ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse186942020.pdf
https://doi.org/10.30534/ijatcse/2020/186942020

Calculating Diagnose Odd Ratio for Thyroid Patients using Different Data Mining Classifiers and Ensemble Techniques

Dhyan Chandra Yadav[1], Saurabh Pal[2]
[1]Dept. of Computer Applications
VBSP University Jaunpur, India, dc9532105114@gmail.com
[2]Head, Dept. of Computer Applications
VBSP University Jaunpur, India, drsaurabhpal@yahoo.co.in

ABSTRACT
Thyroid hormones disorder is very common among people now-a-days. Various data mining techniques are used to identify thyroid problems. Machine learning provides help in multiple ways by different algorithms which are analyzed dataset and also generate different patterns. In this research paper we use three classifiers "Generalized Linear Model, Neural Network and Boosted Tree". Two ensemble techniques Stacking and Random Forest are used to combine the results of three data mining classifiers. Each classifier provides different patterns of measuring the performance with thyroid dataset. The ensemble model behaves as a major classifier in which Random Forest is give more accuracy. All the experiment performed on three features of thyroid dataset Triiodothyronine (T3), Thyroxin (T4) and Thyroid Stimulating Hormone (TSH) different values. In this paper, we have proposed a new technique to find the pattern using Diagnose Odd Ratio (DOR) to find the patients that needs further treatment or not. Diagnose Odd Ratio (DOR) provides a value if this value is high then patient needs treatment and if its value is low patient does not need further treatment. Finally thyroid dataset has been analyzed and the Positive Likelihood Ratio (LR+), Negative Likelihood Ratio (LR-) and Diagnostic Odd Ratio (DOR) with the help of Sensitivity and Specificity are measure.

Key words: Generalized Linear Model, Neural Network, Boosted Tree, Random Forest and Ensemble Methods.

1. INTRODUCTION

The diagnosis of thyroid disease fully depends on hormones. Generally doctors use medical history in diagnosis but it is not sufficient because without physical exam and medical hormonal test does not diagnose clearly. In any case if pituitary gland not properly works means T3, T4 and TSH functions are not properly work. Thyroid function inter relate with every function in human body. If human body function not properly work then some symptoms overcome in human body as like fatigue, weight gain, mood issue, irregular period, muscle pain cold hand, dry and cracking skin neck swelling etc. Generally thyroid problems are two types' hypothyroidism, and hyperthyroidism. These two different problems have different own symptoms in human body. By the help of Iodine easily maintain thyroid gland hormones because thyroid gland converts Iodine in thyroxin (T4) and triiodothyroxine (T3). Many thyroid cells are in the human body. These cells absorb Iodine and amino acid tyrosine for creation T3 and t4, t3 and T4 control metabolism of the body. T3, T4 control and manage oxygen and calories and create it into energy. So it is very clear energy cell directed by thyroid hormones and they make continuous flow of metabolism. The pituitary gland manages by another gland known as process TSH releasing hormone (TRH). It is very important in human that T3 and T4 levels must be always in balancing order. If T3 and T4 levels are low then it is required to improve the production of thyroid hormones so pituitary gland releases more TSH. If T3 and T4 levels are high then no need to improve high thyroid hormone production so pituitary gland release low TSH .In this research paper tries to find the combination of T3, T4 and TSH for good health [1].

2. RELATED WORK

Cohen, Jasonet al., discussed about urinary tract infection. They observed and measure a urine culture without antibiotics. They analyzed all the dataset and presented uncomplication of urinary tract infection by attributes represented symptoms uti and also presented highly therapy in women for urinary tract infection [2].

Alexander, Kennedy et al., discussed about thyroid nodules with indeterminate cytology. After all the experiment they find in term of machine learning sensitivity (0.9), specificity (0.52) and positive predictive value (0.37) and negative predictive value (0.94) and find the highest value of odd diagnosis ratio was (9.89) [3].

Borowczyk, Martyna et al., discussed about thyroid. They observed the attribute of thyroid cancer follicular for thyroid diagnosis. After all the observation they measured sensitivity, specificity, accuracy and also measured Gene expression [4].

Harrell, Bimston et al., discussed about thyroid cytology. After all the observation they find sensitivity (0.91), specificity (0.25) and positive predictive value (0.70) and negative predictive value (0.60) and finally find the highest log likelihood ratio values (3.36) [5].

Rossi, Martina et al., discussed about thyroid therapy and check the correlation of TSH. They managed all the subcategory under in observation in thyroid therapy. They presented more accurate result for the thyroid therapy [6].

Mclver, Castro et al., discussed about evaluation of cytological determinate thyroid nodules. After all the observation they find sensitivity and specificity (0.83) and (0.10) respectively. Finally they find out (0.54) log likelihood ratio value [7].

Megda, Bartosz et al., discussed about thyroid image reporting and data system. They observed and measured and measured machine learning function for likelihood positive ratio and likelihood negative ratio and finally they calculated diagnostic odd ratio for thyroid disease and also compared sensitivity and specificity in all iterations of observation[8].

Noureldine, Olson et al., discussed about surgical decision making process for patients with thyroid nodules. After all the experiment they find (0.99) and (0.09) values of sensitivity and specificity respectively. They find log likelihood ratio value (3.24) [9].

Gao, Lu-Ying et al., discussed about thyroid ultrasound. They predicted all the thyroid set attributes and prepared different types of classes which are represented as target variable [10].

Abeykoon, Mueller et al., discussed about thyroid nodules with indeterminate cytology. After all the observation they find highest sensitivity (1.0) and specificity (0.00) and log likely hood ratio (1.18)[11].

Dobruch, Katarzym et al., discussed about thyroid image reporting and data system. They observed and analyzed all the feature of thyroid cancer attributes by machine learning algorithms and find the highest accuracy and misclassified data from multi faces by ultrasounds [12].

Wu, Lam et al., discussed about classifier testing for indeterminate thyroid nodules surgery. They calculated (0.96) sensitivity and (0.28) specificity and (9.5) log likelihood ratio [13].Baser, Husniyeet al., discussed about thyroid image reporting and data system. They observed and analyzed all the prediction of malignancy and subcategory of all the thyroid dataset [14].

Shrestha, Evasovich et al., discussed about correlation between histological diagnoses. They observed all the experiment and find in machine learning sensitivity (0.86), specificity (0.68) and find all log likelihood ratio (2.59) [15].

Zhang, Lei et al., discussed about thyroid dataset. They analyzed STE features and predict the high sensitivity and specificity for new ultra sound diagnostic thyroid [16].

Valderrabano, Khazai et al., evaluated that thyroSeqv2 performance in thyroid nodules. They find result sensitivity (0.7) and specificity (0.77) machine learning algorithm. After all the observation they find (7.79) log likelihood ratio for thyroid patients [17].

Furlan, Antonio et al., investigated about arthritis attributes. They predict all dataset of arthritis dataset. In all the observations they find good sensitivity and specificity is not very high. They apply machine learning on arthritis patients in which they observed and evaluated by monographersreport [18].

Taye, Gurciullo et al., performed evaluation thyroid nodules. After all the performance they find (8.25) log likelihood ratio with highest sensitivity (0.91) and specificity (0.45)[19].

Liminwang, FangYuan et al., discussed about thyroid disease diagnosis. They used 12 UCI dataset, Naïve Bayes and K-dependence Bayesian classifier and compare 4 medical dataset and detect the advantage of accuracy [20].

Eagles-Smith, Collin A.,discussed about thyroid hormonces variations level. They taked blood sample for total mercury and thyroid hormonces. They measured in all the observation variation between the exposed group copared to the none exposed also compare the values of T3, T4 and TSH [21].

Dhyan Chandra and Saurabh Pal discussed about women thyroid prediction using data mining techniques. They used two ensemble techniques. The first ensemble technique generated by decision tree and second was generated by bagging and boosting techniques. They observed dataset for thyroid symptom and find better accuracy results [22].

Naomi,Gronich et al. discussed about hypothyroidism risk factor for diabetes patients. Thy evaluated results in two phase first high throughput to identify risk factor and in second phase observed time dependent Poisson regression multivariable models to access rate ratio [23].

Machine learning classifier measure and compare the entire hidden pattern by true positive, true negative, false positive, false negative, predictive positive value and negative predictive value from thyroid dataset. Predict all the dataset in three experimental setup and take the instances of T3, T4 and TSH. All the propose works are described as:

1- Take T3, T4, TSH attributes with many varies instances and compare maximum value of sensitivity and specificity during one experiment to another experiment for thyroid patients.
2- Select the values of T3, T4, TSH and compare the values of positive prediction values and negative predictive values and its measure its variation during change the instances of thyroid patients.
3- To select variation of instances and compare positive log likelihood values (LR+) with negative log likely hood (LR-) by many machine learning classifiers.
4- To diagnostic odd ratio (DOR) by sensitivity and specificity due to varies instances values.
5- Finally measure and compare all the desired result by of machine learning classifiers and generated new ensemble model to each other and from 2012-2018 review article/paper.

3. METHODOLOGY

Our research paper followed the some previous year research papers. Collect some research information from previous year and collect clinical attributes of databases from 2012 to 2018. These research paper included the research criteria on the basis of T3, t4 and TSH thyroid patients.

3.1. Data Description

Table.1 Variables representations for Thyroid Dataset

Source	https://github.com/mikeizbicki/datasets/blob/master/csv/uci/new-thyroid.names	
Sample Size	1699= Total: 728 =Hyperthyroidism, 637= Euthyroid State and 334= Hypothyroidism	
Dependent Variables		
Hyperthyroidism	Problem in maintain metabolism(<HP)	
Hypothyroidism	Problem in maintain metabolism (>HP)	
Euthyroidism	Balance(=HP)	
Independent Variables		
T3	(60-200)ng/dl	Triiodothyronine Stimulates the metabolism
T4	(4.5-12.0)µg/dl	Thyroxin produced by thyroid gland
TSH	(6.3-5.5)µl/ml	Thyroid Stimulating Hormone pituitary hormone

HP stands for hormones production. In this investigation we have used dataset from UCI machine learning repository. All the hormones (T3, T4 and TSH) measured in ng/dl, µg/dl and µl/ml. The hormone details mentioned in above table.1.

3.2. Algorithms Description

In this research paper, we examine diagnostic attributes by various machine learning algorithms:

3.2.1. Generalized Linear Model

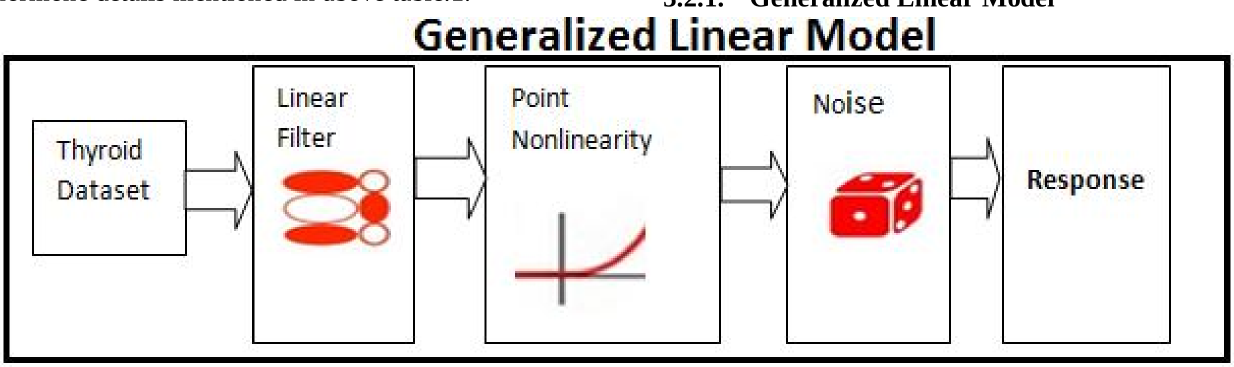

Figure 1: Representation of GLM for thyroid disease [https://images.app.goo.gl/vaLfXsA6S39wdDuV9].

Generalized linear model is used to observe the repeated T3, T4 and TSH in 2012 to 2019 in thyroid dataset. Generalized linear model describe dependent and independent variable for T3, T4 and TSH. It explain regular variable for most analysis of T3, T4, TSH in dataset.

3.2.2. Neural Network Model

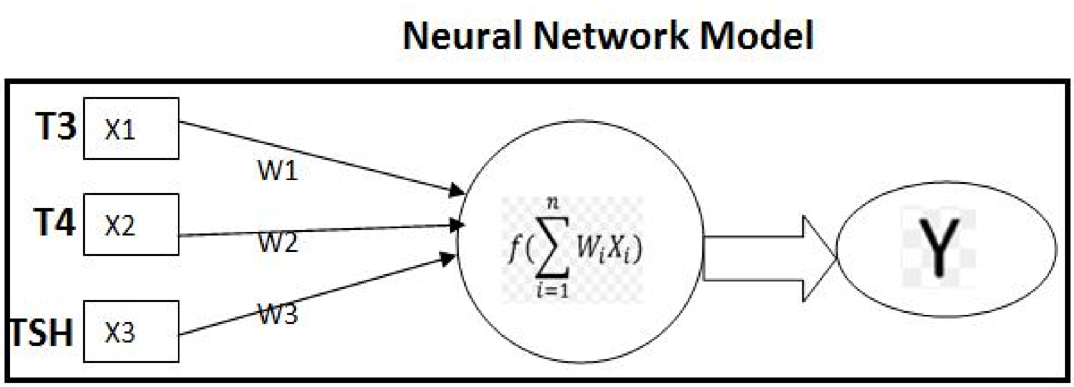

Figure 2. Neural Network Model for thyroid disease [https://images.app.goo.gl/R9wTrZvRokY3DZrNA].

Neural network is a knowledge discovery system for data process in which multiple small processes run in continuous and parallel way to solve the problem. All the small process takes idea by human brain. To take the help physicians in diagnose the lowest, highest and normal values for T3, T4 and TSH by use of neural network method in thyroid disease.

3.2.3. Boosted Tree Model

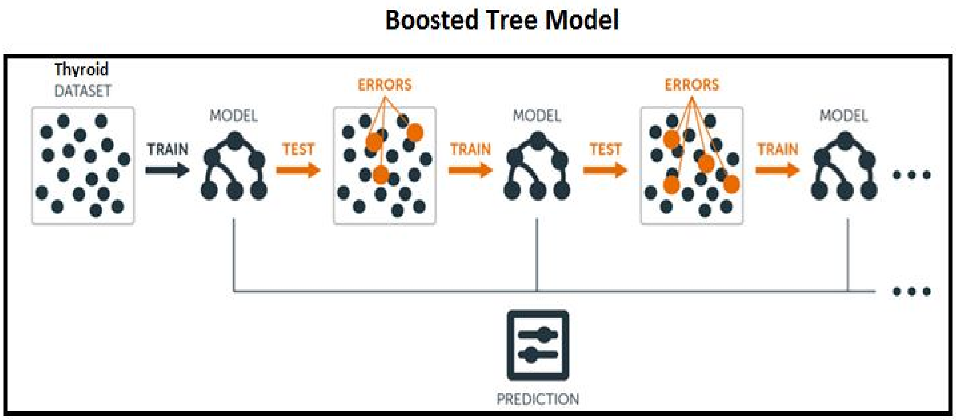

Figure 3: Representation of BT Model for thyroid disease [https://images.app.goo.gl/cKzyoV9w8BD46ucq5].

Boosted tree model generates an additional model in term of train and test for thyroid dataset by the help of this progressive model easily minimise residual error in the sequence of sum of all thyroid train and test previous model.

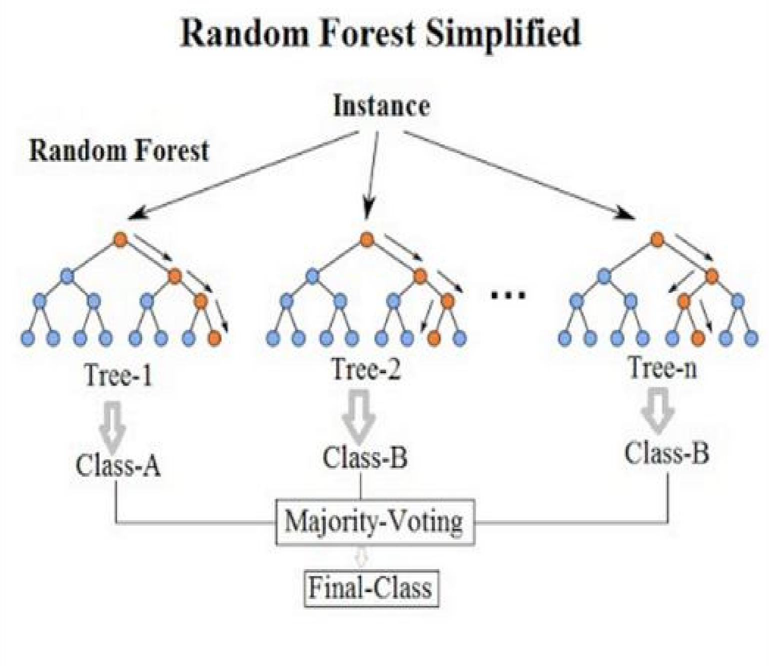

Figure 4: Representation of R F for thyroid disease [https://images.app.goo.gl/jxvQKFF4NC4Rce2u6].

3.3.3. Random Forest Model

This method use in thyroid dataset to train multiple decision trees. This method is used as a forest for multiple thyroid dataset trees. In this research paper used with ensemble model to provide more accurate estimation. This method provides predict more accurate ensemble with stacking model.

3.3.4. Ensemble Model

In this research paper create ensemble model by many trained models in which use different algorithms generalized linear model, neural network, random forest and boosted tree. All the algorithms find different types error in dataset according to his pattern. So Ensemble models more success because it is combined different algorithms. All the other algorithms create as a input set for ensemble model. Machine learning classifier measure and compare the entire hidden pattern by true positive, true negative, false positive, false negative, predictive positive value and negative predictive value from thyroid dataset. Finally analyzed and measure the positive likelihood ratio (LR+), negative likelihood ratio (LR-) and diagnostic odd ratio (DOR) by sensitivity and specificity.

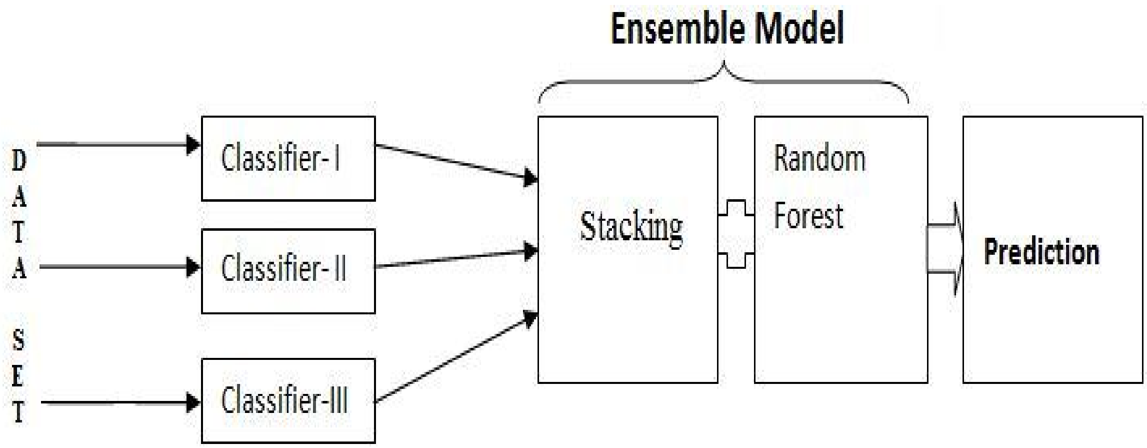

Figure 5: Representation of Ensemble Model for thyroid disease

4. RESULTS

Experiment: I

Table 2: Representation of computational model by classifiers for thyroid instances (499)

Algorithms	Sensitivity	Specificity	PPV	NPV	LR+	LR-	DOR
Generalized Linear Model	0.83	0.29	0.54	0.64	1.18	0.56	2.10
Neural Network	0.70	0.65	0.46	0.63	2.04	0.45	4.49
Boosted Tree	0.48	0.49	0.44	0.53	0.97	1.03	0.94
Ensemble Model	0.96	0.17	0.74	0.63	1.16	0.19	5.96

$$.DR(Positive) = [TP/(TP+FN)]/[(1-(TN/(TN+FP))] \ldots \ldots (1)$$

$$DR(Negative) = [1-(TP/(TP+FN))]/[TN/(TN+FP)] \ldots \ldots (2)$$

$$DOR = DR(Positive)/DR(Negative) \ldots \ldots (3)$$

$$SPE. = [(1-(TN/(TN+FP))] \ldots \ldots (4)$$

$$SEN. = [TP/(TP+FN)] \ldots \ldots (5)$$

The evaluation matrices represented by equations (1-5) [https://www.karger.com/]. In the equations (1-3) DR represents diagnostic ratio (positive/negative) while DOR stands from diagnostic odd ratio for more accurate combination of test of thyroids dataset and in the equations (4-5), SPE., and SEN., represents specificity and sensitivity respectively. The highest (0.96) measurement of the positive test as sensitivity and negative test probability for thyroid patients without disease. Find the highest the positive prediction value (0.74) and corresponding negative predictive values (0.63). These values represent healthy position test results and healthy negative position test result. Finally sensitivity and specificity combined and represents the disease and no disease thyroid patients as subjects. (LR+) test give the highest (1.16) information of diagnostic and (LR-) give the highest(0.19) information of negative test of non disease of thyroid. Finally observed the highest (5.96) information of diagnostic test by positive and negative likelihood ratio of thyroid patients.

Experiment: II

Table 3: Representation of computational model by classifiers for thyroid instances (1099)

Algorithms	Sensitivity	Specificity	PPV	NPV	LR+	LR-	DOR
Generalized Linear Model	0.70	0.37	0.51	0.57	1.12	0.79	1.42
Neural Network	0.88	0.11	0.42	0.54	0.99	1.00	0.99
Boosted Tree	0.78	0.37	0.59	0.60	1.26	0.57	2.20
Ensemble Model	0.63	0.84	0.51	0.58	4.17	0.43	9.64

The experiment-II observed that for more accurate combination of test of thyroid dataset. The (0.63) measurement the sensitivity of Ensemble model and DOR values (9.64). The highest sensitivity of Neural Network is (0.88) but DOR value (0.99). Find the highest the positive prediction value (0.59) of Boosted tree. These values represents healthy position test results and healthy negative position test result.

Experiment: III

Table 4: Representation of computational model by classifiers for thyroid instances (1699)

Algorithms	Sensitivity	Specificity	PPV	NPV	LR+	LR-	DOR
Generalized Linear Model	0.73	0.11	0.14	0.67	0.83	2.30	0.36
Neural Network	0.81	0.26	0.31	0.72	1.11	0.67	1.66
Boosted Tree	0.77	0.19	0.25	0.70	0.96	1.13	0.84
Ensemble Model	0.59	0.91	0.31	0.74	6.79	0.44	15.24

The experiment-III observed that for more accurate combination of test of thyroid dataset. The highest measurement of the positive test (0.81) of Neural Network and (0.59) of Ensemble model. Find the highest values of LR+ (6.795548) Ensemble model and (1.11) of Neural Network . LR+ test give the highest information of diagnostic and LR- give the highest information of negative test of non disease of thyroid. Finally observed the highest information of diagnostic test by positive and negative likelihood ratio of thyroid patients.

5. DISCUSSION

After all the experiments we compared and found that ensemble model give the better result compared to all other classifiers. From the table.5 and figure.6. It is clear that the diagnose odd ratio are always high compare to other classifiers. The positive predictive values are always high. In the variation of instances find the ensemble model always increase the values of diagnose odd ratio. Finally we compare with (2012-2018) previous research paper works and find better results by new Ensemble model. New Ensemble model give the better result compare to other classifiers "Generalized linear model, Neural network, Boosted Tree". The Random forest model with stacking model give better results in prediction. These both techniques work together as an Ensemble model [24-27].

Table 5: Representation of computational model for (2012-18) previous research papers

Study	Year	Sensitivity	Specificity	PPV	NPV	LR+	LR-	DOR
Alexander et al.	2012	0.9	0.52	0.37	0.94	1.88	0.19	9.89
Harrell et al.	2013	0.91	0.25	0.70	0.60	1.21	0.36	3.36
McIver et al.	2014	0.83	0.10	0.16	0.75	0.92	1.7	0.54
Noureld. et al.	2015	0.97	0.09	0.42	0.83	1.07	0.33	3.24
Abeykoon et al.	2016	1	0.00	0.86	0.00	1	1	1.18
Wu et al.	2016	0.96	0.28	0.62	0.86	1.33	0.14	9.5
Shreshta et al.	2016	0.86	0.68	0.63	0.88	2.69	1.04	2.59
Valderr. et al.	2017	0.7	0.77	0.42	0.91	3.04	0.39	7.79
Taye et al.	2018	0.91	0.45	0.27	0.96	1.65	0.2	8.25

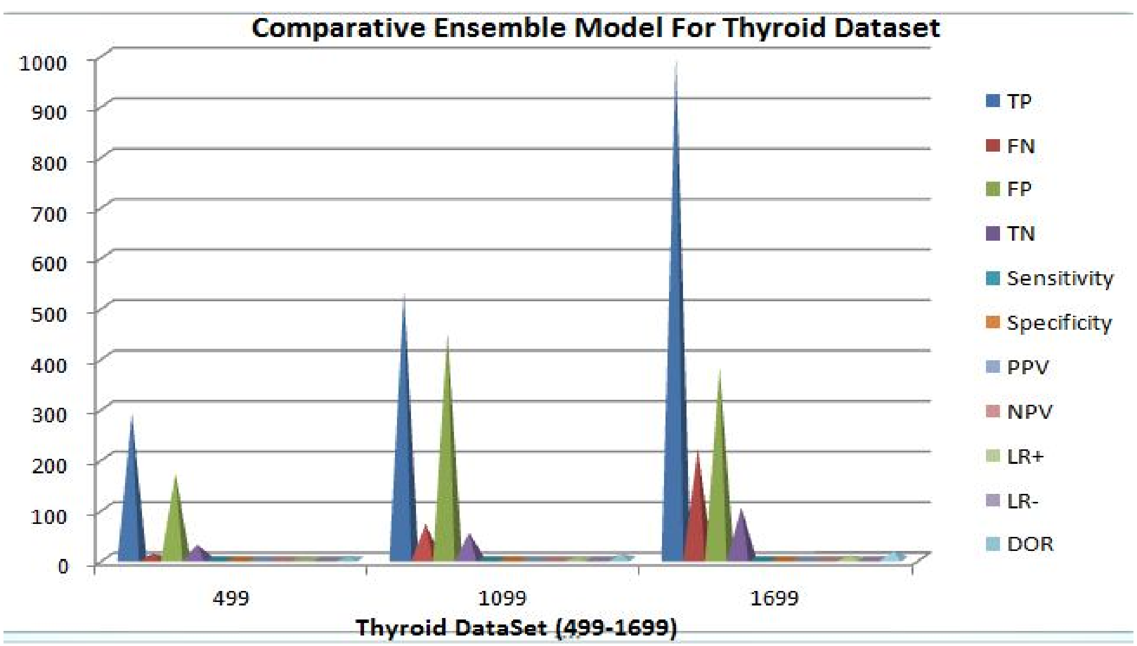

Figure 6: Comparative Ensemble Model for Thyroid Dataset

6. CONCLUSION

In three different experiments take three different size dataset for thyroid dataset as like 499, 1099 and 1699. By the help of machine learning techniques predicts TP, FP,TN, FN, PPV, NPV, Sensitivity, Specificity, LR+, LR- and diagnosis of odd ratio. In this paper use four different types' classifier algorithms GLM, NN, Boosted Tree, Random Forest and Ensemble Model for thyroid dataset. Finally find ensemble method performance is better in all condition compare to all other classifiers and the result evaluated .The performance of classifier methods are varies in increasing the instances in different DOR experiments. The performance of positive and negative prediction values is varies when number of instances varies. The performance of Random Forest with stacking model as an ensemble method is better in all different experiment. For the future work measure and summarize the performance RF and GLM measure with variant depth of AdaBoost and GLM with new features.

CONFLICT OF INTEREST

Authors have no conflict of Interest.

ACKNOWLEDGEMENT

The author is grateful to Veer Bahadur Singh Purvanchal University Jaunpur, Uttar Pradesh, for providing financial support to work as Post Doctoral Research Fellowship.

REFERENCES

1. http://www.endocrineweb.com, Oct 24. 2018.
2. Cohen JE, Yura EM, Chen L, Schaeffer AJ. **MP71-13 Prior Culture Data Can Significantly Enhance Prediction Of Negative Culture And Thus Eliminate Unnecessary Antimicrobial Therapy In Women With Possible Uncomplicated Urinary Tract Infections**. The Journal of Urology. 2019 Apr;201(Supplement 4):e1050-1.
3. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, Friedman L, Kloos RT, LiVolsi VA, Mandel SJ, Raab SS. **Preoperative diagnosis of benign thyroid nodules with indeterminate cytology.** New England Journal of Medicine. 2012 Aug 23;367(8):705-15. https://doi.org/10.1056/NEJMoa1203208
4. Borowczyk M, Szczepanek-Parulska E, Olejarz M, Więckowska B, Verburg FA, Dębicki S, Budny B, Janicka-Jedyńska M, Ziemnicka K, Ruchała M. **Evaluation of 167 Gene Expression Classifier (GEC) and ThyroSeq v2 Diagnostic Accuracy in the Preoperative Assessment of Indeterminate Thyroid Nodules: Bivariate/HROC Meta-analysis.** Endocrine pathology. 2018 Dec 27:1-8.
5. Harrell R, Bimston D. **Surgical utility of Afirma: effects of high cancer prevalence and oncocytic cell types in patients with indeterminate thyroid cytology.** Endocrine practice. 2013 Nov 18;20(4):364-9.
6. Rossi M, Lupo S, Rossi R, Franceschetti P, Trasforini G, Bruni S, Tagliati F, Buratto M, Lanza G, Damiani L, DegliUberti E. **Proposal for a novel management of indeterminate thyroid nodules on the basis of cytopathological subclasses.** Endocrine. 2017 Jul 1;57(1):98-107. https://doi.org/10.1007/s12020-016-1105-4
7. McIver B, Castro MR, Morris JC, Bernet V, Smallridge R, Henry M, Kosok L, Reddi H. **An independent study of a gene expression classifier (Afirma) in the evaluation of cytologically indeterminate thyroid nodules.** The Journal of Clinical Endocrinology & Metabolism. 2014 Nov 1;99(11):4069-77.
8. Migda B, Migda M, Migda MS, Slapa RZ. **Use of the Kwak Thyroid Image Reporting and Data System (K-TIRADS) in differential diagnosis of thyroid nodules: systematic review and meta-analysis.** European radiology. 2018 Jun 1;28(6):2380-8.
9. Noureldine SI, Olson MT, Agrawal N, Prescott JD, Zeiger MA, Tufano RP. **Effect of gene expression classifier molecular testing on the surgical decision-making process for patients with thyroid nodules.** JAMA Otolaryngology–Head & Neck Surgery. 2015 Dec 1;141(12):1082-8.
10. Gao LY, Wang Y, Jiang YX, Yang X, Liu RY, Xi XH, Zhu SL, Zhao RN, Lai XJ, Zhang XY, Zhang B. **Ultrasound is helpful to differentiate Bethesda class III thyroid nodules: A PRISMA-compliant systematic review and meta-analysis.** Medicine. 2017 Apr;96(16).
11. Abeykoon JP, Mueller L, Dong F, Chintakuntlawar AV, Paludo J, Mortada R. **The effect of implementing Gene Expression Classifier on outcomes of thyroid nodules with indeterminate cytology.** Hormones and Cancer. 2016 Aug 1;7(4):272-8.
12. Dobruch-Sobczak KS, Krauze A, Migda B, Mlosek K, Słapa RZ, Bakuła-Zalewska E, Adamczewski Z, Lewiński A, Jakubowski W, Dedecjus M. **Integration of Sonoelastography Into the TIRADS Lexicon Could Influence the Classification.** Frontiers in endocrinology. 2019;10.
13. Wu JX, Lam R, Levin M, Rao J, Sullivan PS, Yeh MW. **Effect of malignancy rates on cost-effectiveness of routine gene expression classifier testing for indeterminate thyroid nodules.** Surgery. 2016 Jan 1;159(1):118-29. https://doi.org/10.1016/j.surg.2015.05.035
14. Baser H, Cakir B, Topaloglu O, Alkan A, Polat SB, Dogan HT, Yazicioğlu MO, Aydin C, Ersoy R. **Diagnostic accuracy of Thyroid Imaging Reporting and Data System in the prediction of malignancy in nodules with atypia and follicular**

lesion of undetermined significance cytologies. Clinical endocrinology. 2017 Apr;86(4):584-90.

15. Shrestha RT, Evasovich MR, Amin K, Radulescu A, Sanghvi TS, Nelson AC, Shahi M, Burmeister LA. **Correlation between histological diagnosis and mutational panel testing of thyroid nodules: a two-year institutional experience.** Thyroid. 2016 Aug 1;26(8):1068-76.
16. Zhang L, Ding Z, Dong F, Wu H, Liang W, Tian H, Ye X, Luo H, Xu J. **Diagnostic performance of Multiple Sound Touch Elastography for differentiating benign and malignant thyroid nodules**. Frontiers in pharmacology. 2018;9:1359.
17. Valderrabano P, Khazai L, Leon ME, Thompson ZJ, Ma Z, Chung CH, Hallanger-Johnson JE, Otto KJ, Rogers KD, Centeno BA, McIver B. **Evaluation of ThyroSeq v2 performance in thyroid nodules with indeterminate cytology.** Endocrine-related cancer. 2017 Mar 1;24(3):127-36.
18. Furlan A, Stramare R. **The thickening of flexor tendons pulleys: a useful ultrasonographical sign in the diagnosis of psoriatic arthritis.** Journal of ultrasound. 2018 Dec 1;21(4):309-14.
19. Taye A, Gurciullo D, Miles BA, Gupta A, Owen RP, Inabnet III WB, Beyda JN, Marti JL. **Clinical performance of a next-generation sequencing assay (ThyroSeq v2) in the evaluation of indeterminate thyroid nodules.** Surgery. 2018 Jan 1;163(1):97-103.
20. Wang L, Cao F, Wang S, Sun M, Dong L. **Using k-dependence causal forest to mine the most significant dependency relationships among clinical variables for thyroid disease diagnosis.** PloS one. 2017 Aug 17;12(8):e0182070.
21. Eagles-Smith CA, Silbergeld EK, Basu N, Bustamante P, Diaz-Barriga F, Hopkins WA, Kidd KA, Nyland JF. **Modulators of mercury risk to wildlife and humans in the context of rapid global change.** Ambio. 2018 Mar 1;47(2):170-97. https://doi.org/10.1007/s13280-017-1011-x
22. Yadav DC, Pal S. **To Generate an Ensemble Model for Women Thyroid Prediction Using Data Mining Techniques.** Asian Pacific journal of cancer prevention: APJCP. 2019 Apr 29;20(4):1275-81.
23. Gronich N, Deftereos SN, Lavi I, Persidis AS, Abernethy DR, Rennert G. **Hypothyroidism is a risk factor for new-onset diabetes: a cohort study.** Diabetes Care. 2015 Sep 1;38(9):1657-64.
24. Yadav DC, Pal S. **Decision Tree Ensemble Techniques To Predict Thyroid Disease .** International Journal of Recent Technology and Engineering (IJRTE), Vol.8, Issue. 3, 2019.
25. Yadav DC, Pal S. **Prediction of thyroid disease using decision tree ensemble method.** Human-Intelligent Systems Integration. 2020 Apr 6:1-7.
26. Verma AK, Pal S, Tiwari BB. **Skin disease prediction using ensemble methods and a new hybrid feature selection technique.** Iran Journal of Computer Science. 2020 Apr 4:1-0.
27. Yadav DC, Pal S. **Thyroid prediction using ensemble data mining techniques.** International Journal of Information Technology. 2019 Nov 28:1-1. https://doi.org/10.1007/s41870-019-00395-7