How to Cite:
Verma, S., Popli, R., Kumar, H., & Srivastava, A. (2022). Classification of thyroid diseases using machine learning frameworks. *International Journal of Health Sciences*, *6*(S1), 7552–7566. https://doi.org/10.53730/ijhs.v6nS1.6603

Classification of thyroid diseases using machine learning frameworks

Shanu Verma
J. C. Bose University of Science and Technology, Faridabad, 121002, India
Corresponding author email: shanu.verma56@gmail.com

Rashmi Popli
J. C. Bose University of Science and Technology, Faridabad, 121002, India

Harish Kumar
J. C. Bose University of Science and Technology, Faridabad, 121002, India

Atul Srivastava
Amity School of Engineering and Technology, AUUP, Lucknow, India

***Abstract*---**Thyroid diseases are instigated due to disparity in production of hormones –TSH, T4 and T3. Most of the patients of thyroid dysfunction go untreated due to late detection or no detection at all. Machine learning based models for detection of thyroid diseases offer a significant assistance to healthcare. The medical history of the patient supplies the features required by machine learning based classification and prediction models for thyroid dysfunction. The aim of this research paper is to acquire a classification model based on machine learning techniques for assessment of euthyroidism, hyperthyroidism, and hypothyroidism among males, females, and children. Different machine learning classification algorithms such as naïve bayes, decision tree, random forest and logistic regression are used for classification of real data. The accuracy of each of the techniques has established using metrics like precision, recall, specificity and sensitivity. A thyroid dataset has been retrieved from two hospitals in Haryana from January 2020 to July 2020 to train the proposed model. The dataset comprises of medical history of 539 thyroid patients including children, men, and women of various ages. Out of 539 patients screened, 163 have irregular TSH, 138 have prevalence of elevated TSH with 376 having minimal TSH elevation. Young girls predominantly suffer more as compared to other strata of patients from Thyroid disorders. This paper classifies thyroid disease and compares machine learning algorithm among each other for predictive thyroid disease and finds the best accuracy among them.

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 18 Feb 2022, Manuscript revised: 09 March 2022, Accepted for publication: 27 April 2022

***Keywords*---**Machine learning, Thyroid Dysfunction, Classification, Hyperthyroidism, Hypothyroidism.

Introduction

The thyroid gland is an imperative organ of humanoid which is located in the throat underneath eminence of thyroid cartilage also named the Adam's apple. The gland of thyroid releases two vital hormones produced from the iodine of the food consumed – Triiodothyroinine (T3) and Thyroxine (T4) which participates in regulating body's metabolism [12]. Too high or too low production of T3 and T4 hormones may cause severe disorders. Thyroid Simulating Hormone (TSH) is produced by hypothalamus, a gland in brain, which signals pituitary gland of the brain to administer the production of T4,T3 hormones. When T4, T3 levels are lowest and production needs to be increased, endocrine gland releases more TSH. When T4,T3 levels are highest the endocrine gland slows down the production by supplying less amount of TSH to thyroid gland.

Abnormality in thyroid gland's functions or structure results in medical attention. The symptoms attribute to increased or decreased concentrations of plasma of hormones, is known as hyperthyroidism and hypothyroidism respectively. Cardiovascular disorders involve an extreme thyroid condition, hypertension, hyper cholesterol levels, clinical depression, and infertility. [6] The second-largest disease in the endocrine world that may lead to a patient is endocrine thyroid gland disorder located on the cervix [4].

Diagnostic revelations regarding thyroid disease are of major concern for medical science. Along with the medical procedures to determine the thyroid disease, machine learning also plays a pivotal role. Machine learning enables probabilistic models to enhance the accuracy of disease determination. Efficiency of machine learning based classification models is directly proportional to the thoroughness of its training using training data set. Medical history of the patient including physical specifications (Age, Weight etc.), symptomatic record (anxiety, racing heart, weight loss, bulging eyes etc.) and diagnostic laboratory outcomes (blood test report, ultrasound report etc.) are commonly used as features for training the model. [5] Trained a machine learning based classification model has following variables as features – Serum, Age, Name, Total thyroxine (T4),Total triiodothyronine (T3), TSH (4th Generation) from a dataset of 539 benign thyroid nodules.

Related Work

Jayatilake et al reviewed and compared machine learning tools for disease prediction in healthcare. The article covers numerous machine learning techniques for decision making which fall in one of different classes of machine learning i.e. reinforcement learning, unsupervised learning and supervised learning. Paper further emphasises and institutes the fact with significant statistical evidences that detection of disease at early stage lessens the treatment overhead and amplifies the recovery rate in various critical diseases like breast cancer, lung cancer, heart diseases, diabetes etc. Even in COVID-19 pandemic situation, artificial intelligence played a vital role at various fronts of COVID management.

Kashyap et al predominantly studied machine learning based prediction and classification models for thyroid diseases. Authors represent a comparative insight of accuracy of Artificial Neural Network (ANN) based models and decision tree based models using confusion matrix. ANN gives better accuracy than decision tree based classifiers.

Thyroid disease fallouts in huge impression on health if not detected in advance [Min Hu et al]. In Japan only 450,000 patients suffering from thyroid are getting treatment against 2.4 million total patients. The classification model based on machine learning proposed by Min Hu et al has mediocre performance with hypothyroidism as compared to hyperthyroidism due to undergoing levothyroxine treatment for patients with hypothyroidism. Levothyroxine treatment may affect the routine laboratory values.

Priyanka et al conducted a survey among young females from rural as well as from urban areas in Bangalore to study thyroid dysfunction. The medical history of women aged 18 to 30 years was acquired from two hospitals – Boring and Lady Cruzon Hospital. Authors used IBM SPSS software to classify the data collected from hospitals and concluded that young women are more prone to suffer from thyroid dysfunction. Table 1 mentions significant work done by several researchers in the field of ML based disease prediction for thyroid dysfunction.

Table 1
Prediction of thyroid disease based on ML techniques

Sr. No	Author details	Publication Journal	Technique Used	Learning and Training Dataset	Result (Classification / Prediction Accuracy)
1	Kashyap et al(2021)	International Journal of Creative Research Thoughts (IJCRT)	o ANN o Decision Tree	Thyroid dataset from UCI machine learning repository	ANN-97.1% DT- 93.7%
2	Duggal et. al (2020)	International Conference on Cloud Computing, Data Science & Engineering	o Support vector machines o Naïve Bayes o Random forest	Thyroid dataset from UCI machine learning repository	SVM-92.92% NB- 74.37% RF-78.21
2	Pal et. al (2020)	Indian Journal of Public Health Research and Development	o Random forest o J48	Thyroid dataset from Github UCI and pathology	RF- 97.59% J48- 99.12%

3	Chaubey et. al (2020)	Natl. Acad. Sci. Lett	o K-nearest neighbours o Decision tree o Logistic regression	Graven Institute in Sydney, Australia, uploaded to the UC-Irvin, knowledge discovery	KNN-96.875% DT- 87.5% LR-81.25%
4	Tyagi et al (2018)	International Conference on Parallel, Distributed and Grid Computing(PDGC-2018)	o Support vector machine (SVM) o K-NN o Decision Trees	Thyroid dataset from UCI machine learning repository	ANN-97.50% SVM - 99.63% KNN-98.62 DT-75.76%

Problem Descriptions and Justification

Based on the study of literature in the field of machine learning in thyroid disease prediction, plausible conclusions can be:

(1) The extent of work done in this direction is significantly less.
(2) Several patients suffering from thyroid dysfunctions may go untreated due to late or no detection. Therefore, early detection of thyroid disease alleviates the burden of treatment along with fastened recovery.
(3) Machine learning based detection models for thyroid disease are rominent add-ons in healthcare.

The above takeaways from the study of previous work in the same field drives to propose a classification model for thyroid dysfunction based on machine learning algorithms – random forest, logistic regression, naïve bayes, support vector machine, k-nearest neighbor, and decision tree.

Thyroid Detection Using Machine Learning

The machine learning deployed in healthcare usually suffers from biases and errors. The objective is to avoid it. Machine learning models are not programmed to do a particular task but it is designed to learn how to do it. Therefore, a resilient dataset for training and testing is required. Depending on the training data, the machine learning algorithm deployed in the model will do classification. The complete process of learning based model for classification at abstract level is as follows:

1. Medical history data of the patient is sustained.
2. The data is processed to adapt it into usable format and unsought things are pruned.
3. Substantial features are selected which will participate in classification.
4. The model is tested on training data.
5. Trained model is tested on testing data and accuracy is estimated.
6. Until the accuracy meets predefined threshold, step 4 and 5 are repeated.

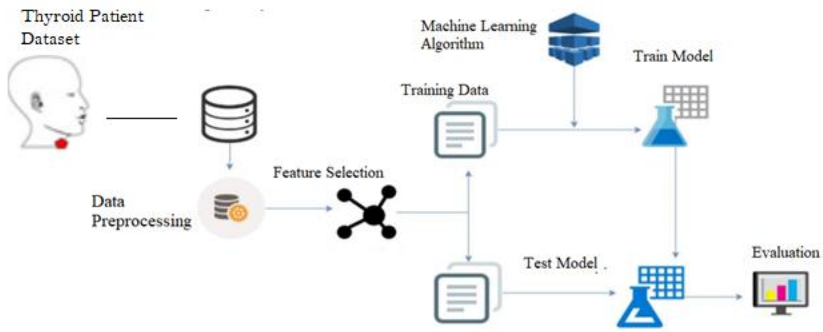

Figure 1: Abstract Machine Learning Based Model

The model is abstract because it does not reveal the algorithm used for classification. Here are several ML based algorithms for classification such as:

Logistic Regression

Logistic regression is part of Machine Learning classification algorithms used to classify disease using probability prediction of categorical target variable. The probability is calculated as follows:

$P(y_i=C_j | X) = 1/(1+e^{(-(\beta_0+\beta_1 x))})$

Where, $P(y_i=C_j | X)$ is the probability of i-th observation belonging to target class C_j, y_i is the target value of i-th observation, X is the matrix of all observations, β_0 and β_1 are learning parameters and e is Euler's number. $\beta_0+\beta_1 x$ is also known as Sigmoid Function.

Decision Tree Algorithm

Decision tree is a classification technique which is basically represented as a binary tree. The process starts at the root node. Each level corresponds to one particular feature. The leaf nodes contain the decision (Class) [25]. The purpose of a Decision Tree is to build a training model that can use to predict the value of the target variable or class by learning simple decision rules inferred from prior data(training data)[26].

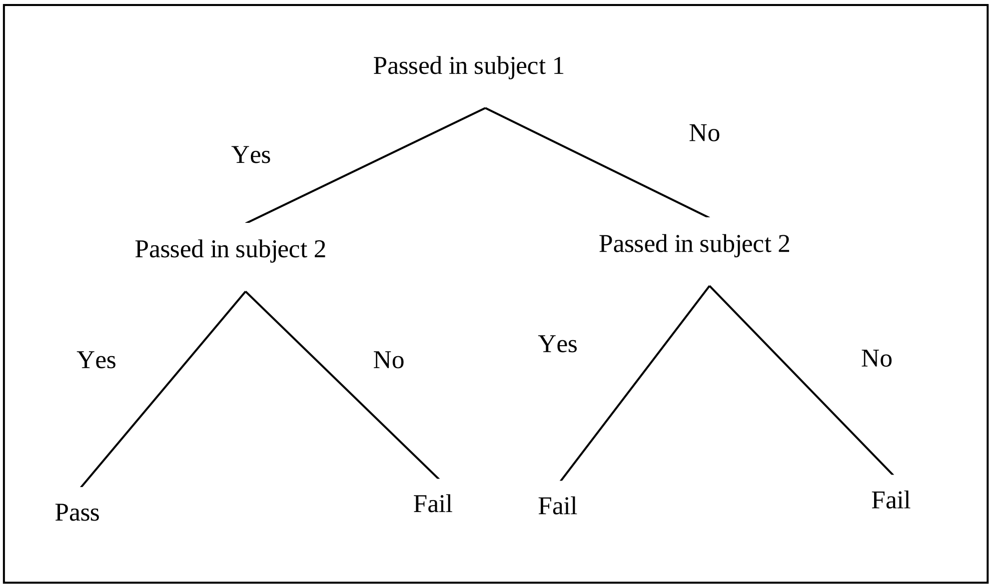

Figure 2: Example of decision tree

Naïve Bayes

Naïve Bayes classifier is suitable for huge dataset with high dimensionality [S Vijayarani]. It is a probabilistic model for classification.
The probability of data record x having the class label C_i is:
$P(C_i | x) = (P(x | C_i)\ P(C_i)) / (P(x))$
In the given input dataset the class label C_i with largest conditional probability value determines the class of the sequence. [27]

Random Forest Algorithm

A random forest algorithm based classifier which conglomerates several decision trees built over sub-samples of the dataset and the decision rests on the average of each decision tree. The accuracy is improved significantly and over-fitting of data is controlled. It is a supervised learning technique and forest refers to collection of trees (decision trees). Random forest has significant usage in the field of healthcare. [2].

Proposed Algorithm

Classification process is carried out on the features extracted from medical history of patient put in a dataset. The dataset is mostly not inapt for direct use in classification model. Some values in the dataset can be noisy, incomplete, ambiguous or missing for some features. Therefore, data cleaning is an integral step in machine learning algorithms. Outcome of data cleaning phase is cleaned dataset of its features bear perfect values but all of them may not be required for classification process. Only most significant and sufficient features are opted out from the dataset which participate in classification process. The dataset is further split into two categories – with treatment and without treatment. The data of patients who went through treatment is put in with treatment set and the model is trained over it. The model is tested and trained over without treatment dataset also.

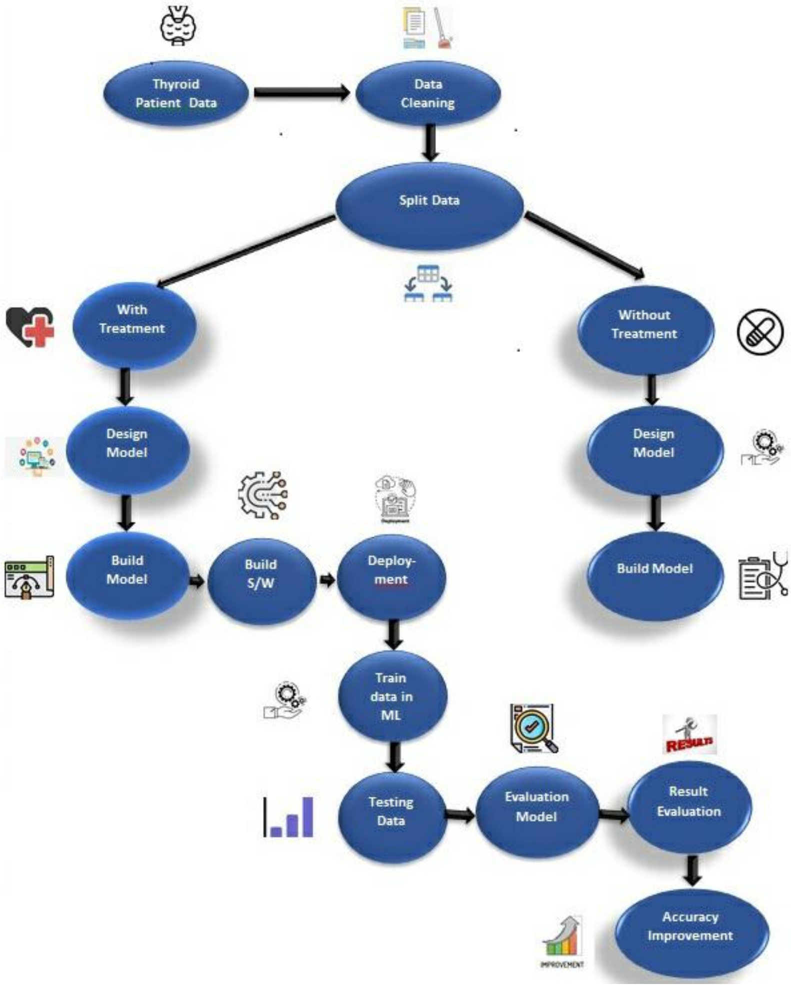

Figure 3: Framework of machine learning

Proposed Classifier Algorithm

- Load the required library files.
- Load Dataset
- Clean Dataset by removing missing values
- Select best features for classification
- The dataset is split into two partitions viz. patients with treatment and patient without treatment.
- Effect of Treatment is calculated as follows: EOT=E[Y_i (1)-Y_i (0)]
- Where, Y_i (1) represents patients with treatment,
- Y_i (0) Represents patients without treatment.
- Fit() Function is used for training data into model.
- Entropy is calculated as follows: $E = -\sum_{i=1} \quad N p_i log_2 p_i$

Where, p_i represents proportion of patients which belong to i-th class.
The above algorithm is developed as a machine learning based model. In the experiments, it is established that the proposed algorithm has training accuracy near to 94%.

Material and Methods

Description of Dataset

Instead of using synthetic dataset available at various online sources like Kaggle, in this paper real dataset thyroid disorder is acquired from two hospitals in Haryana from January 2020 to July 2020 to train the proposed model. The dataset comprises of medical history of 539 thyroid patients included children, men, and women of various ages. Out of 539 patients screened, 163 are having irregular TSH, 138 are having prevalence of elevated TSH with 376 having minimal TSH elevation. TSH levels were low in 25 of the participants in the study. Table 2 gives Meta information about the dataset and dependent/independent variables. Age, TSH, T3, TT4, and T4U were constant independent variables in the thyroid dataset, which had 8 independent variables. Table 3 represent selected features for our model.

Table 2
Representation of thyroid dataset variable

Source	*Description*
Sample Size	Total= 539 Hypothyroidism= 138 Hyperthyroidism= 25 Euthyroidism= 376
Dependent Variables	
Hypothyroidism	Production of Hormone is to little
Hyperthyroidism	Production of Hormone is too much
Euthyroidism	The state of normal thyroid function
Independent Variables	
T3	Triiodothyronine Stimulates the metabolism
T4	Thyroxin produced by thyroid gland
TSH	Thyroid Stimulating hormone pituitary hormone

Table 3
Features and domain of their values

Feature	*Values*
Name	Nominal
Age	Continuous
Sex	0,1
Married	Continuous
TSH	Continuous
T3	Continuous
T4	Continuous

Type	1,2,3
Class	3,2,1

```
th.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 417 entries, 0 to 416
Data columns (total 9 columns):
Name       417 non-null object
Married    417 non-null int64
Age        417 non-null int64
Sex        417 non-null int64
T3         417 non-null float64
T4         417 non-null float64
TSH        417 non-null float64
Type       417 non-null object
Class      417 non-null int64
dtypes: float64(3), int64(4), object(2)
memory usage: 29.4+ KB
```

Figure 4: Dataset attributes

Inclusion and Exclusion Criteria

Men, women, and children who were able to participate without medical complications were included in the analysis. Men, women, and children who were severely sick were not permitted to partake.

Performance Measure

The predicted class of an object fall into various categories – False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN). Table 4 Represents a confusion matrix to represent the understanding of TP, TN, FP and FN. False positive implies that the patient does not have thyroid disease, but has thyroid disease.A false negative means that the patient is unlikely to have thyroid disease and does not have thyroid disease. True positive implies that the patient has predicted thyroid disease and actual thyroid disease.
True negative means that the patient does not have thyroid disease and does not predict thyroid disease.

Table 4
Confusion matrix

		Actual Class	
		Has thyroid disease	Does not have thyroid disease
Predicted Class	Has thyroid disease	*True Positive*	*False Positive*
	Does not have thyroid disease	*False Negative*	*True Negative*

Accuracy: Accuracy of classification is the paramount expectation of from the model. Accuracy simply denotes the extent of correct assessments out of total assessments. Equation ___ specifies accuracy.

Accuracy=(Correct assessments)/(Total Assessments)=(TP+TN)/(TP+TN+FP+FN) (__)

Precision: Precision is the proportion of positive comments predicted to be accurately predicted is the ratio of positive comments to total predicted. Precision (a.k.a. positive predicted value) is a quantity of the proportion of patients identified by classifier to have the disease, actually have had the disease. Precisely, 'the probability that a patient being actually sick if diagnosed as sick by the classifier'.

Accuracy=TP/(TP+FP)

For example, if precision=1, it means that all patients diagnosed by the classifier really had the disease.

Recall (Sensitivity): Recall is the proportion of correctly predicted positive observations out of total positive observations. In general, it points the finger towards the positive cases classified as negative (False Negative) cases by the classifier. Precisely, 'Probability of a sick person being identified as sick by the classifier'.

Recall=TP/(TP+FN)

Specificity: Similar to recall, 'Probability of a healthy person being identified as healthy by the classifier'.   .

Specificity=TN/(TN+FP)

F1 score: The harmonic mean of Precision and Recall is F1-score.

F1 Score=2/(1/(Precision (P))+1/(Recall (R)))=2*(P*R)/(P+R)

Result and Discussion

The proposed algorithm is implemented using python 3 (libraries used Skylearn, pandas, Numpy). Recall, specificity, precision, accuracy, and F1 scores are used as performance metrics to compare efficiency of proposed algorithm with other trivial classification algorithms mentioned in section 4. This research categorizes these diseases into appropriate categories such as hypothyroidism, hyperthyroidism, and normally based on the value of T4, T3, and TSH. The proportion of dataset used for testing and training are 20% and 80% respectively. This proportion applies to both categories of dataset.

The performance measure metrics mentioned in previous section are used to determine the effectiveness of the proposed algorithm. As compared with traditional classification algorithms the results are– Random Forest, Decision Tree, Naïve Bayes, and Logistic Regression.

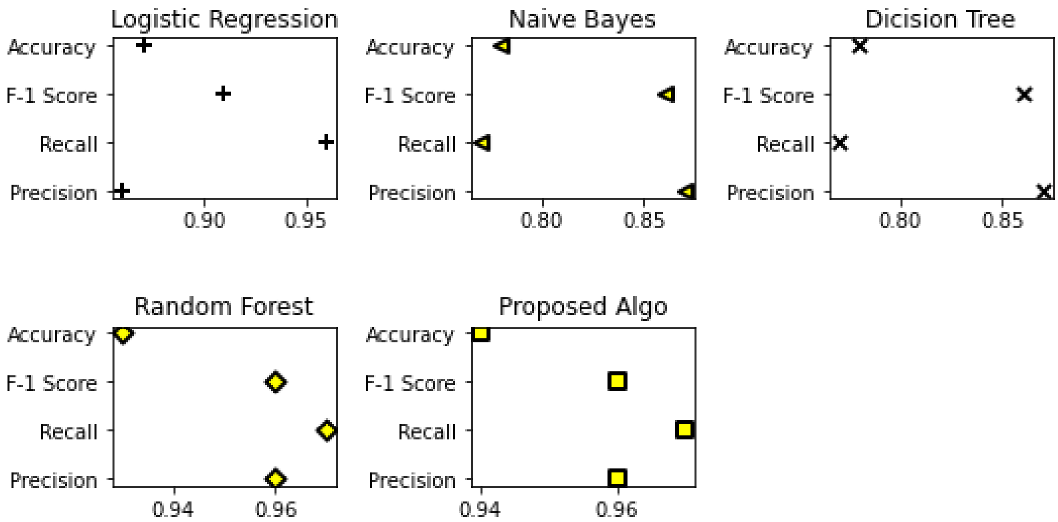

Figure 5: Performance metrics of various classifiers

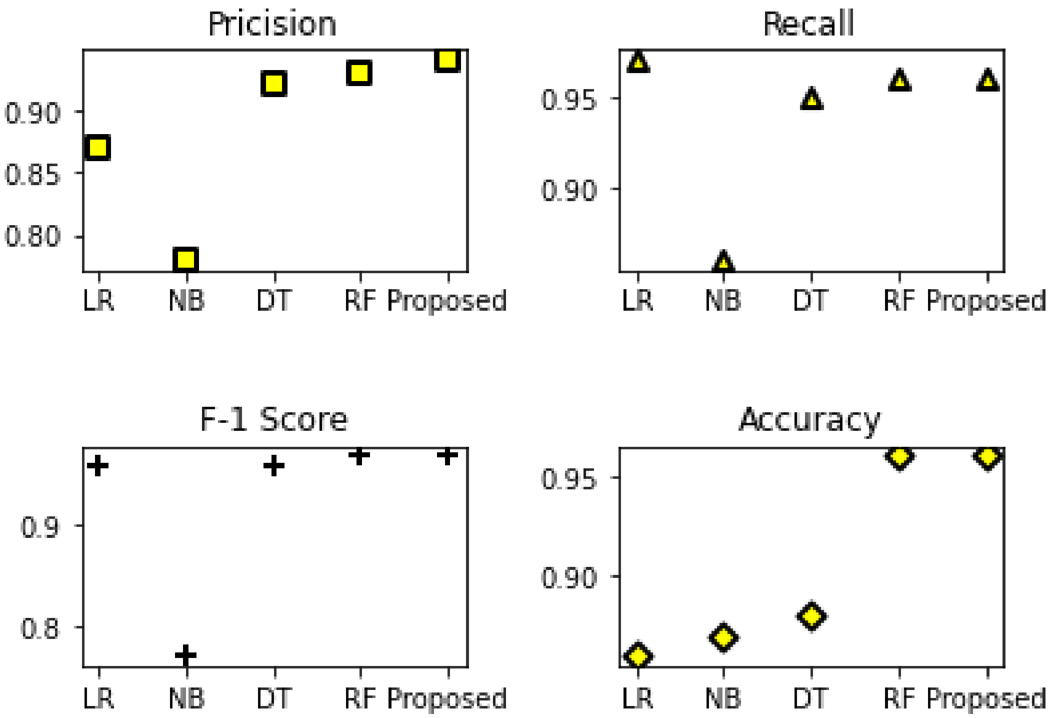

Figure 6: presents a performance of various classifiers projected through performance metrics

It is obvious from the above figure that the proposed algorithm has precision almost 86% which is better than Decision Tree, Logistic Regression and Naïve bayes and is equal to Random Forest. Recall represents the probability of an object being put in the proper class. It is evident from the figure 6 that the proposed algorithm is better than Random Forest Classifier, Naïve Bayes, Logistic Regression, and equals Decision Tree.

F-1 score of the proposed algorithm is 0.96 which is better than Random Forest Classifier, Naïve Bayes, Logistic Regression, and is equal to that of Decision Tree. The proposed algorithm beats all four traditional classifiers in terms of accuracy. By the above analysis it is established that the proposed algorithm successfully classifies the patients to their correct classes with the accuracy of 94%.

Conclusion

Thyroid disease diagnose primarily revolves around three hormones T3, T4 and TSH. Disordered balance of these hormones results in various types of thyroid diseases. Along with diagnostic data corresponding to T3, T4 and TSH, the medical history of the patient which contains symptomatic data is also used by proposed algorithm for diagnostic classification. The performance of proposed algorithm is compared with well-established traditional classifiers like Random Forest Classifier, Logistic Regression, Naïve Bayes and Decision Tree. The algorithm has been tested on the real data collected from various hospitals and through surveys. Less but significant features are chosen for classification by the proposed algorithm so that the patient does not have to go for too many check-ups. Precision, recall, F-1 score and Accuracy are used as performance metrics. The proposed algorithm beats LR, NB and DT classifier in precision (0.96). Recall (0.97) of proposed algorithm is better than LR, NB and DT classifier. F-1 score (0.96) of the proposed algorithm is better than LR, NB and DT classifier. The proposed algorithm accuracy is 94% which is improved than all four traditional algorithms.

Acknowledgement: I would like to thank my research supervisors Dr. Rashmi Popli and co-supervisors Dr. Harish Kumar for providing me guidance throughout this research.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Suresh Kumar Kashyap, Dr. Neelam Sahu A Comparative Study Of Machine Learning Based Model For Thyroid Disease Prediction International Journal of Creative Research Thoughts (IJCRT)
2. Mohammad Savargiv ,Behrooz Masoumi ,(2021) A New Random Forest Algorithm Based on Learning Automata Hindawi Computational Intelligence and Neuroscience Volume 2021, Article ID 5572781, 19 pages https://doi.org/10.1155/2021/5572781
3. Min Hu, Chikashi Asami 2021 Thyroid dysfunction diagnosis from routine laboratory tests based on machine learning doi: https://doi.org/10.1101/2021.03.30.21254605;

4. Liyong Ma , 1 Chengkuan Ma,1 Yuejun Liu,2 and Xuguang Wang Thyroid Diagnosis from SPECT Images Using Convolutional Neural Network with Optimization Volume 2019, Article ID 6212759, 11 pages https://doi.org/10.1155/2019/6212759
5. Jordi L. Reverter , 1,2 Irene Rosas-Allende,3 Carlos Puig-Jove,1,2 Carles Zafon,2,4 Ana Megia,2,5 Ignasi Castells,2,6 Eduarda Pizarro,2,7 Manel Puig-Domingo,1,2 and M. Luisa Granada2 Prognostic Significance of Thyroglobulin Antibodies in Differentiated Thyroid Cancer Volume 2020, Article ID 8312628, 6 pages https://doi.org/10.1155/2020/8312628
6. Gyanendra Chaubey1, Dhananjay Bisen1, Siddharth Arjaria1 ,Vibhash Yadav Thyroid Disease Prediction Using Machine Learning Approaches DOI: 10.1007/s40009-020-00979-z
7. Farzad PAKDEL 1 , Roghayeh GHAZAVI 2 (2019) Effect of Selenium on Thyroid Disorders: Scientometric Analysis Iran J Public Health, Vol. 48, No.3, Mar 2019, pp.410-420
8. Barbara L Parry*, and Daniel F Kripke (2020) Antidepressant and Stabilizing Effects of Thyroid Hormone Augmentation in Women's Mood Disorders Volume 5, Issue 3,March 2020
9. Shaik Razia, P. Swathi Prathyusha, N. Vamsi Krishna, N. Sathya Sumana (2018) A Comparative study of machine learning algorithms on thyroid disease prediction International Journal of Engineering & Technolog vol 7 issue 2 .8 (2018) pp 315-319
10. Rahul Sindhwani, Punj Lata Singh (2019) Agile System in Health Care: Literature Review Springer Nature Singapore Pte Ltd. 2019 , https://doi.org/10.1007/978-981-13-6412-9_61 pp 643-652.
11. Saurabh Bilgaiyan, Santwana Sagnika 2017 A Systematic Review on Software Cost Estimation in Agile Software Development, Journal Of Engineering Science And Technology Review, pp 51-64 , DOI: 10.25103/jestr.104.08
12. Dhyan Chandra Yadav1 , Saurabh Pal 2020 Discovery of Hidden Pattern in Thyroid Disease by Machine Learning Algorithms Indian Journal of Public Health Research and Development DOI Number: 10.37506/v11/i1/2020/ijphrd/193785 pp 61-66
13. Vaishnavi Kannan, Mujeeb A Basit 2019 User stories as lightweight requirements for agile clinical decision support development Journal of the American Medical Informatics Association, Volume 26, Issue 11, November 2019, Pages 1344–1354, https://doi.org/10.1093/jamia/ocz123
14. Agile Estimation Techniques: A True Estimation In An Agile Project 2020 https://www.softwaretestinghelp.com/agile-estimation-techniques
15. Murad Ali, Zubair A Shaikh , Eaman Ali 2015 Estimation of Project Size Using User Stories DOI: 10.2991/racs-15.2016.9 International Conference on Recent Advances in Computer Systems (RACS 2015)
16. Andrew L. Beam 2018 Big Data and Machine Learning in Health Care American Medical Association doi:10.1001/jama.2017.18391
17. Charan R, Akash Yadav M, Aprameya N Katti (2019) Prediction of Thyroid Disease Based on Classification Using Hierarchical Structure Journal of Emerging Technologies and Innovative Research (JETIR) ISSN-2349-5162
18. Michael T. Biochemical Testing of the Thyroid: TSH is the Best and, Oftentimes, Only Test Needed – A Review for Primary Care doi:10.3121/cmr.2016.1309 6 Marshfield Clinic Health System

19. Jamot and Pettersson (2016). Agile challenges within regulated healthcare environments, Journal of Advanced Nursing, 48(5), 454–462
20. Grantina Modern (2020) Correlates of diarrhea and stunting among under-five children in Ruvuma, Tanzania; a hospital-based cross-sectional study Scientific African https://doi.org/10.1016/j.sciaf.2020.e00430 2
21. Olayinka Ayodele Jongboa Development of an ensemble approach to chronic kidney disease diagnosis https://doi.org/10.1016/j.sciaf.2020.e00456
22. K. Shailaja Machine Learning in Healthcare: A Review Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1
23. Priyanka Duggal Prediction Of Thyroid Disorders Using Advanced Machine Learning Techniques International Conference on Cloud Computing, Data Science & Engineering
24. Priyanka. A Nov 2019 Prevalence Of Thyroid Dysfunction Among Young Females Of Urban And Rural Population In And Around Bangalore Volume-9 | Issue-11 | November - 2019 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar Indian Journal of Applied Research
25. R. Vijaya Kumar Reddy1 , K. Prudvi Raju 2016 Prediction of Heart Disease Using Decision Tree Approach International Journal of Advanced Research in Computer Science and Software Engineering
26. S Radhika S Ramiya Shree(2020) Symptoms Based Disease Prediction Using Decision Tree and Electronic Health Record Analysis European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 4, 2020
27. Mostafa Langarizadeh and Fateme Moghbeli (2016) Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review doi: 10.5455/aim.2016.24.364-369
28. S Vijayarani 2014 Naïve Bayes Classification for Predicting Diseases in Haemoglobin Protein Sequences International Journal of Computational Intelligence and Informatics, Vol. 3: No. 4, January - March 2014.
29. Gangwar, Krishanveer and Kumar, Vimal and Singh, Ajay Kr. and Sharma, Vijay Kr., “Classification of Image Dataset using Convolutional Neural Network” Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019. http://dx.doi.org/10.2139/ssrn.3349594
30. Md Iqbal, Vimal Kumar and Vijay Kumar Sharma, “Krishi Portal: Web Based Farmer Help Assistance” International Journal of Advanced Science and Technology, 29(06), 4783 – 4786, 2020.http://sersc.org/journals/index.php/IJAST/article/view/19399
31. Vidhi Tyagi, Shivam Arora, Sattyam Gupta, Vijay Kr. Sharma, Vimal Kumar, “Architecture of an IoT-based Women Safety System” International Journal of Advanced Science and Technology, 29(05), 3670 – 3676, 2020. http://sersc.org/journals/index.php/IJAST/article/view/12258
32. Puneet Gupta, Vijay Kumar Sharma, Naman Mittal, Raghav Bansal, Himanshu Gupta, “Ai Enabled Virtual Environment Simulator” International Journal of Advanced Science and Technology, 29(3), 9604 – 9611, 2020.http://sersc.org/journals/index.php/IJAST/article/view/26920
33. Durgesh Tiwari, Vijay Kumar Sharma, “A Review on Conventional and Lightweight Security Techniques in Mobile and IoT Devices” Journal of

Information and Computational Science, 2020. http://www.joics.net/images/full_pdf/1585903931_B862.pdf
34. Vijay Kumar Sharma, Swati Sharma, “Sarcasm Sentiment Discernment towards Deep Learning” JOURNAL OF XI'AN UNIVERSITY OF ARCHITECTURE & TECHNOLOGY, 12(IV), 66 – 76, 2020. https://www.xajzkjdx.cn/gallery/9- nov2020.pdf
35. Vijay Kumar Sharma, Vimal Kumar, Md. Iqbal, Sachin Tawara, Vishal Jayaswal, “Virtual Mouse Control using Hand Class Gesture” GIS SCIENCE JOURNAL, 7(X), 454-458, 2020. http://gisscience.net/VOLUME-7-ISSUE-12-2020/
36. Vijay Kumar Sharma, Vimal Kumar, Shashwat Pathak, Naman Malik, Rachit Arora, “Image Web Crawler towards Machine Learning”, Integrated Intelligence Enable Networks & Computing (IIENC – 2020), September, 2020.
37. Vijay Kumar Sharma, Naman Malik, Riddhi Jain, Rachit Arora, Prachi Gupta, “American Sign Language Translator using Machine Learning”, JOURNAL OF XI'AN UNIVERSITY OF ARCHITECTURE & TECHNOLOGY, Volume XIII, Issue III, 368 - 371, 2021.https://www.xajzkjdx.cn/Current-Issue/
38. Vijay Kumar Sharma, Swati Sharma, Umang Arora, Anant Gupta, “Finely-grained Real-Time Facial Emotion Recognition Towards Neural Network”, Journal of Bioscience Biotechnology Research Communications, Volume 14, Issue 5, 176-181, 2021. https://bbrc.in/wp-content/uploads/2021/06/Volune-14-No-5-Special-Issue-2021-E-Copy.pdf
39. Shubham Shukla, Ajay Kumar Singh, Vijay Kumar Sharma, “Survey on Importance of Load Balancing for Cloud Computing ” 3rd IEEE (ICAC3N-21) [DOI: 10.1109/ICAC3N53548.2021.9725442] https://ieeexplore.ieee.org/document/9725442
40. Shreya Arora, Stuti Rastogi, Tushar Bhatia, Vijay Kumar Sharma, “Comparison Of Various Classifiers For Movie Data” 3rd IEEE (ICAC3N-21) [DOI: 10.1109/ICAC3N53548.2021.9725515] https://ieeexplore.ieee.org/document/9725515
41. Swati Sharma, Vijay Kumar Sharma, “Machine Learning Application: Sarcasm Detection Model” for the upcoming book ‘Artificial Intelligence for a Sustainable Industry 4.0’ by Springer. https://www.springer.com/gp/book/9783030770693