtual record framework and it contains the name node, secondary node, data node, job tracker and task tracker.

HBase: HBase useful for read or write access to Big Data. The primary point of HBase is to have immense tables with billions of lines and a huge number of segments.

Mahout: provides three major features 1.A simple and able to be made longer or more complex programming condition and structure for building adaptable calculations 2. Give a wide range of premade algorithms for Scala Apache Spark 3. Samsara gives a vector math test environment with R linguistic structure which works at scale.

Predicted result: Naive Bayes hypothesis- the nodes used to construct the classification model, Naive Bayes Predictor - the nodes used to consent the model, Scorer – the node reports a confusion matrix and the going with quality measures in its view.

IV. RESULTS

We tested the dataset with the Naïve Bayes Classification method which derives the mean Standard deviation, the sum of weights and precision of sample hypo thyroid dataset.

In order to apply our classifier and assess its execution, we apply the 10-fold cross validation test, which is a system that parts the original set into a training test to prepare the model, and a test set to assess it. Subsequent to applying the pre-processing and planning strategies, we attempt to break down the information visually and make sense of the dissemination of qualities as far as execution and precision of the model.

Naive Bayes: This arrangement depends on Bayes hypothesis. This categorization calculation utilizes restrictive autonomy, implies it except that property estimation on a given class is free on the estimation of different characteristics.

The Bayes theorem is as follows:
Let,
A={a1, a2, ….....an} be an arrangement of n attributes.

In this hypothesis, A is regarded as proof and H is some theorem mean assessment, the information of A is a subset of class C.

We need to assess P (H|A), the probability that Theory H holds given proof i.e. information sample A.
As per Bayesian hypothesis P (H|A) is given as,

$$P(H|A) = P(A|H)\,P(H) / P(A)$$

Table 4.1
Assessment dimensions for Naive Bayes categorization useful on Thyroid training set for all samples

Attribute (A)	CH (0.05)	Negative (0.92)	PH (0.03)	SH (0)
Age				
Mean	52.324	51.8081	50.3348	41.9457
Std. dev.	19.529	20.3449	18.7207	2.4674
Weight sum	6	3480	95	2
Precision	194	4.9348	4.9348	4.9348
	4.9348			
Sex				
F	146.0	2266.0	70.0	2.0
M	43.0	1078.0	23.0	2.0
[Total]	189.0	3344.0	93.0	4.0
TSH measured				
T	195.0	3113.0	96.0	3.0
F	1.0	370.0	1.0	1.0
[Total]	196.0	3483.0	97.0	4.0
TSH				
Mean	13.7361	1.8496	91.7202	1.8531
Std. dev.	13.1891	4.7925	111.5062	1.8531
Weight sum	194	3112	95	2
Precision	1.8531	1.8531	1.8531	1.8531
T3 measured				
F	38.0	724.0	10.0	1.0
T	158.0	2759.0	87.0	3.0
[Total]	196.0	3483.0	97.0	4.0
T3				
Mean	1.757	2.0632	0.967	1.8618
Std. dev.	0.6581	0.8149	0.6615	0.6206
Weight sum	157	2758	86	2
Precision	0.1551	0.1551	0.1551	0.1551

CCC	3553	94.1941 %
ICC	219	5.8059 %
KS		0.4881
MAE		0.0355
RMSE		0.1554
RAE		48.7715 %
RRSE		81.6037 %
TNC	3772	

Where CCC (appropriately Classified cases), ICC (wrongly Classified cases), KS (Kappa statistic), MAE(Mean absolute error), RMSE(Root mean squared error), RAE(Relative absolute error), RRSE(Root relative squared error), TNC(the total Number of cases)

TP Rate	FP Rate	Precis	Recall	F-Measu	MCC	F-Measu	PRC Area	Class
0.211	0.013	0.461	0.211	0.290	0.288	0.938	0.480	CH
0.993	0.553	0.955	0.993	0.974	0.591	0.967	0.995	-ve
0.589	0.003	0.848	0.589	0.696	0.701	0.985	0.782	PH
0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	SH
Weigh Avg	0.942	0.511	0.927	0.942	0.931	0.578	0.965	0.963

Confusion Matrix for the training set

a	b	c	d
41	149	4	0
19	3456	6	0
29	10	56	0
0	2	0	0

<-- classified as
| a = compensated-hypothyroid
| b = negative
| c = primary-hypothyroid
| d = secondary-hypothyroid

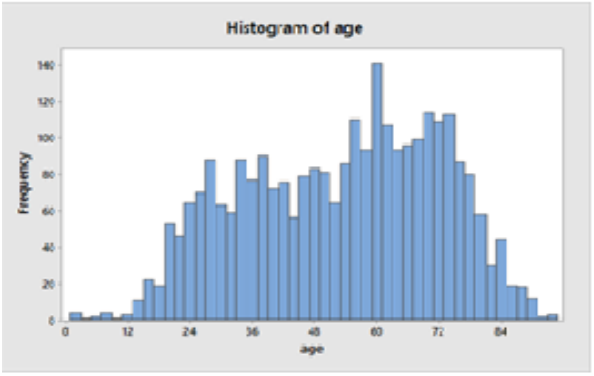

Fig 4.1. Histogram of age

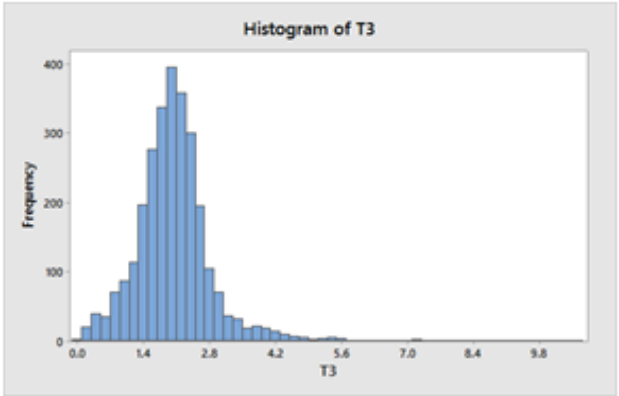

Fig 4.2. Histogram of T3

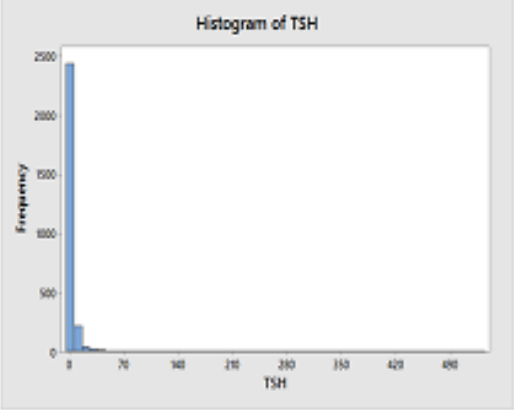

Fig 4.3. Histogmfor TSH

Data after removing the records of null values: 2790 (out of 3772). In our experiment analysis utilizing 10-fold cross validation for model training set

Table 4.2. Assessment dimensions for Naive Bayes categorization useful on Thyroid training set for samples after removing the null values

Attribute	Class CH (0.05)	negative (0.91)	PH (0.03)	SH (0)
Age				
Mean	53.268	52.6531	51.3554	43.9167
Std. dev.	17.5127	19.0697	16.2862	2.5833
Weight sum	151	2554	83	2
Precision	1.0333	1.0333	1.0333	1.0333
Sex				
M	38.0	872.0	23.0	2.0
F	115.0	1684.0	62.0	2.0
[Total]	153.0	2556.0	85.0	4.0
TSH measured				
T	152.0	2555.0	84.0	3.0
[Total]	152.0	2555.0	84.0	3.0
TSH				
Mean	13.3094	1.7949	89.8861	1.9343
Std. dev.	13.2392	4.8536	100.9455	1.9343
Weight sum	151	2554	83	2
Precision	1.9343	1.9343	1.9343	1.9343
T3 measured				
T	152.0	2555.0	84.0	3.0
[Total]	152.0	2555.0	84.0	3.0
T3				
Mean	1.7477	2.0554	0.9668	1.9182
Std. dev.	0.6721	0.8061	0.6635	0.6394
Weight sum	151	2554	83	2
Precision	0.1598	0.1598	0.1598	0.1598
TT4 measured				
T	151.0	2551.0	84.0	3.0
F	2.0	5.0	1.0	1.0
[Total]	153.0	2556.0	85.0	4.0

CCC	2616	93.7634 %
ICC	174	6.2366 %
KS		0.5044
MAE		0.0375
RMSE		0.1587
RAE		47.1461 %
RRSE		79.7852 %
TNC	2790	

TP Rate	FP Rate	Preci -sion	Re -call	F-Measu re	MCC	ROC Area	PRC Area	Class
0.219	0.016	0.446	0.219	0.293	0.286	0.929	0.463	CH
0.992	0.530	0.953	0.992	0.972	0.606	0.955	0.993	-ve
0.602	0.003	0.862	0.602	0.709	0.714	0.988	0.822	PH
0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.001	SH
Weigh -ted Avg	0.938	0.486	0.922	0.938	0.927	0.591	0.954	0.958

a	b	c	d
33	115	3	0
16	2533	5	0
25	8	50	0
0	2	0	0

<-- classified as
| a = compensated-hypothyroid
| b = negative
| c = primary-hypothyroid
| d = secondary-hypothyroid

V. CONCLUSION

Our work analyzed the Hypo Thyroid prediction system using big data framework. Our methodology gives big data for both predictive modeling and information retrieving with more efficiently. In this work we utilized a Naïve Byes hypothesis to predict patients with Hypo Thyroid. The classifier utilized demonstrated its execution in foreseeing with the best outcomes in provisions of accuracy and least execution time. We study the effectiveness and performance analysis of our proposed system with an experiment set, consisting of scalability and quality.

VI. FUTURE SCOPE

The future scope of our research aims at giving big data infrastructure by using different methodologies to calculate similarities between the patient's symptoms among hazard computation devices, for more advanced designing prediction models and element extraction methods and extending our proposed framework to predict other clinical risks with clinical suggestion for preventing thyroid cancer with other diagnosis parameters

VII. REFERENCES

[1]. Lewiński A, Sewerynek E, Karbownik M**:** Aging processes and the thyroid gland. In Aging and Age-Related Diseases: The Basics. Edited by: Karasek M. New York: Nova Science Publishers, Inc; 2006:131–172.Google Scholar

[2]. Faggiano A, Del Prete M, Marciello F, Marotta V, Ramundo V, Colao A: Thyroid diseases in elderly. Minerva Endocrinol 2011, **36:** 211–231.PubMedGoogle Scholar

[3]. Papaleontiou M, Haymart MR: Approach to and treatment of thyroid disorders in the Elderly**.** Med Clin North Am 2012, **96:** 297–310. 10.1016/j.mcna.2012.01.013

[4]. Bahn, R., Burch, H, Cooper, D, et al. Hyperthyroidism and Other Causes of Thyrotoxicosis: Management Guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. Endocrine Practice. Vol 17 No. 3 May/June 2011.

[5]. Braverman, L, Cooper D. Werner & Ingbar's the Thyroid, 10th Edition. WLL/Wolters Kluwer; 2012.

[6]. Dr. Rishitha Banu et. Al., "Predicting thyroid disease using data mining Technique" International Journal of Modern Trends in Engineering and Research on 11 October 2016, pg: 666-670.

[7]. Senthilkumar et al., " Classification of Multi-dimensional Thyroid Dataset Using Data Mining Techniques: Comparison Study" Advances in Natural and Applied Sciences, 9(6) Special 2015, Pages: 24-28.

[8] Rasitha Banu, Baviya "A study on Thyroid disease using Data Mining Technique", IJTRA Journal, aug 2015.

[9] Banu, et al "Predicting Thyroid Disease using Linear Discriminant Analysis (LDA) Data Mining Technique", Communications on Applied Electronics (CAE) –Volume 4– No12, January 2016.

[10] Ebru turanoglu-beka R et al., "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.

