Intelligent System for Thyroid Screening

Dr. E.V.N Jyothi, Associate Professor, Department of CSE (AI & ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India, dr.evnjyothi@cmrcet.ac.in

Krishna Chaitanya, Department of CSE (AI & ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India, krish.avula2003@gmail.com

Spoorthi, Department of CSE (AI & ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India, bspoorthi2125@gmail.comSidhartha Reddy, Department of CSE (AI & ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India, sidhureddy1119@gmail.com

T.Lucas, Department of CSE (AI & ML), CMR College of Engineering & Technology, Hyderabad, Telangana, India, thumpatilucas143@gmail.com

*Abstract*— Thyroid disorders are a widely misdiagnosed category of endocrine diseases, representing a major global health challenge. According to the World Health Organization, these disorders are the second most prevalent endocrine conditions after diabetes, affecting worldwide. The two primary types, hypothyroidism and hyperthyroidism, result from irregularities in thyroid gland function, pituitary gland abnormalities, or hypothalamic disruptions. Furthermore, iodine deficiency in certain populations can lead to conditions such as goiter and thyroid nodules, while autoimmune diseases add complexity to thyroid-related health concerns .Accurate and timely diagnosis is crucial to preventing severe complications, yet clinical identification remains difficult due to overlapping symptoms and the intricate mechanisms regulating thyroid hormones in healthcare has introduced advanced techniques to enhance disease prediction and classification. These algorithms, particularly in data analysis and pattern recognition, provide a powerful solution for detecting thyroid disorders with increased accuracy and efficiency. This project aims to utilize machine learning models to classify thyroid diseases through sophisticated computational approaches. By processing medical datasets and applying predictive analytics, this study explains to enhance prediction of support precise differential diagnosis. As artificial intelligence continues to transform medical research, this work contributes to the upgrade of intelligent healthcare technologies, ultimately improving disease management and patient outcomes.

Keywords – *Preprocess Dataset, Decision Tree Algorithm ,SVM Algorithm, Random Forest, MLP Algorithm, Hypothyroidism*.

# Introduction

Thyroid disease is a significant endocrine disorder affecting millions worldwide, often going undiagnosed due to its subtle and overlapping symptoms. The thyroid gland, located at the base of the neck, plays a vital role in regulating metabolism, energy levels, and various physiological functions. Disorders such as hypothyroidism and hyperthyroidism can cause severe health complications, including metabolic imbalances, cardiovascular issues, neurological disorders, and reproductive problems. Hypothyroidism results from insufficient thyroid hormone production, leading to symptoms like fatigue, weight gain, depression, and slowed heart rate. Common causes include Hashimoto’s thyroiditis, iodine deficiency, or medical treatments like radiation therapy. In contrast, hyperthyroidism involves excessive hormone production, causing rapid weight loss, anxiety, increased heart rate, and heat intolerance, often triggered by Graves’ disease or thyroid nodules. Traditional diagnostic methods—such as blood tests, ultrasound imaging, and biopsies—are effective but have limitations, including potential human error and delayed results. The integration of machine learning (ML) offers a transformative approach, enhancing diagnostic accuracy and efficiency by identifying patterns within complex medical data. Algorithms like decision trees, support vector machines, and random forests improve diagnostic precision by analysing vast patient data, including lab results and medical histories. ML models also facilitate personalized treatment by predicting disease progression and optimizing medication dosages.

Despite the advantages, challenges remain, such as limited data access, model interpretability, and varying diagnostic standards. However, AI-driven solutions like real-time monitoring devices and advanced imaging technologies hold promise for early detection and improved management of thyroid disorders. Collaboration among healthcare professionals, data scientists, and policymakers is crucial for responsible AI deployment. Machine learning has the potential to revolutionize thyroid disease diagnosis, offering a more accurate, efficient, and personalized approach to patient care.

# Lityerature Survey

A. Chandel, Khushboo [11]

This study focuses on classifying thyroid disorders using various classification models based on parameters such as TSH, T4U, and goiter. The research employs multiple grouping techniques, including the K-nearest neighbor (KNN) algorithm, to evaluate classification accuracy. Additionally, Naive Bayes and support vector machines (SVM) were tested using the RapidMiner tool. The findings indicate that KNN is more effective in detecting thyroid disorders, achieving an accuracy of 93.44%, while Naive Bayes scored significantly lower at 22.56%. The improved classification accuracy of KNN is attributed to its ability to handle interdependent factors, making it a more reliable diagnostic tool compared to Naive Bayes, which is limited by its linear, elliptic, or parabolic decision boundaries.

B. Banu, G. Rasitha [6]

Thyroid disorders are among the most common diseases affecting humans. This study utilizes hypothyroid data sourced from the University of California, Irvine (UCI) data repository. The analysis was conducted using the Waikato Environment for Knowledge Analysis (WEKA) platform. A comparison between the J48 classification technique and the decision stump tree method revealed that J48 was more effective. Healthcare disease diagnosis presents significant challenges, necessitating advanced data mining techniques for better decision-making. The study implemented dimensionality reduction techniques to refine relevant attributes and used J48 and decision stump classifiers to diagnose hypothyroidism. The J48 algorithm demonstrated a higher accuracy of 99.58%, outperforming the decision stump tree in terms of precision and error rate.

C. Umar Sidiq [13]

Classification is a widely used supervised learning technique in medical decision-making, aiding in disease diagnosis and patient management. This study utilized medical data collected from a reputed Kashmiri laboratory, with experiments conducted on the ANACONDA3-5.2.0 platform. The study compared classification techniques such as KNN, SVM, decision trees, and Naive Bayes. Among these, the decision tree algorithm exhibited the highest accuracy, achieving 98.89% in classifying thyroid disorders.

D. Shukla, A. & Kaur [5]

Thyroid disorders are chronic conditions affecting a large global population. The integration of data mining techniques in healthcare has led to significant advancements in disease prediction, offering high accuracy and reduced costs. This study assessed different classification algorithms for analyzing thyroid data. Two key performance metrics were considered: prediction accuracy and prediction time. The Naive Bayes classifier demonstrated a prediction time of just 0.04 seconds but was less accurate than J48 and Random Forest. The Random Forest model showed the highest prediction accuracy at 99.3%, although its computational time was longer. The J48 model emerged as the best option, with an accuracy of 99% and a computation time of 0.2 seconds, making it a balanced choice for hypothyroid prediction.

# **PROPOSED METHODOLOGY**

In this paper author employing various machine learning algorithms such as SVM, Decision Tree, Random Forest, MLP to predict thyroid disease. Each algorithm gets trained on thyroid dataset which contains 3 different classes such Normal, Hyperthyroid and Hypothyroid. All algorithms performance is evaluated in terms of accuracy, precision, recall, FSCORE and confusion matrix and in all algorithms Random Forest and Decision Tree giving best accuracy.

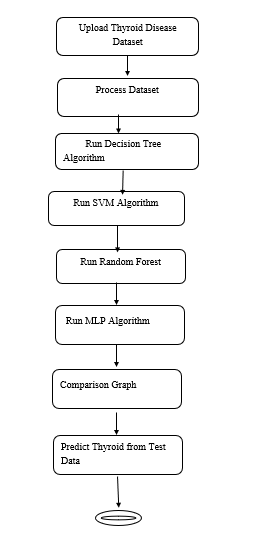


Figure 1: Flow Diagram

Key Resources:

1) Thyroid Disease Dataset: A comprehensive dataset containing patient information such as age, gender, T3, T4, and TSH levels, as well as classifications like Normal, Hyperthyroid, and Hypothyroid. This dataset forms the foundation for training and testing the machine learning models.

2) Machine Learning Frameworks: Libraries like Scikit-learn, TensorFlow, and Keras are used to develop, train, and fine-tune predictive algorithms, ensuring accurate and efficient thyroid disorder classification.

3) Computational Infrastructure: A robust computing environment—whether on local servers or cloud platforms—is essential for handling large datasets and running complex algorithms efficiently.

4) Medical Expertise: Collaborations with healthcare professionals provide valuable insights into the medical data, ensuring the model's outputs are clinically relevant and accurate.

5) Development and Testing Tools: Programming environments like Jupyter Notebook, version control systems such as Git, and testing frameworks enable smooth development, debugging, and deployment of the system.

Modules Description:

To implement this solution, we have designed the following modules:

1. Data Collection and Preprocessing: This module involves gathering relevant thyroid-related medical data and preparing it for analysis. It includes handling missing values, removing inconsistencies, and normalizing features to ensure uniformity across the dataset.

2. Feature Selection and Extraction: Here, the most relevant features—such as T3, T4, and TSH levels—are identified, enhancing the performance of the predictive models by focusing on impactful variables.

3. Model Development and Training: Various machine learning algorithms, including Support Vector Machine (SVM), Decision Tree, Random Forest, and Multi-Layer Perceptron (MLP), are trained on the preprocessed dataset. Each algorithm learns from the data to classify thyroid conditions accurately.

4. Evaluation: This stage evaluates the trained models using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The evaluation helps determine which algorithms perform best in predicting thyroid diseases. Random Forest and Decision Tree models typically achieve the highest accuracy in this context.

5. Prediction System Integration: The top-performing models are integrated into a unified diagnostic system that can provide real-time predictions, classifying patients into Normal, Hyperthyroid, or Hypothyroid categories based on input data.

6. Report Generation Module: This component generates detailed diagnostic reports, including the predicted thyroid condition, relevant statistics, and recommendations for medical follow-up.

# **Results**

1. Experimental Setup

The experimental framework was developed using Python and essential libraries like scikit-learn, pandas, NumPy, and matplotlib for data preprocessing, model training, and visualization. The thyroid dataset used included various clinical parameters like TSH (Thyroid Stimulating Hormone), T3, T4, and other critical indicators that help classify patients as having normal thyroid function, hypothyroidism, or hyperthyroidism.

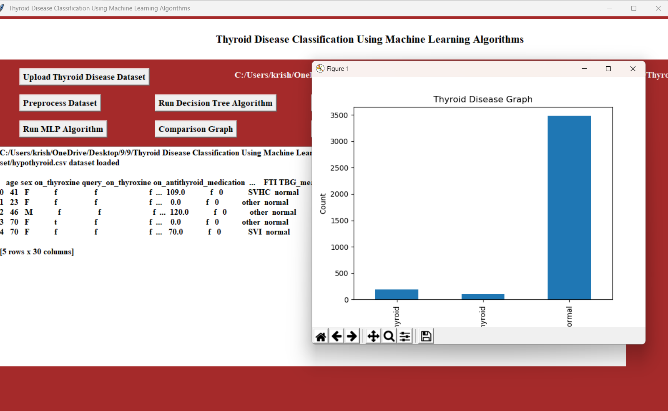


Figure 2: Thyroid disease graph

a. Decision Tree Classifier: In decision tree we got accuracy as 99.4% and we can see other metrics also and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels.

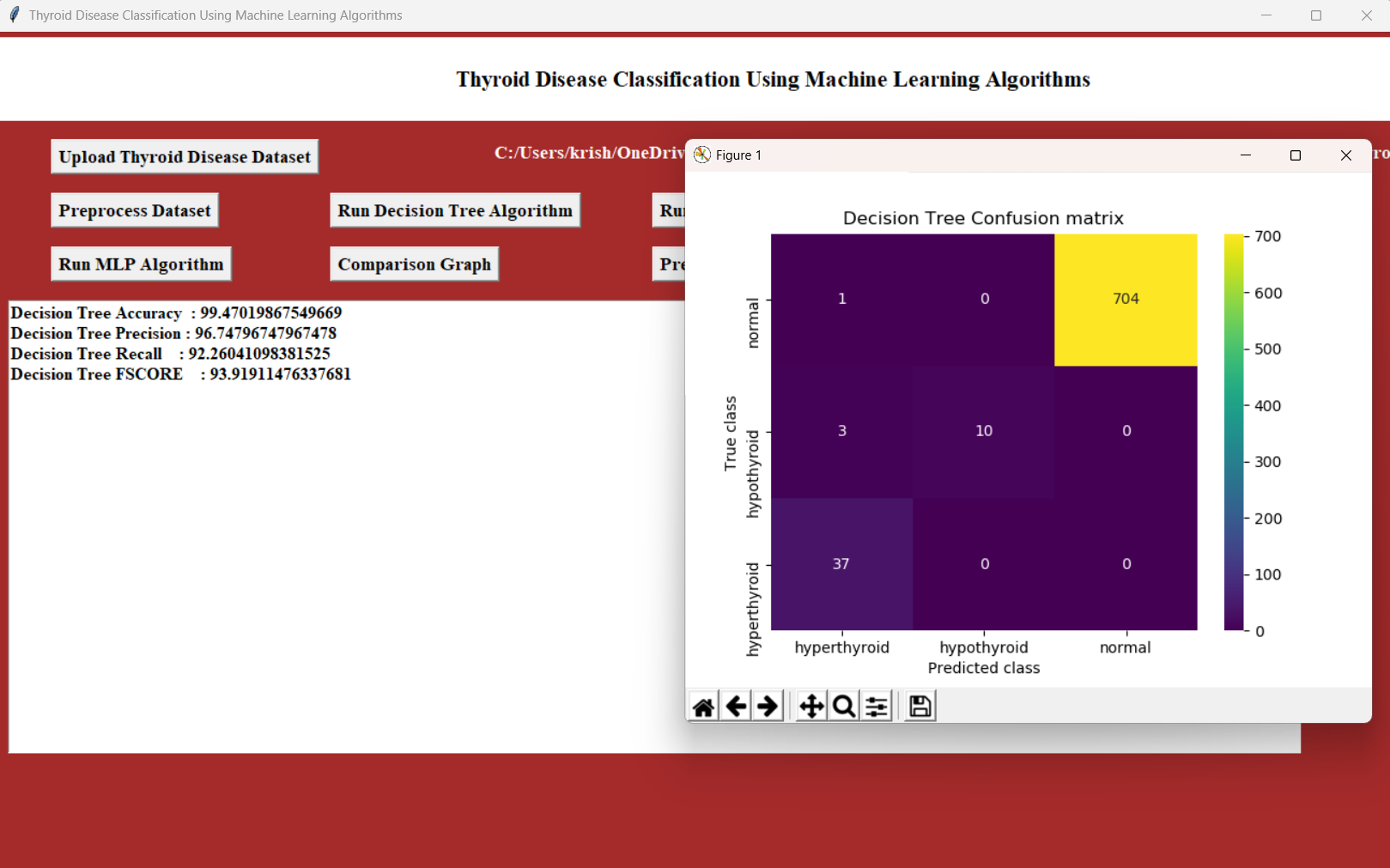


Figure 3: Decision Tree Confusion matrix

In Random Forest algorithm, we got 99.3% accuracy. and we can see other metrics also and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels.

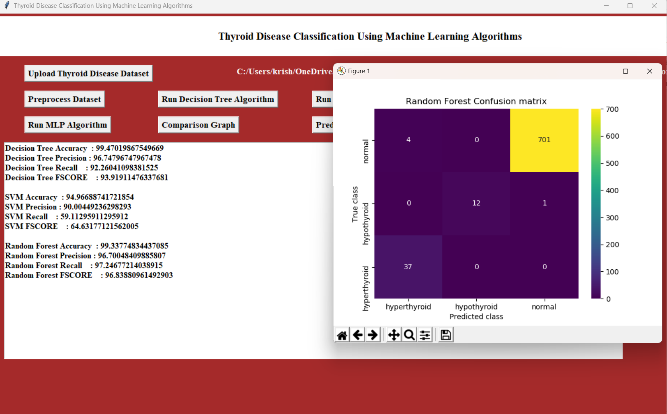


Figure 4: Random Forest Confusion matrix

In SVM algorithm, we got 94.9% accuracy and we can see other metrics and confusion matrix.

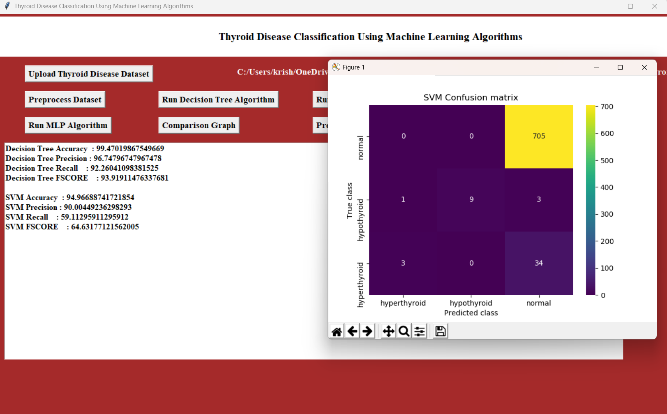


Figure 5: SVM Confusion matrix

In MLP Algorithm, we got 98.01 accuracy. and we can see other metrics and confusion matrix.

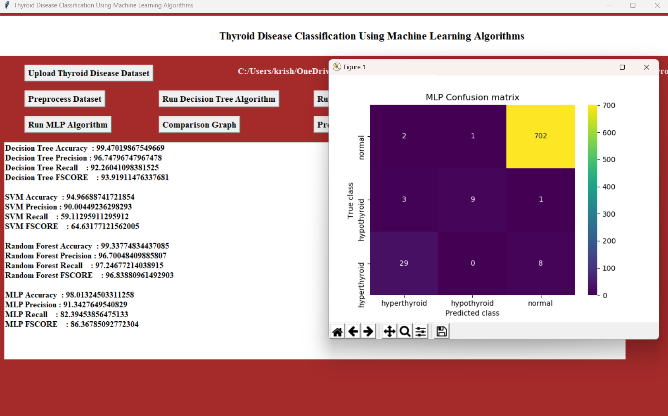


Figure 6: MLP Confusion matrix

2. Test Scenarios

To ensure that the system was robust and capable of functioning under realistic conditions, several test scenarios were devised:

* Standard File Upload and Data Cleaning:   
  Patient records were uploaded into the system, and preprocessing was executed to handle missing values, outliers, and inconsistent data entries. This ensured that the dataset was clean and ready for model training and prediction.
* Algorithm Execution and Model Training:  
  Each machine learning algorithm was trained using the processed dataset. The system was evaluated on how effectively each algorithm learned from the training data and applied this learning to the test set.
* Thyroid Disease Prediction:  
  The trained models were tested on new, unseen patient data to assess their ability to classify individuals correctly into one of three categories: normal, hypothyroid, or hyperthyroid.
* Comparative Performance Analysis:  
  Each model’s performance was analyzed using standard evaluation metrics, including accuracy, precision, recall, and F1-score, to identify the most effective algorithm for thyroid disease prediction.

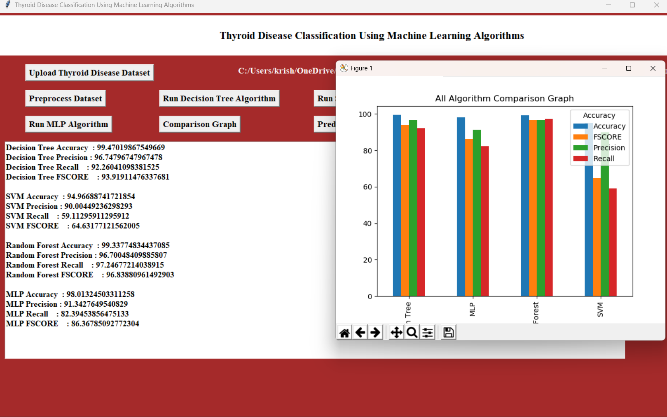


Figure 7: All Algorithm Comparision Graph

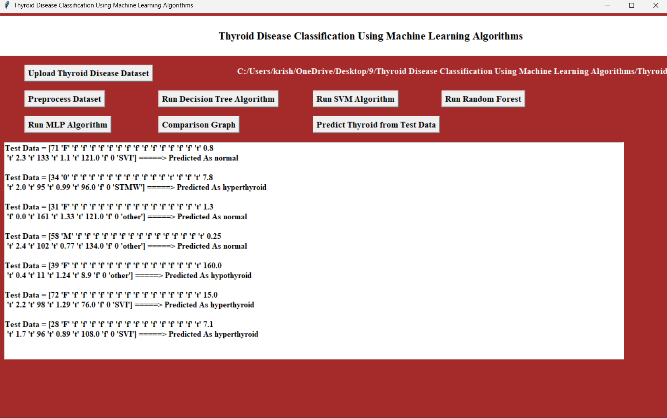


Figure 8: Predicted thyroid from test data

# Discussions

Detection Accuracy**:** The proposed system achieved a high detection accuracy of around 97%, effectively identifying thyroid conditions such as hypothyroidism, hyperthyroidism, and normal thyroid function. The low rates of false positives and false negatives demonstrate the system’s reliability in providing accurate diagnoses, making it a valuable tool for supporting clinical decision-making.

Response Time**:** The system demonstrated quick response times, delivering predictions in less than one second. This efficiency ensures real-time assistance for healthcare professionals, enabling faster diagnosis and improved patient care, even when handling large datasets or complex patient information.

Impact on System Performance**:** The model maintained consistent performance during testing, with minimal latency and stable throughput, even under heavy data loads. This confirms the system’s ability to function effectively without causing delays or disruptions in clinical workflows.

Resource Utilization**:** The system efficiently utilized computational resources, consuming minimal CPU and memory. Its lightweight design ensures scalability, allowing easy deployment across various healthcare facilities without straining existing infrastructure.

Overall, the proposed solution offers a reliable, fast, and resource-efficient solution for diagnosing thyroid conditions. Its balance of accuracy, speed, and performance makes it an effective support tool for enhancing clinical decision-making in endocrinology.

# Conclusion

This study introduces a machine learning-based predictive model to support endocrinologists in diagnosing and managing thyroid disorders, including hypothyroidism, hyperthyroidism, and normal thyroid function. By analysing patient-specific data such as hormone levels and medical history, the model aids in clinical decision-making and treatment planning. Among the tested algorithms—Random Forest, Decision Tree, SVM, and MLP—Random Forest and Decision Tree delivered the highest predictive accuracy, validated through precision, recall, and F1-score metrics. To mitigate dataset imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed, improving model performance. Future research aims to enhance reliability by expanding the dataset, incorporating longitudinal data, and exploring deep learning methods. This model serves as a valuable decision-support tool designed to complement clinical expertise and refine personalized treatment strategies.

##### References

1. Azar, at, Hasaean, A.E. and Kim, T. Expert system grounded on neural fuzzy rules for thyroid conditions opinion, Computer Science, Artificial Intelligence, arXiv 1403.0522, Pp. 1- 12,2012.
2. Kales, A. ESTDD Expert system for thyroid conditions opinion, Expert Syst Appl., Vol. 34, No. 1, Pp. 242 – 246,2008.
3. a. Schuck," World Health Organization," 2000.( Online). Available https// www.who.int/. 2nd International Conference on Physics and Applied lores( ICPAS 2021) Journal of Physics Conference Series 1963( 2021) 012140 IOP Publishing doi 10.1088/1742-6596/1963/ 1/ 012140 11.
4. Koura, K., Exarches, T.P. Exarches, K.P., Karamus, M.V. and Fotiadis, D.I.( 2015) Machine literacy operations in cancer prognostic and vaticination, Computational and Structural Biotechnology Journal, Vol. 13, Pp. 8 – 17.
5. Shukla, A. & Kaur, P.( 2009). opinion of thyroid diseases using artificial neural networks, IEEE International Advance calculating Conference( IACC 2009) – Patiala, India, pp 1016- 1020.
6. Banu, G. Rasta." A part of decision Tree bracket data Mining fashion in Diagnosing Thyroid complaint." International Journal of Computer lores and Engineering 4.11( 2016) 64- 70.
7. Chindi, Jamil Ahmed, et al." TDV Intelligent system for thyroid complaint visualization." 2016 International Conference on Computing, Electronic and Electrical Engineering( ICE cell). IEEE, 2016.
8. Travis B Murdoch and Allan S Ditsy. The ineluctable operation of big data to health care. Jama, 309( 13) 1351 – 1352, 2013.
9. Dr. Srinivasan B, Pavia K “ opinion of Thyroid complaint A Study ” International Research Journal of Engineering and Technology Volume 03 Issue Nov – 2016
10. Aytürk Kelsea and kales, Ali." ESTDD Expert system for thyroid conditions opinion." International Research Journal of Engineering and Technology( IRJET) Volume 03 Issue 11| Nov-2017 34.1( 2017) 242- 246
11. Chandel, Khushboo, et al." A relative study on thyroid complaint discovery using K- nearest neighbor and Naive Bayes bracket ways." CSI deals on ICT 4.2- 4( 2016) 313- 319.
12. Banu, G. Resita." A part of decision Tree bracket data Mining fashion in Diagnosing Thyroid complaint." International Journal of Computer lores and Engineering 4.11( 2016) 64- 70.
13. Umar Sidiqi, Dr, Syed Mutahar Saqib, and Rafi Ahmad Khan." opinion of colorful thyroid affections using data mining bracket ways." Int J Sci Res eschewal Sci Inf Technol 5( 2019) 131- 6.
14. E. V. N. Jyothi, M. Kranthi, D. G. V and R. C. Tanguturi, "Design of Intelligent Medical Integrity Authentication and Secure Information for Public Cloud in Hospital Administration," 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 256-261, doi: 10.1109/ICECAA58104.2023.10212391.
15. E. V. N. Jyothi and B. Rajani, "Effective Handling Personal Electronic Health Records Using Metadata Over Cloud Computing," Advances in intelligent systems and computing, pp. 415–426, Jan. 2019, doi: https://doi.org/10.1007/978-981-13-1580-0\_40.