MEROPOLITAN UNIVERSITY

ARTIFICIAL INTELLIGENCE

RESEARCH PAPER

Prediction of Thyroid Disease

Author:
MD MURSHED UL HOQUE
AFRA IBNAT KHAN
SYED EHSAN UL KARIM

Supervisor:
RAHATARA FERDOUSI
LECTURER,
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING.

November 14, 2019

Abstract

The use of artificial intelligence methods in medical diagnosis is increasing gradually. The effectiveness of classification and recognition systems has improved in a great deal to help medical experts in diagnosing diseases. Artificial Immune Systems (AIS) is a new but effective branch of artificial intelligence. This study aims at diagnosing thyroid disease with Artificial Immune Recognition System (AIRS). Thyroid disease diagnosis is an important classification problem. The thyroid data employed in this study is available from the UCI repository site. This data set is a very commonly used data set in the literature relating the use of classification systems for thyroid disease diagnosis and it was used in this study to compare the classification performance of AIRS with regard to other studies. We obtained a classification accuracy of 96.23%*, which is one of the highest accuracies reached so far. This result ensured that AIRS would be helpful in diagnosing thyroid function based on laboratory tests, and would open the way to various ill diagnoses support by using the recent clinical examination data, and we are actually in progress.*

Keywords: *Hormones, clinical, HypoThyroid, HyperThyroid, Treatment, patients, Risk Prediction , weka .*

1 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.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).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 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.An overactive thyroid can bring about unpredictable pulse and even harm to the heart muscle.Otherwise, misclassification may result in a healthy patientthat endures unnecessary treatment. Hence, the factuality of predictingany disease in conjunction with the thyroid disease is of supreme cardinality.

The thyroid gland is an endocrine gland in the neck.It erects in the lessened part of the human neck, beneath the Adam's apple which aids in the secretion of thyroid hormones and that basically influences the rate of metabolism and protein synthesis.To control the metabolism in the body, thyroid hormones are useful in many ways, counting how briskly the heart beats and how quickly the calories are burnt.The composition of thyroid hormones by the thyroid gland helps in the domination of the body's metabolism. The thyroid glands are composed of two active thyroid hormones, levothyroxine (abbreviated T4) and triiodothyronine (abbreviated T3).To regulate the temperature of the body these hormones are imperative in the fabrication and also in the comprehensive construction and supervision. Specifically,thyroxin (T4) and triiodothyronine (T3) are the two types of active hormones that are customarily composed by the thyroid glands. These hormones are decisive in proteinmanagement; dissemination in the body temperature, along with the energy-bearing and transmission in every part of the body.For these two thyroid hormones i.e. (T3 and T4), iodine is considered as

the main building chunk of the thyroid glands and are prostrated in a few specific problems, some of which are exceptionally prevalent. Insufficiency of thyroid hormones elements to hypothyroidism as well as an excessive thyroid hormones element to Hyperthyroidism. There are many origins related to hyperthyroidism and underactive thyroids. There are various kinds of medications.Thyroid surgery is liable to ionizing radiation, continual tenderness of the thyroid, deficiency of iodine and lack of enzyme to make thyroid hormones.

2 Related Work

Most of the people are not willing to spend time and money to know the prediction for thyroid disease.Rasitha Banu et al [5] 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.

The Process of aging affects both the prevalence and clinical presentation of hypo- and hyperthyroidism. Importantly, subclinical disturbances of thyroid function are more frequent than overt diseases in general population, as well as in elderly people . Consistently, the prevalence of subclinical hypothyroidism, which is characterized by normal free thyroxine (FT4) and elevated thyrotropin (TSH) levels, increases with aging and ranges from 3 to 16% in individuals aged 60 years and older [1].

Although it is known that overt thyroid disorders negatively affect physical and cognitive function in elderly people – for example, overt hypothyroidism is associated with the impairment of attention,concentration,memory, perceptual functions,language,and executive functions [2], subclinical hypothyroidism is not associated with impairment of physical and cognitive function or depression in individuals aged 65 years and older,as compared to euthyroidism.Also Park et al.[3]have demonstrated that subclinical hypothyroidism in elderly subjects is neither associated with cognitive impairment, depression, poor quality of life nor with metabolic disturbances.

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, sleeplessness, heart rate becomes 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.

There are many origins related to hyperthyroidism and underactive thyroids.There are various kinds of medications.Thyroid surgery is liable to ionizing radiation, continual tenderness of the thyroid,deficiency of iodine and lack of enzyme to make thyroid hormones [11].

Thyroid diagnosis is identified by clinical tests leading to the decision [7], including thyroxine and triiodothyronine 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 [5] proposed LDA data mining categorization techniques to classifying the Hypo Thyroid disease by K-fold cross validation.
The composite characteristics and the curative procedures that are being used in the thyroid disorders cater an ample clustering of intricate and assorted data and hence, a propitious framework for the formulation of machine learning models [7].

3 Experimental Analysis

3.1 Decision Tree

To evaluate classifier quality, we can use Decision Tree as a fast and simple technique where cross-validation is 10 and the percentage split is 80%.Consider this algorithm, for the data set:

Correctly Classified Instances 97.0822 %

Incorrectly Classified Instances 2.9178 %

Kappa statistic 0.7227

Mean absolute error 0.0449

Root mean squared error 0.1479

Relative absolute error 40.5756 %

Root relative squared error 65.1762 %

In the confusion matrix the total instances is 754 where A row contain 713 and B row contain 41.

	A	B	Total
A	701	12	713
B	10	31	41
Total			754

figure 1: Confusion matrix for Decision algorithm.

Let TPA be the number of true positives of class A, TPB be the number of true positives of class B.
Accuracy=(TPA+TPB)/Total =(701+31)/754 = 97.0822%

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.983	0.244	0.986	0.983	0.985	0.723	0.973	0.997	negative
	0.756	0.017	0.721	0.756	0.738	0.723	0.973	0.694	positive
W.Avg.	0.971	0.232	0.972	0.971	0.971	0.723	0.973	0.981	-

Table1 : Detailed accuracy by class

3.2 J48

To evaluate classifier quality, we can use J48 as a fast and simple technique where cross-validation is 10 and the percentage split is 80%.Consider this algorithm, for the data set:

Correctly Classified Instances 98.6737 %

Incorrectly Classified Instances 1.3263 %

Kappa statistic 0.8648

Mean absolute error 0.0155

Root mean squared error 0.1087

Relative absolute error 14.0107 %

Root relative squared error 47.914 %

In the confusion matrix the total instances is 754 where A row contain 713 and B row contain 41.

Confusion Matrix:

	A	B	Total
A	710	3	713
B	7	34	41
Total			754

Let TPA be the number of true positives of class A, TPB be the number of true positives of class B. Accuracy=(TPA+TPB)/Total =(710+34)/754 = 98.6737%

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.996	0.171	0.990	0.996	0.993	0.866	0.937	0.990	negative
	0.829	0.004	0.919	0.829	0.872	0.866	0.937	0.878	sick
W.Avg.	0.987	0.162	0.986	0.987	0.986	0.866	0.937	0.984	-

Table : Detailed accuracy by class

3.3 NaiveBayes

To evaluate classifier quality, we can use NaiveBayes as a fast and simple technique where cross-validation is 10 and the percentage split is 80%.Consider this algorithm, for the data set:

Correctly Classified Instances 91.7772 %

Incorrectly Classified Instances 8.2228 %

Kappa statistic 0.477

Mean absolute error 0.1037

Root mean squared error 0.2493

Relative absolute error 93.6374 %

Root relative squared error 109.8596 %

In the confusion matrix the total instances is 754 where A row contain 713 and B row contain 41.

Confusion Matrix:

	A	B	Total
A	659	54	713
B	8	33	41
Total			754

Let TPA be the number of true positives of class A, TPB be the number of true positives of class B. Accuracy=(TPA+TPB)/Total =(659+33)/754 = 91.7772%

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.924	0.195	0.988	0.924	0.955	0.518	0.906	0.983	negative
	0.805	0.076	0.379	0.805	0.516	0.518	0.906	0.596	sick
W. Avg.	0.918	0.189	0.955	0.918	0.931	0.518	0.906	0.962	-

Table : Detailed accuracy by class

3.4 Logistic

To evaluate classifier quality, we can use Logistic as a fast and simple technique where cross-validation is 10 and the percentage split is 80%.Consider this algorithm, for the data set:

Correctly Classified Instances 96.817 %

Incorrectly Classified Instances 3.183 %

Kappa statistic 0.6503

Mean absolute error 0.0487

Root mean squared error 0.1536

Relative absolute error 43.984 %

Root relative squared error 67.6997 %

In the confusion matrix the total instances is 754 where A row contain 713 and B row contain 41.

Confusion Matrix:

	A	B	Total
A	706	7	713
B	17	24	41
Total			754

Let TPA be the number of true positives of class A, TPB be the number of true positives of class B.
Accuracy=(TPA+TPB)/Total =(706+24)/754 = 96.817%

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.990	0.415	0.976	0.990	0.983	0.657	0.923	0.993	negative
	0.585	0.010	0.774	0.585	0.667	0.657	0.923	0.748	sick
W.Avg.	0.968	0.393	0.965	0.968	0.966	0.657	0.923	0.980	-

Table : Detailed accuracy by class

3.5 IBK

To evaluate classifier quality, we can use IBK as a fast and simple technique where cross-validation is 10 and the percentage split is 80%.Consider this algorithm, for the data set:

Correctly Classified Instances 96.817 %

Incorrectly Classified Instances 3.183 %

Kappa statistic 0.6503

Mean absolute error 0.0487

Root mean squared error 0.1536

Relative absolute error 43.984 %

Root relative squared error 67.6997 %

In the confusion matrix the total instances is 754 where A row contain 713 and B row contain 41.

Confusion Matrix:

	A	B	Total
A	706	7	713
B	17	24	41
Total			754

Let TPA be the number of true positives of class A, TPB be the number of true positives of class B.
Accuracy=(TPA+TPB)/Total =(706+24)/754 = 96.817%

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.990	0.415	0.976	0.990	0.983	0.657	0.923	0.993	negative
	0.585	0.010	0.774	0.585	0.667	0.657	0.923	0.748	sick
W.Avg.	0.968	0.393	0.965	0.968	0.966	0.657	0.923	0.980	-

Table : Detailed accuracy by class

4 Data

In this experiment, we used medical data for Thyroid. We have more than three thousand five hundred data. We classified these data into 25 Attributes. The main twenty attributes in our data sets are age, sex, on thyroxine, on antithyroid medication, pregnant, thyroid surgery, I131 treatment, query hypothyroid, query hyperthyroid, lithium, goiter, tumor, hypopituitarism, psych, TSH measured, T3 measured, TT ₋ measured, T4U measured, FTI measured, TBG measured. The TSH variable measured the TSH measured attribute. T3 variable measured the T3 measured attribute. TT4 variable measured the TT4 measured attribute. T4U the variable measured the T4U ₋ measured attribute. FTI variable measured the FTI measured attributes.

For this purpose, we have used WEKA (Waikato Environment for Knowledge Analysis). It is a collection of machine learning for data mining tasks, written in Java, developed at the University of Waikato, New Zealand. The Weka suite contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. For the analysis of our data set, we chose Logistic Regression, Decision Tree, Naive Bayes, J48 and IBK algorithm in WEKA.

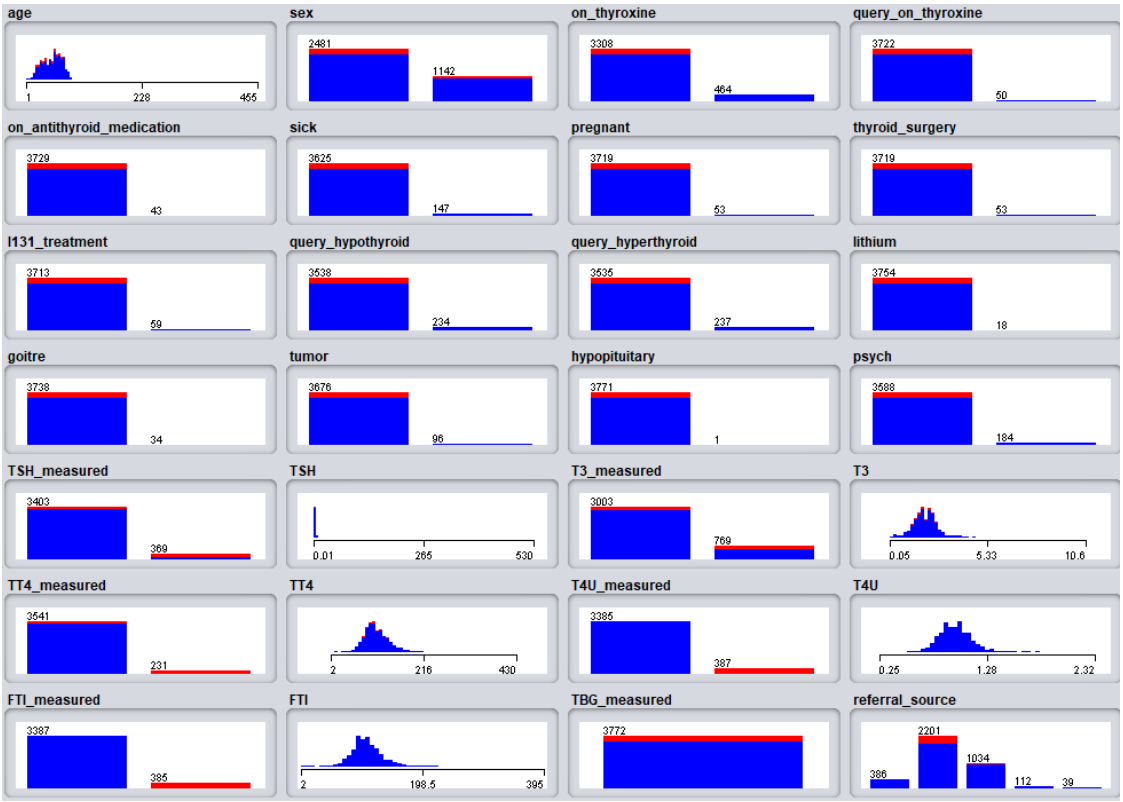

Figure 1: Visualizing all attribute for 3772 instances by using WEKA.

5 Discussion

In this section, we will show the comparison for the result of the different data mining algorithms which we have used for our data set and the comparison graph of different algorithms on the base of accuracy and error.

Algorithm	Accuracy	Error
Decision Tree	97.0822	2.9178
J48	98.6737	1.3263
NaiveBayes	91.7772	8.2228
Logistic	96.817	3.183
IBK	96.817	3.183

Table: Discussion

Here the J48 algorithm shows that the best result where accuracy is 98.6\% and error 1.4\% and the lowest accuracy is 91.7 \% and error is 8.3 \% for NaiveBayes but in the middle Logistic and IBK are equal.The average accuracy for all algorithm is 96.23\% and the error is 3.76\% .

6 Future Scope

In this research work, we have presented a new approach to detect Thyroid. For this purpose, we have collected medical data. We have classified our data set into some attributes. We applied five algorithms: Logistic Regression, Decision Tree, Naive Bayes,IBK and J48. Then we have got the best accuracy of 97.0822\% for the J48 algorithm algorithms. If we can predict thyroid earlier then we can reduce money in treatment, the patient will get treatment in the earlier stage and will suffer less.

In the future, we can use this data to predict cancer. Tree and Support Vector Machine algorithms. Here we have used three thousand and seven hundred instances for this research paper. In the future, a large data set will also be contemplated.

7 References

[1] Lewin´ski 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] Kramer C, von Muhlen D, Kritz-Silverstein D, Barrett-Connor E. Treated hypothyroidism, cognitive function, and depressed mood in old age: the Rancho Bernardo Study. Eur J Endocrinol. 2009;161:917–921.
doi: 10.1530/EJE-09-0606.

[3] Park YJ, Lee EJ, Lee YJ, Choi SH, Park JH, Lee SB, Lim S, Lee WW, Jang HC, Cho BY, Woo JI, Kim KW. Subclinical hypothyroidism (SCH) is not associated with metabolic derangement, cognitive impairment, depression or poor quality of life (QoL) in elderly subjects. Arch Gerontol Geriatr. 2010;50:e68–e73. doi:
10.1016/j.archger.2009.05.015.

[4] www.wikipedia.org.

[5] Dr .G .Rasitha Banu, Baviya " predicting Thyroid disease using Data Mining Technique ", IJMTER journal, March 2015.

[6] Pandey, Rohit Miri , Tandan "Diagnosis And Classification Of Hypothyroid Disease Using Data Mining", International Journal of Engineering Research and Technology (IJERT),Vol. 2 Issue 6, June – 2013

[7] L. Verma, S. Srivastava, P.C. Negi, A hybrid data mining model to predict coronary artery disease cases using noninvasive clinical data, J. Med. Syst. 40 (7) (2016) 1–7

[8] K. Saravana Kumar, Dr. R. Manicka Chezian "Support Vector Machine And K- Nearest Neighbor Based Analysis For The Prediction Of Hypothyroid", International Journal of Pharma and Bio Sciences.,0ct.2014.www.ijpbs.net

[9] Prerana, Parveen Sehgal, Khushboo Taneja "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network", International Journal of Research in Management, Science and Technology (EISSN: 2321- 3264) Vol. 3, No. 2, April 2015.

[10] J. Han and M. Kamber, Data Mining Concepts and Techniques, Elsevier, Australia, 2006

[11] G. Zhang, L.V. Berardi, "An investigation of neural networks in thyroid function diagnosis," Health Care Management Science, 1998, pp. 29-37.

[12] L. Ozyılmaz and T. Yıldırım, "Diagnosis of thyroid disease using artificial neural network methods," in: Proceedings of ICONIP'02 9th international conference on neural information processing (Singapore: Orchid Country Club, 2002) pp. 2033–2036

[13] K. Polat, S. Sahan and S. Gunes, "A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis," Expert Systems with Applications,vol. 32, 2007, pp. 1141-1147

[14] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM," in 3rd International Conference on Bioinformatics andBiomedicalEngineering, 2009. ICBBE 2009

[15] G. Zhang, L.V. Berardi, "An investigation of neural networks in thyroid function diagnosis," Health Care Management Science, 1998, pp. 29-37