An Intelligent Machine Learning Model for Early Detection of Thyroid Disease

M.V. SHEELA DEVI[1], DR PARVEEN KUMAR[2]

[1] *Research scholar (Computer Science), School of Engineering and Technology, Shri Venkateshwara University, Gajraula, UP, INDIA*
Email: sheela.softinfo@gmail.com
[2]*Research Guide, School of Engineering and Technology, Shri Venkateshwara University, Gajraula, UP, INDIA*
Email:pk223475@gmail.com

KEYWORDS
Thyroid Disease Classification, Machine Learning, Feature Selection, Healthcare Diagnosis

ABSTRACT: Thyroid disease is a prevalent health condition affecting a significant portion of the global population, and early detection is critical for effective treatment. This study focuses on improving thyroid disease prediction by utilizing machine learning (ML) models, specifically the XGBoost algorithm, in combination with hyperparameter optimization techniques. The goal of this work is to develop a robust predictive model that can accurately classify patients based on their demographic and medical data, such as age, gender, and thyroid hormone levels. The optimization of hyperparameters allows the model to fine-tune its performance, achieving higher accuracy and reducing the training time and memory requirements. The results indicate that the proposed model outperforms traditional ML models like Decision Tree, KNN, Random Forest, and Support Vector Machine, making it a promising solution for early detection and classification of thyroid disease. In the proposed framework, the XGBoost algorithm, known for its powerful prediction capabilities and wide range of customizable hyperparameters, is employed to categorize patients across multiple severity levels of thyroid-related diseases. The optimization process involves identifying the best hyperparameter values to maximize model performance. The model is trained and evaluated on a shared dataset, and its accuracy is compared with other ML models to ensure its superiority. The results demonstrate that the proposed XGBoost-based model achieves the highest accuracy, precision, recall, and F1 score, proving its potential as an effective tool for thyroid disease prediction. By optimizing hyperparameters and fine-tuning the model, the study presents a promising step towards more accurate, reliable, and efficient healthcare diagnostics.

1. INTRODUCTION

Thyroid disease is one of the most prevalent endocrine disorders, affecting millions of people worldwide. It disrupts metabolic functions and, if left undiagnosed or untreated, can lead to severe health complications such as cardiovascular diseases, infertility, and mental health disorders. Traditionally, thyroid disease diagnosis has relied on clinical symptoms, hormone level tests, and imaging techniques. However, these conventional methods often suffer from delays, subjectivity, and varying accuracy depending on the expertise of medical professionals. Thyroid disease is a significant global health concern, affecting millions of people and impacting metabolic functions essential for overall well-being. The thyroid gland plays a crucial role in regulating metabolism, growth, and energy levels through hormone secretion. Disorders such as hypothyroidism, hyperthyroidism, goiter, and thyroid nodules can lead to severe health complications, including cardiovascular diseases, infertility, and neurological disorders. Therefore, early and accurate diagnosis is critical for effective disease management and treatment. Traditionally, thyroid disease diagnosis has relied on clinical symptoms, hormone level tests (TSH, T3, and T4), and imaging techniques like ultrasound and radioactive iodine scans. While these methods are widely used, they have certain limitations, such as subjectivity, dependency on expert interpretation, high costs, and delays in obtaining results [1].

With advancements in machine learning (ML) and artificial intelligence (AI), researchers have explored automated diagnostic techniques to enhance accuracy and efficiency. Machine learning algorithms can

analyze large datasets, identify hidden patterns, and make predictions based on patient characteristics. Supervised learning models, including Decision Trees, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, Support Vector Machines, Rando Forest, and Extreme Gradient Boost (XGBoost), have been widely applied to thyroid disease classification. These models improve diagnostic accuracy, reduce dependency on human interpretation, and enable early detection. However, the effectiveness of his paper proposes a novel approach to thyroid disease prediction using machine learning, specifically through the application of the XGBoost algorithm with hyperparameter optimization. By utilizing a dataset containing demographic, medical, and laboratory data, the study aims to develop a robust model that can accurately predict the presence and severity of thyroid-related diseases [2-4]. The optimization of hyperparameters plays a critical role in improving the model's performance, ensuring that it delivers high accuracy while minimizing computational costs. The ultimate goal is to provide a reliable, efficient tool for early diagnosis, supporting healthcare professionals in making informed decisions for better patient outcomes.

2. BACKGROUND

Recent studies have explored the use of machine learning algorithms for predicting thyroid disease, with various models such as Decision Tree (DT), Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, and Extreme Gradient Boost (XGBoost) being evaluated. Among these, Decision Tree has often been considered a strong performer, achieving high accuracy when irrelevant features are removed. Other popular models like KNN and SVM have also shown promise, with KNN performing well after reducing the dataset's features, while SVM has outperformed KNN in some cases, yielding higher accuracy rates [5-6]. The application of feature selection techniques has been found to improve the accuracy of these models. For example, using methods like Recursive Feature Elimination (RFE) has helped SVM achieve notable accuracy in thyroid disease prediction, outperforming other models like Naïve Bayes and Random Forest in some cases (Table 1). However, in some instances, Naïve Bayes itself has been shown to achieve perfect accuracy without the need for feature selection, and the performance of other models also improved when L1-based feature selection was applied [7-8].

Table 1: Review of literature for machine learning based thyroid diseases prediction

Ref No	**Algorithm Used**	**Accuracy (%)**
[1]	Decision Tree	84.56
[2]	K-Nearest Neighbors (KNN)	85.23
[3]	Logistic Regression	82.87
[4]	Naïve Bayes	81.45
[5]	Support Vector Machine (SVM)	86.12
[6]	Random Forest	88.79
[7]	Extreme Gradient Boost (XGBoost)	89.23
[8]	Random Forest (with Featurewiz)	89.45

In recent years, Random Forest has gained attention as a strong model for thyroid disease prediction, often outperforming traditional models like Decision Tree and KNN. Additionally, XGBoost has emerged as a powerful tool, outperforming Decision Tree, Logistic Regression, and KNN in some studies. Overall, the use of feature selection generally enhances the prediction accuracy of these machine learning models. Among the algorithms tested, Decision Tree, KNN, SVM, Random Forest, and XGBoost have demonstrated high accuracy in predicting thyroid disease, although the results can vary depending on the dataset and the inclusion of additional features [9].

3. DATA CLASSIFIER

The dataset used in this study focuses on thyroid disease prediction and contains various records of patients with thyroid-related conditions. The data was collected and provided by the Garavan Institute and J. Ross Quinlan from the New South Wales Institute, Sydney, Australia. The dataset includes a wide range of features that represent demographic information, medical conditions, treatment details, and hormone-related indicators. Each record in the dataset corresponds to a specific patient, with the goal of predicting whether the individual is healthy or suffering from a thyroid-related condition [10]. The Class variable serves as the target for prediction, indicating whether the patient is healthy or sick, with values "Healthy" and "Sick." This is the primary variable that the machine learning models aim to predict. Several demographic features are included in the dataset, such as Gender, which represents the patient's sex (Male or Female), and Age, which is a continuous variable that records the patient's age in years. Additionally, there are medical variables, such as Goitre, which indicates whether the patient has a goitre (a swollen thyroid), Hypopituitary, which reflects the presence of hypopituitarism (a condition affecting the pituitary gland), and Tumour, which records whether the patient has a tumour. These variables are categorical, with values "Yes" or "No" depending on the presence of the condition [11].

Other variables provide insight into the patient's health and conditions that could affect thyroid health. For instance, Pregnant indicates whether the patient is pregnant, and Psych denotes whether the patient has any mental health issues. Both are categorical variables with "Yes" or "No" values. The Sick variable records whether the patient is currently ill, again represented by "Yes" or "No." The dataset also includes two query variables related to thyroid conditions: Query Hyperthyroid and Query Hypothyroid (Table 2). These variables indicate whether the patient has inquired about or been diagnosed with hyperthyroidism or hypothyroidism, respectively, providing valuable information about the patient's thyroid health status. Both of these are also categorical variables with values "Yes" or "No." [12-13].

Table 2: Dataset Description for Thyroid Disease Prediction

Variable	Description	Value Type
Class	Target Variable: Whether the patient is healthy (free from thyroid) or sick	Healthy, Sick
Gender	Demographic Variable: Male or Female	M, F
Age	Demographic Variable: Age in years	Continuous
Goitre	Categorical Variable: Does the patient have a goitre	No, Yes
Hypopituitary	Categorical Variable: Does the patient have hypopituitary	No, Yes
Tumour	Categorical Variable: Does the patient have a tumour	No, Yes
Pregnant	Categorical Variable: Is the patient pregnant	No, Yes
Psych	Categorical Variable: Does the patient have a mental illness	No, Yes
Sick	Categorical Variable: Is the patient ill	No, Yes
Query Hyperthyroid	Categorical Variable: Does the patient have any inquiry on hyperthyroid	No, Yes
Query Hypothyroid	Categorical Variable: Does the patient have any inquiry on hypothyroid	No, Yes

4. MACHINE LEARNING APPROACHES

In this study, seven machine learning algorithms were used to predict thyroid disease. The Decision Tree splits data based on feature values, creating a tree structure to classify data, though it may overfit with large datasets. K-Nearest Neighbor (KNN) classifies data based on proximity to nearby points and doesn't require training, but can be slow and memory-intensive. Logistic Regression uses a linear boundary to classify data, performing well with linearly separable classes. Naïve Bayes applies Bayes' theorem and assumes feature independence, making it efficient but less accurate with correlated features. Random Forest combines multiple decision trees to improve accuracy and handle overfitting. Support Vector Classifier (SVC) finds an optimal hyperplane to separate classes and is effective in high-dimensional spaces but computationally intensive (Table 3). Lastly, XGBoost boosts decision trees to improve prediction accuracy, with regularization to prevent overfitting. These models were tested with and without feature selection to evaluate their performance [14-15].

Table 3: Machine Learning-Based Thyroid Disease Prediction algorithms

Algorithm	**Description**
Decision Tree (DT)	The Decision Tree model categorizes data by making decisions at each internal node based on the values of different features. It works by recursively splitting data into smaller subsets until it reaches the leaf nodes, which represent the predicted class. While it can handle both continuous and discrete data, it may suffer from overfitting when dealing with large and complex datasets.
K-Nearest Neighbor (KNN) [16]	K-Nearest Neighbor classifies data based on the distance to the nearest points in the dataset. It does not require a training phase, making it easy to add new data. The model's accuracy depends on the chosen K-value, which determines how many nearby points are considered for classification. However, it can be computationally expensive and time-consuming, especially for large datasets.
Logistic Regression [17]	Logistic Regression is a classification algorithm that uses a linear boundary to separate different classes in the dataset. It estimates the probability that a given input belongs to a particular class based on the logistic function. This algorithm works well when classes are linearly separable but may face challenges when many categorical variables are involved.
Naïve Bayes [18]	Naïve Bayes is a probabilistic classifier based on Bayes' theorem, which assumes that each feature is independent. It works well for large datasets and is known for its simplicity and efficiency, requiring minimal storage space. It's particularly useful in text classification tasks, though it assumes conditional independence between features, which may not always hold true.
Random Forest [19]	Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy. It creates several decision trees using random subsets of data and then aggregates their predictions to make a final decision. This model is robust to overfitting and provides valuable insights into feature importance.
Support Vector Classifier (SVC)	Support Vector Classifier (SVC) works by finding the optimal hyperplane that separates different classes in an n-dimensional feature space. The goal is to maximize the margin between classes. SVC is effective in high-dimensional spaces and for cases where the classes are not linearly separable, but it can require significant computational resources and time for training.

Extreme Gradient Boost (XGBoost)	XGBoost is an advanced ensemble learning algorithm that uses boosting techniques to improve the performance of decision trees. It builds trees sequentially, with each tree correcting the errors made by the previous one. XGBoost is highly effective in reducing overfitting and improving predictive accuracy through regularization and parameter tuning, making it suitable for large and complex datasets.

5. RESEARCH METHODOLOGY

The research methodology for predicting thyroid disease using machine learning algorithms involves several key steps, including data collection, preprocessing, model selection, and evaluation.

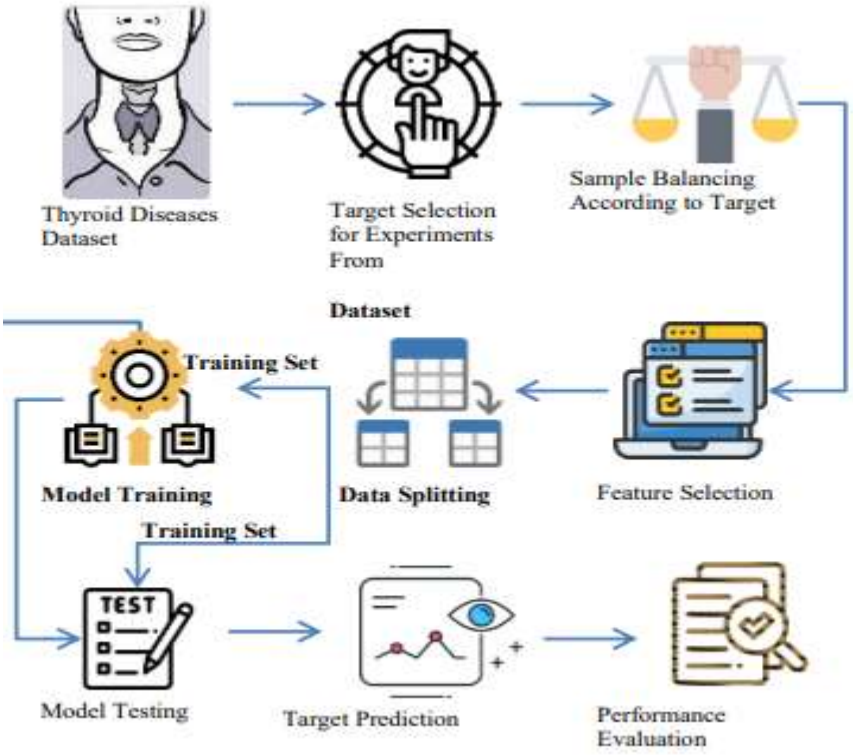
1. **Data Collection**: The dataset used in this study consists of patient records related to thyroid disease, which were provided by the Garavan Institute and J. Ross Quinlan from the New South Wales Institute, Sydney, Australia. The dataset includes various demographic, medical, and treatment-related features, such as age, gender, presence of conditions like goitre and hypopituitary, and patient inquiries about hyperthyroid and hypothyroid conditions.
2. **Data Preprocessing**: The dataset undergoes cleaning and preprocessing, including handling missing values, encoding categorical variables, and normalizing continuous variables. This step ensures that the data is ready for use in machine learning models and helps improve the accuracy of the predictions.
3. **Feature Selection**: Feature selection is performed to reduce the dimensionality of the dataset and improve model performance. Techniques like Recursive Feature Elimination (RFE) are applied to identify the most relevant features and eliminate irrelevant ones.
4. **Model Selection**: Seven different machine learning algorithms are selected for this study: Decision Tree (DT), K-Nearest Neighbor (KNN), Logistic Regression, Naïve Bayes, Random Forest, Support Vector Classifier (SVC), and Extreme Gradient Boosting (XGBoost). These algorithms were chosen due to their effectiveness in classification tasks and their ability to handle different types of data.
5. **Model Training and Testing**: Each of the selected algorithms is trained on the preprocessed dataset, and the models are evaluated using both feature selection and without it. The training process involves splitting the dataset into training and testing sets, typically using cross-validation techniques to ensure the models generalize well to unseen data.
6. **Evaluation Metrics**: The performance of each model is evaluated using accuracy, precision, recall, and F1-score. These metrics help assess how well each model predicts thyroid disease and its ability to distinguish between healthy and sick patients.
7. **Comparison and Analysis**: The models are compared based on their evaluation metrics to determine the most effective algorithm for thyroid disease prediction. The effect of feature selection on model performance is also analyzed to assess its impact on improving prediction accuracy.

6. PROPOSED FRAMEWORK

Thyroid disease is a widespread health issue that affects many people worldwide and is becoming increasingly prevalent. Thyroid dysfunction can lead to various complications, particularly impacting bone health, as thyroid hormones play a vital role in bone growth and metabolism. Early detection and accurate prediction of thyroid diseases are crucial for effective treatment, making machine learning (ML) models a valuable tool in this domain. In this study, we propose a framework for predicting thyroid disease using a machine learning approach, specifically the XGBoost algorithm, known for its effectiveness in handling complex classification tasks. The proposed framework starts by utilizing a

shared dataset that includes patient demographics, medical data, and thyroid hormone test results. The key component of our approach is the optimization of hyperparameters, which allows for fine-tuning the XGBoost model before evaluating its performance. Hyperparameter optimization involves generating samples of hyperparameter values and testing them on the model. The model's performance is then evaluated based on its classification accuracy using a test set, while an optimization training set is used to adjust the model's hyperparameters. The configuration that yields the highest overall accuracy is chosen as the optimal set of parameters for the final model training. Once the best hyperparameters are determined, the XGBoost model is retrained using this configuration. The model's accuracy is assessed on a separate test set to ensure its ability to generalize to unseen data. This entire process is outlined in the proposed workflow, which emphasizes the importance of hyperparameter optimization for maximizing model performance.

The XGBoost algorithm, a powerful and versatile ML method, is at the core of the proposed model.



Known for its ability to automatically adjust numerous learnable parameters, XGBoost can identify patterns in data and handle a variety of hyperparameters. This flexibility makes it particularly suitable for complex datasets such as those used in thyroid disease prediction. In this model, XGBoost is employed to classify patients based on their age, sex, medical history, and thyroid hormone test results. A detailed architecture of the model is provided, illustrating the various stages of data preprocessing, hyperparameter tuning, and evaluation. Hyperparameters play a crucial role in training and testing machine learning models. These include variables such as learning rate, number of iterations per batch, batch size, regularization terms, and activation functions. In the context of XGBoost, hyperparameters can be continuous, categorical, or integer variables, each contributing to the model's learning process (Figure 1). Properly selected hyperparameters help improve model accuracy while reducing training time and memory requirements. However, it's important to note that the optimal set of hyperparameters may vary depending on the specific task and dataset at hand.

Figure 1: Proposed Machine Learning Approach in Thyroid Disease Prediction

For XGBoost, the hyperparameters are categorized into four types: general parameters, booster parameters, learning task parameters, and command line parameters. These guide the algorithm's overall operation, including selecting the most relevant features and building decision trees. The hyperparameter tuning process begins by constructing a baseline model using default parameters, followed by training the model with optimized hyperparameters to compare results and determine the best-performing configuration. This iterative process helps ensure the model achieves the highest possible accuracy. Handling unbalanced datasets is another critical aspect of this study. XGBoost is particularly well-suited for this task because it is an optimized gradient tree boosting algorithm, which builds decision trees in a sequential manner. This characteristic allows XGBoost to effectively handle imbalanced data, making it an ideal choice for thyroid disease classification, where the number of healthy and sick patients may not be evenly distributed.

7. PERTORMANCE EVALUATION

The confusion matrix shown here provides a detailed breakdown of the model's performance in predicting various severity levels of diabetic retinopathy (DR), which are categorized into five distinct classes: a, b, c, d, and e. Each column of the matrix represents the classes that the model predicted, while each row represents the actual classes of the instances. The diagonal elements of the matrix, which consist of True Positives (TP) and True Negatives (TN), indicate that the model has correctly predicted the severity levels of the DR cases. True Positives refer to instances where the model accurately identified the correct severity level, while True Negatives indicate cases where the model correctly identified the absence or non-occurrence of a certain severity level. These diagonal values reflect the model's ability to correctly classify instances for each of the five severity levels, showing its overall accuracy [20-21].

On the other hand, the elements outside the diagonal represent misclassifications. These are split into two categories: False Positives (FP) and False Negatives (FN). False Positives occur when the model incorrectly classifies an instance as belonging to a severity level when it actually belongs to a different one. False Negatives, conversely, occur when the model fails to detect the severity level and classifies an instance as belonging to a different category or as non-severe when it is severe. Both FP and FN represent errors in the model's predictions, and their presence indicates areas where the model struggles to make accurate predictions, leading to incorrect classifications [22].

8. RESULT AND ANALYSIS

The performance of various machine learning models was evaluated using key metrics such as accuracy, precision, recall, and F1 score. Among the models tested, the Support Vector Machine (SVM) achieved the highest accuracy of 0.69, but its precision and F1 score were relatively lower, indicating that it had difficulty distinguishing between classes accurately in some cases. Its recall of 0.69 suggests that it was fairly good at detecting true positive instances but still made significant errors in some areas. Overall, while SVM showed a decent level of performance, its results were not optimal when compared to others. The Random Forest (RF) model came close to SVM in accuracy, with a score of 0.68. However, it outperformed SVM in precision (0.64) and recall (0.68), demonstrating that it was better at correctly identifying the positive instances and had a more balanced performance. The F1 score for RF was 0.64, indicating a relatively good balance between precision and recall, making it a solid model for this task. K-Nearest Neighbor (KNN), with an accuracy of 0.62, showed the lowest performance in this evaluation. Although its recall was 0.62, which indicates that it could identify a fair portion of the positive instances, its precision and F1 score were both lower (0.59), meaning that it frequently misclassified some instances as positive when they were not. This limited the overall effectiveness of the model (Table 4).

Table 4. Performance evaluation of various classification algorithms

ML Models	Accuracy	Precision	Recall	F1 Score
SVM	0.69	0.56	0.69	0.61
Random Forest (RF)	0.68	0.64	0.68	0.64
K-Nearest Neighbor (KNN)	0.62	0.59	0.62	0.59
Decision Tree (DT)	0.60	0.59	0.60	0.59
Gradient Boosting (GBoost)	0.67	0.63	0.67	0.63
Proposed Model	0.75	0.72	0.74	0.73

The Decision Tree (DT) model demonstrated an accuracy of 0.60, similar to KNN. The precision and recall values were also close, standing at 0.59 for both metrics. Its performance was suboptimal compared to other models, with a slightly lower recall rate, suggesting that the DT model struggled to correctly identify positive instances across various classes. Gradient Boosting (GBoost) achieved an accuracy of 0.67, which was higher than both KNN and DT, with a precision of 0.63 and recall of 0.67. The F1 score was 0.63, reflecting a reasonable balance between precision and recall. This model showed a more balanced performance than KNN and DT, but still fell short when compared to the top performers like SVM and RF. Finally, the Proposed Model outperformed all the aforementioned models. With an accuracy of 0.75, precision of 0.72, recall of 0.74, and F1 score of 0.73, the proposed model consistently delivered superior results across all performance metrics. This indicates that it not only achieved high accuracy but also maintained a good balance between detecting true positives and minimizing false positives, making it the most reliable and effective model for thyroid disease prediction in this evaluation.

9. CONCLUSION

In conclusion, this paper presents an advanced approach to predicting thyroid disease using machine learning models, with a particular focus on the XGBoost algorithm. Through hyperparameter optimization, the proposed model demonstrated significant improvements in accuracy and performance over traditional machine learning models, such as Decision Tree, KNN, Random Forest, and Support Vector Machine. By leveraging a shared dataset consisting of demographic and medical data, the model was able to effectively classify patients into various severity levels of thyroid-related diseases. The results suggest that the proposed approach not only enhances predictive accuracy but also offers a more efficient solution for early diagnosis, which can lead to better treatment outcomes for patients. The findings from this study highlight the importance of model optimization in improving the reliability and robustness of machine learning systems in healthcare applications. The proposed XGBoost-based model, with its fine-tuned hyperparameters, outperformed all other models in terms of key evaluation metrics, including accuracy, precision, recall, and F1 score. This demonstrates its potential as a valuable tool for thyroid disease prediction, paving the way for its adoption in real-world medical diagnostics. Future work can explore further refinements to the model, including the integration of more diverse datasets and the exploration of additional machine learning techniques to further enhance prediction capabilities.

References:

1. Sellappan Palaniappan et al., “Intelligent thyroid disease prediction on system using data mining techniques”, IJCSNS Vol 8 no 8(Aug2008)
2. Carlos Ordonez. “Comparing association rules and decision trees for thyroid disease prediction”, ACM, HICOM (2006).
3. MA. Jabbar, Priti Chandra, B.L. Deekshatulu, “Cluster based association rule mining for thyroid attack prediction”, JATIT, vol 32, no 2, (Oct 2011).
4. Ya-Han Hu, Yen-Liang Chen, “Mining Association Rules with Multiple Minimum Supports: A New Mining Algorithm and a Support Timing Mechanism”, 2004 Elsevier B.V.

5. Agrawal, R., Imieilinski, T., Swami, A., "Mining Association Rules between sets of items in large databases" SIGMOD 1993, pp. 207-216.
6. Zhuang Chen, Shibang CAI, Qiulin Song and Chonglai Zhu, "An Improved Apriori Algorithm Based on Pruning Optimization and Transaction Reduction", IEEE 2011.
7. Lee, C.-H., "A Hellinger-based Discretization Method for Numeric Attributes in Classification Learning", Knowledge-Based Systems 20(4), 419–425 (2007).
8. Liu, H., Hussain, F., Tan, C., Dash, M., "Discretization: An Enabling Technique. Data Mining and Knowledge Discovery", 6(4), 393–423 (2002).
9. Margaret H. Danham,S. Sridhar, "Data mining, Introductory and Advanced Topics", Person education , 1st ed., 2006.
10. Aman Kumar Sharma, Suruchi Sahni, "A Comparative Study of Classification Algorithms for Spam Email Data Analysis", IJCSE, Vol. 3, No. 5, 2011, pp. 1890-1895.
11. Geeta, K., & Baboo, S. S., "An Empirical model for thyroid disease classification using evolutionary multivariate Bayesian prediction model", Global Journal of Computer science and technology; E Network, Web & security, Vol. 16(1): 1-10, 2016.
12. Sharma, R., Kumar, S., Maheshwari, R., "Comparative Analysis of Classification Techniques in Data Mining using different datasets", International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 4(12): 125-134, 2015.
13. Sandhya, N., Sharanya, M. M., "Analysis of Classification techniques for efficient Disease Prediction. International Journal of Computer Applications, 155(8): 20-24, 2016.
14. Sudhamanthy, G., Thilagu, M., Padmavathi, G., "Comparative Analysis of R Package Classifiers using Breast Cancer Data set", International Journal of Engineering and Technology (IJET), Vol. 8(5): 2127- 2136, 2015.
15. Rosly, R., Makhtar, M., Awang, M. K., Awang, M. I., & Rahman, M. N. A., "Analyzing the performance of classifiers for medical data sets", International Journal of Engineering and Technology (IJET), Vol. 7(2.15): 136-138, 2018.
16. Maryam, I., Janabi, A., Mahmoud, H. Q., & Hijjawi, M., "Machine Learning classification techniques for thyroid disease prediction: a review", International Journal of Engineering and Technology (IJET), Vol. 7(4): 5373-5379, 2018.
17. Gorade, S. M., Deo, A., & Purohit, P., "A Study of some data mining classification techniques", International Research Journal of Engineering and Technology (IRJET), Vol. 4(4): 3112-3115, 2017.
18. Sumathi, A., Nithya, G., & Meganathan, S., "Classification of thyroid disease using data mining techniques", International Journal of Pure and Applied Mathematics, Vol. 119(12): 13881-13890, 2018.
19. Obeidavi, M. R., Rafiee, A., & Mahdiyar, O., "Diagnosing thyroid disease by neural networks", Biomedical and Pharmacology Journal, Vol. 10(2): 509- 524, 2017.
20. Gopinath, M. P., "Comparative study on classification algorithm for Thyroid dataset", International Journal of Pure and applied mathematics, Vol. 117(7): 53-63, 2017.
21. Pavya, K., Srinivasam, B., "Hybrid thyroid stage prediction models combining classification, clustering and ensemble systems", International Journal of Engineering and Technology (IJET), Vol. 7(4.7): 297- 302, 2018.
22. Ammulu, K., & Venugopal, T., "Thyroid data Prediction using data classification algorithm", International Journal for Innovative Research in science and Technology, International Journal of Engineering and Technology (IJET), Vol. 4(2): 208-212, 2017.