IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

Machine Learning Models for Predicting Hypothyroidism: Utilizing Synthetic Data for Improved Accuracy

Vinutha N Sajjan[1], Varsha S[2], Sheela S[3]
Undergraduate Students, Department of Computer Science and Technology[1,2]
Assistant Professor, Department of Computer Science and Technology[3]
Global Academy of Technology, Bengaluru, India

Abstract: *This study presents a novel approach to early hypothyroidism detection by integrating synthetic data generation with machine learning (ML) techniques. Facing the challenge of limited and imbalanced healthcare datasets, we employ synthetic data to augment training sets, ensuring a richer and more diverse data pool for ML application. Key indicators of early hypothyroidism are distilled through feature selection, optimizing ML model inputs. We test various ML classifiers, including Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBM), demonstrating enhanced diagnostic accuracy with our approach. Initial outcomes suggest that combining synthetic data with ML significantly boosts early detection capabilities, offering a promising direction for overcoming traditional data scarcity in medical diagnostics.*

Keywords: Decision tree, Hypothyroidism Prediction, Machine learning, NB, Random Forest, SVM

I. INTRODUCTION

Thyroid diseases, which include a broad range of conditions impacting the thyroid gland, are among the most common endocrine disorders worldwide. They can take on different forms, the most frequently encountered of which are hypothyroidism and hyperthyroidism, which are defined, respectively, by the underproduction and overproduction of thyroid hormones. Given that these hormones are necessary to regulate growth, development, and metabolism, any imbalance can result in a variety of health problems. Thyroid diseases must be identified early to be able to improve patient outcomes and avoid serious complications.

However, there are numerous challenges for early detection of thyroid disorders.. Thyroid disease frequently presents with mild, ambiguous initial symptoms that that are prone to error for those of other illnesses. Thyroid diseases may not always be accurately detected in their early stages utilizing traditional diagnostic methods, which are mainly based on clinical examination and hormone level assessments. Consequently, there is a rising need for more sophisticated and accurate diagnostic instruments.

In this regard, machine learning (ML) shows itself to be a potent ally in the fight against thyroid conditions. ML algorithms present a promising way to improve early detection and diagnosis owing to their ability to examine and decipher complicated patterns in sizable datasets. Large volumes of data, such as patient demographics, clinical symptoms, biochemical test results, and imaging data, can be processed by these algorithms to ascertain potential thyroid disease indicators that conventional approaches might miss.

This article's primary focus is on the application of machine learning techniques in the early diagnosis of thyroid disorders..We look at how thyroid disease diagnostics have evolved., the shortcomings of existing methods, and machine learning's capacity to get past these barriers. Our goal is to build predictive framework that can more precisely identify people who early on are susceptible to thyroid disorders by integrating cutting-edge ML models into the diagnostic process. These developments could lead to tailored treatment regimens and improve clinical decision-making, which would ultimately raise the bar for treating patients with thyroid disorders.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

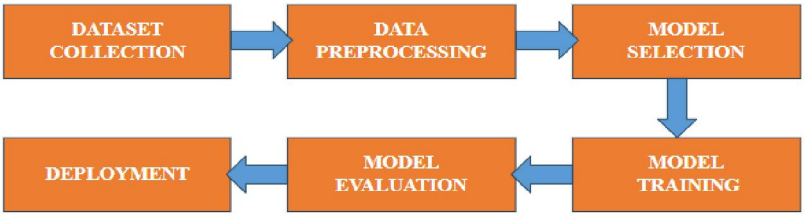

Figure 1: Working

II. LITERATURE REVIEW

According to Pan, Qiao, Yuanyuan Zhanget al. [1],The paper introduces a new method for classifying thyroid diseases using a random forest-based ensemble classification algorithm. The motivation is the increasing prevalence of thyroid diseases and the need for accurate classification in clinical treatment. The proposed method utilizes Principal Component Analysis (PCA) for feature selection, and Rotation Transformation to enhance the diversity of base classifiers, addressing the accuracy-diversity dilemma. The algorithm is compared with Bagging, Random Forest, and AdaBoost, showing superior performance on UCI machine learning repository dataset and a more complex real clinical medical dataset.The classification method involves three stages: data preprocessing, feature dimension reduction using PCA, and classification/prediction using a C4.5 decision tree as the base classifier and random forest ensemble algorithm. Experimental results demonstrate the potency of the proposed method, achieving high accuracy on both standard and real clinical datasets. The paper concludes by discussing future work, including considering additional factors for thyroid diagnosis and addressing imbalanced classification issues. The research is supported by grants from Shanghai Science and Technology Innovation Action Plan.Key components of the proposed method include the use of random forest, PCA for dimension reduction, and Rotation Transformation to enhance classifier diversity. The experimental results highlight the method's superiority over existing techniques, emphasizing its potential for accurate and efficient thyroid disease classification.

The statement highlighted byPavya, K., and B. Srinivasanet al.[2], discusses the importance of early and accurate diagnosis of thyroid diseases in the healthcare industry, especially considering their symptoms overlapping with other conditions. It focuses on second step of an Automatic Thyroid Disease Computer Aided Diagnosis (ATD-CAD) system, which involves feature selection to improve disease detection accuracy.The study compares the effectiveness of two feature selection methods: filter-based (using F-Score) and wrapper-based (using Recursive Feature Elimination) algorithms. Principal Component Analysis (PCA) is also evaluated for dimensionality reduction. Four classifiers (MultiLayer Perceptron, SVM, Back Propagation Neural Network and Extreme Learning Machine) are utilized to analyze the selected algorithms.Experimental results from a dataset of 215 patients with thyroid disease show that both F-Score and Recursive Feature Elimination algorithms enhance disease diagnosis performance. Among them, the wrapper-based approach shows the highest efficiency, achieving a maximum accuracywith the Extreme Learning Machine classifier. The study concludes that integrating filter- and wrapper-based feature selection algorithms improves the display of thyroid disease diagnosis systems and suggests future research directions in enhancing feature selection and classifier performance.

This research paper by Tyagi, Ankita, Ritika Mehra, and Aditya Saxena.et al.[3], focuses on utilizing machine learning algorithms for predicting thyroid disease on data obtained from the machine learning repository at UCI. Thyroid diseases, such as hyperthyroidism and hypothyroidism, are challenging to diagnose, and the study employs data cleansing techniques to prepare the data for analysis.The paper discusses the importance of thyroid hormones in controlling metabolism and highlights the significance of accurate disease prediction to avoid unnecessary treatments. Numerous machine learning techniques, such as Support Vector Machine (SVM), KNN, and Decision Trees, are utilized to predict the risk of patients developing thyroid disease.The authors propose a hybrid model for solving complex medical diagnosis tasks and provide an overview of the architecture of the thyroid prediction system. The methodology involves supervised learning, and attributes such as age, sex, and thyroid disease diagnosis is based on thyroid hormone levels..The performance study of the proposed algorithms reveals high accuracy for SVM and KNN, indicating their effectiveness in predicting thyroid disease. The paper concludes with the suggestion to develop

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

algorithms that require a minimal number of parameters for diagnosing thyroid diseases, aiming for cost-effective and time-efficient solutions. The authors express gratitude to DIT University for research support.

The authors Shahid, Afzal Hussainet al. [4],focuses comparing Random Forest (RF), SVM, and K-Nearest Neighbors (K-NN). The thyroid gland produces hormones crucial for regulating metabolism. Hyperthyroidism and hypothyroidism occur when there is an excess or deficiency of thyroid hormones. The study aims to assess the efficiency of RF, SVM, and K-NN classifiers in diagnosing thyroid disease based on attributes such as TSH, TT4, T3, FT4, and T4U.his study focuses on the diagnosis of thyroid disease using machine learning classifiers. The dataset used is obtained from machine learning repository at UCI and includes 7200 instances with 21 attributes, categorized into hyperthyroidism, hypothyroidism, and normal subjects.The report offers information on each class's performance metrics. (hyperthyroidism, hypothyroidism, normal subjects), including precision, recall, specificity, accuracy, and F1-score.The dataset's imbalanced nature is addressed by splitting it into training and test datasets (80/20 ratio). For best outcomes, specific parameters are adjusted for every classifier.. The study concludes that RF demonstrates superior performance in diagnosing thyroid diseases compared to SVM and K-NN. The study makes recommendations for possible uses of machine learning in medical diagnosis and creates new research directions, including the evaluation of metaheuristic algorithms on this dataset.

The authors Rao, Amulya R.et al. [5],explains how classification-based machine learning techniques are used to diagnose thyroid conditions, highlighting the significance of early detection for successful treatment. Traditional methods involve thorough inspection and blood tests, but the study proposes a predictive model using DT ID3 and Naive Bayes algorithms. The dataset is obtained from Kaggle and includes attributes like age, gender, T3, T4, and TSH values.The thyroid gland's role in maintaining metabolism and overall health is highlighted, and the paper addresses the prevalence of thyroid disorders in India. Model aims to predict thyroid disease at early stages, contributing to timely and accurate diagnosis.The dataset includes patient information and past clinical history, aiding accurate predictions. The proposed model suggests that Decision Tree's output triggers Naive Bayes for stage prediction. The study claims an accuracy rate of 95% with over 3000 training dataset attributes.The conclusion highlights the potential benefits of the proposed model in early identification of thyroid diseases, reducing redundant data, and simplifying treatment.

The authors Mir,Yashir et al. [6],used variety of performance metrics, including accuracy, specificity, sensitivity, precision, recall, and receiver optimization curve (ROC), to evaluate the classifiers' performance.The primary dataset was collected from Sawai Man Singh (SMS) hospital in Jaipur, India, through a pre-designed questionnaire. It consisted of 1464 instances with 21 attributes, including both pathological and serological parameters.The study highlighted the limitations of using the UCI dataset, emphasizing its outdated nature and the presence of missing values. The authors proposed a new dataset that includes both pathological observations and serological tests from real-world patients.The experiments were conducted using WEKA, and a 10-fold cross-validation method was employed to ensure fair and reliable results.The results suggest that different machine learning algorithms may perform better under specific experimental conditions. Bagging and SVM showed promising results in the overall prediction, while SVM excelled in predicting TD based on pathological observations, and J48 performed well with serological parameters.

The authors Riajuliislam, Md, Khandakar Zahidur Rahim et al. [7],addresses the prevalence of thyroid disease, particularly hypothyroidism, among females in Bangladesh, emphasizing the need for early detection to prevent it from becoming a critical health issue. The authors employ machine learning techniques to predict hypothyroidism in the primary stage, utilizing feature selection methods such as Recursive Feature Selection (RFE), Univariate Feature Selection (UFS), and Principal Component Analysis (PCA). Classification algorithms including Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), and Naive Bayes (NB) are applied to evaluate the effectiveness of the feature selection techniques.The research is based on a dataset collected from a diagnostic center in Dhaka, Bangladesh, comprising 519 instances with nine attributes. The authors highlight the gender-based susceptibility to thyroid disease, with females at a higher risk, and note the lack of awareness among a significant portion of the affected population. The thyroid gland's role in regulating various bodily functions is discussed, with a focus on the hormones T3 and T4.The conclusion highlights that, when compared to UFS and PCA, RFE exhibits superior accuracy and is most operational feature selection technique for predicting hypothyroidism.. The study acknowledges limitations in data collection due to pandemic and expresses the intention to explore larger datasets in the future. Overall, the research contributes to the understanding and predicting thyroid diseases in the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

context of Bangladesh, particularly among females.All things considered, the research offers a comprehensive examination of various models with an emphasis on precision and effectiveness in the classification of hypothyroidism, which contributes significantly to medical diagnostics.

The authorsChaubey, Gyanendra, Dhananjay Bisen et al. [8],compares the execution of three widely used algorithms: logistic regression, decision trees, and k-nearest neighbor (KNN). The research employs a dataset from the UC Irvine Knowledge Discovery in Databases archive, specifically the "new-thyroid" dataset containing attributes like total serum thyroxin (T4) and total serum triiodothyronine (T3).The introduction emphasizes the prevalence of thyroid disease in India, particularly among women aged 17–54. The paper highlights the importance of early and accurate diagnosis, and it suggests that ML methods can provide valuable insights into causes, affected age groups, and relevant treatments.The three classification algorithms (logistic regression, decision trees, and KNN) are studied in detail, outlining the steps involved in training, validation, and testing for each method. Logistic regression is characterized by a linear decision boundary and a sigmoid function, while decision trees involve nodes and leaves, with a top-down approach to develop the tree. KNN is described as a lazy learner that classifies test tuples based on the majority class of their k-nearest neighbors.The research methodology involves using the "new-thyroid" dataset, dividing it into training, validation, and test sets, and applying each algorithm. The results and analysis section presents the visualization of the training dataset and evaluates the performance of each algorithm. The performance of the models is measured using confusion matrices.The paper concludes that KNN demonstrates superior performance for this dataset in predicting thyroid disease. The accuracy comparison with previous works shows that the proposed algorithms outperform some studies but may be outperformed by others based on different datasets.The conclusion also suggests future work, considering larger datasets and exploring additional machine learning and deep learning algorithms such as SVM, Naïve Bayes, autoencoders, ANNs, and CNNs for improved results in thyroid disease prediction.

The authors Chaganti, Rajasekhar, Furqan Rustam et al. [9],focuses on advancing thyroid disease prediction by explaining a novel methodology that emphasizes feature selection for both ML and deep learning models. The study addresses limitations found in existing approaches, such as binary classification, small datasets, and insufficient validation. Through a comprehensive methodology, incorporating Forward Feature Selection (FFS), Backward Feature Elimination (BFE), Bi-Directional Feature Elimination (BiDFE), and Machine Learning Feature Selection (MLFS) using extra tree classifiers, the authors aim to predict five distinct thyroid diseases.The evaluation involves assessing the accuracy of variety machine learning models (Random Forest, Logistic Regression, Support Vector Machine, AdaBoost, and Gradient Boosting Machine) and deep learning models (Convolutional Neural Network, Long Short-Term Memory Network, and CNN-LSTM). The findings show that in terms of accuracy and computational complexity, machine learning models perform more effectively than deep learning models.. Notably, the Random Forest model, when combined with MLFS, achieves a remarkable accuracy of 0.99.To systematically address the limitations of existing thyroid disease prediction studies, the authors utilize a well-known UCI dataset, employing data balancing and feature selection. The study provides an in-depth performance comparison of machine learning and deep learning models, contributing to the improvement of healthcare practices for thyroid disease detection. Emphasizing the significance of advanced technologies and feature engineering, the research highlights the superiority of specific feature selection techniques and ML models in optimizing thyroid disease prediction.

The writers Guleria, Kalpna, Shagun Sharmaet al. [10],addresses the significant prevalence of thyroid disorders in India, the study evaluates various MLmodels, such as DT, random forests, naive Bayes, and multiclass classifiers, along with a deep learning-based Artificial Neural Network (ANN), known for its proficiency in handling text data. Results indicate that decision trees and random forests achieve the highest accuracyin predicting hypothyroidism. The deep learning-based ANN also demonstrates competitive accuracy. The study concludes by highlighting the potential impact of these models on hypothyroidism detection, early diagnosis, and subsequent reduction in mortality rates. Additionally, it proposes future research avenues involving the expansion of datasets and exploration of additional machine learning or deep learning-based prediction models for enhanced hypothyroidism detection. Overall, the research makes a valuable contribution to medical diagnostics, offering a thorough analysis of diverse models with a focus on accuracy and efficiency in hypothyroidism classification.The method's superiority over current techniques is further demonstrated by the experimental results, illustrating its potential for reliable and efficient thyroid disease classification.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

The authors Devi, Munisamy Shyamala et al. [11],discusses aboutthe Blunge Calibration Intelligent Feature Classification Model, for the early prediction of hypothyroid disease. With collaborative efforts from institutions in India, Romania, and Pakistan, the study addresses the significant prevalence of abnormal thyroid functioning, affecting 12% of the Indian population. The primary focus is on overcoming the challenges of early detection when symptoms may not be apparent, emphasizing the critical role of timely intervention in preventing health issues. The study utilizes a dataset of 3163 patient details obtained fromUCI machine-learning repository, employing preprocessing techniques, ANOVA analysis, and various classification algorithms to enhance prediction accuracy.To optimize accuracy, the research employs feature-selection algorithms for dimensionality reduction. Two distinct models, the Blunge Calibration Regression Model (BCRM) and Blunge Calibration Classifier Model (BCCM), are developed, achieving impressive accuracies, in predicting hypothyroidism. These models incorporate soft blending based on the sum of predicted probabilities of classifiers, showcasing their uniqueness in updating estimators at runtime. The implementation setup involves Python running on an NVidia Tesla V100 GPU server, with 30 training epochs and a batch size of 64.The study concludes by highlighting the significance of the BCCM and BCRM models in efficient early prediction of hypothyroidism, contributing to the field of machine learning in medical diagnosis. The research article is published under an open-access license, emphasizing its accessibility to the scientific community. The comprehensive literature review covers various aspects of machine learning applications in medical diagnosis, underscoring the importance of accurate predictions in healthcare scenarios. Overall, the study provides valuable insights into the prediction of hypothyroidism, offering efficient models that can potentially revolutionize early diagnosis in medical practice.

The authors Hossain, Md Bipul, Anika Shama et al. [12],focuses on predicting thyroid diseases, specifically hypothyroidism and hyperthyroidism, using ML algorithms. To create a predictive model, the study uses data gathered from multiple medical sectors and various machine learning algorithms.. The goal is to identify the most significant features that can aid in detecting thyroid diseases more accurately. The authors use algorithms such as Decision Tree, Random Forest, Gradient Boosting, Naive Bayes, Logistic Regression, K-Nearest Neighbor, and Support Vector Machine to predict thyroid diseases. Results indicate that Random Forest provides the highest evaluation score across all sectors in the dataset, outperforming other algorithms such as Naive Bayes. The findings suggest that machine learning, particularly Random Forest, coupled with feature selection, can be effective in detecting and diagnosing thyroid diseases. The study concludes by emphasizing the potential impact of the research on the scientific community and healthcare professionals, providing insights for real-time applications in thyroid disease analysis. The use of explainable artificial intelligence (XAI) is also advocated for better understanding and accountability of machine learning models in healthcare. The provided text discusses the evaluation of machine learning models for predicting thyroid disease. The evaluation is based on various performance metrics such as accuracy, precision, recall, and F1-score. The text outlines the preprocessing steps for handling missing values, converting categorical attributes, and balancing the dataset.The evaluation involves the comparison of seven different machine learning algorithms: Decision Tree Classifier, Random Forest Classifier, Naive Bayes Classifier, Gradient Boosting Classifier, Logistic Regression Classifier, K-Nearest Neighbor, and Support Vector Machine. The analysis is performed on three different datasets: one with all attributes, one with 14 features selected using feature importance, and one with 14 features selected using univariate feature selection.

According to the study, the authors Sa, Ri, Taiyu Yang et al. [13], aims to develop machine-learning models predicting treatment response to radioiodine (131I) therapy and thyrotropin (TSH) suppression therapy in patients with differentiated thyroid cancer (DTC) without structural disease based on pre-treatment information. The study includes 597 and 326 patients randomly assigned to training cohorts for predicting 131I therapy and TSH suppression therapy responses, respectively. Six supervised algorithms are employed, with Random Forest (RF) showing the best performance. Stimulated and suppressed thyroglobulin (Tg) and radioiodine uptake contribute to 131I therapy response, while thyroid remnant and TSH are crucial for Tg decline under TSH suppression therapy. RF demonstrates superior performance in predicting effective response (ER) to 131I therapy and biochemical remission (BR) to TSH suppression therapy. Factors such as sex, TNM-N, risk stratification, RAIU%, Tgon, TSHoff, Tgoff, and courses of 131I therapy before the current course significantly influence non-ER to 131I therapy. This study highlights the effectiveness of machine learning models, particularly the RF algorithm, in predicting treatment responses to 131I therapy and TSH

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

suppression therapy in DTC patients without structural disease. The integration of pre-treatment variables enhances prediction accuracy, offering valuable insights for decision-making in patient managementtherapy and TSH suppression therapy, such as genetic factors or detailed histopathological features, was not considered in this study. Additionally, the study might benefit from a more diverse population to enhance the generalizability of the machine learning models.

The writers Naeem, Awad Bin et al. [14],focuses on hypothyroidism, emphasizing the importance of accurate diagnosis for effective treatment. It explores the application of ensemble approaches in medical datasets to improve diagnostic accuracy. Three machine learning algorithms—KNN, Naive Bayes, and SVM—are employed and compared. The study uses a dataset from Kaggle with 33 attributes related to hypothyroidism

symptoms.The introduction provides background on thyroid disorders, their symptoms, and the significance of early diagnosis. It highlights the role of machine learning algorithms in classification and the integration of IoT and AI in healthcare. The interdisciplinary nature of data science and its contribution to

predictive analyses are discussed.The methodology involves data collection from Kaggle, population sampling, and the use of RapidMiner with KNN, Naive Bayes, and SVM. The dataset includes patient information and symptoms. The research aims to assess the ratio of hypothyroidism cases, improve public health policies, and compare model findings.The results indicate that SVM outperforms KNN and Naive Bayes with an accuracy of 84.72%. The study emphasizes the importance of accurate diagnosis for hypothyroidism and suggests potential benefits for medical practitioners and researchers.In conclusion, the research suggests that SVM is more effective in diagnosing hypothyroidism based on the dataset used. Future work could involve exploring new models like deep learning and fuzzy learning to further enhance accuracy. The study was a collaborative effort with contributions from multiple authors, each involved in different aspects of the research.

The authors Akash, K. Thirumala et al. [15],delves into the critical realm of thyroid disorders, underscoring the pivotal significance of early detection to ameliorate outcomes. It accentuates the integration of artificial intelligence (AI) and machine learning (ML), particularly through Convolutional Neural Networks (CNNs), as a potent avenue for enhancing the precision and efficiency of early thyroid disorder detection.The various applications of AI in thyroid disorder detection are elucidated, spanning image analysis, blood test analysis, symptom analysis, and natural language processing. An essential aspect highlighted is the significance of feature selection, acknowledging its role in refining predictive models by discerningly choosing pertinent features.The conclusion resonates with the ongoing evolution of AI in thyroid disorder detection, advocating for further research to delineate its efficacy. In essence, the abstract provides a comprehensive exploration of early detection's paramountcy, AI's pivotal role, and multifaceted facets concerning thyroid disorders.All things considered, the research offers insightful information about hypothyroidism prediction, along with effective models that have the potential to completely transform medical practice's use of early diagnosis

Table 1: Literature Review Overview: Papers, Methodologies, and Results

SI. NO	Title of the Paper and Authors	Year	Methodology/Algorithm	Results
1	Improved ensemble classification method of thyroid disease based on random forest. Pan, Qiao, Yuanyuan Zhang, Min Zuo, Lan Xiang, and Dehua Chen	2016	random forest, using Principal Component Analysis and Rotation Transformation	The suggested method's efficacy is demonstrated by the experimental results, which show high accuracy on both standard and real clinical datasets.
2	Feature selection algorithms to improve thyroid disease diagnosis. Pavya, K., and B. Srinivasan.	2017	Filter-based and wrapper-based feature selection algorithms, along with Principal Component Analysis	Recursive Feature Elimination and F-Score enhanced the ability to diagnose thyroid disease using a wrapper-based algorithm
3	Interactive thyroid disease prediction system using	2018	ML-KNN,SVM,DT. DL-ANN.	High accuracy for SVM and KNN is found. In order to find quick and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

	machine learning technique. Tyagi, Ankita, Ritika Mehra, and Aditya Saxena.			affordable solutions, the paper's conclusion recommends creating algorithms that diagnose thyroid disorders with the fewest possible parameters.
4	A study on label TSH, T3, T4U, TT4, FTI in hyperthyroidism and hypothyroidism using machine learning techniques. Shahid, Afzal Hussain, Maheshwari .	2019	RF, SVM, KNN	According to the study's findings, RF performs better than SVM and K-NN at diagnosing thyroid disorders.
5	A machine learning approach to predict thyroid disease at early stages of diagnosis. Rao, Amulya R., and B. S. Renuka.	2020	Decision Tree ID3 and Naïve Bayes algorithms.	SVM demonstrated exceptional performance in predicting TD based on pathological observations, while J48 demonstrated strong performance with serological parameters. Bagging and SVM yielded encouraging overall prediction results.
6	Thyroid disease prediction using hybrid machine learning techniques: An effective framework. Mir, Yasir Iqbal, and Sonu Mittal.	2020	NB, SVM, J48	Based on pathological observations, SVM was very good at predicting TD, and J48 did well when using serological parameters.
7	Prediction of thyroid disease (hypothyroid) in early stage using feature selection and classification techniques. Riajuliislam, Md, Khandakar	2021	Using feature selection techniques (RFE, UFS, PCA) and classification algorithms (SVM, DT, RF, LR, NB)	RFE feature selection methodology proved to be the most accurate of the four classification algorithms on a consistent basis.
8	Thyroid disease prediction using machine learning approaches. Chaubey, Gyanendra, Dhananjay Bisen, Siddharth Arjaria, and Vibhash Yadav.	2021	Logistic Regression, Decision Tree,KNN.	In order to achieve better results in thyroid disease prediction, the conclusion recommends future work that takes into account larger datasets and investigates additional machine learning and deep learning algorithms like SVM, Naïve Bayes, autoencoders, ANNs, and CNNs.
9	Thyroid disease prediction using selective features and machine learning techniques. Chaganti, Rajasekhar, Furqan Rustam, Isabel .	2022	RF, SVM, LR, ADA,GBM.	The findings show that in terms of accuracy and computational complexity, machine learning models perform better than deep learning models. Notably, MLFS and the Random Forest model together achieve remarkably high accuracy.
10	Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep	2022	ML-Naive Bayes, Decision Tree Random Forest, Multiclass Classifier DL-ANN	The results show that when it comes to predicting hypothyroidism, decision trees and random forests perform the best. Competitive

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

	learning. Guleria, Kalpna, Shagun Sharma, Sushil Kumar, and Sunita Tiwari.			accuracy is also demonstrated by the deep learning-based ANN.
11	A Novel Blunge Calibration Intelligent Feature Classification Model for the Prediction of Hypothyroid Disease. Devi, Munisamy Shyamala, Venkatesan Dhilip Kumar, Adrian Brezulianu, Oana Geman, and Muhammad Arif.	2023	Blunge Calibration Classifier Model and Blunge Calibration Regressor Model.	The study uses feature-selection algorithms for dimensionality reduction in order to maximize accuracy. The Blunge Calibration Regression Model (BCRM) and the Blunge Calibration Classifier Model (BCCM) are two separate models that are developed to predict hypothyroidism with impressive accuracy.
12	An Explainable Artificial Intelligence Framework for the Predictive Analysis of Hypo and Hyper Thyroidism Using Machine Learning Algorithms. Hossain, Md Bipul, Anika Shama.	2023	DT,RF,GB,NB,LR,SVM.	The findings suggest that machine learning, particularly Random Forest, coupled with feature selection, can be effective in detecting and diagnosing thyroid diseases
13	Random Forest for Predicting Treatment Response to Radioiodine and Thyrotropin Suppression Therapy in Patients With Differentiated Thyroid Cancer But Without Structural Disease. Sa, Ri, Taiyu Yang, Zexu Zhang, and Feng Guan.	2023	Random Forest.	It was discovered that the Random Forest model was a helpful tool for predicting treatment response in patients with differentiated thyroid cancer who do not have structural disease to radioiodine and thyroid hormone suppression therapy.
14	Hypothyroidism disease diagnosis by using machine learning algorithms. Naeem, Awad Bin, Biswaranjan Senapati.	2023	KNN, NB, SVM	Depending on the dataset used, the research indicates that SVM is more helpful in determining the hypothyroidism diagnosis . Future research may investigate novel models such as fuzzy and deep learning in an effort to increase accuracy even more.
15	Predicting Thyroid Dysfunction Using Machine Learning Techniques. Akash, K. Thirumala, F. Mohammed Usman, T. Nitesh Kumar.	2023	SVM, LR, DT, RF.	The conclusion supports further research to make clear how well AI detects thyroid disorders and is in line with the continuous development of this technology.

III. CONCLUSION

In recent years, the investigation of alternative machine learning techniques for the diagnosis of thyroid disease has revealed a variety of efficient and advanced tools.. Evidence from various studies underscores the superiority of neural networks in achieving higher accuracy rates compared to other methods. However, support vector machines and decision trees have also been recognized for their commendable performance. This surge in technological advancements has undeniably propelled the global medical community towards significant strides in detecting and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

managing thyroid conditions. Nonetheless, it is advised to streamline the diagnostic process by minimizing the number of required criteria. A reduction in the diagnostic parameters would alleviate the burden on patients by diminishing the need for extensive and often expensive testing, thereby making the diagnostic process more efficient and cost-effective. Consequently, the need for improved algorithms to be developed is urgent. and predictive models that leverage a limited set of variables to diagnose thyroid diseases efficiently. Such innovations promise to enhance patient care by offering quicker, more affordable diagnostic solutions.

REFERENCES

[1] Pan, Qiao, Yuanyuan Zhang, Min Zuo, Lan Xiang, and Dehua Chen. "Improved ensemble classification method of thyroid disease based on random forest." In 2016 8th International conference on information technology in medicine and education (ITME), pp. 567-571. IEEE, 2016.

[2] Pavya, K., and B. Srinivasan. "Feature selection algorithms to improve thyroid disease diagnosis." In 2017 International conference on innovations in green energy and healthcare technologies (IGEHT), pp. 1-5. IEEE, 2017.

[3] Tyagi, Ankita, Ritika Mehra, and Aditya Saxena. "Interactive thyroid disease prediction system using machine learning technique." In 2018 Fifth international conference on parallel, distributed and grid computing (PDGC), pp. 689-693. IEEE, 2018.

[4] Shahid, Afzal Hussain, Maheshwari Prasad Singh, Rahul Kumar Raj, Rashmi Suman, Drakhshan Jawaid, and Muqtadir Alam. "A study on label TSH, T3, T4U, TT4, FTI in hyperthyroidism and hypothyroidism using machine learning techniques." In 2019 International conference on communication and electronics systems (ICCES), pp. 930-933. IEEE, 2019.

[5] Rao, Amulya R., and B. S. Renuka. "A machine learning approach to predict thyroid disease at early stages of diagnosis." In 2020 IEEE international conference for innovation in technology (INOCON), pp. 1-4. IEEE, 2020.

[6] Mir, Yasir Iqbal, and Sonu Mittal. "Thyroid disease prediction using hybrid machine learning techniques: An effective framework." International Journal of Scientific & Technology Research 9, no. 2 (2020): 2868-2874.

[7] Riajuliislam, Md, Khandakar Zahidur Rahim, and Antara Mahmud. "Prediction of thyroid disease (hypothyroid) in early stage using feature selection and classification techniques." In 2021 International conference on information and communication technology for sustainable development (ICICT4SD), pp. 60-64. IEEE, 2021.

[8] Chaubey, Gyanendra, Dhananjay Bisen, Siddharth Arjaria, and Vibhash Yadav. "Thyroid disease prediction using machine learning approaches." National Academy Science Letters 44, no. 3 (2021): 233-238.

[9] Chaganti, Rajasekhar, Furqan Rustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez, and Imran Ashraf. "Thyroid disease prediction using selective features and machine learning techniques." Cancers 14, no. 16 (2022): 3914.

[10] Guleria, Kalpna, Shagun Sharma, Sushil Kumar, and Sunita Tiwari. "Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning." Measurement: Sensors 24 (2022): 100482.

[11] Devi, Munisamy Shyamala, Venkatesan Dhilip Kumar, Adrian Brezulianu, Oana Geman, and Muhammad Arif. "A Novel Blunge Calibration Intelligent Feature Classification Model for the Prediction of Hypothyroid Disease." Sensors 23, no. 3 (2023): 1128.

[12] Hossain, Md Bipul, Anika Shama, Apurba Adhikary, Avi Deb Raha, KM Aslam Uddin, Mohammad Amzad Hossain, Imtia Islam, Saydul Akbar Murad, Md Shirajum Munir, and Anupam Kumar Bairagi. "An Explainable Artificial Intelligence Framework for the Predictive Analysis of Hypo and Hyper Thyroidism Using Machine Learning Algorithms." Human-Centric Intelligent Systems (2023): 1-21.

[13] Sa, Ri, Taiyu Yang, Zexu Zhang, and Feng Guan. "Random Forest for Predicting Treatment Response to Radioiodine and Thyrotropin Suppression Therapy in Patients With Differentiated Thyroid Cancer But Without Structural Disease." The Oncologist 29, no. 1 (2023): e68-e80.

[14] Naeem, Awad Bin, Biswaranjan Senapati, Alok Singh Chauhan, Mukta Makhija, Arpita Singh, Meghna Gupta, Pradeep Kumar Tiwari, and Wael MF Abdel-Rehim. "Hypothyroidism disease diagnosis by using machine learning algorithms." International Journal of Intelligent Systems and Applications in Engineering 11, no. 3 (2023): 368-373.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

[15] Akash, K. Thirumala, F. Mohammed Usman, T. Nitesh Kumar, Mohammed Riyaz Ahmed, and Raveendra Gudodagi. "Predicting Thyroid Dysfunction Using Machine Learning Techniques." In 2023 12th International Conference on Advanced Computing (ICoAC), pp. 1-8. IEEE, 2023.