IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 May 2024, accepted 20 June 2024, date of publication 24 June 2024, date of current version 3 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3418974

RESEARCH ARTICLE

An Improved Framework for Detecting Thyroid Disease Using Filter-Based Feature Selection and Stacking Ensemble

GEORGE OBAIDO[1,2], (Member, IEEE), OKECHINYERE ACHILONU[3], BLESSING OGBUOKIRI[4], CHIMEREMMA SANDRA AMADI[5], LAWAL HABEEBULLAHI[6], TONY OHALLORAN[7], CHIDOZIE WILLIAMS CHUKWU[8], EBIKELLA DOMOR MIENYE[9], MIKAIL ALIYU[10], OLUFUNKE FASAWE[10], IBUKUNOLA ABOSEDE MODUPE[11], EREPAMO JOB OMIETIMI[12], AND KEHINDE ARULEBA[13]

[1]Berkeley Institute for Data Science, University of California at Berkeley, Berkeley, CA 94720, USA
[2]Center for Human-Compatible Artificial Intelligence, University of California at Berkeley, Berkeley, CA 94720, USA
[3]School of Public Health, University of the Witwatersrand at Johannesburg, Johannesburg 2017, South Africa
[4]Department of Computer Science, Brock University, St. Catharines, ON L2S 3A1, Canada
[5]Department of Information Technology, Federal University of Technology, Owerri (FUTO), Owerri 460113, Nigeria
[6]Department of Computer Science, Summit University at Offa, Offa 250101, Nigeria
[7]School of Computer Science, National University of Ireland, Galway, H91 TK33 Ireland
[8]Department of Mathematics, Wake Forest University, Winston-Salem, NC 27106, USA
[9]College of Business and Economics, University of Johannesburg, Johannesburg 2006, South Africa
[10]School of Public Health, University of California at Berkeley, Berkeley, CA 94704, USA
[11]Department of Computer Science, Vaal University of Technology, Vanderbijlpark 1900, South Africa
[12]Department of Geology, University of Pretoria, Pretoria 0028, South Africa
[13]School of Computing and Mathematical Sciences, University of Leicester, LE1 7RH Leicester, U.K.

Corresponding author: Blessing Ogbuokiri (bogbuokiri@brocku.ca)

ABSTRACT In recent years, machine learning (ML) has become a pivotal tool for predicting and diagnosing thyroid disease. While many studies have explored the use of individual ML models for thyroid disease detection, the accuracy and robustness of these single-model approaches are often constrained by data imbalance and inherent model biases. This study introduces a filter-based feature selection and stacking-based ensemble ML framework, tailored specifically for thyroid disease detection. This framework capitalizes on the collective strengths of multiple base models by aggregating their predictions, aiming to surpass the predictive performance of individual models. Such an approach can also reduce screening time and costs considering few clinical attributes are used for diagnosis. Through extensive experiments conducted on a clinical thyroid disease dataset, the filter-based feature selection approach and the ensemble learning method demonstrated superior discriminative ability, reflected by improved receiver operating characteristic-area under the curve (ROC-AUC) scores of 99.9%. The proposed framework sheds light on the complementary strengths of different base models, fostering a deeper understanding of their joint predictive performance. Our findings underscore the potential of ensemble strategies to significantly improve the efficacy of ML-based detection of thyroid diseases, marking a shift from reliance on single models to more robust, collective approaches.

INDEX TERMS Artificial intelligence, healthcare, machine learning, filter-based stacking ensemble learning, thyroid disease.

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

I. INTRODUCTION

Approximately 40% of the global population suffers from iodine deficiencies, leading to thyroid-related diseases that affect over 200 million people worldwide [1], [2], [3], [4]. The manifestation of thyroid diseases is largely influenced by dietary iodine, an essential component of thyroid hormones [1], [5], [6], [7], [8]. An imbalance in thyroid hormone production can lead to various thyroid diseases, which constitute a significant global health issue. These diseases notably impair the physical and psychosocial well-being of affected individuals, particularly during early life due to their impact on cognition and growth. Common thyroid diseases, including hypothyroidism, hyperthyroidism, thyroid nodules, goiter, and thyroid cancer, which are all influenced by hormonal imbalances [9], [10], [11], [12].

The incidence of thyroid cancer, the most prevalent endocrine cancer globally, has seen a significant increase in recent years [13]. Thyroid cancer develops in the thyroid gland, a butterfly-shaped gland located at the front of the neck (Figure 1). This type of cancer occurs when cells within the gland begin to proliferate uncontrollably, leading to the formation of tumors [14], [15], [16]. The thyroid is critical to the endocrine system, producing hormones that regulate metabolism, heart rate, and body temperature. Thyroid cancer is staged from I to IV, with the stage indicating the tumor's aggressiveness and spread. Despite its increasing prevalence, the mortality rate for thyroid cancer remains relatively stable [17], [18], [19], [20], [21], [22]. In many Asian countries, thyroid cancer is one of the top three contributors to Disability-Adjusted Life Years (DALYs), a measure reflecting the overall disease burden [13]. Nevertheless, with timely diagnosis and proper treatment, a significant number of patients can fully recover and lead healthy lives.

The quest for improved diagnostic methods for thyroid diseases is driven by the recognition of the disease's healthcare challenges and the limitations of existing diagnostic approaches. Current research has highlighted both the strengths and weaknesses of these methods, exploring various imaging modalities, laboratory tests, and clinical assessments in diagnosing thyroid diseases [24], [25], [26]. Presently, diagnosis heavily depends on human assessment, such as interpreting medical images and evaluating fine-needle aspiration biopsies, which can be subjective and vary in accuracy [27], [28], [29], [30]. Despite advancements in medical technology, thyroid diseases pose significant challenges, including the differentiation of benign from malignant thyroid nodules, early detection of thyroid cancer, and timely identification of thyroid dysfunction. Additionally, managing thyroid diseases requires a well-balanced treatment approach, where precise diagnosis and prognosis are essential for customized patient care.

Machine Learning (ML), a key branch of artificial intelligence (AI), employs a variety of algorithms to learn from data, continuously improving its performance through learning and adjustments [31], [32], [33], [34], [34], [35], [36], [37], [38]. ML has shown effectiveness in numerous fields, including healthcare, where it is primarily used for disease diagnosis. Thyroid disease diagnosis, in particular, has greatly benefited from ML advancements [39], [40], [41], [42]. Other examples include employing artificial neural networks (ANN) and other models as classifiers [43], using selective features [44], applying random forest (RF) models [45], [46], adopting decision tree ensemble approaches [47], utilizing boosting ensemble methods [48], employing feature selection methods alongside support vector machines (SVM) [49], and integrating decision trees and k-nearest neighbor (KNN) techniques [50].

In particular, Islam et al. [43] developed a predictive model for thyroid disease using a range of ML algorithms and found that the ANN classifier surpassed others in performance, achieving an accuracy of 0.9587. This was closely followed by the CatBoost and XGBoost classifiers, with accuracies of 95.38% and 95.33%, respectively. Chaganti et al. [44] applied feature engineering methods, such as forward, backward, and bidirectional feature selection, alongside ML and deep learning models. This approach aimed to predict various types of thyroid conditions more accurately and reliably. Their study suggested that careful feature selection, combined with ML models, significantly improves predictive capabilities for thyroid disease detection. Duggal and Shukla [49] employed feature selection techniques, including univariate selection, recursive feature elimination, and tree-based feature selection, together with classification techniques like Naïve Bayes, SVM, and RF for diagnosing thyroid diseases. They discovered that the SVM, paired with the recursive feature elimination method, achieved a notable accuracy rate of 92.92%.

Alyas et al. [45] utilized various ML algorithms, including decision trees, RF, K-NN, and ANN, to classify and promptly detect thyroid diseases using ultrasound images. The RF algorithm was particularly effective, highlighting its potential for automating and enhancing the diagnostic process in medical practice, especially for thyroid diseases. In a similar vein, Sonuç et al. [46] employed the RF model alongside other ML models to categorize thyroid disease into hyperthyroidism, hypothyroidism, and normal. Their study focused on a cohort of Iraqi individuals, including those with overactive and underactive thyroid glands. Mishra et al. [51] enhanced the RF model by adding the sequential minimal optimization (SMO), decision table, and K-star classifier, aiming to improve hypothyroidism diagnosis. Chaubey et al. [50] experimented with algorithms like logistic regression, decision trees, and K-NN for thyroid disease prediction. Their findings suggested that the K-NN classifier was the most effective in their specific study context, offering a promising approach for thyroid disease prediction.

Other studies have utilized ensemble learning techniques for thyroid disease diagnosis. For example, Yadav and Pal [47] proposed tree-based ensemble methods for the early detection of thyroid diseases, including severe conditions like thyroid cancer. Their study indicated that this ensemble method could significantly improve thyroid disease

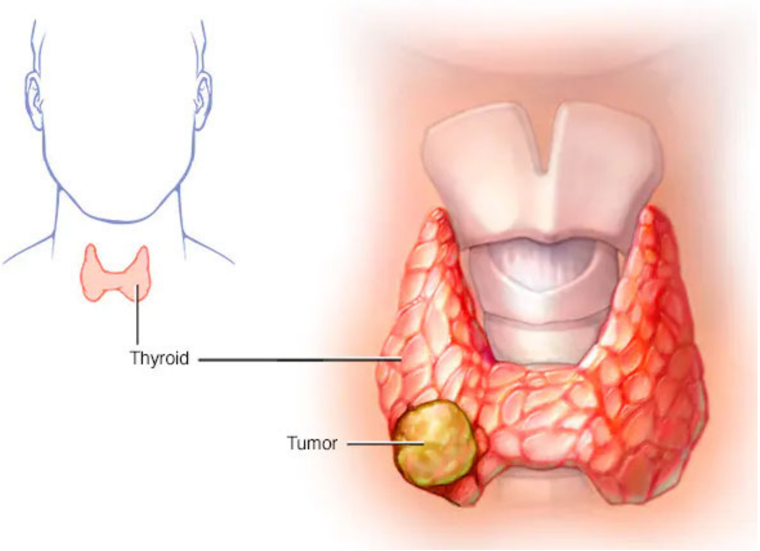


FIGURE 1. **Thyroid cancer originates from abnormal growth of cells in the thyroid, a butterfly-shaped gland situated at the base of the neck, just beneath the Adam's apple [23].**

prediction. Awujoola et al. [52] utilized the bagging ensemble method, combining J48 and SimpleCart models, to enhance the accuracy of thyroid disease prediction. This approach leverages the strengths of both algorithms within a bagging framework, aiming to enhance predictive accuracy for thyroid conditions. Agilandeeswari et al. [53] developed a voting ensemble technique that combines decisions from various regression and classification algorithms to predict thyroid diseases. Additionally, Akhtar et al. [54] extended homogeneous ensembling, utilizing a layered ensemble approach combined with multiple feature selection techniques to enhance thyroid disorder detection. This method effectively integrates several ensemble models, improving their collective predictive power for more accurate thyroid case identification. Ciaburro [48] explored AdaBoostM1, a boosting ensemble ML algorithm, demonstrating its practical application and theoretical benefits in diagnosing thyroid disease. Alshayeji [55] applied data mining and ensemble strategies using Bayesian optimization to enhance early diagnosis of thyroid diseases. This approach aimed to improve the accuracy and efficiency of thyroid disease detection by leveraging advanced optimization techniques. Haitham [56] applied deep learning and ensemble methods to elevate diagnostic accuracy and reliability in thyroid nodule detection. This study emphasized integrating advanced AI techniques with traditional medical practices, potentially transforming how thyroid diseases are diagnosed and managed.

In this study, a filter-based feature selection and stacking-based ensemble framework is introduced, specifically designed for thyroid disease detection. By eliminating features with minimal contributions, filter-based selection strategies enhance the predictive accuracy of models. This improvement occurs because the removal of noise and irrelevant data helps prevent the model from learning spurious patterns that do not generalize to unseen data, thereby reducing the risk of overfitting. Through extensive experiments conducted with a real-world thyroid disease dataset, we consistently demonstrated the superior performance of our ensemble approach across various metrics. Our proposed framework not only enhances predictive accuracy but also provides insightful revelations into the strengths of different base models, thereby enriching our understanding of their combined efficacy. These findings highlight the importance of moving beyond single-model approaches and adopting ensemble strategies, which significantly improve the effectiveness of ML in thyroid disease detection. The following are the contributions of the study:

1) The study introduces an improved ML technique for thyroid disease detection, demonstrating the effectiveness of filter-based strategies and ensemble methods.
2) The filter-based feature selection strategy, by removing irrelevant or redundant features, plays a crucial role in enhancing the overall effectiveness and efficiency of machine learning models.
3) The stacked ensemble model facilitates a more personalized diagnostic approach by leveraging the strengths of multiple model predictions, potentially identifying unique or rare thyroid conditions that might be overlooked by individual models.
4) The comparison of our study with existing approaches underscores its effectiveness in thyroid disease detection.

The remainder of this paper is structured as follows: Section II explores related work in the field. Section III describes the methodology employed in this study, while Section IV details the experimental setup and procedures.

Section V presents the findings of our research, and Section VI thoroughly discusses the study and its implications. Lastly, Section VII summarizes the conclusions drawn from this study and outlines potential directions for future research.

II. BACKGROUND

A. DATASET

The dataset contains 1232 samples and 19 features, providing demographic information on patients from between 2010 and 2012 [57]. Table 1 presents a detailed list of features related to thyroid examinations, which are essential for diagnosing and managing thyroid conditions. It includes the age and gender of the patient, with age being a risk factor for thyroid conditions and malignancies, and women being diagnosed more frequently than men. The dataset also measures levels of thyroid-related hormones and antibodies such as Free Triiodothyronine (FT3), Free Thyroxine (FT4), Thyroid Stimulating Hormone (TSH), Thyroid Peroxidase Antibodies (TPO), and Thyroglobulin Antibodies (TGAb), which are pivotal for assessing thyroid function and detecting autoimmune thyroiditis. Additionally, it covers anatomical and morphological details such as the site, size, shape, and echogenicity patterns of the thyroid or nodules, which can indicate the severity or nature of the condition, such as the likelihood of malignancy based on patterns like multifocality, irregular margins, specific calcifications, and blood flow characteristics observed in ultrasound imaging. Other attributes include echo strength, which refers to the intensity of echogenicity, and composition, describing whether a nodule is solid, cystic, or a mix of both. The table also includes a designation of nodules as benign or malignant (‘mal’) and whether changes are bilateral (‘multilateral’), affecting both thyroid lobes.

Healthcare datasets frequently show an imbalance, characterized by a substantial discrepancy in the distribution of various classes or outcomes under investigation [58], [59], [60]. Such imbalances present difficulties in creating precise predictive models and performing effective data analysis in the healthcare field.

B. MACHINE LEARNING MODELS

This section presents the ML models explored in this study, namely LR, SVM, KNN, DT, and ANN.

1) LOGISTIC REGRESSION

Logistic regression (LR) models are used to investigate the associations between risk factors and a target event [61], [62], [63]. LR is versatile and finds application in a wide range of classification and regression problems, including binary and multi-class scenarios [64], [65]. As a statistical tool, LR models the likelihood of specific outcomes based on input variables. In medical research, LR has been pivotal in predicting the onset of diseases, confirming or refuting diagnoses based on symptoms and test results, gauging the effectiveness of new treatments, identifying high-risk patients for particular conditions, and classifying patients for tailored care plans. The mathematical representation of LR is given as:

$$\sigma(t) = \frac{1}{1+e^{-t}} \tag{1}$$

In the context of logistic regression, this is the input to the sigmoid function before transformation. t is the weighted sum of the input features and the weights. Given in matrix notation:

$$t = \mathbf{x}_i^T \mathbf{W} \tag{2}$$

Expanding it out for d features:

$$t = w_0 + w_1 x_{i,1} + w_2 x_{i,2} + \ldots + w_d x_{i,d} \tag{3}$$

where $\mathbf{W}$ is the weight vector, including the bias term w_0. $\mathbf{x}_i$ is the feature vector for the i-th instance. $x_{i,1}, x_{i,2}, \ldots, x_{i,d}$ are the individual features of the i-th instance. $w_0, w_1, \ldots, w_d$ are the weights corresponding to each feature, with w_0 being the bias term.

2) SUPPORT VECTOR MACHINE

Support Vector Machines (SVMs) are robust methods used for classification and regression tasks [66], [67], [68]. The fundamental concept behind SVMs is to distinguish between classes by maximizing the margin between them, particularly in the training set [69]. For a binary classification problem with two classes, the SVM tries to find the optimal hyperplane that maximizes the margin between the two classes. The hyperplane can be described by the equation:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{4}$$

where $\mathbf{w}$ is the weight vector, which is normal to the hyperplane, $\mathbf{x}$ is an input vector, and b is the bias term.

3) K-NEAREST NEIGHBOR

The k-Nearest Neighbour (K-NN) algorithm is a non-parametric method used for classification and regression [70], [71]. K-NN classifies data points based on their proximity to query points. A key feature of K-NN is its distance metric used to determine the similarity between data points. Commonly, the Euclidean distance is employed, but other metrics like Manhattan, Minkowski, or Hamming distance can be used depending on the nature of the data. For an Euclidean distance for continuous variables:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{m} (x_{i,l} - x_{j,l})^2} \tag{5}$$

where $d(x_i, x_j)$ is the distance between points x_i and x_j. m is the number of features. $x_{i,l}$ and $x_{j,l}$ are the l-th features of points x_i and x_j, respectively.

TABLE 1. Attributes and descriptions of thyroid dataset.

No.	Attribute	Description	Category	Scale
at1	age	The age of the patient.	Numerical	Years
at2	gender	The gender of the patient.	Categorical	male, female
at3	FT3	Free triiodothyronine level.	Numerical	pmol/L
at4	FT4	Free thyroxine level.	Numerical	pmol/L
at5	TSH	Thyroid Stimulating Hormone level.	Numerical	µIU/mL
at6	TPO	Thyroid peroxidase antibodies level.	Numerical	IU/mL
at7	TGAb	Thyroglobulin antibodies level.	Numerical	IU/mL
at8	site	Location or region of the thyroid or thyroid nodule.	Categorical	N/A
at9	echo_pattern	Echogenicity pattern seen on ultrasound.	Categorical	N/A
at10	multifocality	Multiple nodules or focal areas of interest.	Categorical	-
at11	size	Size of the thyroid nodule.	Numerical	cm
at12	shape	Shape of the nodule on ultrasound.	Categorical	-
at13	margin	Refers to the edges or boundary of the nodule.	Categorical	-
at14	calcification	Presence of calcifications within the nodule.	Categorical	-
at15	echo_strength	Echogenicity strength or intensity on ultrasound.	Numerical	Arbitrary units
at16	blood_flow	Blood flow characteristics of the nodule on Doppler ultrasound.	Categorical	-
at17	composition	Nodule is solid, cystic, or a mix of both.	Categorical	-
at18	mal	Benign vs. malignant classification of the nodule.	Categorical	-
at19	multilateral	Nodules or thyroid changes present on both sides (lobes) of the thyroid.	Categorical	-

4) DECISION TREES

Decision trees (DTs) are non-parametric methods that use a hierarchical, tree-like model of decisions composed of a root node, branches, internal nodes, and leaf nodes [72], [73], [74]. DTs employ a divide-and-conquer strategy to identify optimal split points within the tree. Popular decision tree algorithms, such as ID3, C4.5, and CART, are used for both classification and regression tasks, determining the best feature to split on at each step [75]. Building a decision tree involves making decisions at each node. This decision-making is based on choosing the best split among all possible splits. The quality of a split is measured using certain criteria:

Entropy of a set:

$$S: \quad E(S) = -p_+ \log_2(p_+) - p_- \log_2(p_-) \tag{6}$$

where p_+ is the proportion of positive examples in S and p_- is the proportion of negative examples in S.

Information Gain based on a split feature:

$$\text{Gain}(S, F) = E(S) - \sum_{v \in \text{Values}(F)} \frac{|S_v|}{|S|} E(S_v) \tag{7}$$

where S_v is the subset of S for which feature F has value v.

Here, S is the current dataset, F is the feature being considered for the split, Values(F) are the possible values of feature F, S_v is the subset of S for which feature F has value v, $|S_v|$ is the number of instances in subset S_v, $|S|$ is the total number of instances in dataset S, $E(S)$ is the entropy of dataset S, and $E(S_v)$ is the entropy of subset S_v. The information gain is calculated as the difference between the entropy of the current dataset S and the weighted sum of the entropies of the subsets after the split.

5) ARTIFICIAL NEURAL NETWORK

The Artificial Neural Network (ANN) is a widely used neural network with the capability to perform function estimation [76], [77], [78]. It is proficient in managing both linear and non-linear data relationships. With its composition of multiple layers of interconnected neurons, the ANN acquires data representations by adjusting weights during the training process. For a neuron k in the output layer:

$$p(y = k|\mathbf{x}) = \frac{e^{z_k}}{\sum_{l=1}^{K} e^{z_l}} \tag{8}$$

where k is the number of output neurons (equal to the number of classes in a classification task). $p(y = k|\mathbf{x})$ is the probability that input $\mathbf{x}$ belongs to class k. The ANN is trained using backpropagation, which adjusts the weights and biases to minimize the difference between the predicted outputs and the actual labels, often employing the cross-entropy loss for classification tasks [79].

These models were selected due to their distinct strengths and complementary characteristics, which together offer a robust approach to addressing the complexities of the dataset. LR was chosen for its simplicity and interpretability, making it ideal for establishing a baseline in binary classification tasks. It provides clear probabilistic outputs, which are valuable for understanding the impact of different features on predictions. SVM was included for its effectiveness in high-dimensional spaces and its capacity to model non-linear decision boundaries through kernel functions. It is beneficial for complex classification problems where the decision surface is not readily apparent. DTs were selected for their intuitive understanding and ability to handle non-linear relationships. Their structure makes it easy to visualize and interpret, which is beneficial for communicating findings to stakeholders who may not have a technical background. Thanks to their deep and flexible architecture, ANN provides exceptional modeling capabilities, especially in capturing intricate patterns in large datasets. This makes them suitable for more complex problems where other models might fail to capture all the nuances in the data.

C. FILTER-BASED FEATURE SELECTION WITH INFORMATION GAIN

Feature selection is an integral part of the data preprocessing step. It involves selecting specific features from the dataset for use in the training process of a learning algorithm. The Information Gain (IG) filter-based feature selection method is primarily used to measure the effectiveness of features in classifying data in decision tree models [80], [81] and can also be broadly applied in other contexts.

IG is based on the concept of entropy from information theory, representing the impurity or uncertainty in a group of examples. Mathematically, the IG between two variables X and Y is formulated as the difference between the initial entropy of X and the entropy of X after observing Y. This can be expressed as:

$$IG(X|Y) = H(X) - H(X|Y) \tag{9}$$

where $H(X)$ is the entropy for variable X and $H(X|Y)$ represents the conditional entropy for X given Y. To compute the IG value for an attribute, calculate the entropy of the target variable for the entire dataset and subtract the conditional entropies for every potential value of that attribute. Furthermore, the entropy $H(X)$ and conditional entropy $H(X|Y)$ are computed as:

$$H(X) = -\sum_{x \in X} P(x) \log_2(x) \tag{10}$$

$$H(X|Y) = -\sum_{x \in X} P(x) \sum_{y \in Y} P(x|y) \log_2(P(x|y)) \tag{11}$$

Therefore, when considering two variables X and Z, a variable Y is deemed to have a stronger correlation with X than with Z if $IG(X|Y) > IG(Z|Y)$. Moreover, IG evaluates each attribute independently and assesses its relevance to the target variable.

D. THE STACKING ENSEMBLE TECHNIQUE

Ensemble learning is a technique that involves combining two or more ML algorithms to create a more effective model [48], [82], [83]. This approach utilizes the strengths and mitigates the weaknesses of individual models, leading to improved performance in various tasks. Ensemble methods can be particularly beneficial in scenarios where a single algorithm might struggle due to limitations like bias or variance. There are several common types of ensemble methods, including bagging, boosting, and stacking. Bagging, short for *bootstrap aggregating*, involves training multiple models in parallel, each on a random subset of the data, and then averaging their predictions. Boosting, on the other hand, trains models sequentially, with each new model focusing on the errors made by the previous ones, thereby improving the overall accuracy.

The stacking ensemble also referred to as ''stacked generalization,'' is based on the concept where multiple models are combined to produce a given prediction [84], [85], [86], [87]. A significant benefit of stacking ensembles is their ability to improve the predictive accuracy of unbalanced datasets [88], [89]. The stacking ensemble has been successfully applied in various domains such as image classification, natural language processing, and financial forecasting. One of the key advantages of a stacking ensemble is its flexibility in incorporating diverse base learners, ranging from simple algorithms like decision trees to complex models like neural networks [90], [91].

Moreover, stacking ensemble can effectively capture the complementary strengths of different models, mitigating the weaknesses of individual learners and leading to enhanced overall performance [93]. Figure 2 illustrates the process flow of the stacked ensemble method. This is achieved through a meta-learner, which learns to combine the predictions of the base models, often yielding more robust and accurate predictions than any single model alone [94]. Furthermore, the versatility of the stacking ensemble allows for the integration of various feature engineering techniques, model hyperparameters, and ensemble strategies, providing ample room for experimentation and optimization [95]. This adaptability makes the stacking ensemble a popular choice among data scientists and machine learning practitioners for tackling a wide range of prediction tasks [92], [96], [97].

III. METHODOLOGY

A. OUR PROPOSED APPROACH

In the stacking ensemble, each model acts as an individual contributor, offering its unique perspective and prediction based on the data (See Figure 3). These individual predictions are then collected and used as inputs for a higher-level model, often known as the meta-model. The role of the meta-model is to synthesize these inputs, discern patterns among the base models' predictions, and produce a more refined and potentially more accurate final prediction [37]. This layered approach allows for a deeper understanding of the data, leveraging the strengths of each base model while compensating for their weaknesses. As a result, stacking ensemble methods can often outperform any single model in the ensemble, especially when there is a diverse set of base models providing varied insights into the data. The implementation of the stacked ensemble classifier is presented in Algorithm 1.

In Algorithm 1, the dataset comprises feature vectors and corresponding labels as input. The dataset D undergoes stratified K-fold cross-validation, a process that divides the data into training sets T_k and validation sets V_k, ensuring each fold is representative of the overall class distribution. This approach enhances consistency and reliability during validation. Base classifiers, each with a unique analytical approach, are trained on the training set T_k. These classifiers, encompassing various ML algorithms, enrich the learning phase by capturing different patterns in the data. Their

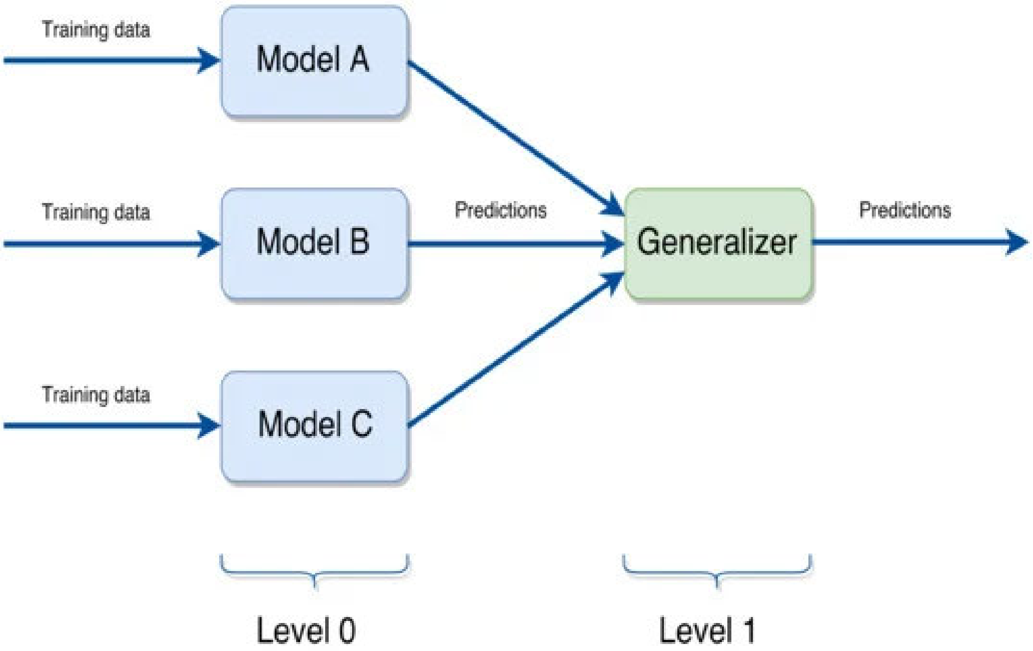


FIGURE 2. A stacked ensemble workflow [92].

Algorithm 1 Stacking Ensemble Classifier for Thyroid Disease Diagnosis

```
Input: Thyroid disease dataset $D$, which consists of feature vectors $X = \{x_1, x_2, \dots, x_{n_i}\}$ and corresponding labels $Y = \{y_1, y_2, \dots, y_{n_i}\}$.
Output: Predictions ($P$) from the ensemble classifier $C_{\text{stacked}}$.
Step 1: Perform stratified K-fold cross-validation on $D$ to create training and validation sets.
for $k = 1$ to $K$ do
    Divide $D$ into training set $T_k$ and validation set $V_k$.
    Step 2: Train base classifiers: LR, SVM, KNN, DT, and ANN on $T_k$.
    for each classifier $C_j$ in {LR, SVM, KNN, DT, ANN} do
        Train classifier $C_j$ on $T_k$.
        Make predictions on $V_k$ to create features for $D^k_{\text{meta}}$.
    end for
    Aggregate predictions from $D^k_{\text{meta}}$ into $D_{\text{meta}}$.
end for
Step 3: Train the meta-classifier (stacking ensemble classifier) on aggregated $D_{\text{meta}}$.
Step 4: Prepare a new, unseen dataset $D_{\text{test}}$.
Step 5: Use the trained meta-classifier to obtain the final predictions $P$ by applying it to $D_{\text{test}}$.
Step 6: return the ensemble predictions $P$.
```

predictions on the validation set V_k are then used as meta-features for D_{meta}. The meta-classifier, trained on these meta-features, synthesizes these insights. It combines the base classifiers' predictions to yield the final predictions. This methodology leverages the collective strengths of diverse classifiers, thereby enhancing predictive accuracy.

B. PERFORMANCE METRICS

Performance metrics are used to measure the performance of ML models [98], [99]. These metrics provide a deeper understanding of various model attributes, including accuracy, precision, sensitivity, specificity, F1-score, and the area under the Receiver Operating Characteristics (ROC) curve (AUC), among others.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

$$Specificity = \frac{TN}{TN + FP} \tag{15}$$

$$F1\ measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{16}$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{17}$$

where:

- True Positives (TP) represent the number of correctly predicted positive instances.
- False Negatives (FN) represent the number of positive instances incorrectly classified as negative.

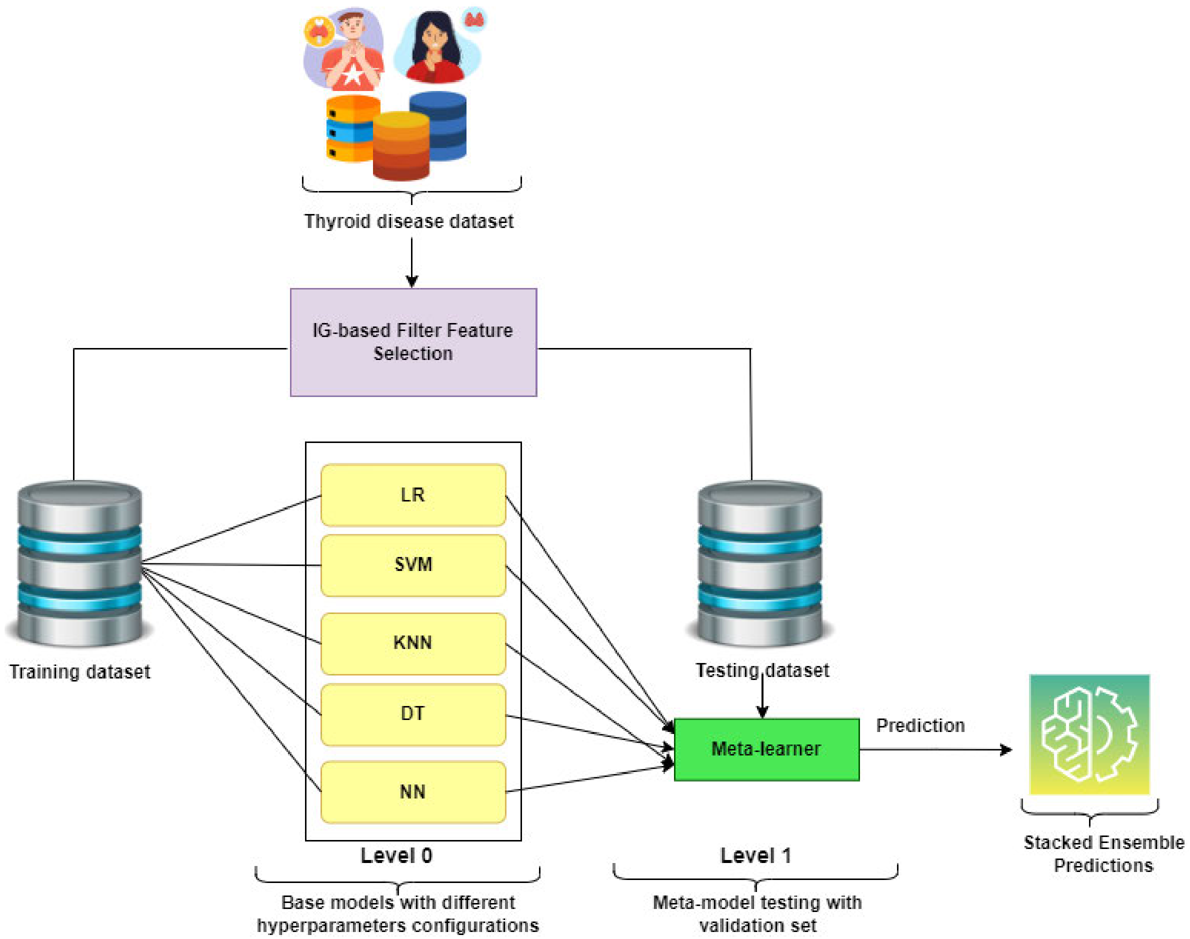

Thyroid disease dataset
IG-based Filter Feature Selection
Training dataset
LR
SVM
KNN
DT
NN
Testing dataset
Meta-learner
Prediction
Stacked Ensemble Predictions
Level 0
Base models with different hyperparameters configurations
Level 1
Meta-model testing with validation set

FIGURE 3. The stacked-based ensemble with IG process.

- True Negatives (TN) represent the number of correctly predicted negative instances.
- False Positives (FP) represent the number of negative instances incorrectly classified as positive.

In the context of diagnosing thyroid diseases, assessing the effectiveness of ML models extends beyond mere accuracy. While accuracy provides an overview of correct predictions made by a model, it may not adequately capture nuances, especially in imbalanced datasets with a higher proportion of patients without the disease. Here, balanced accuracy becomes essential, offering insights into the model's performance in correctly identifying both the presence and absence of the disease. Precision is also vital; an incorrect positive diagnosis could subject a patient to unnecessary treatments. Similarly, high sensitivity is paramount to avoid missing a thyroid disease diagnosis, which can have serious health implications. Specificity is equally important to ensure that those without the disease are not falsely diagnosed, thus preventing unnecessary treatments. The F1-score, which balances precision and sensitivity, provides a comprehensive view of a model's performance, especially when the cost of false positives and false negatives is high. Lastly, the ROC-AUC metric is utilized to assess the discriminative performance of the models. This metric is crucial in medical diagnostics as it helps in effectively identifying true positives (sensitivity) while minimizing false positives (specificity).

IV. EXPERIMENTAL ANALYSIS

In this study, the construction of the models was carried out with rigor and methodical precision, ensuring that every phase of the process was optimized for accuracy and sensitivity, as depicted in Figure 3. In the initial stages, the dataset underwent an information gain feature extraction process and later on a scaling procedure. This step ensured that each feature was standardized to the same scale, a particularly significant aspect for models sensitive to variations in feature scales, such as SVM and K-NN. The goal was to both accelerate the convergence speed of the algorithms and boost their overall performance.

We acknowledge the challenges presented by imbalanced datasets, which can introduce biases and potentially lead to overfitting. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) [100], a well-established method was utilized. SMOTE is valuable for

TABLE 2. Hyperparameters for the classifiers.

Classifier	Values
LR	$C = 1.0$
SVM	$C = 1.0$ Kernel='rbf' $\gamma = 0.1$
KNN	$k = 7$ Metric='euclidean'
DT	Max Depth=5 Min Samples/Leaf=2 Criterion='gini'
ANN	hidden layer sizes=50 alpha=0.001

its capability to generate synthetic samples, ensuring a balanced representation of both minority and majority classes. By employing SMOTE, our goal was to foster the development of models that were more generalized and robust. Another crucial component of our evaluation process was hyperparameter tuning, as shown in Table 2. For each model, the hyperparameter grids were defined and a wide range of parameter combinations was explored. This exhaustive approach enabled the identification and selection of the best parameters to optimize the model's performance.

To enhance the robustness of the model evaluations, a stratified k-fold cross-validation technique was employed. This method ensured that each fold of the cross-validation contained a proportion of samples from each class that mirrored the complete dataset. A 10-fold stratified cross-validation was chosen, balancing computational efficiency with the need for robust and reliable performance estimates. A distinctive aspect of our methodology was integrating the stacked ensemble model, wherein predictions from base models served as input for another model, the meta-model, ultimately producing the final prediction. LR was selected as the meta-model, with an impressive maximum iteration limit of 1000 set to ensure the best fit for the data. The objective was to leverage the distinct strengths of each model, aiming to create an ensemble with the potential to exceed the performance of any single model in terms of accuracy.

V. RESULTS

In the assessment of the models, varying levels of performance were observed, as outlined in Table 4 and Table 5. These results provide a comprehensive overview of the models' performance, essential for evaluating their effectiveness in diagnosing the thyroid condition.

A. PERFORMANCE OF THE CLASSIFIERS WITHOUT FEATURE SELECTION

For the task without the feature selection task, as shown in Table 3, the Stacked Ensemble model outperformed others in terms of accuracy, with a score of 84.9%. This model also achieved the highest AUC value of 90%, sensitivity value of 81%, and specificity of 87%. This is followed closely by the LR model at 80.2% with AUC of 84%, sensitivity at 80.3%, and specificity at 83%. Interestingly, the K-NN model, despite its accuracy of 78.4%, displayed commendable precision, closely with the SVM, both scoring 84.1% and 84.4% respectively. The ANN and Stacked Ensemble models exhibited a close match in terms of balanced accuracy, precision, and specificity, with the latter slightly edging out in most metrics. The classifiers' AUC scores are shown in Figure 4.

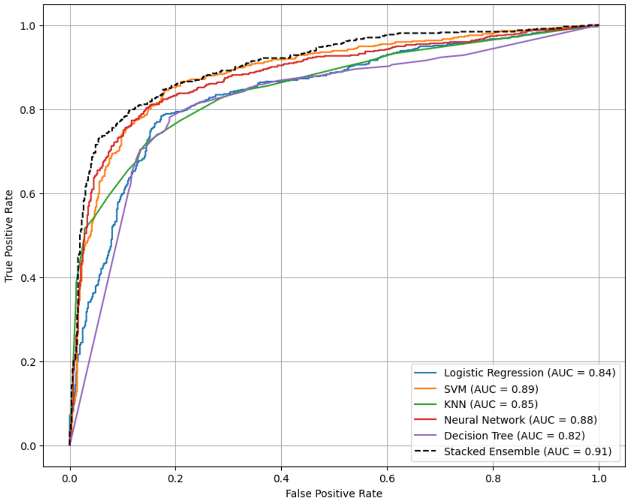


FIGURE 4. AUC of the classifiers on the entire feature set.

B. PERFORMANCE OF THE CLASSIFIERS AFTER FEATURE SELECTION

In this section, the performance of the classifiers after the feature selection task is presented. Table 4 shows the IG feature ranking, where features are ranked from highest to lowest IG value, suggesting the relative importance of each feature in the model. The table reveals that the feature named size holds the highest IG value at 0.173, placing it in the 10th position in terms of the ordering of the features, but it ranks highest in terms of its information gain. This is followed by calcification with an IG of 0.142, and age with an IG of 0.100, indicating their significant roles in the model. On the other hand, features such as FT3, FT4, and TGAb have an IG of 0.000, indicating that they contribute no informational value to the outcome of the model according to the measure used.

After computing the Information Gain (IG) values for various features in a dataset, the next step involves establishing a benchmark for feature selection. The standard deviation is calculated to serve as the threshold value for this task. The standard deviation is widely used because it effectively expresses the diversity of the IG value distribution [101]. The standard deviation of the IG values listed in the table is 0.0498. Of the 18 features, 7 features have IG values greater than the threshold of 0.0498, and 11 features have IG values less than the threshold. Consequently, features such as size, calcification, age, multilateral, site, blood flow, and shape have IG values above this threshold, indicating that they are

TABLE 3. Performance evaluation of the models on the entire feature set.

Model	Acc	Balanced Acc	Precision	AUC	F1	Sensitivity	Specificity
LR	80.2	81.7	82.2	84.0	80.1	80.3	83.0
SVM	82.1	82.1	84.4	89.0	80.0	79.2	85.0
KNN	78.4	78.5	84.1	89.0	82.0	70.0	87.0
DT	78.4	78.6	81.2	75.2	78.2	75.0	82.1
ANN	82.4	82.7	86.1	88.0	82.0	78.1	87.3
Stacked Ensemble	84.9	84.0	86.7	90.0	83.4	81.0	87.0

TABLE 4. Information gain feature ranking.

No.	Feature	IG Value
10	size	0.173
13	calcification	0.142
0	age	0.100
17	multilateral	0.092
7	site	0.090
15	blood_flow	0.062
11	shape	0.055
12	margin	0.046
16	composition	0.033
1	gender	0.029
8	echo_pattern	0.023
9	multifocality	0.018
14	echo_strength	0.017
5	TPO	0.012
4	TSH	0.001
6	TGAb	0.000
3	FT4	0.000
2	FT3	0.000

particularly informative in the context of the dataset. This suggests that these features are strong predictors or are highly associated with the outcome being studied. Features such as margin, composition, gender, echo pattern, and others that fell below the threshold may be less critical in predicting the condition of interest within this specific dataset. However, this does not mean these features are clinically unimportant; rather, they might not differentiate well between different states or outcomes in this particular analysis.

The reduced feature set was used to train the models. Table 5 presents the performance metrics of the reduced feature set. SVM, LR, DT, ANN, and the Stacked Ensemble all achieved improved accuracies. For instance, the Stacked Ensemble exhibited exceptional performance across all metrics, scoring 99.9% in accuracy, balanced accuracy, precision, sensitivity, and 99.9% AUC, along with a 99.8% specificity and a 99.7% F1 score. Following closely, the KNN model demonstrated impressive performance, achieving an 89.0% accuracy and a higher AUC at 97%. DT also performed well with an AUC of 94%, and ANN recorded an AUC of 93.0%. Finally, SVM also showed good performance with an AUC of 90%.

The Stacked Ensemble consistently demonstrated superior performance across both tasks, closely followed by the ANN, SVM, and DT models (refer to Figure 4 and Figure 5). The performance of the LR and K-NN models varied, highlighting their dataset-specific effectiveness. This comparison underscores the strength and efficacy of ensemble methods, particularly in contrast with single-model approaches. These findings suggest that the Stacked Ensemble approach can effectively integrate the strengths of individual models, leading to more robust and improved predictions. This reinforces the value of ensemble methods in complex ML tasks where single models may not consistently deliver optimal results.

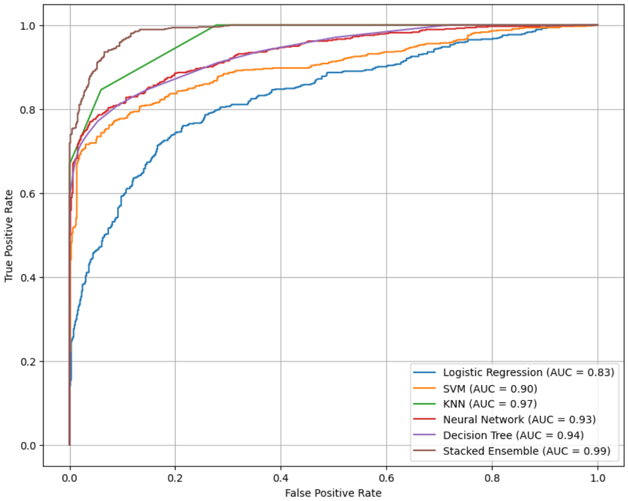


FIGURE 5. AUC of the classifiers on the reduced feature set.

VI. DISCUSSION

Thyroid diseases are a major healthcare concern globally, significantly affecting individuals' quality of life and health. The study applied an approach that combines the filter-based feature selection method and the stacking ensemble method to investigate thyroid diseases. The findings offer valuable insights into thyroid disease detection for clinicians and researchers. Traditional diagnostic methods for thyroid diseases vary in precision and efficacy, influenced by several factors. These include differences in physicians' diagnostic approaches, the challenges of consolidating diverse diagnostic data from various healthcare providers for comprehensive assessments, the importance of early diagnosis in slowing the progression of thyroid cancer and reducing mortality rates, and the difficulty in identifying rare thyroid cancer subtypes with unique characteristics. These factors significantly diminish the accuracy of thyroid disease diagnoses, presenting obstacles to improving patient care and developing tailored diagnostic and treatment options. Consequently, there is a crucial need for data-driven approaches like ML to enhance clinical decision-making [102], [103].

One of the distinctive strengths of our study was the application of the stacking ensemble and filter-based method for diagnosing thyroid diseases, incorporating predictions from various ML models. This approach significantly improved

TABLE 5. Performance evaluation of the models on the reduced feature set.

Model	Acc	Balanced acc	Precision	AUC	F1	Sensitivity	Specificity
LR	82.0	82.5	82.0	83.0	78.0	82.0	83.0
SVM	84.0	84.0	87.0	90.0	83.0	80.0	88.0
KNN	89.0	84.0	93.0	97.0	89.0	93.0	94.0
DT	86.0	86.5	94.0	94.0	85.0	78.0	95.0
ANN	86.0	86.5	94.0	93.0	86.0	80.0	93.0
Stacked Ensemble	99.9	99.9	99.9	99.9	99.7	99.9	99.8

TABLE 6. Comparison with other existing studies.

Reference	Method	Accuracy	Sensitivity	Specificity
[48]	Boosting Ensemble	99.7	-	-
[111]	Self-Stack Ensemble	99.5	99.9	-
[112]	Stacked Ensemble	99.5	99.5	-
[53]	Voting Ensemble	97.6	97.9	-
[54]	Homogenous Ensembling	99.3	-	-
[113]	Decision Tree Ensemble	99.2	-	-
[52]	Bagging Ensemble	99.6	-	-
[55]	Data Mining and Bayesian Optimization Ensemble	99.5	99.4	99.6
[56]	Deep Learning Ensemble Approach	95.1	-	-
Our Study	**Stacked Ensemble with IG Feature Selection**	**99.9**	**99.9**	**99.8**

the diagnosis of thyroid diseases, with the stacking ensemble model demonstrating better performance. The technique achieved accuracy, sensitivity, and specificity levels of 99.9%, 99.9%, and 99.8%, respectively, surpassing some previous works in this domain, as detailed in Table 6. The stacking ensemble model's suitability for predicting thyroid diseases is both consistent and promising. Our findings also reveal that in cases of imbalanced datasets, the stacking ensemble approach effectively enhances diagnostic accuracy, as supported by Yan et al. [88]. Thus, our study suggests that the stacked ensemble classifier is a superior method compared to single models in addressing thyroid disease classification challenges, maximizing diagnostic accuracy. Additionally, the analysis of IG values within a dataset not only aids in refining the focus of clinical investigations but also enhances the efficiency and effectiveness of patient care. By identifying which features are most and least predictive, clinicians and researchers can develop more targeted diagnostic algorithms and treatment protocols, ultimately leading to improved patient outcomes and more efficient use of healthcare resources.

Our findings not only underscore the importance of the size and calcification of thyroid nodules but also highlight how these characteristics are critical indicators of potential malignancy, aligning with observations in similar studies [104], [105], [106]. Larger nodules are more likely to be biopsied because their size often correlates with an increased risk of cancer [107], [108]. Guidelines from various thyroid associations suggest that nodules larger than a certain threshold (often around 1 cm in diameter) warrant a finer assessment, including ultrasound and possibly fine needle aspiration, depending on other coexisting features [109], [110]. This approach is aimed at early detection of thyroid cancers, which are typically more treatable when identified early. This approach is aimed at early detection of thyroid cancers, which are typically more treatable when identified early. The presence of calcifications, especially specific types like microcalcifications or peripheral calcifications, enhances the specificity of ultrasound in predicting malignancy.

While our study represents an advancement in thyroid disease research through the utilization of the stacking ensemble method and filter-based feature selection strategy, it is crucial to acknowledge its limitations. Variability in data quality and availability may have introduced bias into our analyses. Future research endeavors should prioritize the introduction of other ensemble methods, such as boosting and bagging, to ensure the robustness and generalizability of our findings. The potential of ML, particularly when combined with ensemble methods, holds immense promise for furthering our understanding of thyroid diseases. Subsequent investigations could explore the integration of genomic, proteomic, and imaging data to unravel the intricate molecular underpinnings of thyroid diseases. The development of predictive models for patient prognosis and treatment response, grounded in individualized data and driven by ensemble insights, may herald a new era of precision medicine in thyroid disease management.

VII. CONCLUSION AND FUTURE DIRECTION

In this study, methodological strategies for thyroid disease classification and prediction have been provided. The performance of five distinct machine-learning base learners and their integration into a stacked ensemble were explored. This approach sets our study apart from prior thyroid disease classification research using ML. The classifiers were applied to a thyroid disease dataset, where the combined predictive power of the base classifiers through the stacking method, together with the filter-based method, consistently surpassed individual model predictions. Our findings highlight the stacking ensemble model's effectiveness in improving thyroid disease detection. However, the study has limitations due to its reliance on secondary data, which constrains control over data availability, quality, and the completeness of information captured. Despite the stacked model's performance, there is room for further enhancement. Future research will explore other approaches by utilizing diverse datasets to predict the severity of thyroid disease conditions.

REFERENCES

[1] L. Aversano, M. L. Bernardi, M. Cimitile, A. Maiellaro, and R. Pecori, "A systematic review on artificial intelligence techniques for detecting thyroid diseases," *PeerJ Comput. Sci.*, vol. 9, p. e1394, Jun. 2023.

[2] G. V. Iyengar and P. P. Nair, "Global outlook on nutrition and the environment: Meeting the challenges of the next millennium," *Sci. Total Environ.*, vol. 249, nos. 1–3, pp. 331–346, Apr. 2000.

[3] S. Keestra, V. H. Tabor, and A. Alvergne, "Reinterpreting patterns of variation in human thyroid function: An evolutionary ecology perspective," *Evol., Med., Public Health*, vol. 9, no. 1, pp. 93–112, 2021.

[4] M. B. Zimmermann, "Iodine deficiency," *Endocrine Rev.*, vol. 30, no. 4, pp. 376–408, 2009.

[5] L. H. Duntas, "Thyroid disease and lipids," *Thyroid*, vol. 12, no. 4, pp. 287–293, Apr. 2002.

[6] P. Farling, "Thyroid disease," *Brit. J. Anaesthesia*, vol. 85, no. 1, pp. 15–28, 2000.

[7] M. Helfand and L. M. Crapo, "Screening for thyroid disease," *Ann. Internal Med.*, vol. 112, no. 11, pp. 840–849, 1990.

[8] S. Prathibha, D. Dahiya, C. R. Rene Robin, C. Venkata Nishkala, and S. Swedha, "A novel technique for detecting various thyroid diseases using deep learning," *Intell. Autom. Soft Comput.*, vol. 35, no. 1, pp. 199–214, 2023.

[9] A. Faggiano, M. Del Prete, F. Marciello, V. Marotta, V. Ramundo, and A. Colao, "Thyroid diseases in elderly," *Minerva Endocrinol.*, vol. 36, no. 3, pp. 211–231, 2011.

[10] G. Mariani, M. Tonacchera, M. Grosso, F. Orsolini, P. Vitti, and H. W. Strauss, "The role of nuclear medicine in the clinical management of benign thyroid disorders—Part 1: Hyperthyroidism," *J. Nucl. Med.*, vol. 62, no. 3, pp. 304–312, Mar. 2021.

[11] M. P. Vanderpump, "The epidemiology of thyroid disease," *Brit. Med. Bull.*, vol. 99, no. 1, pp. 39–51, 2011.

[12] J. P. Walsh, "Managing thyroid disease in general practice," *Med. J. Aust.*, vol. 205, no. 4, pp. 179–184, Aug. 2016.

[13] Y. Deng, H. Li, M. Wang, N. Li, T. Tian, Y. Wu, P. Xu, S. Yang, Z. Zhai, and L. Zhou, "Global burden of thyroid cancer from 1990 to 2017," *JAMA Netw. Open*, vol. 3, no. 6, 1990, Art. no. e208759.

[14] I. K. Kang, C. K. Jung, K. Kim, J. Park, J. S. Kim, and J. Bae, "Papillary thyroid carcinoma in a separate pyramidal lobe mimicking thyroglossal duct cyst carcinoma: A case report," *J. Endocrine Surg.*, vol. 22, no. 4, p. 138, 2022.

[15] K. Lee, C. Anastasopoulou, C. Chandran, and S. Cassaro, "Thyroid cancer," in *StatPearls [Internet]*. St. Petersburg, FL, USA: StatPearls Publishing, 2017.

[16] G. Seifert, K. Hennings, and J. Caselitz, "Metastatic tumors to the parotid and submandibular glands: Analysis and differential diagnosis of 108 cases," *Pathol.-Res. Pract.*, vol. 181, no. 6, pp. 684–692, 1986.

[17] J. Albores-Saavedra, D. E. Henson, E. Glazer, and A. M. Schwartz, "Changing patterns in the incidence and survival of thyroid cancer with follicular phenotype—Papillary, follicular, and anaplastic: A morphological and epidemiological study," *Endocrine Pathol.*, vol. 18, no. 1, pp. 1–7, May 2007.

[18] A. Bikas and K. D. Burman, "Epidemiology of thyroid cancer," in *The Thyroid and Its Diseases: A Comprehensive Guide for the Clinician*. Cham, Switzerland: Springer, 2019, pp. 541–547.

[19] L. Davies, L. G. T. Morris, M. Haymart, A. Y. Chen, D. Goldenberg, J. Morris, J. B. Ogilvie, D. J. Terris, J. Netterville, R. J. Wong, and G. Randolph, "American association of clinical endocrinologists and American college of endocrinology disease state clinical review: The increasing incidence of thyroid cancer," *Endocrine Pract.*, vol. 21, no. 6, pp. 686–696, Jun. 2015.

[20] L. Davies and H. G. Welch, "Current thyroid cancer trends in the United States," *JAMA Otolaryngol. Head Neck Surg.*, vol. 140, no. 4, p. 317, Apr. 2014.

[21] M. D. McCradden, J. A. Anderson, E. A. Stephenson, E. Drysdale, L. Erdman, A. Goldenberg, and R. Zlotnik Shaul, "A research ethics framework for the clinical translation of healthcare machine learning," *Amer. J. Bioethics*, vol. 22, no. 5, pp. 8–22, May 2022.

[22] F. M. Salman and S. S. Abu-Naser, "Thyroid knowledge based system," *Int. J. Academic Eng. Res.*, vol. 3, no. 5, pp. 11–20, May 2019.

[23] M. Clinic. (2021). *Thyroid Cancer—Symptoms and Causes*. Accessed: Dec. 10, 2023. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/thyroid-cancer/diagnosis-treatment/drc-20354167

[24] U. Feldt-Rasmussen and M. Klose, "Clinical strategies in the testing of thyroid function," in *Endotext [Internet]*, Nov. 2020.

[25] S. Kalra, S. Aggarwal, and D. Khandelwal, "Thyroid dysfunction and type 2 diabetes mellitus: Screening strategies and implications for management," *Diabetes Therapy*, vol. 10, no. 6, pp. 2035–2044, Dec. 2019.

[26] S.-B. Soh and T.-C. Aw, "Laboratory testing in thyroid conditions–pitfalls and clinical utility," *Ann. Lab. Med.*, vol. 39, no. 1, pp. 3–14, Jan. 2019.

[27] I. Iakovou, E. Giannoula, and C. Sachpekidis, "Imaging and imaging-based management of pediatric thyroid nodules," *J. Clin. Med.*, vol. 9, no. 2, p. 384, Feb. 2020.

[28] J. Patel, J. Klopper, and E. E. Cottrill, "Molecular diagnostics in the evaluation of thyroid nodules: Current use and prospective opportunities," *Frontiers Endocrinol.*, vol. 14, Feb. 2023, Art. no. 1101410.

[29] I. D. Mienye, P. Kenneth Ainah, I. D. Emmanuel, and E. Esenogho, "Sparse noise minimization in image classification using genetic algorithm and DenseNet," in *Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS)*, Mar. 2021, pp. 103–108.

[30] X. Zheng, S. Yu, J. Long, Q. Wei, L. Liu, C. Liu, and W. Ren, "Comparison of the clinical characteristics of primary thyroid lymphoma and diffuse sclerosing variant of papillary thyroid carcinoma," *Endocrine Connections*, vol. 11, no. 1, Jan. 2022, Art. no. e210364.

[31] K. Aruleba, G. Obaido, B. Ogbuokiri, A. O. Fadaka, A. Klein, T. A. Adekiya, and R. T. Aruleba, "Applications of computational methods in biomedical breast cancer imaging diagnostics: A review," *J. Imag.*, vol. 6, no. 10, p. 105, Oct. 2020.

[32] R. T. Aruleba, T. A. Adekiya, N. Ayawei, G. Obaido, K. Aruleba, I. D. Mienye, I. Aruleba, and B. Ogbuokiri, "COVID-19 diagnosis: A review of rapid antigen, RT-PCR and artificial intelligence methods," *Bioengineering*, vol. 9, no. 4, p. 153, Apr. 2022.

[33] J. Feng, R. V. Phillips, I. Malenica, A. Bishara, A. E. Hubbard, L. A. Celi, and R. Pirracchio, "Clinical artificial intelligence quality improvement: Towards continual monitoring and updating of AI algorithms in healthcare," *NPJ Digit. Med.*, vol. 5, no. 1, p. 66, May 2022.

[34] H. Bhatt, N. K. Jadav, A. Kumari, R. Gupta, S. Tanwar, Z. Polkowski, A. Tolba, and A. S. Hassanein, "Artificial neural network-driven federated learning for heart stroke prediction in healthcare 4.0 underlying 5G," *Concurrency Comput., Pract. Exper.*, vol. 36, no. 3, Feb. 2024, Art. no. e7911.

[35] K. Kumar, P. Kumar, D. Deb, M.-L. Unguresan, and V. Muresan, "Artificial intelligence and machine learning based intervention in medical infrastructure: A review and future trends," *Healthcare*, vol. 11, no. 2, p. 207, Jan. 2023.

[36] J. M. Meier and T. Tschoellitsch, "Artificial intelligence and machine learning in patient blood management: A scoping review," *Anesthesia Analgesia*, vol. 135, no. 3, pp. 524–531, 2022.

[37] I. D. Mienye, G. Obaido, K. Aruleba, and O. A. Dada, "Enhanced prediction of chronic kidney disease using feature selection and boosted classifiers," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2021, pp. 527–537.

[38] S. Poudel, "A study of disease diagnosis using machine learning," *Med. Sci. Forum*, vol. 10, no. 1, p. 8, 2022.

[39] H. Abbad Ur Rehman, C.-Y. Lin, Z. Mushtaq, and S.-F. Su, "Performance analysis of machine learning algorithms for thyroid disease," *Arabian J. Sci. Eng.*, vol. 46, no. 10, pp. 9437–9449, Oct. 2021.

[40] M. A. Asif, M. M. Nishat, F. Faisal, M. F. Shikder, M. H. Udoy, R. R. Dip, and R. Ahsan, "Computer aided diagnosis of thyroid disease using machine learning algorithms," in *Proc. 11th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2020, pp. 222–225.

[41] L. Aversano, M. L. Bernardi, M. Cimitile, M. Iammarino, P. E. Macchia, I. C. Nettore, and C. Verdone, "Thyroid disease treatment prediction with machine learning approaches," *Proc. Comput. Sci.*, vol. 192, pp. 1031–1040, Jan. 2021.

[42] A. Tyagi, R. Mehra, and A. Saxena, "Interactive thyroid disease prediction system using machine learning technique," in *Proc. 5th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Dec. 2018, pp. 689–693.

[43] S. S. Islam, M. S. Haque, M. S. U. Miah, T. B. Sarwar, and R. Nugraha, "Application of machine learning algorithms to predict the thyroid disease risk: An experimental comparative study," *PeerJ Comput. Sci.*, vol. 8, p. e898, Mar. 2022.

[44] R. Chaganti, F. Rustam, I. De La Torre Díez, J. L. V. Mazón, C. L. Rodríguez, and I. Ashraf, "Thyroid disease prediction using selective features and machine learning techniques," *Cancers*, vol. 14, no. 16, p. 3914, Aug. 2022.

[45] T. Alyas, M. Hamid, K. Alissa, T. Faiz, N. Tabassum, and A. Ahmad, "Empirical method for thyroid disease classification using a machine learning approach," *BioMed Res. Int.*, vol. 2022, pp. 1–10, Jun. 2022.

[46] K. Salman and E. Sonuç, "Thyroid disease classification using machine learning algorithms," *J. Phys., Conf. Ser.*, vol. 1963, no. 1, Jul. 2021, Art. no. 012140.

[47] D. C. Yadav and S. Pal, "Prediction of thyroid disease using decision tree ensemble method," *Hum.-Intelligent Syst. Integr.*, vol. 2, nos. 1–4, pp. 89–95, Dec. 2020.

[48] G. Ciaburro, "An ensemble classifier approach for thyroid disease diagnosis using the adaboostm algorithm," in *Machine Learning, Big Data, and IoT for Medical Informatics*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 365–387.

[49] P. Duggal and S. Shukla, "Prediction of thyroid disorders using advanced machine learning techniques," in *Proc. 10th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2020, pp. 670–675.

[50] G. Chaubey, D. Bisen, S. Arjaria, and V. Yadav, "Thyroid disease prediction using machine learning approaches," *Nat. Acad. Sci. Lett.*, vol. 44, no. 3, pp. 233–238, Jun. 2021.

[51] S. Mishra, Y. Tadesse, A. Dash, L. Jena, and P. Ranjan, "Thyroid disorder analysis using random forest classifier," in *Intelligent and Cloud Computing: Proceedings of ICICC 2019*, vol. 2. Cham, Switzerland: Springer, 2021, pp. 385–390.

[52] A. O. J., F. Ogwueleka, and P. O. Odion, "Effective and accurate bootstrap aggregating (Bagging) ensemble algorithm model for prediction and classification of hypothyroid disease," *Int. J. Comput. Appl.*, vol. 176, no. 39, pp. 40–48, Jul. 2020.

[53] L. Agilandeeswari, I. Khatri, J. Advani, and S. M. Nihal, "An efficient thyroid disease detection using voting based ensemble classifier," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2021, pp. 1395–1405.

[54] T. Akhtar, S. Arif, Z. Mushtaq, S. O. Gilani, M. Jamil, Y. Ayaz, and S. I. Butt, "Ensemble-based effective diagnosis of thyroid disorder with various feature selection techniques," in *Proc. 2nd Int. Conf. Smart Syst. Emerg. Technol. (SMARTTECH)*, May 2022, pp. 14–19.

[55] M. H. Alshayeji, "Early thyroid risk prediction by data mining and ensemble classifiers," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 3, pp. 1195–1213, Sep. 2023.

[56] N. Amgad, H. Haitham, M. Alabrak, and A. Mohammed, "Enhancing thyroid cancer diagnosis through a resilient deep learning ensemble approach," in *Proc. 6th Int. Conf. Comput. Informat. (ICCI)*, Mar. 2024, pp. 195–202.

[57] I. Balikçi Çiçek and Z. Küçükakçali, "Machine learning approach for thyroid cancer diagnosis using clinical data," *Middle Black Sea J. Health Sci.*, vol. 9, no. 3, pp. 440–452, Aug. 2023.

[58] I. D. Mienye and Y. Sun, "Effective feature selection for improved prediction of heart disease," in *Proc. Pan-Afr. Artif. Intell. Smart Syst. Conf.* Cham, Switzerland: Springer, 2021, pp. 94–107.

[59] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *J. Big Data*, vol. 5, no. 1, pp. 1–30, Dec. 2018.

[60] U. Mgboh, B. Ogbuokiri, G. Obaido, and K. Aruleba, "Visual data mining: A comparative analysis of selected datasets," in *Proc. Int. Conf. Intell. Syst. Design Appl.*, 1351th ed., A. Abraham, V. Piuri, N. Gandhi, P. Siarry, A. Kaklauskas, and A. Madureira, Eds. Cham, Switzerland: Springer, 2021, pp. 377–391.

[61] T. G. Nick and K. M. Campbell, "Logistic regression," in *Topics Biostatistics*. Springer, 2007, pp. 273–301.

[62] G. Obaido, B. Ogbuokiri, C. W. Chukwu, F. J. Osaye, O. F. Egbelowo, M. I. Uzochukwu, I. D. Mienye, K. Aruleba, M. Primus, and O. Achilonu, "An improved ensemble method for predicting hyperchloremia in adults with diabetic ketoacidosis," *IEEE Access*, vol. 12, pp. 9536–9549, 2024.

[63] G. Obaido, B. Ogbuokiri, I. D. Mienye, and S. M. Kasongo, "A voting classifier for mortality prediction post-thoracic surgery," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2022, pp. 263–272.

[64] B. Ogbuokiri, A. Ahmadi, N. L. Bragazzi, Z. M. Nia, B. Mellado, J. Wu, J. Orbinski, A. Asgary, and J. Kong, "Public sentiments toward COVID-19 vaccines in South African cities: An analysis of Twitter posts," *Frontiers Public Health*, vol. 10, Aug. 2022, Art. no. 987376.

[65] B. Ogbuokiri, A. Ahmadi, Z. M. Nia, B. Mellado, J. Wu, J. Orbinski, A. Asgary, and J. Kong, "Vaccine hesitancy hotspots in Africa: An insight from geotagged Twitter posts," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 1, Feb. 2024.

[66] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020.

[67] G. Obaido, B. Ogbuokiri, T. G. Swart, N. Ayawei, S. M. Kasongo, K. Aruleba, I. D. Mienye, I. Aruleba, W. Chukwu, F. Osaye, O. F. Egbelowo, S. Simphiwe, and E. Esenogho, "An interpretable machine learning approach for hepatitis b diagnosis," *Appl. Sci.*, vol. 12, no. 21, p. 11127, Nov. 2022.

[68] Y. Zhang, "Support vector machine classification algorithm and its application," in *Proc. Int. Conf. Inf. Comput. Appl.*, Chengde, China. Cham, Switzerland: Springer, Jun. 2012, pp. 179–186.

[69] B. Ogbuokiri, A. Ahmadi, B. Mellado, J. Wu, J. Orbinski, A. Asgary, and J. Kong, "Can post-vaccination sentiment affect the acceptance of booster jab?" in *Intelligent Systems Design and Applications*, A. Abraham, S. Pllana, G. Casalino, K. Ma, and A. Bajaj, Eds. Cham, Switzerland: Springer, 2023, pp. 200–211.

[70] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers—A tutorial," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–25, Jul. 2022.

[71] J. J. Valero-Mas, A. J. Gallego, P. Alonso-Jiménez, and X. Serra, "Multilabel prototype generation for data reduction in K-nearest neighbour classification," *Pattern Recognit.*, vol. 135, Mar. 2023, Art. no. 109190.

[72] R. Ahuja, A. Chug, S. Gupta, P. Ahuja, and S. Kohli, "Classification and clustering algorithms of machine learning with their applications," in *Nature-Inspired Computation in Data Mining and Machine Learning*. Cham, Switzerland: Springer, 2020, pp. 225–248.

[73] Y. Izza, A. Ignatiev, and J. Marques-Silva, "On tackling explanation redundancy in decision trees," *J. Artif. Intell. Res.*, vol. 75, pp. 261–321, Sep. 2022.

[74] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261–283, Apr. 2013.

[75] S. Singh and P. Gupta, "Comparative study ID3, cart and C4. 5 decision tree algorithm: A survey," *Int. J. Adv. Inf. Sci. Technol.*, vol. 27, no. 27, pp. 97–103, 2014.

[76] E. Heidari, M. A. Sobati, and S. Movahedirad, "Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN)," *Chemometric Intell. Lab. Syst.*, vol. 155, pp. 73–85, Jul. 2016.

[77] T. Sathish, P. Sunagar, V. Singh, S. Boopathi, R. Sathyamurthy, A. M. Al-Enizi, B. Pandit, M. Gupta, and S. S. Sehgal, "Characteristics estimation of natural fibre reinforced plastic composites using deep multilayer perceptron (MLP) technique," *Chemosphere*, vol. 337, Oct. 2023, Art. no. 139346.

[78] H. Taud and J. Mas, "Multilayer perceptron (MLP)," in *Geomatic Approaches for Modeling Land Change Scenarios*. Cham, Switzerland: Springer, 2018, pp. 451–455.

[79] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using MLP classifier," *Appl. Acoust.*, vol. 180, Sep. 2021, Art. no. 108077.

[80] B. Azhagusundari and A. S. Thanamani, "Feature selection based on information gain," *Int. J. Innov. Technol. Exploring Eng.*, vol. 2, no. 2, pp. 18–21, 2013.

[81] I. D. Mienye and Y. Sun, "A machine learning method with hybrid feature selection for improved credit card fraud detection," *Appl. Sci.*, vol. 13, no. 12, p. 7254, Jun. 2023.

[82] R. Polikar, "Ensemble learning," in *Ensemble Machine Learning: Methods and Applications*. New York, NY, USA: Springer, 2012, pp. 1–34.

[83] G. I. Webb and Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 8, pp. 980–991, Aug. 2004.

[84] J. Dou, A. P. Yunus, D. T. Bui, A. Merghadi, M. Sahana, Z. Zhu, C.-W. Chen, Z. Han, and B. T. Pham, "Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan," *Landslides*, vol. 17, no. 3, pp. 641–658, Mar. 2020.

[85] A. Gupta, V. Jain, and A. Singh, "Stacking ensemble-based intelligent machine learning model for predicting Post-COVID-19 complications," *New Gener. Comput.*, vol. 40, no. 4, pp. 987–1007, Dec. 2022.

[86] M. G. Meharie, W. J. Mengesha, Z. A. Gariy, and R. N. N. Mutuku, "Application of stacking ensemble machine learning algorithm in predicting the cost of highway construction projects," *Eng., Construct. Architectural Manage.*, vol. 29, no. 7, pp. 2836–2853, Aug. 2022.

[87] B. Pavlyshenko, "Using stacking approaches for machine learning models," in *Proc. IEEE 2nd Int. Conf. Data Stream Mining Process. (DSMP)*, Aug. 2018, pp. 255–258.

[88] J. Yan and S. Han, "Classifying imbalanced data sets by a novel RE-sample and cost-sensitive stacked generalization method," *Math. Problems Eng.*, vol. 2018, pp. 1–13, Sep. 2018.

[89] S. Rajagopal, P. P. Kundapur, and K. S. Hareesha, "A stacking ensemble for network intrusion detection using heterogeneous datasets," *Secur. Commun. Netw.*, vol. 2020, pp. 1–9, Jan. 2020.

[90] Y. Kim and E. Riloff, "Stacked generalization for medical concept extraction from clinical notes," in *Proc. BioNLP*, 2015, pp. 61–70.

[91] T. Aboneh, A. Rorissa, and R. Srinivasagan, "Stacking-based ensemble learning method for multi-spectral image classification," *Technologies*, vol. 10, no. 1, p. 17, Jan. 2022.

[92] F. Divina, A. Gilson, F. Goméz-Vela, M. G. Torres, and J. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," *Energies*, vol. 11, no. 4, p. 949, Apr. 2018.

[93] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Mach. Learn.*, vol. 36, pp. 105–139, Jul. 1999.

[94] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992.

[95] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.

[96] M. Liang, T. Chang, B. An, X. Duan, L. Du, X. Wang, J. Miao, L. Xu, X. Gao, L. Zhang, J. Li, and H. Gao, "A stacking ensemble learning framework for genomic prediction," *Frontiers Genet.*, vol. 12, Mar. 2021, Art. no. 600040.

[97] R. Sikora, "A modified stacking ensemble machine learning algorithm using genetic algorithms," in *Handbook Res. Organizational Transformations Through Big Data Analytics*. Hershey, PA, USA: IGI Global, 2015, pp. 43–53.

[98] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," 2018, *arXiv:1809.03006*.

[99] A. Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdiscipl. J. Inf., Knowl., Manage.*, vol. 14, pp. 45–76, Jan. 2019.

[100] M. Hao, Y. Wang, and S. H. Bryant, "An efficient algorithm coupled with synthetic minority over-sampling technique to classify imbalanced PubChem BioAssay data," *Analytica Chim. Acta*, vol. 806, pp. 117–127, Jan. 2014.

[101] M. I. Prasetiyowati, N. U. Maulidevi, and K. Surendro, "Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest," *J. Big Data*, vol. 8, no. 1, p. 84, Dec. 2021.

[102] H. M. Almahshi, E. A. Almasri, H. Alquran, W. A. Mustafa, and A. Alkhayyat, "Hypothyroidism prediction and detection using machine learning," in *Proc. 5th Int. Conf. Eng. Technol. Appl. (IICETA)*, May 2022, pp. 159–163.

[103] M.-H. Tsai, J. T. C. Chang, H.-H. Lu, Y.-H. Wu, T.-H. Pao, Y.-J. Cheng, W.-Y. Zheng, C.-Y. Chou, J.-H. Lin, T. Yu, and J.-H. Chiang, "Development and validation of a machine learning model of radiation-induced hypothyroidism with clinical and dose–volume features," *Radiotherapy Oncol.*, vol. 189, Dec. 2023, Art. no. 109911.

[104] Y. Endo, "Tumor size and presence of calcifications on ultrasonography are pre-operative predictors of lymph node metastases in patients with papillary thyroid cancer," *Ultrasound Quart.*, vol. 28, no. 2, p. 129, Jun. 2012.

[105] G. R. Kim, M. H. Kim, H. J. Moon, W. Y. Chung, J. Y. Kwak, and E.-K. Kim, "Sonographic characteristics suggesting papillary thyroid carcinoma according to nodule size," *Ann. Surgical Oncol.*, vol. 20, no. 3, pp. 906–913, Mar. 2013.

[106] K. Kobayashi, T. Fujimoto, H. Ota, M. Hirokawa, T. Yabuta, H. Masuoka, M. Fukushima, T. Higashiyama, M. Kihara, Y. Ito, A. Miya, and A. Miyauchi, "Calcifications in thyroid tumors on ultrasonography: Calcification types and relationship with histopathological type," *Ultrasound Int. Open*, vol. 4, no. 2, pp. E45–E51, Apr. 2018.

[107] S. C. Kamran, E. Marqusee, M. I. Kim, M. C. Frates, J. Ritner, H. Peters, C. B. Benson, P. M. Doubilet, E. S. Cibas, and J. Barletta, "Thyroid nodule size and prediction of cancer," *J. Clin. Endocrinol. Metabolism*, vol. 98, no. 2, pp. 564–570, 2013.

[108] M. K. Gould, J. Donington, W. R. Lynch, P. J. Mazzone, D. E. Midthun, D. P. Naidich, and R. S. Wiener, "Evaluation of individuals with pulmonary nodules: When is it lung cancer: Diagnosis and management of lung cancer: American college of chest physicians evidence-based clinical practice guidelines," *Chest*, vol. 143, no. 5, pp. e93S–e120S, 2013.

[109] M. A. Knox, "Thyroid nodules," *Amer. Family Physician*, vol. 88, no. 3, pp. 193–196, 2013.

[110] R. B. Valentini, B. M. D. Macedo, R. F. Izquierdo, and E. L. S. Meyer, "Painless thyroiditis associated to thyroid carcinoma: Role of initial ultrasonography evaluation," *Arch. Endocrinol. Metabolism*, vol. 60, no. 2, pp. 178–182, Apr. 2016.

[111] S. Ji, "SSC: The novel self-stack ensemble model for thyroid disease prediction," *PLoS ONE*, vol. 19, no. 1, Jan. 2024, Art. no. e0295501.

[112] M. Karmeni, E. B. Abdallah, K. Boukadi, and M. Abed, "Towards an accurate stacked ensemble learning model for thyroid earlier detection," in *Proc. IEEE/ACS 19th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Dec. 2022, pp. 1–8.

[113] D. C. Yadav and S. Pal, "Decision tree ensemble techniques to predict thyroid disease," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 8242–8246, Sep. 2019.

GEORGE OBAIDO (Member, IEEE) received the master's and Ph.D. degrees in computer science from the University of the Witwatersrand, Johannesburg, South Africa. He is currently a Berkeley Institute for Data Science (BIDS)-Center for Human-Compatible Artificial Intelligence (CHAI) Postdoctoral Research Fellow with the University of California at Berkeley, Berkeley, CA, USA. His research interest includes machine learning to find solutions to problems of societal importance.

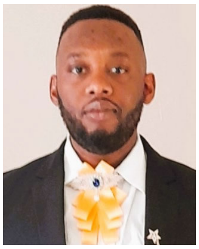
OKECHINYERE ACHILONU received the M.Sc. degree in statistics and actuarial science and the Ph.D. degree in biostatistics from the University of the Witwatersrand, Johannesburg, South Africa. She is currently a Biostatistics Lecturer with the University of the Witwatersrand, Johannesburg. She has several years of experience in clinical analysis, statistical support, and teaching. She is passionate about developing statistical and machine-learning models from structured and unstructured clinical and epidemiological research data. Her research interests include developing and assessing analytical frameworks for knowledge discovery toward improving predictive and personalized medicine, focusing on non-communicable diseases, such as cancer, chronic kidney failure, and diabetes.

BLESSING OGBUOKIRI received the Ph.D. degree in computer science from the University of the Witwatersrand, Johannesburg, South Africa. He is currently an Assistant Professor with the Department of Computer Science, Brock University, Canada. He was previously a Postdoctoral Fellow and an Instructor with The Africa-Canada Artificial Intelligence and Data Innovation Consortium Laboratory, Department of Mathematics and Statistics, York University, Toronto, Canada. He collaborates with researchers from various fields. His research interests include the intersection of AI and health, helping communities, and governments tackle infectious diseases. He was the Co-Chair of the NeurIPS 2023 Conference (Affinity Workshops) and organized the Black in AI Workshop. This demonstrates his commitment to making AI inclusive and sharing knowledge within the community. Through his efforts, he is using AI to address real health issues effectively.

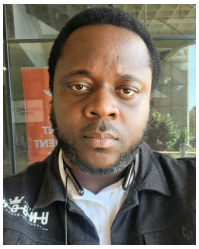
CHIMEREMMA SANDRA AMADI received the bachelor's and master's degrees from the Federal University of Technology, Owerri (FUTO), Nigeria. She is currently an Information Technology (IT) Lecturer with FUTO. Her research pursuits revolve around harnessing the power of machine learning to address pressing societal challenges. She is deeply committed to exploring real-world applications of machine learning, particularly in areas where its impact can be transformative and where the challenges are most formidable.

LAWAL HABEEBULLAHI received the B.Sc. degree in computer science from Usmanu Danfodiyo University Sokoto, Nigeria. He is currently a Research Fellow with the Innovation and Advance Science Research Group, Summit University, Offa, Nigeria. His research interests include machine learning, data science, and drone building and piloting reflect a diverse and multidisciplinary approach to problem-solving and innovation.

TONY OHALLORAN received the bachelor's degree in mathematical sciences from University College Cork. He is currently pursuing the master's degree in artificial intelligence with the National University of Ireland, Galway. He is also an Analytics Specialist at Evervault. He is passionate about using technology to improve people's lives. His research interests include the intersection of AI and healthcare.



CHIDOZIE WILLIAMS CHUKWU received the master's and Ph.D. degrees in applied mathematics from the University of Johannesburg, South Africa. He is currently a Visiting Assistant Professor with Wake Forest University, Winston-Salem, NC, USA. His research interests include mathematical biology and the application of machine learning to predict future trends in the emergence and resurgence of epidemics.



EBIKELLA DOMOR MIENYE received the bachelor's and master's degrees in accountancy. He is currently pursuing the Ph.D. degree in taxation with the College of Business and Economics, University of Johannesburg, South Africa, demonstrating his commitment to advancing his expertise and knowledge. His research interests include tax compliance and taxation and the fourth industrial revolution.

MIKAIL ALIYU received the master's degree in public health, focusing on global health, from the University of Leeds, U.K., and the Doctor of Public Health (DrPH) degree from the University of California at Berkeley, Berkeley, CA, USA. His work has provided health systems with essential tools and strategies to address both communicable and non-communicable diseases. His primary research interests include enhancing health systems' capabilities for prompt disease detection, optimizing treatment methods, and establishing resilient health infrastructure.

OLUFUNKE FASAWE is currently pursuing the Doctor of Public Health (DrPH) degree with the School of Public Health, University of California at Berkeley, Berkeley, CA, USA. She was with Clinton Health Access Initiative (CHAI)-based in Nigeria where she was the Senior Director for Primary Health Care Global Strategy and the Director of Programs and Lead for the Sexual, Reproductive, Maternal, Newborn, and Child Health Portfolio of programs for CHAI Nigeria office. She has more than ten years of experience in global health working on policy, program design and planning, implementation, grant management, monitoring, and evaluation at the country level.

IBUKUNOLA ABOSEDE MODUPE is currently a Lecturer with the Vaal University of Technology. She has many years of experience in the industry and academia. Her primary research interests include machine learning, computational linguistics, natural language processing (NLP), neural machine translation, low-resource language, and zero-short learning

EREPAMO JOB OMIETIMI received the B.Sc. degree in geology from the University of Benin, the M.Sc. degree in petroleum geology and sedimentology from the University of Ibadan, Nigeria, and the Ph.D. degree in geology from the University of Pretoria, South Africa. His research interests include machine learning, environmental geochemistry, paleoenvironment, and climate studies.

KEHINDE ARULEBA received the Ph.D. degree in computer science from the University of the Witwatersrand, Johannesburg, South Africa. He is currently a Lecturer with the University of Leicester, U.K. Before that, he was a Postdoctoral Fellow with Walter Sisulu University, South Africa. His expertise extends beyond computer science. He actively collaborates with researchers across disciplines. His research interests include machine learning, data ethics, and ICT4D.

• • •