Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 23s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Weighted Fuzzy-Based Model for Thyroid Prediction in Distributed Environment

Amutha C[1], Dr. P. Periyasamy[2]

[1]Research Scholar, Department of Computer Science, Parks College (Autonomous), Chinnakarai, Tripur, Affiliated to Bharathiar University, Coimbatore, Tamilnadu, India,

Email ID: arjammusan24@gmail.com

[2]Associate Professor, Department of Computer Science, Parks College (Autonomous), Chinnakarai, Tripur, Affiliated to Bharathiar University, Coimbatore, Tamilnadu, India,

Email ID: drperi1975@gmail.com

Cite this paper as: Amutha C, Dr. P. Periyasamy, (2025) A Weighted Fuzzy-Based Model for Thyroid Prediction in Distributed Environment. *Journal of Neonatal Surgery*, 14 (23s), 377-391

ABSTRACT

Thyroid disorder is the most widespread endocrine disorders globally. Thyroid gland plays a crucial role in regulating human metabolism, its dysfunction poses significant health concerns. An automated, dependable, and precise machine learning (ML) system for thyroid disease identification is essential to improve diagnostic efficiency and minimize errors. The proposed model seeks to address several limitations in existing approaches, including insufficient feature analysis, inadequate visualization, and the need for enhanced prediction accuracy and reliability. This study utilizes a publicly available thyroid infected dataset from the UCI-ML repository, which includes twenty-nine clinical features. These features were instrumental in developing a weighted fuzzy model capable of predicting thyroid disease by evaluating initial signs and symptoms and eliminating the need for manual assessment of these characteristics. The feature assessment and classification help in identifying the contribution of each feature to thyroid disease prediction. To overcome the issue of over-fitting, the model employs data normalization. The use of weighted fuzzy further strengthens the system's reliability by evaluating multiple classifiers in the decision-making process. The suggested framework achieved an impressive sensitivity of 99%, accuracy of 99.5%, and specificity of 99.9% demonstrating it's potential for real-time CAD systems, thereby facilitating timely recognition and appropriate treatment in its early stage..

Keywords: *thyroid, mining approaches, prediction, accuracy, evaluation*.

1. INTRODUCTION

The significant increase in human metabolism, reproductive activity, and neuronal development is primarily regulated by thyroid hormones [1]. When the thyroid gland is not capable to consistently produce optimal hormone levels, regular bodily functions in humans are affected. This condition is referred to as thyroid dysfunction. Medically, thyroid problems can manifest as thyroid cancer or thyroiditis. The two main thyroid-related disorders are hypothyroidism and hyperthyroidism [2]. Worldwide, the prevalence of thyroid disorders is rising, with up to thirty to forty percent of patients at endocrinology clinics being affected. In the United States, approximately sixty percent of individuals are unaware they have a thyroid condition, which is estimated to impact 20 million people [3]. Given that the symptoms of thyroid disease can be difficult to differentiate from those of other conditions, diagnosing the illness can be a intricate and requires more time. Traditionally, diagnosing thyroid disorders involves multiple blood tests and thorough examination by medical professionals. However, the primary challenge is in accurately diagnosing the condition in its early stages [4].

The accuracy and efficiency of diagnosing thyroid illnesses have been greatly enhanced through the use of modern computational techniques. By analyzing vast amounts of patient data, these tools can identify patterns and anomalies that traditional diagnostic methods might overlook. These techniques enable a comprehensive thyroid function evaluation by processing diverse datasets, including patient histories, ultrasound images, and blood test results [5]. Their ability to distinguish between multiple thyroid illness, such as hyperthyroidism and hypothyroidism, facilitates more precise and personalized treatment for patients. These computational methods are capable of processing large, complex datasets, and they improve their predictive capabilities over time by understanding from historical data. Additionally, they can integrate data from various sources like genetic information, lifestyle factors, and environmental influences—to provide a holistic view of a patient's health [6]. This integrative approach can detect thyroid disorders before symptoms appear, enabling early diagnosis and intervention [7].

The application of these advanced computational techniques represents a breakthrough in medical diagnostics, offering patients with thyroid disease faster, more accurate, and personalized healthcare solutions [8]. Achieving high levels of recall, accuracy, and precision in thyroid diagnostics requires a multimodal approach that addresses both model performance and practical, real-world application. Advanced algorithms and rigorous validation of diverse datasets reflecting various thyroid conditions are necessary to ensure robust accuracy [9]. Furthermore, optimizing for faster computation times is critical, as is employing efficient preprocessing techniques and algorithms without compromising accuracy. To ensure the model's generalisability across different datasets, a scalable architecture capable of handling diverse data distributions and demographic variability must be developed [10]. In addition, the model's usability for healthcare professionals is enhanced by incorporating user-friendly interfaces and interpretability tools, promoting informed decision-making and improving diagnostic outcomes for thyroid analysis [11]–[12]. This research seeks to advance the diagnosis of thyroid disease, paving the way for significant progress in medical practice. The study's primary objective is to accurately detect various thyroid disorders using feature-based classification techniques, allowing for timely and targeted medical interventions [13]. Furthermore, to facilitate practical implementation in clinical environments—where prompt diagnosis is crucial—the research focuses on optimizing computational efficiency to minimize execution times. By assessing the infrastructure requirements for deploying such models in hospitals, this study aims to improve patient outcomes and drive forward innovations in healthcare delivery [14]–[15]. The primary contributions of this work are outlined as follows:

1) To propose a novel weighted fuzzy model for thyroid prediction based on weighted rule analysis and fuzzification;

2) A weighted fuzzy model enhances the reliability of the prediction model by comparing multiple classifiers, rather than relying on a single machine learning algorithm.

The research is summarized as follows: section 2 compares existing methods. The methodology used is eloborated in section 3, and numerical findings are given in section 4. The outcomes are reviewed in section 5.

2. RELATED WORKS

A vital classification challenge is an ML method for accurately assessing thyroid disease. An integral part of our bodies is the thyroid gland. It aids in controlling our metabolism. Thyroid hormone deficiency results in hypothyroidism while hyperthyroidism results as an increase in the production of thyroid hormone. The author aimed to develop a machine learning (ML)-based classification framework for classifying people with thyroid disease from a publicly available data repository. With several explanatory criteria, the subjects were divided into healthy and thyroid disease categories [16]. A class balancer known as the SMOTE was employed to account for the minority class (thyroid infection) in the dataset. Azure ML tools were integrated with filter-based feature selection techniques to develop a prediction model, particularly mutual knowledge applied to a two-class NN classification algorithm. The proposed two-class NN technique, constructed using SMOTE and specified variables, performed better than other ML approaches in terms of performance metrics [17]. Thyroid issues are a worldwide health concern that requires timely and precise assessments. Three-phase learning on ML algorithms for thyroid illness detection was carried out by [18]. First, the SMOTE is used to adjust for the imbalance in the dataset. The prediction accuracy of various algorithms, including DT, RF, SVM, K-NN, and NB, is then assessed. Each method's efficacy is evaluated using several performance metrics. The outcomes contribute to improving trustworthy clinical decision-making tools in healthcare by providing crucial insights into the optimal machine-learning technique for thyroid ailment diagnosis [19].

Although many studies have been conducted on DM in data mining, there is currently no standardized procedure for assessing the efficacy of current approaches. Therefore, by integrating DM algorithms with less preprocessing, the author in [20] wishes to suggest a novel method for creating RIs, with the efficacy of five strategies assessed. Two records were collected from individuals who underwent physical examinations. The Test data set employed the Hoffmann, Bhattacharya, EM, Kosmic, and refineR algorithms and the two-step data preparation process to generate RIs for thyroid-related hormone compounds. Conventional RIs calculated from the reference statistic set, which only comprised individuals who fulfilled strict eligibility and exclusion criteria, have been compared to algorithm-derived RIs. They use the BR matrix. Hormone resistance intervals associated with the thyroid have been identified [21]. The EM approach may not work well for another hormone but consistently establishes TSH RIs with conventional TSH RIs. The RIs for free and total triiodo-thyronine and free and total thy, calculated by Hoffmann, Bhattacharya, and refined approaches, are comparable to and enhance the usual values [22]. Based on the BR matrix, an effective technique for precisely evaluating the method's efficacy is created [23].

When combined with less initial processing, the EM technique can handle highly skewed data, but it has drawbacks in other situations. The other four methods perform well when dealing with data with a Gaussian or near-Gaussian distribution [24] – [25]. The strategy that is most appropriate for the information distribution should be used. One of the least prevalent endocrinopathies in the world is thyroid disease. People should be concerned about thyroid disease since the gland that makes the thyroid hormone controls a person's metabolism. An autonomous, reliable, and effective thyroid identification machine learning framework is needed to save time and prevent errors. The author in [26] study aims to improve forecast accuracy and dependability while addressing current work limits, including a lack of in-depth characteristic analysis and presentation. The UCI-ML database used 29 healthcare variables from a public thyroid illness dataset. The healthcare factors helped us

create a machine learning algorithm that could identify thyroid disease by evaluating the early signs and symptoms and replacing manual inspection of these parameters [27]. We may better understand the significance of qualities in thyroid prognostic activities with the use of feature evaluation and representation [28].

Moreover, a 5-fold CV method and data normalization with the SMOTE were used to eliminate the over-fitting problem. The predicted result model is more dependable in combined learning since prognostic judgments are made using various classifiers. The boosting technique produced the proposed model, which might be applied to real-time CAD systems to promote timely intervention and ease of assessment. In medical science, thyroid prediction is a complex assumption [29]. The healthcare industry can now handle enormous volumes of medical records thanks to ML algorithms that are becoming more and more potent and small. Researchers can use a range of data values to generate predictions with ML approches. Data cleaning strategies are used in the dataset and produce trustworthy results. Data preparation techniques are applied to deal with missing and noisy values. The author used the Adaboost and Bagging algorithms to categorize thyroids. The most effective technique for thyroid forecasting is identified by testing the ways and comparing the results [30].

3. METHODOLOGY

This section considers the online dataset for thyroid prediction where the sample variables are considered for prediction purposes. The variables have different ranges, so it's important to adjust the data to bring them to a similar scale. This process is called normalization, and it is done using the following Eq. (1):

$$X_{normalized} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

In this process, $X_{normalized}$ is the new, adjusted value. The smallest score for each feature is declared as zero, the largest score is declared as one, and all other values are changed to numbers between zero and one.

3.1. Association Rule Mining Algorithm

Based on the user-defined minimal support with less confidence responsibility from the transaction database D, the association rules were retrieved from the association rule mining. The procedure to be followed during the mining operation consists of the next two steps. First, we identify every item set frequently with a threshold higher than the available one. Second, a comparable strong association rule should be created based on the frequent item sets obtained. Support and confidence, respectively, are used to produce the rule. Furthermore, interest can be used to eliminate unnecessary or useless association rules and extract valuable ones. The second pace of the two steps above is relatively simple; it is necessary to document all manageable association rules that are reliant on the usual thing sets that have been discovered; the help edge and certainty edge control the performance measure; and the affiliation rules both satisfy the help limit and confidence saturation to produce association rules that meet the requirements.

3.2. Associating weighted rule model

An important topic in intelligence-based database discovery is association rules, which show the intriguing relationships between data items in transaction databases. The attributes of the database are uniform and equal. This means there is no distinction between importance and subordination among the items in the database; instead, the frequency of occurrence is used to determine each item's importance. The elements in the database are distributed uniformly, meaning they occur at the same or comparable frequency. However, the items in the transaction dataset with actual transactions are frequently dispersed unevenly, and in some cases, the frequency may vary significantly. These will cause the database's data objects to be distributed unevenly. The least amount support is set high or low will be an issue when this occurs. Setting it too high could prevent the association rules from involving objects with lower frequencies; setting it too low could result in many unimportant association rules and cause a combination explosion, decreasing the algorithm's effectiveness. To reflect the varying relevance of the items in the database, we have developed the idea of the weighted item which allows us to assign different items to different weights and, in turn, enhance the issue method, also known as the weighted association protocols.

3.3. Associating keyword-based model

The organization of the document vector is the term that appears most frequently. The frequently occurring keywords are identified based on the document's frequent information. The noun often determines how the error is generated. Words from high-frequency documents are commonly used with little regard for the source's content. For instance, information mining algorithms use bunching strategies to combine similar terms.

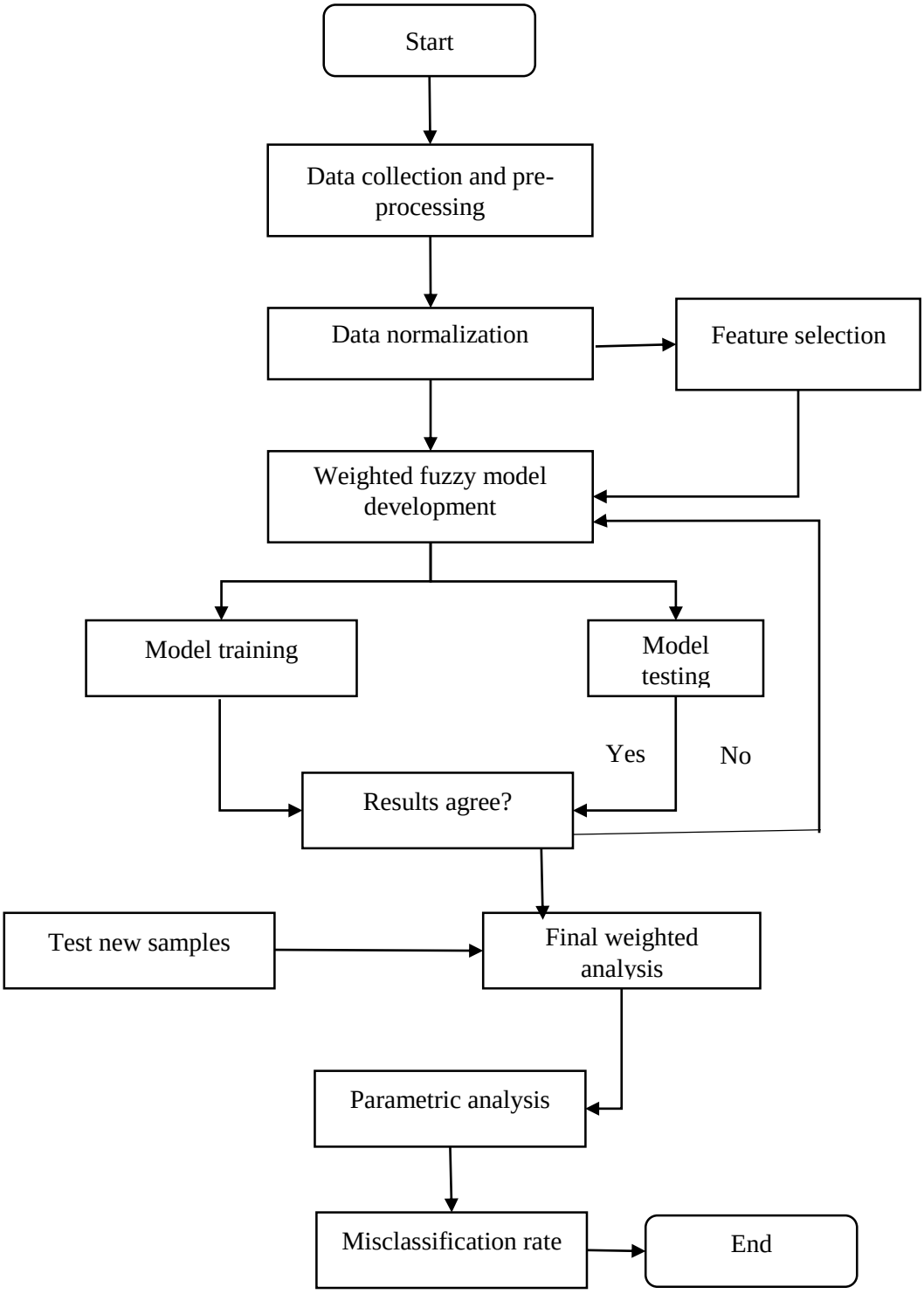

Fig 1. Proposed flow diagram

The category focuses on comparison articles. It establishes a document vector based on the weight and motivation from the terms recurrence and document recurrence, and it views all words except stop words as indicative of the document. The terms and their weight scores relate to a document, and the sets are essentially the exceptional parts of the document vector. The document frequency is used to construct the vector, and the word frequencies serve as essential building blocks. Each term's weight estimates suggested a positioning score. It is a key factor in shaping the document. This served as an evaluation of the phrase as a stopword or watchword for the research. The weighing scheme algorithm has two points: a) A discriminative value distinguishes the documents from each other. b) A stop word or keyword is a crucial metric. Data mining

techniques discover the relationship between documents. Frequency-based term weighting is a statistical method for determining the relationship between words. The process above is the most commonly used strategy to differentiate and describe the document from others. The aforementioned data mining methods work effectively in the application of document classification. Good retrieval systems and clustering approaches are the most widely utilized classification methods. To define the document, the scheme is assessed using a criterion known as an evaluation standard. The measure is frequency statistics and the aforementioned evaluation measures will no longer be the most effective way to describe the phrases.

The keyword-based term weighting (KBW) approach is an additional technique. This approach separates the document using a keyword value factor. The method is analytically based since it is subjected to analytical analysis. This method retrieves the document's keyword list. One of the essential elements is used to measure the term. It was created as a document's keyword. The technique employs a keyword ranking score to rank the measured values from the elements that have been pooled. The results of the document are used to calculate the measured ranking scores using these weighting approaches. The keyword-based weighting strategy has addressed the many drawbacks of the frequency-based weighting approach. There are several keywords in the document, each taken from the results. The keyword positioning strategy uses several criteria, including the report, the region, and the function of the words in a sentence or section. A record's representative phrases are its substance expressions. By separating the catchphrase from the type of text and testing the text's substance, substance words are eliminated from the content. The exploration document occasionally uses keywords as evidenced by the abstract, experiment, title, body, conclusion, and catchphrases in the title or the first section of the text. There are several ways to choose a catchphrase, and they can be categorized as word level, passage level, sentence level, and content dimension aspects. The situation job data and word-level grammatical features are the most important highlights. Compound nouns, numerals, common nouns, and formal persons, places, or things are the most commonly used grammatical forms. The sentence region, express or proviso, and sentence type are the most often used for syntactic or sentence-level highlights. The way a sentence is constructed and the expressive manner of the words used in it affect the meaning of the statement and the terminology employed. When a term in the subjective clause has a weighting scheme different from that of the auxiliary clause or modifying clause, the two are unequal. The type of term determines the weight created as a feature that appears in the text.

The weight score of a term in a text is used to determine the characteristic of a sentence, phrase, term, or expression and it is taken from every single document. Each data archive characterizes the first stage of catchphrases and processes their weight assessments. The document's words and keywords are stored in an inverted file structure. This framework provides a dependable method for cluster enlargement. The original cluster k is where documents are added. Predicting thyroid is the most challenging task in the medical world and data mining techniques are used to make this prediction. One of the primary uses of data mining is thyroid prediction. It uses vast healthcare information; computerized instruments are being gathered from scientific organizations and clinical information. According to the results of the DM approaches, the researchers can find and investigate new patterns and characterize the links among a vast number of variables with the help of the current simulated automated medical instrument. The past datasets are used to anticipate the outcomes.

3.4. Data fuzzification

We have created a unique fuzzification technology feature to enhance data interpretability during the last processing stage. This final step is essential because a large portion of the data from the system is difficult for consumers to understand and represent. After all, it is a continuous data with measurements that are frequently difficult to comprehend. Fuzzifying these data can enhance the outcomes that the mining algorithms discover while making the results more comprehensible. According to the parameters in the dataset or expert-supplied data, we suggest a fuzzification technique that automatically processes data values based on their distribution (refer to the kinds of input supplied to the method). To achieve this author generated a distributed algorithm in Spark that adheres to the MapReduce concept. This enables the researcher to handle massive volumes of data that are kept by many healthcare centers within the healthcare system. Algorithm 1 describes the overall procedure. Dataset are the inputs for the algorithm. The distributions and descriptions of the parameters the experts have established are kept in the dictionary. However, for parameters that consumers have not specified, a default amount of labels is applied, and they will be generated randomly according to their distribution. We will give the algorithm a dictionary with the labels and the validation function used for the dataset variable if the values depend on the dataset parameters. The entire procedure will be handled in a distributed fashion. For distributed processing, MATLAB 2020a automatically separates the data into pieces. This has been defined by utilizing the terms DC (distribute computing) and S_i to symbolize each data point. The method that distributes the computation through MapReduce uses the cluster's global variable. Furthermore, the process invokes the fuzzification function outlined. There are various sections to this function. To generate fresh fuzzified parameters, it initially determines whether the variable's name appears in the Intervals hash list. If it does, it uses the names of the labels that Intervals specify and its configuration that pertains to the defined space. An automatic process that splits the variable's values into multiple intervals described in DefaultIntervals based on the variable's percentiles is employed if the variable cannot be located in the database. An example with DefaultIntervals set to three is displayed in The percentiles utilized in this instance were 25 and 37.5 for establishing the first label's trapezoidal pattern and the left portion of the 2nd label, and 62,5 and 75 for determining the second label's right portion and the third label. Thus, using the appropriate percentiles, the GenerateIntervals procedure splits the collection into k equally distributed fuzzy sets.

For example, when the value of k is set to 4, the considered percentiles are calculated as follows: $\left\{\frac{100}{k+1}, \frac{100}{k+1} + \frac{100}{(k+1)(k-1)}, \frac{2-100}{k+1} + \frac{100}{(k+1)(k-1)}, \frac{2-100}{k+1} + \frac{200}{(k+1)(k-1)}, \frac{2-100}{k+1} + \frac{2-100}{(k+1)(k-1)}, \frac{3-100}{k+1} + \frac{3-100}{(k+1)(k-1)}\right\}$ which generates $\{p_{20}, p_{26.6}, p_{46.6}, p_{53.3}, p_{73.3}, p_{80}\}$. Conversely, the global variable Intervals' specified ranges are used by the FuzzyDivision procedure. The difficulty of this computing approach is $O(n/c)$, where n represents the transaction count and c represents the distributed computation nodes. This intricacy results from the algorithm's need to process the data set's elements and convert them into various fuzzy labels. The experts established various intervals for producing the fuzzy labels for the use case that was being studied as in Fig 2.

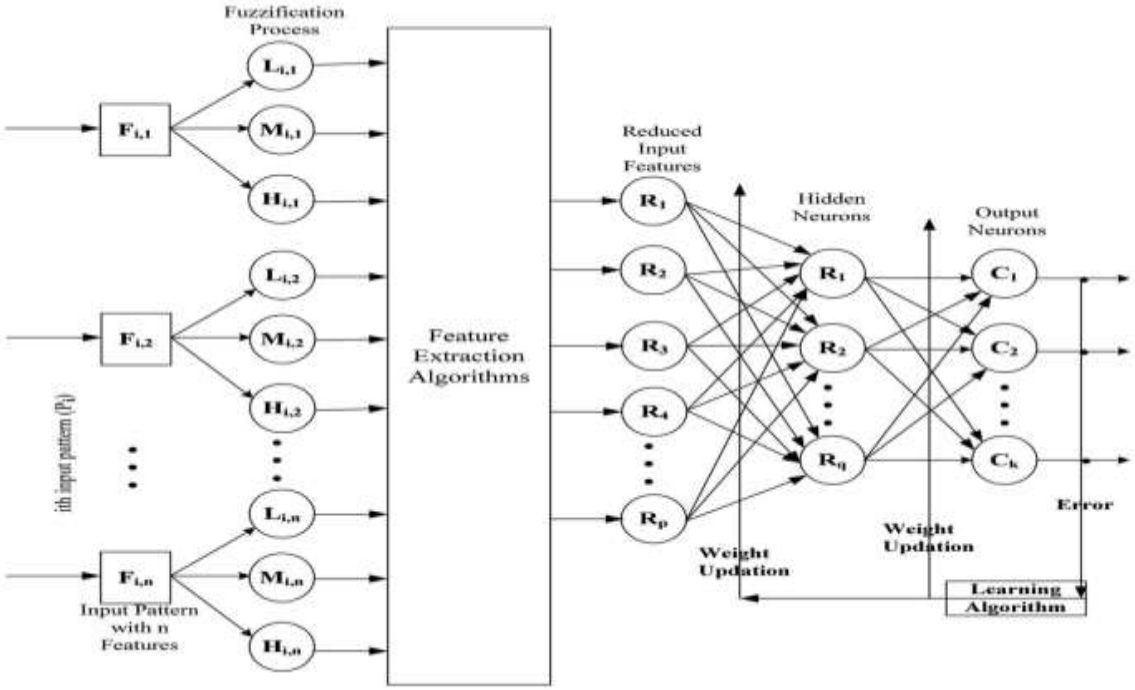

Fig 2. Weighted Fuzzified model

After this step, a reduced matrix (X) is created, which contains only the important information from the original features needed for classification. Next, this suppressed matrix is given to the network to predict diseases. The weighted fuzzy uses a model called back-propagation for classification. The reduced matrix, which has been processed (fuzzified), serves as the input to this classifier to help classify diseases. In the network, all the connections between the input and the hidden layers are fully linked, as well as the connections between hidden layers and the output layer. At first, all these connections are allotted with arbitrary weights between zero and one. The units in the input layer match the number of properties in the minimized matrix. There are number of units in the output layer which is equal to the different disease classes in the repository. The blocks in the hidden layer is determined by using an equation (Eq. 2), which calculates the units count based on the input, hidden, and output layers.

$$hidden_{nodes} = [\sqrt{input_{nodes}} * output_{nodes}] \quad (2)$$

In the feed-forward phase, the system is trained using the information from the reduced fuzzified matrix. The net input is calculated by adding up the weighted inputs and including a bias. This process can be represented numerically in Equation 3 where n represents the hidden layer neurons.

$$U_n = B_n + \sum_{j=1}^{r} W_{i,j} * X_{i,j} \quad (3)$$

In this case, B_n is the bias for the nth neuron, X_i represents the input pattern from the reduced fuzzified matrix (which includes values like $x_{i,1}, x_{i,2}, \ldots, x_{i,r}$), and W_i refers to the connection weights for the n^{th} neuron (such as $w_{i,1}, w_{i,2}, \ldots, w_{i,r}$). U_n represents the network input. Similarly, the total input for each layer is calculated, and a sigmoid activation function is employed to calculate the result produced between various layers. The final output from the output layer is determined by using the sigmoid activation function, which is given in Equation (4), where φ is the sigmoid function and O_n represents the

neuron output.

$$O_n = \varphi(U_n) = \frac{1}{1 + e^{-U_n}} \quad (4)$$

In the back-propagation step, the error is calculated by taking away the actual findings from the target (expected) result. This error is represented mathematically in Eq. (5).

$$Error(i) = Target_{output}(i) - actual_{output}(i) \quad (5)$$

In this case, $i = 1, 2, \ldots, m$ are the data points in the dataset. The RMSE is calculated using a specific formula and shows how far off the system's predictions are from the actual values, on average. The smaller the RMSE, the better the model's predictions match the true values.

$$RMSE = \sqrt{error(i)^2} \quad (6)$$

In a similar way, errors are calculated, and the weights and biases are updated during the learning procedure. The weights between the different layers are adjusted by calculating the necessary changes to reduce the overall error in the network. This is represented in Eq. (7), where α (alpha) is the learning rate, which falls within the range of [0, 1]. The learning rate regulates how much the weights are adjusted with each update.

$$\Delta weight = -\alpha * \frac{\partial error}{\partial output} * \frac{\partial output}{\partial net} * \frac{\partial net}{\partial weight} \quad (7)$$

The updated bias and weights of the system can be calculated using the formulas in Eq. (8) and Eq. (9). These equations help adjust the weights and bias after each learning step to reduce the error and improve the system's performance.

$$New_{weight} = old_{weight} + \Delta weight \quad (7)$$

$$New_{bias} = old_{bias} + \Delta bias \quad (8)$$

This procedure is frequently executed to reduce the RMSE of the system, or until a stopping condition is met. In the same way, any dataset can be used to build a model for disease prediction. Algorithm 1 shows the models weighted fuzzification approach.

Algorithm 1

Input: thyroid dataset $t_k = \{item_1, \ldots, item_m\}$ //kaggle dataset, labels, intervals, variables, default intervals

Output: Fuzzy data analysis

//Initialize

1. Feature = thyroid_dataset. Namefeatures ()
2. Variable (features)
3. Analyse chunk data;
4. $i = 0$;
5. do
6. if variables exist in provided list
7. if feature $[i] \in intervals$ then
8. Interval = Intervals $[feature\ [i][0]]$
9. Labels = Intervals $[feature\ [i][1]]$
10. else
11. Interval = ProduceIntervals (defaultintervals, data [feature [i]])
12. Labels = ProduceIntervals (defaultintervals)
13. endif
14. for $j = 0, j < labels; j++$ do

15. Fuzzydata[labels] = fuzzification (Interval[j], Interval[j+1], type)
16. type = linear, logarithmic and exponential;
17. $i++$
18. end for
19. while $[feature] > i$;
20. return fuzzified data

4. RESULTS ANALYSIS

The effectiveness and utility of various categorization algorithms for recognizing thyroid disorders have been assessed in this work using model development. A model's efficacy is evaluated with a confusion matrix and all pertinent measurements, such as the TP, TN, FP, FN, F1-score, accuracy, ROC-AUC score, precision, and recall.

Confusion matrix: It represents one of the simplest ways to compute the accuracy and effectiveness of the framework. This matrix is a table with the 2D, "Actual class" and "Predicted class," in each dimension. The real thyroid disease categories are shown in rows and the anticipated ones are in columns. The dataset has two classes: Class 0 and Class 1.

True Positives are examples where the data point's anticipated and actual classes are correct.

True Negatives are instances in which both the target and actual classes of the data point are incorrect.

False Positives are instances in which the expected class of a data point is True, but its real class is incorrect.

False Negatives are situations in which the expected class of the data point is False while the actual class is correct.

Accuracy: It is determined by dividing the cardiovascular disease forecasts / by the number of correct diagnoses. The efficacy of the four classification methods serves as the basis for the accuracy assessment.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

Precision: It indicates the percentage of a positive class's predictions that are positive for cardiovascular disease. High precision means that repeated reading results are acquired or the measurements yield consistent results. Because of the inadequate precision, the measurement's value fluctuates.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Recall: The capacity of a test to identify a person with thyroid problems as positive is known as recall. Because there are fewer FN findings from a highly sensitive test, a smaller number of cardiovascular disease patients are overlooked, and the True Positive rate (TPR) is another name for it.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

Specificity: It is computed by dividing the entire amount of negative findings about cardiovascular disease by the total amount of accurate negative forecasts. Another name for it is True Negative Rate (TNR).

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

F1-score: It is the harmonic mean of precision and recall. If the F1 score is more then it represents exceptional precision and recall of the suggested framework.

$$F1 - score = 2 * \frac{Recall * precision}{Recall + precision} \tag{13}$$

AUC-ROC: It is categorization efficiency metric. The AUC-ROC statistic provides information about the system's capability to distinguish between classes. The greater the AUC indicates the model is accurate. Plotting TPR against FPR at various thresholds allows for its mathematical production.

4.1. Implementation analysis

Several mining approaches are used in this work to predict thyroid. The efficacy of these methods has been evaluated, and

an analysis between them has been finished. Before going on to the part on performance assessment for different methodologies, the data is analyzed, and the findings are presented in the next section. Exploratory data analysis (EDA) is a technique for data analysis that employs fundamental charting applications to understand the information before going on to effective ML. It uses a variety of methodologies (mainly graphical), from statistics to linear algebra. EDA, or "a first glimpse at the data," is a critical step in evaluating the outcomes of an experiment. Enhancing features and generating accurate, suitably interpreted outcomes are utilized to understand and condense the dataset's information.

4.2. Feature analysis

The model shows the association between every feature and the desired value. The target and different traits have different connections. The heat-map of the relationship between numerical results is shown. There is a strong correlation between the two variables. Thus, either feature should be depreciated. However, based on the relationship with the target and all other metrics, feature_1 relates more with the target. Feature_2 has been eliminated from the original data set. There are frequently reported gender variations in the start and rate of thyroid in older adults. The impacted rates were 21.62 percent for women and 78.38 percent for men is shown. Men are more likely than women to have thyroid disease on average. Our dataset assigns every individual to a specific number of potential professions. The association between work and cardiovascular disease is weak. Genetic background likewise has minimal bearing on the incidence of thyroid. Lack of physical activity, or idleness, is the 4th most crucial risk factor for morbidity globally, accounting for 6% of all fatalities. Coronary illness is 67.57% higher in those without physical activity. The body's reaction to hazardous situations is stress. An important risk factor for thyroid is previous experience of discomfort. Here, 71 percent of those who had previously had thyroid had been identified as having disease. The research aims to determine the most effective method for classifying thyroid disorders. This section presents the top performers according to several performance metrics and summarizes all the research results. After the performance is determined, the best classification is found using a dataset that includes all 19 variables. Second, considering 14 characteristics that correlate significantly with the desired value will yield the best predictor for disease. Lastly, this work has compared several performance metrics of different machine-learning methods for both datasets.

4.3. Benchmarking features

Nineteen features in the dataset can be used to predict thyroid disease. The efficacy of each unique AI technique in every feature dataset has been evaluated using a variety of assessment measures including F1 score, accuracy, ROC-AUC, precision, and recall value. Regarding the F1 score, accuracy, ROC-AUC, precision, and recall score, RF scores are better than the remaining models shown in Table 1 and Fig 3. RF achieves the highest accuracy of 89.9%, with precision, ROC-AUC, recall, and F1-score scores of 84%, 88%, 100%, and 91%, respectively. As a result, it has been determined that for the entire set of 19 attributes, RF outperforms the six methods used for classification. 75%, 77.5%, 75%, 75%, and 75% accuracy, precision, recall, F1-score, and ROC-AUC scores, respectively, have all been attained with KNN. With 87.50%, 63.64%, 73.68%, 76.26% of precision, recall, F1-score, and ROC-AUC score, NB has 75 percent accuracy rate. Although NBs is more precise than K-NN, K-NN and NB perform well. Compared to KNN, Naïve Bayes has a lesser recall rate and F1 score; nevertheless, the ROC-AUC value is more important. The accuracy of LR is 70%. SVM provides 70 percent accuracy and 72 percent F1-score. Decision trees offer an accuracy of 70 percent and an average of 75% of the F1-score. As a result, DT and LR adopt similar positions on race. However, MLP fails to perform well in comparison to other approaches. Thus, it can be concluded that RF is the most effective classification technique for the whole dataset and that multilayer perceptron (MLP) is less successful than other algorithms. The proposed weighted model gives promising outcomes. Fig 3 shows the performance study of AI approaches using every attribute in a graphical format.

4.4. Feature selection

After minimizing the set of features to 14, this study assessed performance metrics for every AI technique. The examination of algorithms in the test dataset using specific characteristics is shown in Table 8 and Fig. 10. 75 percent accuracy, 75 percent precision, 81.83 percent recall, 78.27 percent F1-score, and 74.25 percent ROC-AUC score are all displayed by LR. 70 percent accuracy, 77.78 percent precision, 63.65 percent recall, 70 percent F1-score, and 70.72 percent ROC-AUC score are all attained via Naïve Bayes. When compared to LR, KNN scores effectively. MLP has the least recall (63.64%) and F1 score (70%) of all the methods. K-NN's accuracy, recall, precision ROC-AUC, and F1-score are 85% for all metrics. Accuracy, F1-score, precision, recall, and ROC-AUC score are 90%, 90.92%, 100%, 90.92%, and 89.90%, respectively, all maximum for RF. Both SVM and decision trees have attained 80 percent accuracy. In this case, SVM and Decision Tree perform approximately equal, with DT slightly outperforming SVM. The accuracy rate of LR is 70 percent, the precision is 66.68 percent, the recall is 76.93 percent, and the ROC-AUC score is 67.69 percent, with a 76.93 percent F1-score. MLP has the least F1-score (67.68%) and precision (66.67%) compared to other methods. As a result, multilayer perceptron is not as effective as other classifiers, and RF is the most suitable classification method for the dataset we chose. The proposed weighted fuzzy model gives 95% accuracy, 96% precision, recall of 96.5%, F1-score of 96% and ROC of 99% respectively for all features. The proposed weighted fuzzy model gives 97% accuracy, 97.5% precision, recall of 98%, F1-score of 98.2% and ROC of 99.9% respectively for chosen features. Fig 4 shows the efficiency assessment of all strategies for selecting attributes in a graphical format.

Table 1. Algorithm evaluation with all thyroid features

Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC (%)
LR	70	69	81	75	68
NB	75	87	64	73	76
k-NN	76	75	75	76	75
SVM	71	72	72	72	69
DT	72	70	76	76	68
RF	89	85	91	91	88
MLP	90	60	69	65	57
Proposed Weighted Fuzzy	95	96	96.5	96	99

Table 2. Algorithm evaluation with selected thyroid features

Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC (%)
LR	75	75	81	78	74
NB	70	77	63	70	70
k-NN	85	85	85	85	85
SVM	80	76	83	83	78
DT	82	81	81	81	79
RF	90	90	92	90	89
MLP	70	66	76	76	67
Proposed Weighted Fuzzy	**97**	**97.5**	**98**	**98.2**	**99.9**

Table 3. Misclassification rate comparison

Algorithms	Misclassification rate
LR	0.123
NB	0.05
k-NN	0.04
SVM	0.08
DT	0.11

RF	0.018
MLP	0.005
Proposed Weighted Fuzzy	0.03

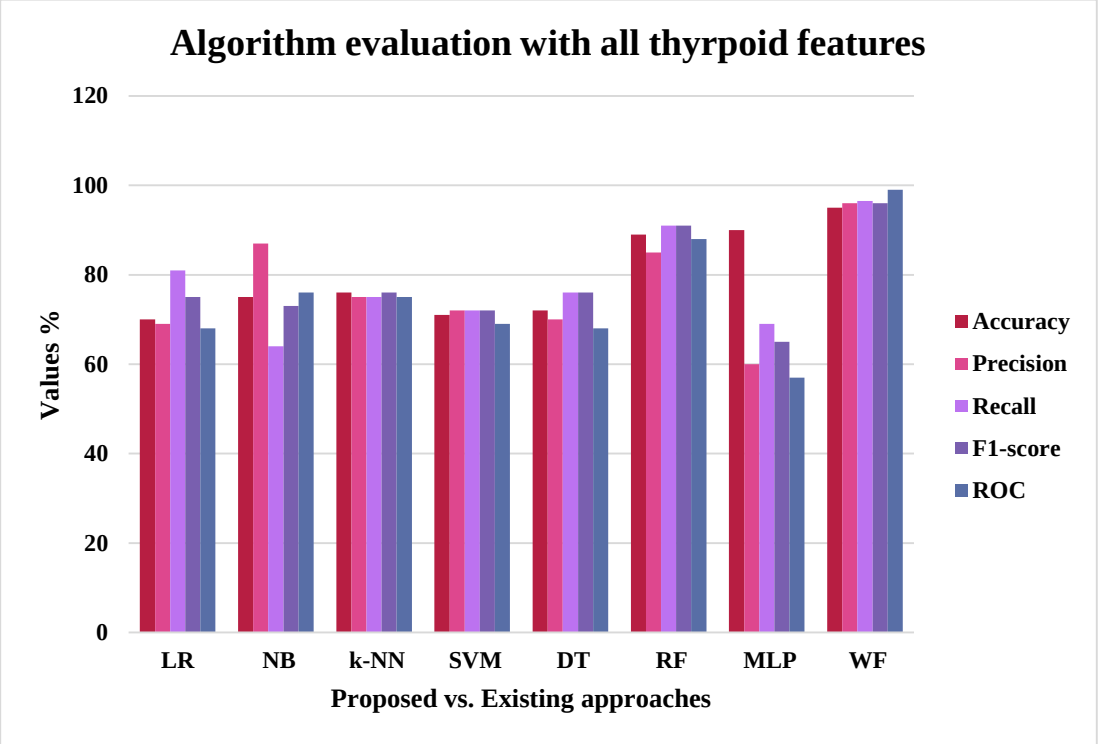

Fig 3. Algorithm evaluation with all thyroid features

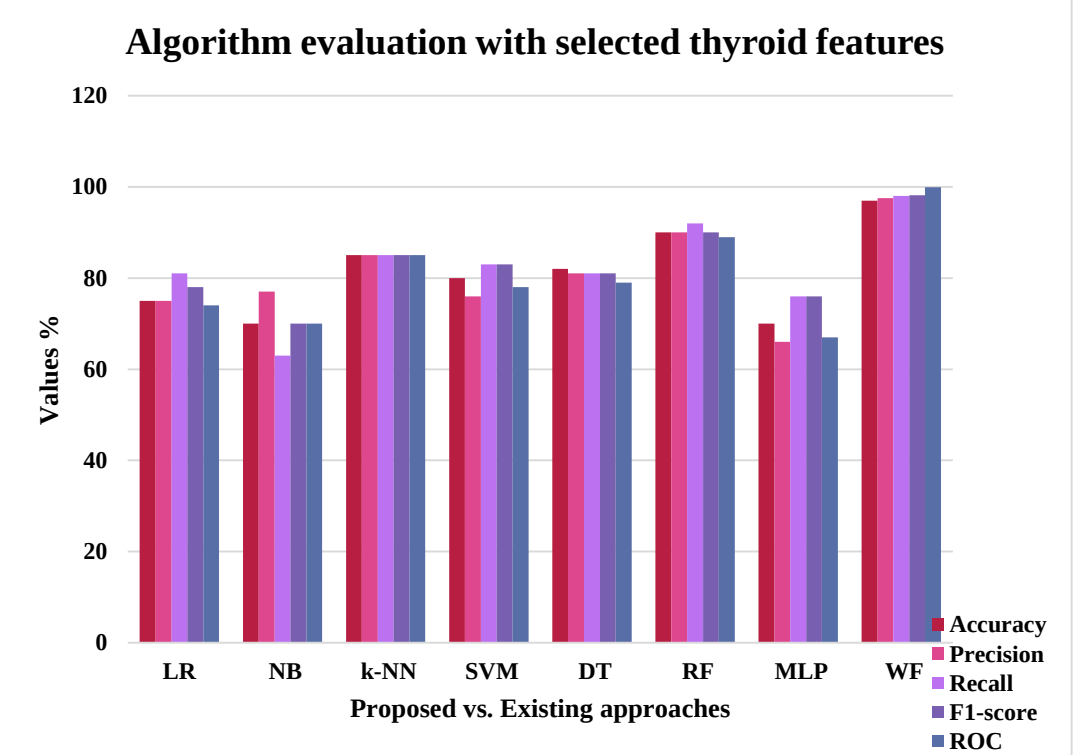

Fig 4. Algorithm evaluation with selected thyroid features

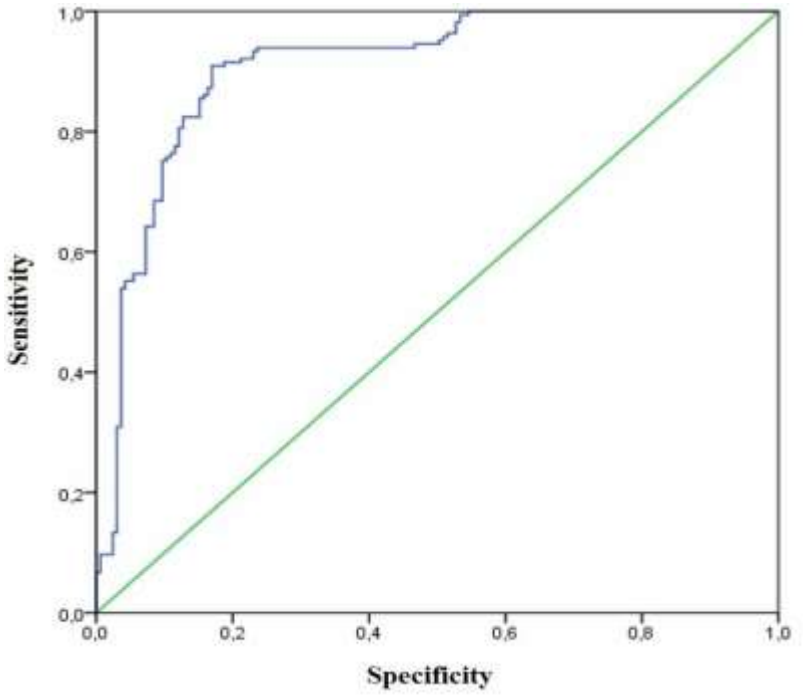

Fig 5. ROC (sensitivity vs. specificity)

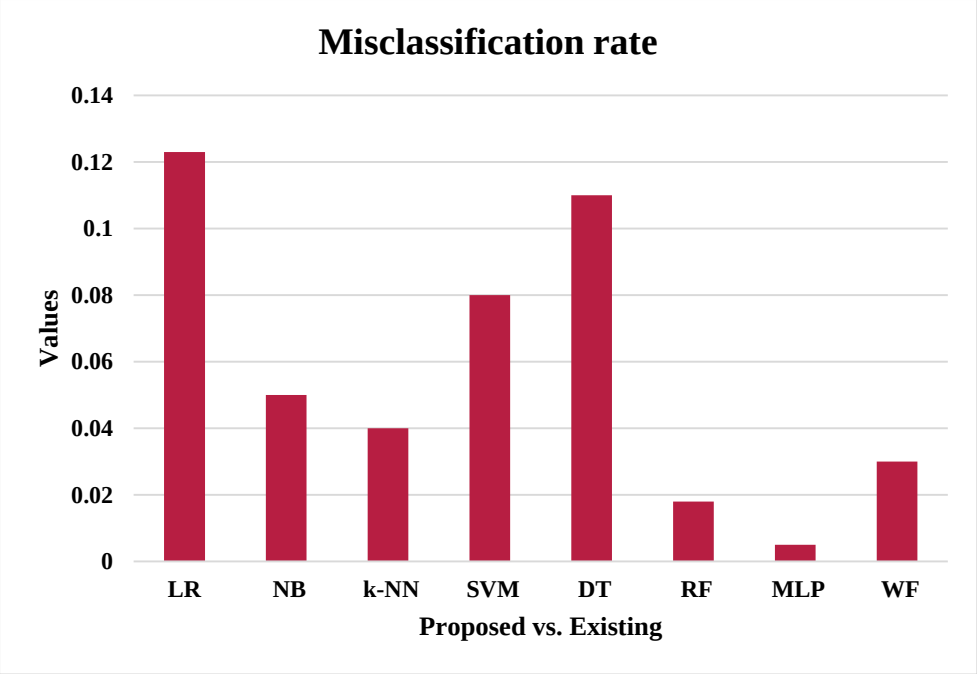

Fig 6. Misclassification rate comparison

4.5. Comparative analysis

There are notable differences between the chosen dataset and the efficiency measures from the findings of the entire dataset. The precision of KNN, MLP, DT, SVM, and LR has increased for the selected dataset. But just as accuracy decreases with Naïve Bayes, so does precision. KNN, SVM, DT, and multilayer perceptron have all improved recall values. However, none of the algorithms' recalls decreased this time. For MLP, DT, K-NN, LR, SVM, and the F1-scores have risen. NB's F1 score has declined. The ROC-AUC score has gone up for every algorithm except Naive Bayes. Thus, based on the data above, it can be said that RF works well for both datasets. NB performs worse in every sector with every feature. By increasing the dataset, NBs' efficiency can be enhanced. MLP also performs better than the all-attributes dataset. Fig 5 and Fig 6 clarifies that the dataset of chosen attributes performs better than the database of all variables.

5. CONCLUSION

A thorough analysis of patient features for the prediction of thyroid has been made possible by this research. A weighted fuzzy model was used to choose the most essential characteristics. It has been found that only 14 features are highly correlated. Lastly, 7 AI techniques (LR, Naïve Bayes, KNN, SVM, DT, RF, and MLP) are compared using the dataset with all characteristics and selected features. Compared to other AI methods, RF has a superior efficiency rate with a 90 percent accuracy rate using chosen features, 90.91 percent precision, hundred percent recall, 90.92 percent F1-score, and 89.91 percent ROC-AUC score except NBs, the dataset of chosen features scores better than the dataset of all features. The NB prediction accuracy has significantly declined due to the absence of further datasets and selective feature sets. The majority of the dataset's characteristics are closely related to others. By methodically examining the effectiveness of the many elements, the clinicians will be assisted in effectively collecting the documents. Instead of documenting and maintaining all the variables, the data executives can preserve the essential features for forecasting cardiovascular disease. We intend to publicly validate our proposed technique as an element of our subsequent work

REFERENCES

[1] Zimek A, Filzmoser P. There and back again: Outlier detection between statistical reasoning and data mining algorithms. Wiley Interdiscip Rev Data Min Knowl Discov. 2018;8 (6):1–26.

[2] Santos-Pereira J, Gruenwald L, Bernardino J. Top data mining tools for the healthcare industry. J King Saud Univ Comput Inf Sci. 2022;34 (8):4968–82

[3] Begum A, Parkavi A. Prediction of thyroid disease using data mining techniques in 2019 5th International conference on advanced computing and communication systems ICACCS 2019; 2019 pp. 342–345.

[4] Zhu Y, Fu Z, Fei J. An image augmentation method using the convolutional network for thyroid nodule classification by transfer learning. In: 2017 3rd IEEE International Conference on Computer Communications

ICCC 2017; 2017, vol. 2018-Janua, pp. 1819–1823.

[5] Ilyas M, et al. Deep learning based classification of thyroid cancer using different medical imaging modalities: a systematic review. VFAST Trans Softw Eng. 2021

[6] Begum A, Parkavi A. Prediction of thyroid disease using data mining techniques. In: 2019 5th International Conference on Advanced Computing and Communication Systems ICACCS 2019, no. August 2016; 2019, pp. 342–345.

[7] Ahmed I, et al. Lithium from breast milk inhibits thyroid iodine uptake and hormone production, which are remedied by maternal iodine supplementation. Bipolar Disord. 2021;23(6):615–25.

[8] Priyadharsini D, Sasikala S. Efcient thyroid disease prediction using features selection and meta-classifiers. In: Proceedings—6th International Conference on Computing Methodologies and Communication ICCMC 2022, no. ICCMC; 2022, pp. 1236–1243

[9] Ahmed I, Mohiuddin R, Muqeet MA, Kumar JA, Thaniserikaran A. Thyroid cancer detection using deep neural network. In: Proceedings— International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022, no. ICAAIC; 2022, pp. 166–169

[10] Jamil M, Wlvvxh WKU. Ensemble-based effective diagnosis of thyroid disorder with various feature selection techniques; 2022, pp. 14–19.

[11] Rashad, N.M.; Samir, G.M. Prevalence, risks, and comorbidity of thyroid dysfunction: A cross-sectional epidemiological study. Egypt. J. Intern. Med. 2020, 31, 635–641.

[12] Mirbabaie, M.; Stieglitz, S.; Frick, N.R.J. Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. Health Technol. 2021, 11, 693–731.

[13] Holzinger, A.; Keiblinger, K.; Holub, P.; Zatloukal, K.; Müller, H. AI for life: Trends in artificial intelligence for biotechnology. New Biotechnol. 2023, 74, 16–24.

[14] Alyas, T.; Hamid, M.; Alissa, K.; Faiz, T.; Tabassum, N.; Ahmad, A. Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach. Biomed. Res. Int. 2022, 2022, 9809932.

[15] Garcia de Lomana, M.; Weber, A.G.; Birk, B.; Landsiedel, R.; Achenbach, J.; Schleifer, K.J.; Mathea, M.; Kirchmair, J. In Silico Models to Predict the Perturbation of Molecular Initiating Events Related to Thyroid Hormone Homeostasis. Chem. Res. Toxicol. 2021, 34, 396–411.

[16] Alnaggar, M.; Handosa, M.; Medhat, T.; Rashad, M.Z.; Author, C.; Alnaggar, M. Thyroid Disease Multi-classMulti-class Classification based on Optimized Gradient Boosting Model. Egypt. J. Artif. Intell. 2023, 2, 1–14.

[17] Sengupta, D.; Mondal, S.; Raj, A.; Anand, A. Binary Classification of Thyroid Using Comprehensive Set of Machine Learning Algorithms. In Frontiers of ICT in Healthcare: Proceedings of EAIT 2022; Springer Nature: Singapore, 2023; pp. 265–276.

[18] Trivedi, N.K.; Tiwari, R.G.; Agarwal, A.K.; Gautam, V. A Detailed Investigation and Analysis of Using Machine Learning Techniques for Thyroid Diagnosis. In Proceedings of the 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 1–3 March 2023; pp. 1–5.

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

[20] Kumari P, Kaur B. Mechanism for disease classification in predicting thyroid disease. In: 2023 1st International conference on circuits, power and intelligent systems (CCPIS); 2023, pp. 1–6.

[21] Zimek A, Filzmoser P. There and back again: Outlier detection between statistical reasoning and data mining algorithms. Wiley Interdiscip Rev Data Min Knowl Discov. 2018;8(6):1–26.

[22] Santos-Pereira J, Gruenwald L, Bernardino J. Top data mining tools for the healthcare industry. J King Saud Univ Comput Inf Sci. 2022;34(8):4968–82

[23] Begum A, Parkavi A. Prediction of thyroid disease using data mining techniques in 2019 5th International conference on advanced computing and communication systems ICACCS 2019; 2019 pp. 342–345.

[24] Zhu Y, Fu Z, Fei J. An image augmentation method using the convolutional network for thyroid nodule classification by transfer learning. In: 2017 3rd IEEE International Conference on Computer Communications ICCC 2017; 2017, vol. 2018-Janua, pp. 1819–1823.

[25] Ilyas M, et al. Deep learning based classification of thyroid cancer using different medical imaging modalities: a systematic review. VFAST Trans Softw Eng. 2021

[26] Begum A, Parkavi A. Prediction of thyroid disease using data mining techniques. In: 2019 5th International

Conference on Advanced Computing and Communication Systems ICACCS 2019, no. August 2016; 2019, pp. 342–345.

[27] Ahmed I, et al. Lithium from breast milk inhibits thyroid iodine uptake and hormone production, which are remedied by maternal iodine supplementation. Bipolar Disord. 2021;23(6):615–25.

[28] Priyadharsini D, Sasikala S. Efcient thyroid disease prediction using features selection and meta-classifiers. In: Proceedings—6th International Conference on Computing Methodologies and Communication ICCMC 2022, no. ICCMC; 2022, pp. 1236–1243

[29] Ahmed I, Mohiuddin R, Muqeet MA, Kumar JA, Thaniserikaran A. Thyroid cancer detection using deep neural network. In: Proceedings— International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022, no. ICAAIC; 2022, pp. 166–169

[30] Jamil M, Wlvvxh WKU. Ensemble-based effective diagnosis of thyroid disorder with various feature selection techniques; 2022, pp. 14–19..

.