iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34381

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

Performance Evaluation of SVM and Random Forest for the Diagnosis of Thyroid Disorder

Shivastuti[1], Haneet Kour[2], Jatinder Manhas[3], Vinod Sharma[4]
[1, 2, 4]*Department of Computer Science and IT, University of Jammu, India*
[3]*Department of Computer Science and IT, Bhaderwah Campus, University of Jammu, India*

Abstract: ***Thyroid Disorder is the most common condition in the medical world, caused by the thyroid gland's failure to secrete hormones. Two separate machine learning (ML) techniques, including support vector machine (SVM) and random forest, were evaluated in this paper for thyroid disorder diagnosis. The experiment was conducted on the Thyroid Dataset, which was obtained from the UCI machine learning repository. These two machine learning techniques were compared using four performance metrics: accuracy, precision, recall, and f-score. SVM is found to be better than random forest for thyroid disorder diagnosis based on the experimental findings.***
Keywords: SVM, Python, Random Forest, Thyroid Disorder

I. INTRODUCTION

Thyroid Disorder is expected to become one of the most well-known endocrine issues on the planet. According to [**1**], hypothyroidism affects one in every 38000 people across the globe. Thyroid disease affects approximately 42 million people in India. Thyroid disorder is diagnosed based on its functional behavior and form, which involves hyperthyroidism and hypothyroidism, respectively. Hyperthyroidism is a disorder that occurs when our bodies produce too much hormone. Hypothyroidism is a condition in which the body produces inadequate thyroid hormone. Both of these conditions are severe and need treatment, and they can affect everyone [**2**]. As a result, early and reliable disease detection is becoming increasingly important in today's world, as the number of health problems increases year after year. In reality, a variety of machine learning methods have been used to determine the seriousness of thyroid disease. These ML techniques were found to be efficient for thyroid disorder diagnosis. The main of the current research work is to evaluate the performance of two most commonly used supervised machine learning algorithms i.e. support vector machine (SVM) and random forest for thyroid disorder diagnosis. In this paper, two thyroid disorders namely hypothyroidism and hyperthyroidism have been considered.
This research paper is organized as follows: Section 2 represents Literature Review related to the current work, Section 3 explains Materials and Methodology, Section 4 shows Results and Discussion, and Section 5 describes Conclusion and Future Scope.

II. LITERATURE REVIEW

The main aim of proposed research is to predict Thyroid Disorder by using machine learning techniques and evaluate their performance in order to find out the better technique for thyroid disease detection. In past, researchers have applied various machine learning techniques for the prediction of Thyroid disorders.
For the prediction of thyroid disorder, *Razia et al.* [**3**] conducted a comparative analysis of machine learning algorithms such as Support Vector Machine (SVM), Multiple Linear Regression, Nave Bayes, and Decision Tree. Thyroid disorder dataset gathered from UCI machine learning was used in the experimental analysis. Performance evaluation was performed using confusion matrix that predicted decision tree to be better approach in the diagnosis of thyroid with accuracy of 99.23%.
Soni and Giri [**4**] applied k-nearest neighbor (KNN), decision tree, and support vector machine for diagnosis of thyroid disease on dataset taken from UCI repository. The experimental analysis concluded support vector machine to be the best technique for thyroid disorder diagnosis as it achieved highest accuracy.
Begum and Parkav [**5**] evaluated KNN with naïve bayes and SVM for the prediction of thyroid disease. The authors also found out the correlation of TSH, T3, and T4 towards hyperthyroidism and hyporthyroidism. Experimental results were evaluated based on speed, accuracy and performance of the model and cost for the treatment. *Raisinghani et al.* [**6**] introduced various ML techniques i.e. decision tree, random forest, SVM, artificial neural network and logistic regression for the prediction of thyroid disorder. The performance comparison of these ML approaches was then evaluated using f1-score, recall, precision and accuracy.

Dharmarajan et al. [**7**] evaluated the performance of SVM, decision tree and naïve bayes for the prediction of thyroid disorders. The experimentation was carried out on the dataset of 500 patients. Decsion tree achieved best accuracy of 97.35%. *Chaubey et al.* **[8]** implemented logistic regression, decision trees and KNN algorithms to predict the thyroid disease. KNN achieved the highest accuracy of 96.88%.

III. MATERIALS AND METHODOLOGY

The main objective of the present research work is to predict thyroid disorder using machine learning techniques. In this study, two ML algorithms i.e. Support Vector Machine (SVM) [**10**] and Random Forest (RF) [**11**] have been taken. The experimental study has been conducted on *Python* using *Spyder* interface. These ML approaches have been implemented using *sklearn* library. The dataset for the experimental work has been collected from UCI machine learning repository [**12**]. This is the benchmark dataset of thyroid disorders; consisting of 7200 instances each with 21 input attributes and one output attribute (i.e. Class Label). The class labels are depicted as: Class 1 represents *normal* case; Class 2 means *hyperthyroidism* and Class 3 means *hypothyroidism*. The detail of thyroid disorder dataset is presented in table 1. The methodology for the presented work is provided in figure 1.

Table 1. Thyroid Dataset

No. of Instances	7200
No. of Input Attributes	21
No. of Output Classes	3
Class Label	1-Normal 2-Hyperthyroidism 3-Hypothyroidism

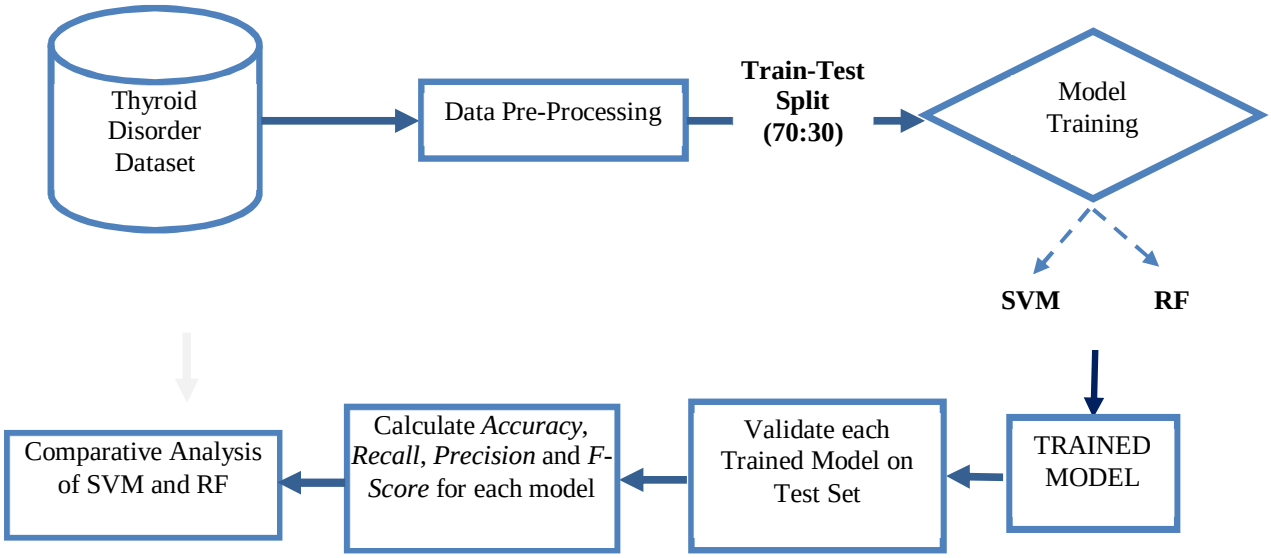


Fig. 1. Methodology of the current work

After records collection, *pre-processing* has been done to convert categorical values into numerical variables, and to detect outliers present within the data using statistical technique *Min-Max normalization* has been executed to transform all the input attributes to the range [0, 1]. The pre-processed data was parted into two sets in 70:30 proportions: *Train Set* and *Test Set*. Thus, *train set* has 5040 samples and *test set* has 2160 samples. Two ML techniques *SVM* and *RF* have been applied on *Train Set* individually with 10-fold cross validation in order to build the model for thyroid disease prediction. Each trained model has been validated with *test set* in order to assess its performance. *Accuracy*, *recall*, *precision*, and *F-score* were used to evaluate the performance of these two ML-based trained models. Based on the experimental findings, a comparison of SVM and RF-based trained models has been performed to determine which technique is the most accurate for Thyroid disorder prediction.

IV. RESULT AND DISCUSSION

The main goal of this study is to compare the success of SVM and random forest in predicting thyroid disorders. Table 2 shows the experimental findings for both SVM and RF techniques in the prediction of thyroid disorders. Both methods can be used to predict thyroid disorders, according to the findings of the experiments. Figure 2 presents comparative analysis of both ML techniques under study. Out of both classifiers under study, SVM showed best results as it achieved higher value for *accuracy*, *precision*, and *recall* being 93%, 89%, and 93% respectively. RF predicted lowest value for *accuracy* of 92%, *precision* of 85%, *recall* and *f1-score* of 92% and 0.88 respectively. In the current study, SVM has been discovered to be better methodology as compared with RF for the diagnosis of thyroid disorder.

Table 2: Experimental Results for SVM and RF in thyroid disorder prediction

ML Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	93%	89%	93%	0.91
RF	92%	85%	92%	0.88

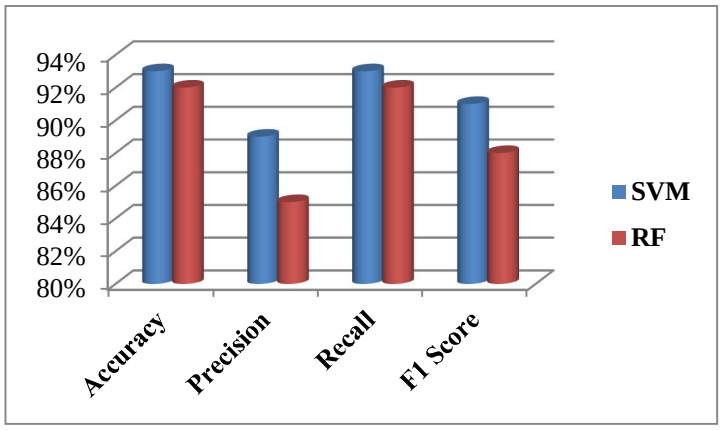


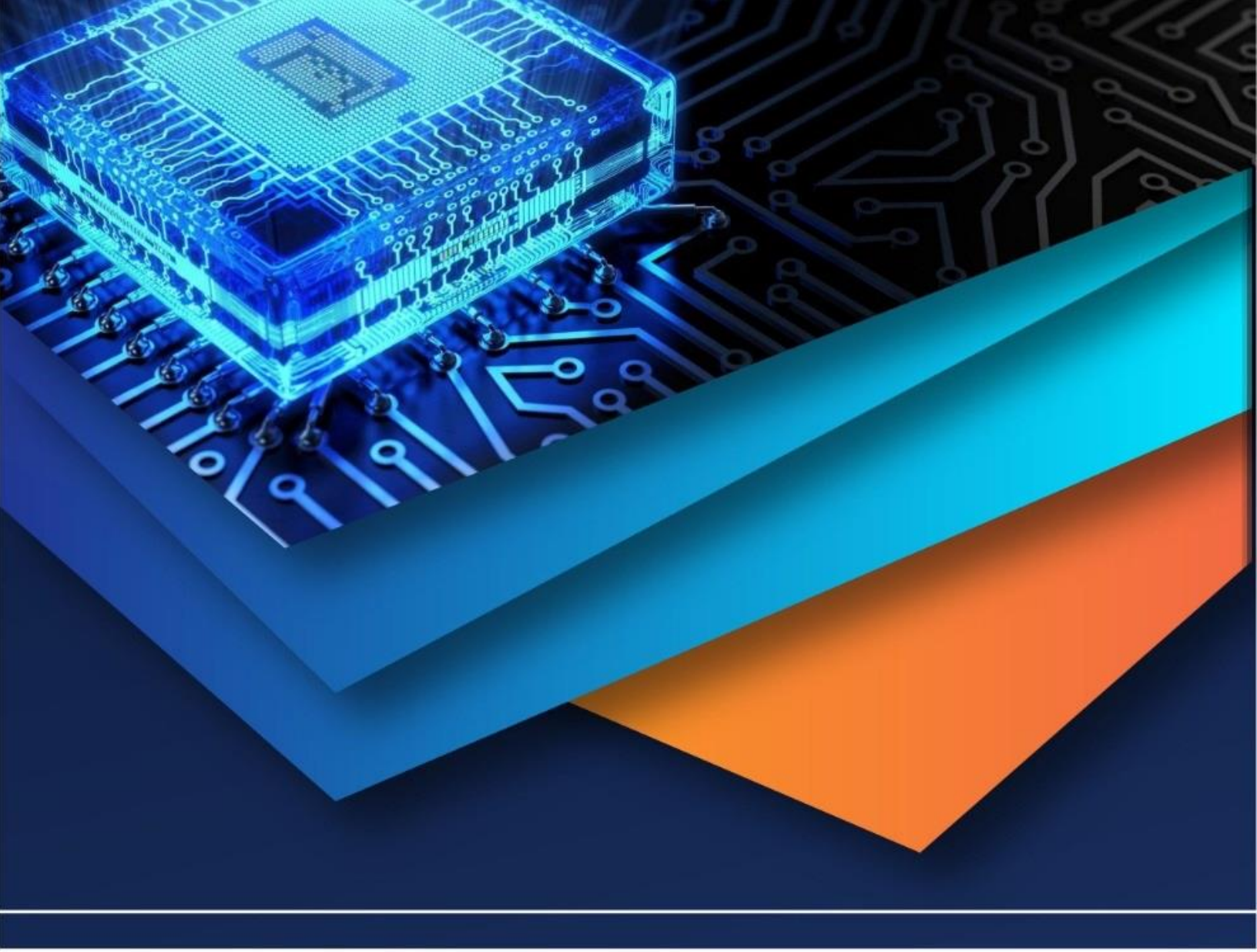
Fig. 2. Comparison of ML techniques under study

V. CONCLUSION AND FUTURE SCOPE

The experimental observations indicate that both the techniques can be utilized as master framework for the Thyroid Disorder prediction. SVM was found to be more effective technique for the Thyroid Disorder prediction. The same study can be applied to other medical disorders. Also, this research can be expanded by incorporating other supervised machine learning methods, as well as clustering techniques, in the diagnosis of thyroid disease.

REFERENCES

[1] Thyroid disorders in India: An epidemiological perspective, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169866/#ref1. [Accessed Online]
[2] Thyroid Disorder, https://www.lalpathlabs.com/blog/thyroid. [Accessed Online]
[3] S. Razia, P. S. Kumar and A. S. Rao, "Machine Learning Techniques for Thyroid Disease Diagnosis: A Systematic Review", In: Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough. Studies in Computational Intelligence, Springer, Cham, vol. 885, 2020. https://doi.org/10.1007/978-3-030-38445-6_15
[4] A. Soni and R. N. Giri, "Diagnosis of Thyroid Disease using Some Machine Learning Technique & Comparatively Analysis", International Journal of Creative Research Thoughts, vol. 6, no. 2, 2018.
[5] Begum and A. Parkavi, "Prediction of thyroid Disease using Data Mining Techniques", IEEE Journal, 2019.
[6] S. Raisinghani , R. Shamdasani , M. Motwani, A. Bahreja , and P. R. N. Lalitha, "Thyroid Prediction using Machine Learning Techniques", Springer Journal, pp. 140–150, 2019.
[7] K. Dharmarajan, K. Balasree, A. S. Arunachalam and K. Abirmai, "Thyroid Disease Classification using Decision Tree and SVM", Indian Journal of Public Health Research & Development, vol. 11, 2020.
[8] G. Chaubey, D. Bisen, and S. Arjaria, "Thyroid Disease Prediction using Machine Learning Approaches", National Academy Science Letters, 2020.
[9] Monkey Learn, Support Vector Machine –"Introduction to Support Vector Machines", https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/. [Accessed Online]
[10] L. Breiman and A. Cutler, "Random Forest", https://www.stat.berkeley.edu/~ breiman/RandomForests/cc_home.htm. [Accessed Online]
[11] http://archive.ics.uci.edu/ml/machine-learning-databases/thyroiddisease/. [Accessed Online]



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY