COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING
e-ISSN: 2582-7464

Computational Intelligence and Machine Learning
Vol-4, Issue-2, October 2023

Machine Learning Fusion Algorithm using for Forecasting Thyroid Disease

K. P. Manikandan [1*], B. Anusha [2], S. Girija [3], K. Harshitha [4], M. Keerthana Royal [5]

[1] Assistant Professor, Department of CSE, Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India

[2, 3, 4, 5] Department of CSE, Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India

*Corresponding Author Email: manikandankp@mits.ac.in

Abstract

This paper proposes several feature selection and classification procedures for thyroid ailment diagnosis, which is one of the most critical classification issues. Two Thyroid disease refers to a set of disorders affecting the thyroid gland, which produces thyroid hormones. Hormones are in charge of controlling the pace of metabolism in the body. Hyperthyroidism and hypothyroidism are two types of thyroid diseases. They are classified. Thyroid disease is a challenging issue to resolve. The process of extracting or choosing a group of features is an important challenge in the field of pattern recognition. This is a step in the pre-processing process. As an example, consider the word sequence. The words "sequence backward selection" and "ahead selection" are used interchangeably. Two well-known heuristic approaches are utilized for feature extraction. selection. Genetics is a science.

In the health system, where there is a huge amount of data and information to manage, machine learning algorithms are essential for dealing with data. Our study on thyroid disease employed machine learning techniques. With the aim of classifying thyroid disease into three groups—hyperthyroidism, hypothyroidism, and normal—we conducted this study using data from Iraqi individuals, some of whom have hyperthyroidism and others who have hypothyroidism.

Keywords

AdaBoost, Decision Tree, Support Vector Machines and XgBoost.

INTRODUCTION

Thyroid illness, a branch of endocrinology, is one of the least understood and recognized diseases [1]. Diabetes is the most common endocrine illness, followed by thyroid issues, according to the World Health Organization [2]. Hypothyroidism and hyperfunctioning hyperthyroidism afflict 2% and 1% of the population, respectively [3]. Men are involved in around one-tenth of all instances. Thyroid dysfunction caused by pituitary gland failure or hypothalamic dysfunction can result in hyper- and hypothyroidism [4]. Goitres, or active thyroid nodules, can occur at a rate of up to 15% in areas where dietary iodine levels are low. Furthermore, a variety of malignancies can form in the thyroid gland [5], making it a potentially dangerous location [6]. Endocrine glands, particularly the thyroid, create and distribute hormones throughout the body [7]. It is placed in the front center of the body. Hormones produced by the thyroid gland govern physiological fluid balance, digestion, and other processes [8] [9]. T3 (triiodothyronine), T4 (thyroid hormone), and TSH (thyroid stimulating hormone) can all be used to treat the thyroid gland [10]. Thyroid problems are classified as hypothyroidism or hyperthyroidism [11]. Data mining is a semi-automated approach for discovering patterns in vast volumes of data [12]. Hyperthyroidism is characterized by an excess of thyroid hormones produced by the thyroid gland [13]. Hyperthyroidism is caused by an increase in thyroid hormone levels [14]. Trembling, dry skin, increased sensitivity to heat, thinning hair, weight loss, an elevated heart rate, high blood pressure, excessive perspiration, neck enlargement, anxiety [15], shorter menstrual cycles, irregular stomach motions, and an expanded neck are some of the symptoms. Hypothyroidism causes the thyroid gland to become underactive [16] [17]. Hypothyroidism is caused by a decrease in thyroid hormone production [18]. In medical words, hypo denotes inadequate or less. Inflammation and thyroid gland injury are the two most prevalent causes of hypothyroidism [19]. Obesity, low heart rate, increased sensitivity to heat, neck swelling, dry skin, numb hands, hair difficulties, heavy menstrual periods, and digestive disorders are also symptoms [20].

PROPOSED MODEL

Several machine learning methods for classifying thyroid problems have been developed, but none fully address the issue of misdiagnosis. Furthermore, comparable research has presented methods to analyses this illness categorization, but they typically overlook the data's magnitude and heterogeneity. As a result, we recommend Support Vector Machines, XGBoost, Decision Trees, and AdaBoost. Performing machine learning-based classifier testing.

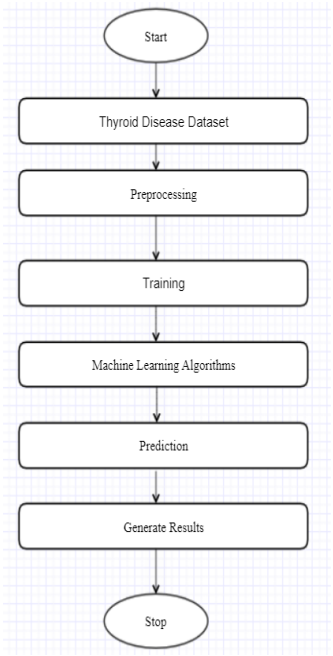

Figure 1. Block diagram of Thyroid Disease

Architecture

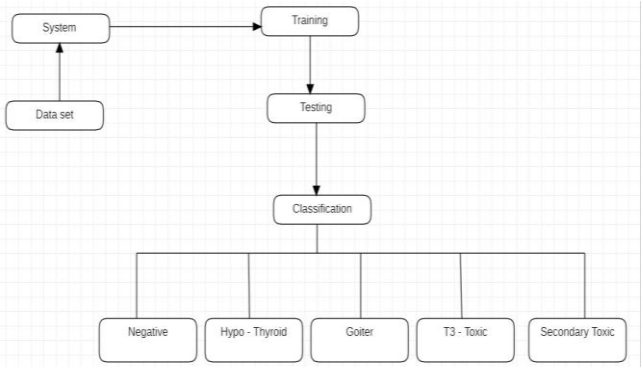

Figure 2. Architecture of Thyroid Disease

Decision Tree

Each core node in a flowchart-like structure represents a feature (or attribute), a branch represents a decision rule, and each leaf node represents the conclusion. A decision tree's root node is the node at the top. It gains the power to divide based on how important particular features are. This type of tree division is known as recursive partitioning. This decision-making method is comparable to a flowchart. It is a flowchart-like image that perfectly represents human level thinking. As a consequence, decision trees are straightforward to comprehend and analyses.

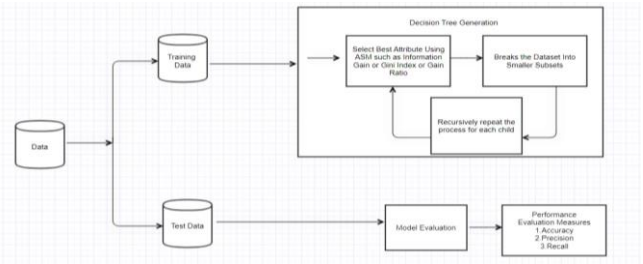

Figure 3. Generating Dataset using Decision Tree

XG Boost

The acronym XGBoost stands for Extreme Gradient Boosting. XGBoost is a distributed gradient boosting library that is exceptionally efficient, versatile, and portable. To develop machine learning algorithms, use the Gradient Boosting framework. allows parallel tree boosting to perform a variety of data science queries quickly.

Adaboost Classifier

In 1996, Yoav Freund and Robert Schapire introduced Ada-Boost, also known as adaptable boosting, as an ensemble boosting classifier. To improve accuracy, combine numerous classifiers. AdaBoost is a method for creating iterative ensembles. The AdaBoost classifier constructs a strong classifier that is extremely accurate by merging a number of weak classifiers. Adaboost's central tenet is to train data samples and create classifier weights at each iteration to ensure accurate prediction of anomalous events. A simple classifier is any machine learning method that gives weights to the training data. Adaboost must meet two requirements:

1. To train the classifier interactively, a variety of weighted training cases should be employed.
2. It attempts to offer a suitable match for these occurrences in each iteration by categorizing training mistake. It works in the following manner:
   i. Adaboost begins by selecting a random member of the education subset.
   ii. It trains the AdaBoost machine learning model repeatedly by picking the training set based on the accuracy of the previous training.
   iii. It provides erroneously classified data additional weight, increasing the possibility that these observations will be categorized in the next cycle.
   iv. Weight is assigned to the trained classifier in each iteration based on its accuracy. The classifier with the highest accuracy will be given more weight.
   v. This method is used repeatedly until the task is completed.

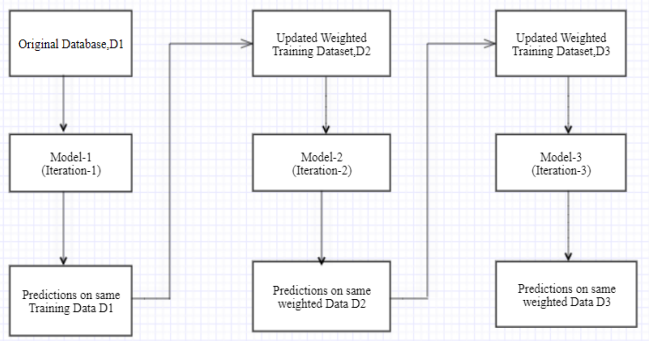

Figure 4. Adaboost Classifier

Support Vector Machine

Support Vector Machines are commonly considered of as a classification approach, despite the fact that they can address both classification and regression issues. It easily handles a large number of continuous and categorical variables. To

discriminate between classes, SVM creates a hyperplane in multidimensional space. To reduce inaccuracy, SVM leverages

SVM's basic principle is to determine the maximum marginal hyperplane (MMH) that optimally divides the dataset into classes. repeatedly created optimum hyperplanes. The classifier with the highest accuracy will be given more weight.

In the following steps, maximize the marginal hyperplane:

1. Use hyperplanes to get the best feasible class separation. The left-hand figure depicts black, blue, and orange hyperplanes. While the black successfully identifies the two groups, the blue and orange exhibit more classification mistakes in this circumstance.
2. Select the hyperplane in the right image that is farthest distant from the two closest data points.

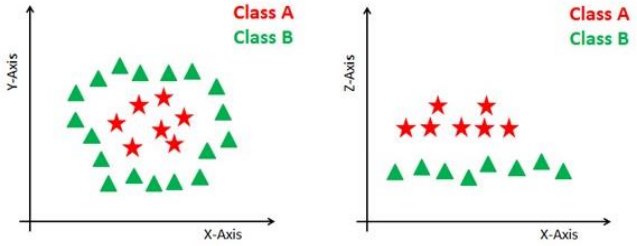

Figure 5. Support Vector Machine Classifier

EXPERIMENTAL RESULTS AND DISCOVERY

The thyroid dataset shown below was utilized to analyses the experimental results. Four algorithms were used in this study: Decision tree, XGBoost, AdaBoost, and Support vector machine.

Dataset

Table 1. Thyroid Disease Dataset

age	Sex	goitre	psych	on_thyroxine
35	F	f	f	f
63	M	f	f	f
25	F	f	f	f
53	F	f	f	f
92	F	f	f	f
67	M	f	f	f
60	F	f	f	f
60	F	f	f	f
48	F	f	f	f
27	F	f	f	f
73	F	f	f	f
19	M	f	f	f
72	F	f	f	f
16	M	f	f	f
54	F	f	f	f
39	F	f	f	t
38	M	f	t	f
33	F	f	f	f
45	F	f	f	f
54	F	f	f	f
21	F	t	f	f
19	M	f	f	f
51	F	F	f	F
63	F	F	f	F
51	F	F	f	F
64	M	F	f	F
19	F	F	f	F
40	F	F	f	F
54	F	F	f	F
19	F	F	t	F
33	F	F	f	F
71	M	F	f	F
49	F	F	f	F
79	F	F	f	F
21	F	F	f	F
20	F	F	f	F
20	F	T	f	F
79	M	F	f	F
64	F	F	f	T
59	F	F	f	F

Figure 6. Thyroid disease Classifier

Figure 7. Information about Thyroid Disease

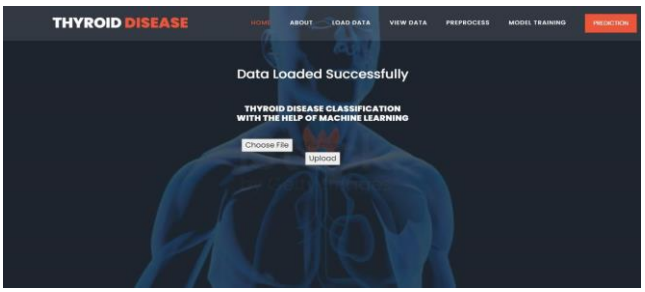

Figure 8. Classifier Data loaded successfully

Figure 9. Classifier dataset displayed

Figure 10. Classifier data splits successfully

Figure 11. Classifier choosing Machine Learning algorithms

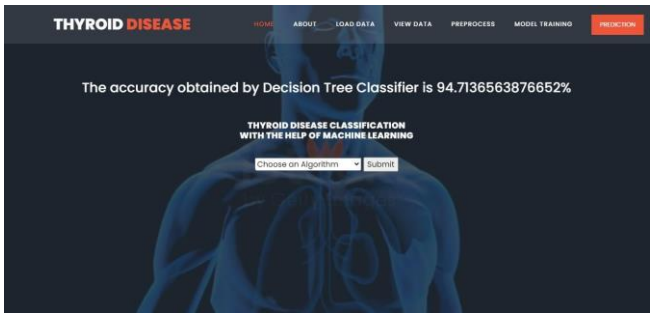

Figure 12. Accuracy obtained by the Classifier

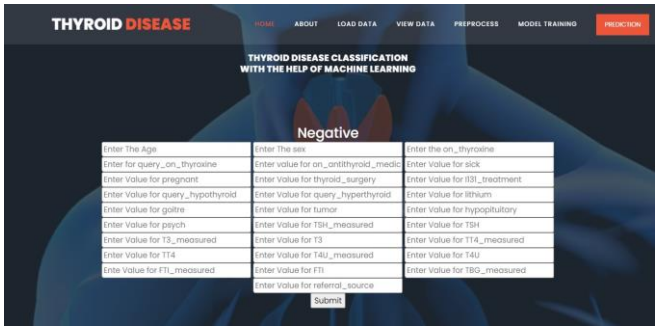

Figure 13. Thyroid Disease Classification with the help of Machine Learning

CONCLUSION

Thyroid disease is one of the ailments that is affecting the worldwide population and is getting more common. Based on medical reports that demonstrate a serious imbalance in thyroid illness, our study investigates the division of thyroid disease into hyperthyroidism and hypothyroidism. The sickness was classified using an algorithm. Machine learning has created two models with promising outcomes using a variety of techniques. All of the characteristics in the first model, which has 16 inputs and 1 output, are captured, and the Ada-boost approach outperforms the other algorithms with a score of 97.35. According to previous study, the second embodiment lacks the following properties. The characteristics 1-query_thyroxine was removed. 2. search for "hypothyroid," 3. search for "hyperthyroid".

REFERENCES

[1] O. Senashova and M. Samuels, “Diagnosis and management of nodular thyroid disease,” Vascular and Interventional Radiology, vol. 25, no. 2, article 100816, 2022.

[2] X. Zhang, V. C. Lee, J. Rong, J. C. Lee, and F. Liu, “Deep convolutional neural networks in thyroid disease detection: a multi-classification comparison by ultrasonography and computed tomography,” Computer Methods and Programs in Biomedicine, vol. 220, article 106823, 2022.

[3] M. Kang, T. S. Wang, T. W. Yen, K. Doffek, D. B. Evans, and S. Dream, “The clinical utility of preoperative thyroglobulin for surgical decision making in thyroid disease,” Journal of Surgical Research, vol. 270, pp. 230–235, 2022.

[4] Chaubey, D. Bisen, S. Arjaria, and V. Yadav, “Thyroid disease prediction using machine learning approaches,” National Academy Science Letters, vol. 3, pp. 128–133, 2021.

[5] Dewangan, A. Shrivas, and P. Kumar, “Classification of thyroid disease with feature selection technique,” International Journal of Engineering & Technology, vol. 2, no. 3, pp. 128–133, 2021.

[6] K. Shankar, S. Lakshmanaprabu, D. Gupta, A. Maseleno, and V. Albuquerque, “Optimal feature-based multi-kernel SVM approach for thyroid disease classification,” The Journal of Supercomputing, vol. 28, no. 76, pp. 1128–1143, 2020.

[7] P. Poudel, A. Illanes, M. Sadeghi, and M. Friebe, “Patch based texture classification of thyroid ultrasound images using convolutional neural network,” in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5828–5831, Berlin, Germany, 2019.

[8] M. Guo and D. Yongzhao, “Classification of thyroid ultrasound standard plane images using ResNet-18 networks,” in 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), pp. 324–328, Xiamen, China, 2019.

[9] Ma, J. Guan, W. Zhao, and C. Wang, “An efficient diagnosis system for Thyroid disease based on enhanced Kernelized Extreme Learning Machine Approach,” in International Conference on Cognitive Computing, pp. 86–101, Cham, 2018.

[10] Aswathi and A. Antony, “An intelligent system for thyroid disease classification and diagnosis,” in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1261–1264,

Coimbatore, India, 2018.
[11] K. Chandel, S. Veenita Kunwar, T. C. Sabitha, and S. Mukherjee, "A comparative study on thyroid disease detection using K-nearest neighbor and naïve Bayes classification techniques," CSI Transactions on ICT, vol. 4, no. 2-4, pp. 313–319, 2018.
[12] K. Shankar, S. Lakshmanaprabu, D. Gupta, A. Maseleno, and V. Albuquerque, "Optimal feature-based multi-kernel SVM approach for thyroid disease classification," The Journal of Supercomputing, vol. 28, no. 76, pp. 1128–1143, 2017.
[13] Dr. Srinivasan B, Pavya K "Diagnosis of Thyroid Disease: A Study" International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov – 2016.
[14] Khushboo Taneja, Parveen Sehgal, Prerana "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of Research in Management, Science & Technology (E-ISSN: 2321- 3264) Vol. 3, No. 2, April 2016.
[15] Kouroua, K., Exarchosa, T.P. Exarchosa, K.P., Karamouzisc, M.V. and Fotiadisa, D.I. (2015) Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal, Vol. 13, Pp.8–17.
[16] S. Prerana and K. Taneja, "Predictive data mining for diagnosis of thyroid disease using neural network," International Journal of Research in Management, Science & Technology, vol. 3, no. 2, pp. 75–80, 2015.
[17] Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. Jama, 309(13):1351–1352, 2013.
[18] Azar, a.T, Hassanien, A.E. and Kim, T. Expert system based on neural fuzzy rules for thyroid diseases diagnosis, Computer Science, Artificial Intelligence, arXiv:1403.0522, Pp. 1-12,2012.
[19] Shukla, A. & Kaur, P. (2009). Diagnosis of thyroid disorders using artificial neural networks, IEEE International Advance computing Conference (IACC 2009)– Patiala, India, pp 1016-1020.
[20] Keles, A. ESTDD: Expert system for thyroid diseases diagnosis, Expert Syst Appl., Vol. 34, No.1, Pp.242–246,2008.