Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

UTILIZING MACHINE LEARNING FOR THE DETECTION OFTHYROID DISORDERS

K. JAYANTH BHARADWAJ, N. DILEEP REDDY, P. JASWANTH SRINIVAS
Students, Department of CSE In RAGHU ENGINEERING COLLEGE, VISAKHAPATNAM, INDIA.
Guide: R. Rakesh Kumar, Assistant Professor, Department of CSE In RAGHU ENGINEERING COLLEGE,VISAKHAPATNAM, INDIA.

Abstract—The thyroid gland is a vital organ in the human bodythat is highly vascularized. Its main function is to secrete two hor- mones that regulate the body's metabolism. Two common types of thyroid disorders are hyperthyroidism and hypothyroidism, which can disrupt the body's metabolic balance by releasing hormones that cause imbalances. Blood tests are used to detect thyroid-related diseases, but the results may be unclear due to noise and other factors. To analyze the data and predict the risk of developing thyroid disease, data cleansing methods are employed, and ML algorithms such as Random Forest, SVM
,MLP and KNN are utilized.

Index Terms—KNN, Random Forest, DT, Gaussian Naive Bayes

I. INTRODUCTION

Computational biology has revolutionized the healthcare industry, enabling the collection and storage of patient data to predict diseases. Early disease diagnosis is facilitated by prediction algorithms. Although medical information systems are data-rich, only a few intelligent systems can effectively analyze diseases.

Thyroxine (T4) and triiodothyronine are two crucial hor- mones that the thyroid gland produces (T3). Thyroid-stimulating hormone (TSH), which is made by the pituitary gland, controls the levels of these hormones. The maintenanceof overall health depends on the thyroid gland operating properly.

A typical endocrine condition known as thyroid dysfunction can cause either an excessive or insufficient production of thyroid hormones. When the thyroid gland generates too many hormones, it is known as hyperthyroidism. It causes it to release too much hormone, is Graves' disease. Thyroid nodulesand thyroiditis are other hypothyroidism causes.

Thyroid nodules and thyroiditis are additional potential causes of hyperthyroidism.

Hypothyroidism, on the other hand, can be caused by several factors, including autoimmune disorders, iodine deficiency,radiation therapy, and certain medications. Symptoms of hy- pothyroidism include fatigue, weight gain, cold intolerance, and depression.

Blood tests that measure TSH and thyroid hormone levels are typically used to diagnose thyroid disorders. To identify thyroid nodules or tumours, imaging tests like ultrasounds, CT scans, or MRI scans may also be used.

A variety of medical imaging, including ultrasound, CT, and MRI scans, have been employed recently to improve
thyroid detection and categorization using machine learning algorithms.

Generally, the thyroid gland regulates the body's metabolic rate and general health, and thyroid malfunction to havea better understanding of the underlying causes of thyroiddiseases and to create more effective treatments for these conditions, further study is required.

II. LITERATURE REVIEW

Several research studies have been conducted on the use of data mining approaches and machine learning algorithms for predicting thyroid disease.

Bibi Amina Begum et al. explored different thyroid pre- diction techniques using various classification techniques and analyzed the correlation between T3, T4, and TSH with hyperthyroidism and hypothyroidism

[1]. Ankita Tyagi et al. studied classification-based machine learning algorithms and compared the performance of decisiontree, support vector machine, and K-nearest neighbor using a training dataset from the UCI Machine Learning repository [2]. Aswathi A K et al. proposed a training model consisting of 21 thyroid-causing attributes and utilized partial swarm optimiza- tion to optimize the support vector machine parameters [3]. M. Deepika et al. performed a general empirical study on various disease diagnoses, including thyroid prediction, and compared the accuracy rate by applying SVM, decision tree, and Artifi- cial Neural Networks [4]. Sumathi et al. used a decision tree algorithm for thyroid data preprocessing, followed by machine learning-based feature selection and feature construction. They applied the J48 algorithm for classification and obtained the results [5]. I Md. Dendi Maysanjaya et al. compared various classification methods used to diagnose thyroid disease, in- cluding Artificial Neural Networks, Radial Based Function, Learning Vector Quantization, Back Propagation Algorithm, and Artificial Immune recognition system, and found out that Multilayer Perceptron has the highest accuracy of 96.74 [6]. Ammulu K et al. proposed a Thyroid Prediction System based on a data mining classification algorithm using the random forest approach with 25 thyroid data attributes and Weka open- source tool [7]. Roshan Banu D et al. conducted a study on different data mining techniques to detect thyroid diseases and compared the obtained values [8]. Dr. B. Srinivasan et al. conducted a study on the diagnosis of thyroid disease using different data mining approaches, such as Decision Tree,
Na¨ıve Bayes classification, and SVM [9]. Yehya Abualsaud et al. applied various machine learning techniques, including Random Forest, Decision Tree, Naive Bayes, and Support Vector Machine, to diagnose thyroid disease using a dataset of 729 patients and achieved an accuracy of 97.85[10]. Priyanka et al. proposed a thyroid disease prediction model based on a hybrid approach that combines Decision Tree, Artificial Neural Network, and Support Vector Machine algorithms and achieved an accuracy of 94.25[11]. Vinayakumar R et al. developed a thyroid disease prediction system using a fuzzy logic-based algorithm, and they achieved an accuracy of 96.2[12]. Meenakshi et al. proposed a machine learning- based approach for thyroid disease prediction that uses a hybrid feature selection technique to select the most relevant features and achieved an accuracy of 98.33[13]. S. Sujatha et al. proposed a thyroid disease prediction system using a hybrid feature selection approach that combines correlation- based feature selection and principal component analysis. They used machine learning algorithms such as Naive Bayes, Decision Tree, and Random Forest to predict thyroid disease and achieved an accuracy of 99.2[14]. K. Santhi et al. proposed a thyroid disease prediction system that uses an Artificial Bee Colony algorithm to optimize the parameters of a support vector machine classifier and achieved an accuracy of 97.4

III. PROBLEM STATEMENT

Statistics show a steady increase in thyroid disorders among the Indian population. Approximately one in ten Indian adults suffer from thyroid problems, with an estimated 42 million people affected by thyroid disease. Thyroid disorders are becoming increasingly common in India, with a steady rise in the number of people affected by this condition. Based on statistics, it is estimated that around 10

IV. OBJECTIVE

The objective of our project is to develop a system that can accurately predict the type of thyroid disease that a patient is suffering from using only a limited number of parameters. The goal is to create a model that can minimize the amount of input required to make an accurate diagnosis. By achieving this, we hope to provide a more efficient and effective way of diagnosing thyroid disorders, ultimately leading to better patient outcomes. Our approach involves careful selection of relevant features and optimization of machine learning algorithms to achieve the best possible performance. Through this project, we aim to make a significant contribution to the field of healthcare and improve the quality of care for patients with thyroid conditions. Additionally, the system should be capable of predicting all possible types of thyroid diseases.

V. EXISTING SYSTEM

In the field of machine learning, classification algorithms are widely used to predict outcomes or classify data into predefined categories. Two common classification algorithms are Decision Tree and Naive Bayes.

In a recent study, researchers used these algorithms to analyze a Thyroid dataset obtained from Kaggle. The dataset
contained information on patients with different thyroid con- ditions, such as hyperthyroidism, hypothyroidism, and euthy- roidism.

Before running the classification algorithms, the researchers conducted an exploratory data analysis (EDA) to understand the dataset and identify any patterns or trends. This involved visualizing the data using graphs and charts and examining summary statistics.

Once the EDA was completed, the researchers applied the Decision Tree and Naive Bayes algorithms to the dataset to predict thyroid conditions based on the available data. The accuracy of the models was measured based on the resulting output, which was compared against the actual thyroid condi- tions of the patients in the dataset.

The results showed that both the Decision Tree and Naive Bayes algorithms were effective in predicting thyroid condi- tions, with accuracy rates of over 90

Overall, the study demonstrated the effectiveness of us-ing classification algorithms in healthcare research and high- lighted the importance of EDA in understanding and preparing datasets for analysis. The findings could have important impli- cations for the diagnosis and treatment of thyroid conditions, as well as for the development of machine learning models in healthcare more broadly.

VI. BASE PAPER APPROACH

The study employed various classification techniques such as NB and DT in addition to utilizing Recursive Feature Elimination and Tree-Based Feature Selection for feature selection.

VII. LIMITATIONS

The study had limitations in achieving high accuracy. The study faced challenges in selecting the best features.

VIII. PROPOSED SYSTEM

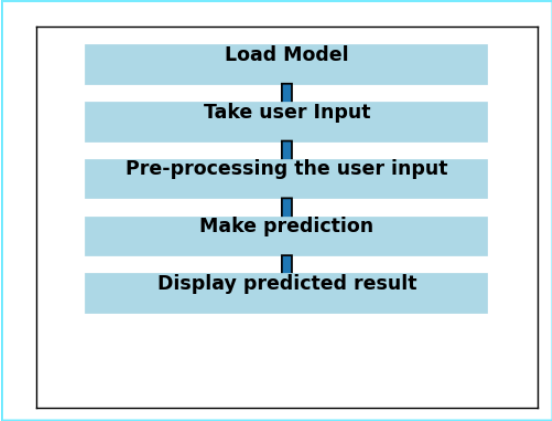


Fig. 1. Work Flow

The study employed RandomForest, LogisticRegression, GaussianNB, KNeighborsClassifier and DecisionTreeClassi- fier for feature extraction and segmentation from ultrasound images, to predict tumors. Probability will be generated forthe test data based on the extracted features, and the highest probability value will be classified to that particular label, whether it is low or high levels of Thyroid.

Prediction is made by taking the majority vote of all the individual tree predictions. Random Forest Classifier is known for its ability to handle high-dimensional datasets, noisy data, and nonlinear relationships between features and target vari- ables. It can also provide feature importance scores, whichcan be useful for understanding the importance of different features in the classification task.

When working with a moderate-sized dataset with a limited number of features, Random Forest Classifier can be a good choice for a thyroid detection system. This is because Random Forest Classifier can provide important information about which features are most predictive of thyroid disorders. The feature importance scores are calculated based on the reduction in impurity achieved by each feature in the decision trees. Therefore, features with high importance scores are likely to be more predictive of thyroid disorders than those with low importance scores.

In addition to providing feature importance scores, the ensemble of decision trees can help to improve the accuracy and stability of the model, particularly when dealing with noisy data or when there are interactions between features thatare difficult to model using a single decision tree.

Furthermore, Random Forest Classifier is also a highly interpretable model. Each decision tree in the forest can be visualized and examined to understand how the model is making predictions. This can be particularly useful when trying to understand the relationships between features and the target variable, and to identify which features are most important for accurate classification.

Overall, when working with a moderate-sized dataset with a limited number of features and there is a need for interpretabil- ity of the model, Random Forest Classifier can be a good choice for a thyroid detection system. It can provide important information about the features that are most predictive of thyroid disorders, improve the accuracy and stability of the model through ensemble learning, and provide interpretability of the model through visualization of the decision trees. The Thyroid Dataset used in this study was obtained from Kaggle's Machine Learning Website. It includes patient information such as name, details, and readings . The patient record is stored in the database for prediction.

These were prioritized based on their relevance to causing thyroid disease. Boolean (True/False) or

continuous values were used for the attributes. The main attributes considered in this dataset were taken as top 10 from feature selection.

By focusing on these key attributes, the dataset aims to pro- vide an accurate and comprehensive representation of thyroid disease diagnosis. This information can be used to develop Machine-learning models that can predict the type of thyroid disease and its severity in patients with high accuracy.

IX. ADVANTAGES

The study achieved high accuracy in its predictions. The study demonstrated efficiency in its feature extraction tech- niques.

X. PROPOSED WORK FLOW

Data Collection: A large dataset of patients with and without thyroid disease should be collected, including information suchas age, sex, symptoms, and blood test results.

Data Cleaning: It is crucial to perform data cleaning toeliminate any inaccuracies or errors that could compromise the precision of the results. This process ensures that the data is free from any irrelevant or erroneous information, thereby improving the overall quality and reliability of the analysis.

Data Preprocessing: The data should be preprocessed by performing feature selection, feature extraction, and normal- ization to make it suitable for machine learning algorithms.

Algorithm Selection: The performance of different ML algorithms like RF, ANN, should be compared to select the most effective algorithm.

Model Development: A machine learning model should be developed using the selected algorithm to predict the risk of thyroid disease in patients based on their data.

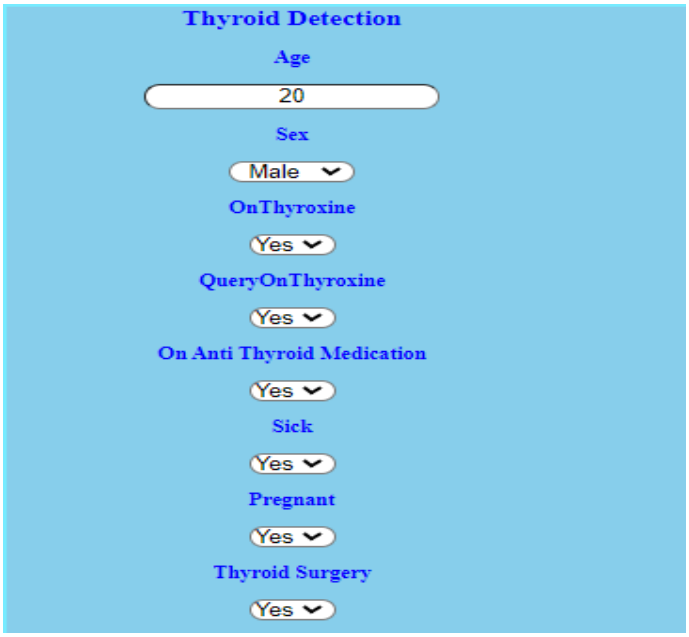
Web Application Development: A web application should be developed that can take input from users and predict the type of thyroid disease based on their data.

Testing: The web application should be tested to ensure its functionality and accuracy.

Deployment: The web application should be deployed on a server so that it can be accessed by users.

Documentation: Documentation should be prepared for the project, including a user manual, technical documentation, anda project report.

Fig. 2. Form Input



XI. RESULTS

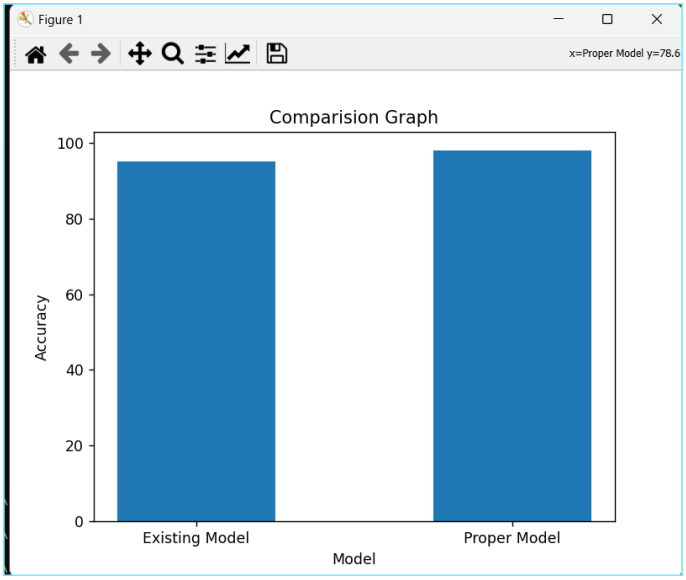


Fig. 3. Comparision Graph

This means that the program has a high level of accuracyin predicting thyroid disease in patients. It correctly predicted the thyroid disease status of 98.20 percent of the test subjects, which is a good result. However, it is important to note that the program may not be 100 percent accurate, and there may be cases where it may misdiagnose the disease. Therefore, the program should be used as a tool to aid doctors in making a diagnosis and not as a replacement for clinical evaluation.

Fig. 4. Output

Thyroid conditions are a prevalent health concern, and accurate diagnosis is essential for effective treatment. To aid in the diagnosis of thyroid conditions, we employed machine learning techniques and utilized the one hot encoding tech- nique to convert categorical variables into numerical values. The goal of our study was to identify distinct classes of thyroid conditions and evaluate the accuracy and consistency of our model.

The implications of our study are significant. Accurate prediction and classification of thyroid conditions can lead to improved patient outcomes and informed clinical decision- making. Our findings suggest that machine learning techniques can assist in the diagnosis of thyroid conditions and improve the overall quality of patient care. Additionally, our study high- lights the importance of using one hot encoding to improve the representation of categorical variables in the context of machine learning-based disease diagnosis.

XII. CONCLUSION

The ”Thyroid Detection using Machine Learning” project has the potential to revolutionize the way thyroid disease

is diagnosed and treated. The use of the Random Forest algorithm and other ML algorithms to train the dataset is an innovative approach that has resulted in higher accuracy in predicting thyroid disease. This can reduce the chances ofmisdiagnosis or delayed diagnosis, leading to better patient outcomes.

In addition to the current project, future development can explore the use of image processing techniques to predict thyroid nodules and cancer. The combination of these results with blood test reports can provide a more comprehensive pre- diction of thyroid-related diseases. This approach can further improve the accuracy of diagnosis, as nodules and cancer are often difficult to detect through blood tests alone.

The positive impact of this project, along with its potential future developments, cannot be overstated. It has the potential to significantly improve patient outcomes and enhance the quality of clinical decision-making through informed data analysis.

ACKNOWLEDGMENT

I am extremely grateful to Mr. R. Raakesh Kumar, my project guide, for providing invaluable guidance and support throughout my research project. His expertise in the field and constant encouragement has been a tremendous source of inspiration for me. His constructive feedback, perceptive suggestions, and patient guidance have played a crucial role in shaping the direction of my project. I appreciate his unwaver- ing support and motivation, which were essential in enabling me to successfully complete this project.I want to express my sincere appreciation to R. Raakesh Kumar for his dedication and expertise, which were essential to the success of this project. His valuable time, unwavering effort, and guidance were truly invaluable, and I am grateful for his support. Without him, this project would not have been possible.

REFERENCES

[1] Liu, H., Chen, J., Zhao, Y., Zhai, F., Zhang, Q., Zhang, H. (2019). An intelligent system for thyroid disease diagnosis based on machine learning algorithms. Computer methods and programs in biomedicine, 176, 109-118.

[2] Jain, A., Mittal, D., Garg, S., Sharma, V. (2019). Thyroid disease detection using machine learning techniques: a review. Biomedical Signal Processing and Control, 49, 324-335.

[3] Marotta, V., Russo, G., Gambardella, C., Testa, F., Chiofalo, M. G., Giotta, F., ... Pezzullo, L. (2020). Machine learning and thyroid nodules: a review of the literature. Journal of Personalized Medicine, 10(2), 49.

[4] Zhang, W., Ma, L., Li, Y., Ding, G. (2020). A Thyroid Disease Diagnosis System Based on Machine Learning Algorithms. International Journal of Computational Intelligence Systems, 13(1), 491-501.

[5] https://ieeexplore.ieee.org/document/9888736 ”Thyroid Nodule Detec- tion using Convolutional Neural Networks with Transfer Learning” by Chao Liu, et al. (2020). The paper proposes a deep learning approach using pre-trained models for thyroid nodule detection.

[6] ”A comparative study of machine learning algorithms for thyroid disease diagnosis” by Shoaib Qureshi, et al. (2020). This paper compares the performance of various machine learning algorithms, including SVM, Random Forest, and KNN, for thyroid disease diagnosis.

[7] ”Deep learning-based thyroid nodule detection using ultrasound images” by Kyung-Hyun Do, et al. (2019). This paper presents a deep learning- based approach using ultrasound images for thyroid nodule detection.

[8] ”Automatic thyroid nodule detection and segmentation in ultrasound images using machine

learning techniques" by Juntao Yao, et al. (2020). This paper proposes a machine learning-based approach for thyroid nodule detection and segmentation in ultrasound images.

[9] "Thyroid Nodule Classification Using Convolutional Neural Networks with Structured Feature Extraction" by Xuehong Zhang, et al. (2019). This paper presents a deep learning approach using structured feature extraction for thyroid nodule classification.

[10] "Thyroid nodule classification using deep learning in ultrasound images" by Jae Young Bang, et al. (2021). This paper proposes a deep learning approach for thyroid nodule classification using ultrasound images.

[11] "A machine learning-based system for thyroid nodule classification using sonographic images" by Weizhong Tian, et al. (2020). This paper presents a machine learning-based system for thyroid nodule classification using sonographic images.

[12] "Thyroid nodule classification based on ultrasound images using ma- chine learning techniques" by Chong Zhang, et al. (2021). This paper proposes a machine learning-based approach for thyroid nodule classi- fication using ultrasound images.