Classification of Various Thyroid Ailments Using Data Mining Techniques

Namratha N [1] and Raghavendra S P[2]

IV Sem MCA, Department of Master of Computer Applications[1]

Assistant Professor, Department of Master of Computer Applications[2]

JNN College of Engineering, Shimoga, India

Abstract: *Data mining is a technique in which we are generating the required useful information by recognizing the pattern present in the large data. There are many techniques to do this like Clustering, Association, Machine Learning, Data Cleaning, Data Visualizing, Classification, Neural Networks, Prediction, and Data Warehousing in which we are using Classification technique to achieve the required result. In Classification we are classifying the predefined dataset. It is mainly used in the health care sector. In Classification technique we are using Logistic regression, K-Nearest Neighbors [KNN] ,Random forest, Naive Bayes , to classify the Various Thyroid Ailments.*

Keywords: Logistic regression [LR], Random forest [RF], Naïve Bayes[NB], K-Nearest Neighbor[KNN], Machine Learning

I. INTRODUCTION

Classification is one of the technique which detects and diagnosis various amount of diseases around the world wide, like skin disease, heart diseases, diabetics, thyroid diseases etc. Among these disease thyroid disease become one of the most common disease now a days so people need to aware about this in order to do so Indian government established Indian thyroid society they have introduce Making India Thyroid Aware[MITA] along with Abbott India, the main moto of the institute to generate the consciousness about the thyroid for the society in order to detect the early stage of thyroid disease.

Thyriod gland is basically a organ which located at the bottom position of the neck. This gland secrete the two important hormones one is Thyroxine(T4) and another one Triiodothyronine(T3).Basically the theses two hormones are helps to maintains the calories in our body and to provide the brain development as well as regulate the bone or muscles strength. If these two respective hormones do not produce in the proper manner that is either in less quantity or in high quantity then the thyroid will occur. If the body is secreting to much of thyroid hormone then this condition is called hyperthyroid and if the body is secreting little too much then it is called hypothyroidism. In this research this work the thyroid aliments are classified by taking the trained data set and by imposing these four algorithms Logistic regression, Random forest, Naïve bayes, K Nearest Neighbor.

II. LITERATURE SURVEY

By considering the literature survey Several works have been done on classifying the thyroid aliments some of them have been listed below:

Lerina Aversanoetal. [7]has used Decision Tree, KNN, Random Forest , extra Tree classifer, MLPC, XGB, Catboast Gradient Boosting classification techniques in order to get the required result. Chandan R et al.[2] has used the KNN, SVM, ANN, Decission Tree classification techniques in order to get the required result. Sumathi A et al.[9] has used Decision tree and J48 classification techniques in order to get the result. Ebru Turanoglu-Bekar et al.[4] has used Decision tree,J48,CART,NBtree,BF tree, LAD tree, REP tree, Random tree, Random forest, LMT ,FT, Decision stump classification techniques in order to get the result.

Shaik Razia et al.[8] has used decision tree classification technique in order to get result. Umar Sidiq et al.[10] has used KNN, Decision tree, Naive Bayes-Fold cross validation classification techniques in order to get the result. Dr. B. Srinivasan et al.[3] has used the Decision tree Naive Bayes, Backpropagation neural network, Support Vector machine classification techniques in order to get the result. Charan R et al.[1] has used the KNN ,Decision tree classification

techniques in order to get result. Haria Viral et al.[5] has used logistic regression, KNN, Naive Bayes, SVM, Decision tree classification techniques in order to get the result. Khalid Salman et al.[6] has used SVM, Random forest, Decision Tree algorithm, Naive Bayces,KNN,Multi layer preceptron, linear discrimiant analysis classification techniques in order to get the result.

III. PROPOSED METHODOLOGY

In this proposed methodology it is mainly focusing on classifying the various types of thyroid aliments it has been implemented using the following techniques Logistic regression[LR], Naive Bayes, K-Nearest Neighbor[KNN], Random forest.

3.1. Algorithm

Step1: Enter the value to attributes Age, Gender, Pregnant,TSH,T3,T4.

Step2: Check the validation for the entered fields

    if it is true then go for next step3

    otherwise go to step1

Step3: Select the appropriate classifier

- Naive Bayce
- Logistic regression
- Random forest
- K-NN

Step4: Predicting the result based on the selected classification module

    if the result is 1 then it is hyper thyroid

    if the result is 2 then it is hypo thyroid

    otherwise, the result will be non-thyroid

Step5: To predict the result using another classification technique

    then go to step 3

    else go to step 6

Step6: Display the accuracy of each classification models.
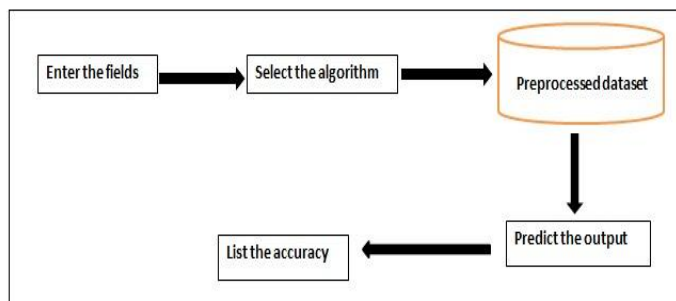
3.2 Block Diagram

Figure 1: The block diagram proposed methodology

As shown in Figure 1 first Enter the values for given attributes for the classification such as Age, gender, Pregnant,TSH,T3 and T4 .later on select the particular classification technique from the given four classification techniques that are Logistic Regression[LR], NavieBayce, KNN and Random forest, it will match values and process the value that are present in the preprocessed dataset and predict the output later on it will calculate the success for the given respective classification techniques.

IV. RESULT ANALYSIS

Authors	Classification Techniques	
	Classifier	Success rate
Chandan R et al.[2]	KNN	93.48%
	SVM	95.38%
	ANN	75.38%
	Decision Tree	92.3%
	Logistic Regression	96.92%
Ebru Turanoglu-Bekar et al.[4]	NB tree	75%
	LAD tree	66.25%
	J48	66.25%
	BFTree	65.25%
	LMT	65%
	Random forest	65%
	DS	41%
	FT	63.75%
	Random tree	62.50%
	REPTree	62.50%
	CART	58.75%
Khalid salman et al.[6]	Decision Tree	90.13%
	Support vector Machine	92.53%
	Random forest	91.2%
	Naive Regression	90.63
	Logistic Regression[LR]	91.73%
	Linear Discriminant Analysis	83.2%
	K-Neighbor Classifier	91.47%
	MLP	96.4%
Proposed work	Logistic Regression[LR]	96.93%
	Random forest	93.18%
	K-Nearest Neighbor	97.56%
	Naive Bayes	78.94%

Table 1: Analysis of experimental result

The Table 1 is the comparison of various classification techniques with the proposed system. On comparing with the above classification technique the proposed system will provide more success rate compare to that of existing system.
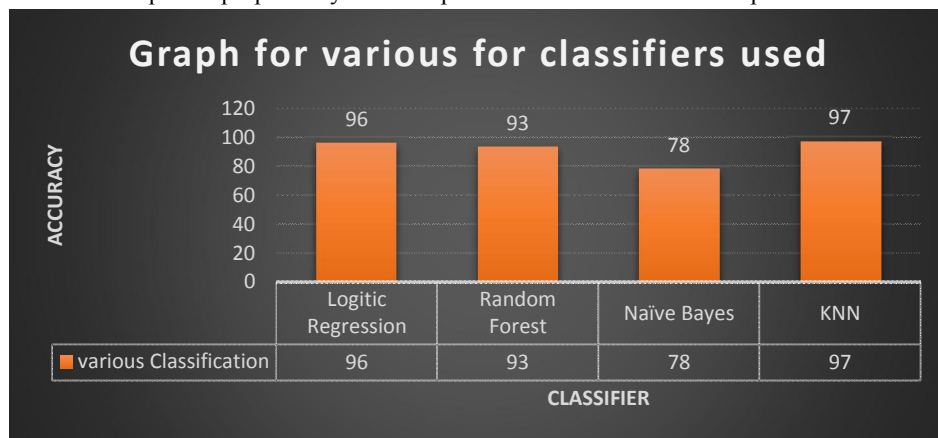

Figure 2: Graph for various for classification techniques

V. CONCLUSION

In India, thyroid disease is one of the common disease. In order to predict and classify the various types of thyroid disease this work has been done by considering the different data mining classification techniques. The classification techniques which are used for this works are Logistic Regression[LR], Random Forest, Naive Bayes and K-Nearest Neighbor classifiers. In this research work the Logistic regression works in the accuracy of 96.93%,the Random forest works in the accuracy of 93.18%,Navie Bayes works with the accuracy of 78.94% and K Nearest Neighbor works with the accuracy of 97.56%.Hence by considering all the accuracy rate Navie Bayes will produce less accuracy of 78.94% for this trained data set and K-Nearest Neighbor works with the highest accuracy of 97.56%.Thus by comparing these classification techniques K Nearest Neighbor is outdistance over other classification techniques. For feature enhancement if this technique combine it with other classification techniques then it would provide improved accuracy rate when compared with this accuracy rate.

ACKNOWLEDGMENT

REFERENCES

[1]. Charan R, Akash Yadav M, Aprameya N Katti, Mohith P, Prediction of Thyroid Disease Based on Classification Using Hierarchical Structure,JETIR,Volume 6,Issue 5,ISSN-2349-5162,2019

[2]. Chandan R Department of ECE DSU, Banglore,Karnatka,India,ChetanVasan department of ECE DSU,Banglore,Karnatka,India,Chethan MS and Devikarani H S THYROID DETECTION USING MACHINE LEARNING International Journal of Engineering Appiled Science and Technology,vol 5,Issue 9,ISSN No.2455-2143,Pages 173-177,January 2021 in IJEAST

[3]. Dr.B.Srinivasan ,K.Pavya, Diagnosis of Thyroid Disease Using Data Mining Techniques: A Study, International Research Journal of Engineering and Technology(IRJET) e-ISSN:2395-0056 p-ISSN:2395-0072

[4]. EbruTuranoglu-Bekar, GozdeUlutagay, Suzan Kantarcı-Savas , Classification of Thyroid Disease by Using Data Mining Models: A Comparison of Decision Tree Algorithms Oxford Jounalsof Intelligent Decission and Data Science 2016 No.2(2016)13-28 Volume 2016 Article ID ojids-0002,16 pages year 2016

[5]. Haria Viral More Suraksha Patel Bijal and Patil Harshali , Thyroid Prediction System using Machine Learning Techniques,Haria Viral al.,IJSRR 2018,7(4),67-681 ISSN:2279-0543.

[6]. https://r.search.yahoo.com/_ylt=Awr9Dt9Q1Kpi.sEA0wlXNyoA;_ylu=Y29sbwNncTEEcG9zAzQEdnRpZA NMT0NVSTAxOF8xBHNlYwNzcg/RV=2/RE=1655391440/RO=10/RU=https%3a%2f%2fwww.theweek.in %2fnews%2fhealth%2f2019%2f07%2f23%2fthyroid-disorders-rise-india.html/RK=2/RS=RWE93Fn1CrXXYmB9SpFd0AHYKNU-

[7]. Khalid salman, EmrullahSonuç, Thyroid Disease Classification Using Machine Learning Algorithms,2nd Intenational Conference On Physics and Applied Sciences(ICPAS 2021) IOP Publishing Journal of Physics:Conference Series 1963(2021)012140 doi:10.1088/1742-6596/1963/1012140.

[8]. LerinaAversanoa, Mario Luca Bernardia, Marta Cimitileb, Martina Iammarinoa, Paolo EmidioMacchiac, Immacolata Cristina Nettorec, Chiara Verdonea Thyroid Disease Treatment prediction with machine learning Lerina Aversano et.al/Procedia Computer Science 192(2021)1031-1040 .

[9]. Shaik Razia and M. R. NarasingaRao, Machine Learning Techniques for Thyroid Disease DiagnosisIndiaJournalofScienceandTechnology,vol9(28),DOI:10.17485/ijst/2016/v9i28/93705,July 2016,ISSN(Print):0974-6846,ISSN(online)0974-5645.

[10]. S P Raghavendra, Ajit Danti, "Recognition of Signature Using Neural Network and Euclidean Distance for Bank Cheque, K. C. Santosh and R. S. Hegadi (Eds.): RTIP2R 2018, CCIS 1037, pp. 228–243, 2019. https://doi.org/10.1007/978-981-13-9187-3_21, Springer Nature Singapore Pte Ltd. 2019

[11]. S P Raghavendra, M.J.Adarsh, Shoieb Ahamed and J. Shree Hari, "Estimation of Human Age and Gender Based on LBP Features Using Two Level Decision by SVM" _ K. C. Santosh and B. Gawali (Eds.): RTIP2R 2020, CCIS 1380, pp. 82– 94, 2021. https://doi.org/10.1007/978-981-16-0507-9_8, _Springer Nature

Singapore Pte Ltd. 2021.

[12]. Sumathi A, Nithya G and Meganathan S Classification of Thyroid Disease Using Data Mining Techniques International Journals of Pure and Applied Mathematics Volume 119 No.12 2018,13881-13890

[13]. Umar Sidiq1, Dr. Syed Mutahar Aaqib, Dr. Rafi Ahmad Khan, Diagnosis of Various Thyroid Ailments using Data Mining Classification Techniques,Intenational Journals of ScienceResearch in Computer Science, Engineering and Information Technology ,2019 IJRCSEIT Volume 5,Issue 1,ISSN-2456-330 DOI:http://doi.org/10.32628/CSEIT195119

BIOGRAPHY

[1]. PYTHON CRASH COURSE E02: A Hands-On, Project-Based Introduction to Programming by Eric Matthes

[2]. Introduction to Machine Learning with Python: A Guide for Data Scientists By Andreas C. Müller, Sarah Guido

[3]. Thyroid Diseases 1st Edition Edited By Fabrizio Monaco