IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14666

Thyroid Detection System using K Means and Fuzzy C Means

Athul V S[1], Ayswariya V J[2]

Student, MSc Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[1]

Assistant Professor, Department of Computer Science and Applications, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[2]

Abstract: This study explores thyroid disease detection using K-Means and Fuzzy C-Means clustering algorithms. By analyzing patient data, the models classify thyroid conditions efficiently. Comparative evaluation highlights accuracy and effectiveness, aiding early diagnosis. The research emphasizes the importance of machine learning in medical diagnostics, enhancing predictive capabilities for thyroid disorders. The data is first pre- processed, selected, extracted and then classified defect or normal class. This algorithm gives best result for overlapped data also. Existing studies focus on the binary classification tasks. This study improves on prior experience with conventional CNN models (i.e., VGG models), the more advanced CNN architecture, the Exception model was implemented and compared to achieve the automatic diagnosis with increased efficiency and accuracy. It is also likely that the established structure can be easily translated to determine the diagnosis of other disease

Keywords: Thyroid Disease, Data Mining, Fuzzy C Means Clustering.

I. INTRODUCTION

Large amount of data set are extracted to identify and analyze the pattern of data using data mining techniques. This technique is used for discovering the knowledge base from the given data set. All the metabolic process in our body is influenced by the hormone produced by the thyroid gland. Human body system is controlled by the hormone produced by the thyroid. This thyroid disease is most wide spreading and has become more common. Its abnormality may be a simple goiter or life-threatening cancer. A simple goiter needs no treatment it is just an enlarged gland. The abnormality is based on the level of thyroid stimulating hormone. There are two cases where the excess amount of thyroid production is hyperthyroidism and less production of hormone is hypothyroidism. A long period of untreated hypothyroidism might lead to myxedema coma. Myxedema coma is very rare but if affected surgery should be taken immediately. It might even lead to weight gain, dry skin and fatigue along with sleepiness. TSH test is done to check if your thyroid glands are working properly in deep sense checkup is done to check for overactive and underactive conditions that are hyperthyroidism and hypothyroidism respectively. They are usually done on the morning. The normal TSH ranges from 0.4(mIU/L) to 5(mIU/L), when the TSH level is lower than the normal level we call it as underactive called hypothyroidism and when the level is above normal it is overactive called hyperthyroidism. The reasons for the lower and higher level of THS are mainly due to too much of iodine content in your body or graves, disease or due to too much consumption of supplement that naturally contains thyroid hormone.

The thyroid gland is responsible for producing hormones that regulate various bodily functions, including metabolism, heart rate, and energy levels. Thyroid diseases, such as hypothyroidism, hyperthyroidism, and thyroid cancer, can lead to significant health issues if not diagnosed and treated correctly.

In this problem, we aim to cluster patients with thyroid diseases based on their medical attributes using unsupervised learning techniques. Specifically, we will employ two popular clustering algorithms—K-Means and Fuzzy C-Means (FCM)—to analyze and group patients with similar characteristics. The goal is to discover hidden patterns and relationships in the data that might aid in early detection or categorization of thyroid disease.

1. **Data Preprocessing**: Clean and preprocess the thyroid disease dataset, including handling missing values, normalizing data, and selecting relevant features.
2. **Clustering using K-Means**: Apply the K-Means clustering algorithm to group patients based on their attributes (e.g., hormone levels, age, gender, etc.). The number of clusters (K) must be determined through appropriate techniques such as the Elbow Method or Silhouette Score.
3. **Clustering using Fuzzy C-Means**: Use Fuzzy C-Means clustering to allow for soft clustering, where each patient may belong to multiple clusters with varying degrees of membership.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14666

4. **Cluster Evaluation**: Assess the quality of the clusters using metrics such as the **Silhouette Score** and **Adjusted Rand Index (ARI)** to measure how well the data points fit within their respective clusters.
5. **Interpretation of Results**: Interpret the clusters to identify potential patterns and associations related to thyroid disease, which can potentially guide future research, diagnostics, or treatment plans.

What is Known:

1. **Thyroid Disease Data**:

- The dataset contains medical records of patients with thyroid-related disorders, including both healthy and diseased individuals.
- The dataset likely includes features such as age, gender, hormone levels (e.g., TSH, T3, T4), and potentially other medical or demographic attributes.

2. ***Clustering Algorithms***:

- K-Means: A well-known clustering algorithm that groups data into a specified number of clusters, where each data point is assigned to exactly one cluster based on the nearest centroid.
- Fuzzy C-Means (FCM): A clustering algorithm that assigns data points to multiple clusters with varying degrees of membership (i.e., a soft assignment), rather than forcing a single hard classification like K-Means.

What is Unknown:

1. **Disease Categorization**:

- The exact number of clusters needed to represent the data effectively. This will be determined through techniques such as the Elbow Method or Silhouette Score inK-Means and chosen based on clustering characteristics in FCM.
- What specific thyroid conditions (e.g., hypothyroidism, hyperthyroidism, thyroid cancer) or disease severity categories exist in the dataset. Unsupervised learning might reveal unknown groupings that aren't labeled explicitly in the data.

2. ***Optimal Features for Clustering***:

- While we have medical attributes like hormone levels, the exact set of relevant features that would be most effective for clustering is unknown. Feature selection and data preprocessing steps will determine which features provide the most meaningful groupings.

3. ***Cluster Interpretability***:

- The exact meaning or interpretation of the clusters produced by K-Means and Fuzzy C-Means is not known in advance. We will need to analyze the clusters to identify patterns or relationships (e.g., what differentiates a cluster of healthy individuals from those with hypothyroidism).

Aim:

The primary aim of this study is to apply unsupervised learning algorithms (K-Means and Fuzzy C-Means) to cluster patients based on their medical attributes, in order to uncover hidden patterns or groupings related to thyroid diseases. These patterns could aid in understanding the underlying structures of thyroid disease, potentially providing insights for diagnosis or treatment.

Specific Objectives:

1. Cluster patients into distinct groups using K-Means and Fuzzy C-Means algorithms based on features such as hormone levels (TSH, T3, T4), age, gender, and other relevant medical or demographic attributes.
2. Evaluate the clustering performance using metrics such as Silhouette Score and Adjusted Rand Index (ARI) to assess the quality and effectiveness of the clustering.
3. Compare the results of K-Means and Fuzzy C-Means clustering to determine which algorithm better captures the underlying structure of thyroid disease in the dataset.
4. Interpret the resulting clusters to identify potential patterns or insights related to different thyroid conditions (e.g., hypothyroidism, hyperthyroidism, healthy).

Hypothesis:

1. Hypothesis 1: Using unsupervised clustering algorithms like K-Means and Fuzzy C- Means, we will be able to identify distinct groups of thyroid disease patients based on their medical features, such as hormone levels and age.
2. Hypothesis 2: The clusters identified by the Fuzzy C-Means algorithm will be more flexible and potentially offer better insights compared to the hard clustering produced by K-Means, as FCM allows for partial membership in multiple clusters.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14666

3. Hypothesis 3: The clustering patterns identified by these algorithms may correspond to existing thyroid disease categories (e.g., healthy, hypothyroidism, hyperthyroidism), or they may uncover new, previously unidentified disease subgroups.
4. Hypothesis 4: The optimal number of clusters (K) for the K-Means algorithm will be determined using the Elbow Method, and Fuzzy C-Means will assign data points with varying degrees of membership across these clusters, leading to more nuanced classification.

II. RELATED WORK

The application of clustering algorithms such as K-Means and Fuzzy C-Means (FCM) for thyroid disease detection has garnered significant attention due to their ability to classify patients based on patterns in large datasets, particularly in the realm of medical imaging and clinical data. These algorithms can segment patients into groups corresponding to different thyroid disorders, such as hypothyroidism, hyperthyroidism, and thyroid cancer, by analyzing features like thyroid hormone levels, ultrasound images, and patient medical history. The study and evaluation of these algorithms involve assessing their performance in terms of classification accuracy, clustering effectiveness, computational efficiency, and clinical applicability.

Study Design and Data Collection

The first step in the study involves the collection of a comprehensive dataset that includes a variety of patient characteristics relevant to thyroid disease. This could include thyroid function test results (e.g., TSH, T3, T4 levels), ultrasound images of the thyroid gland, patient demographics, and clinical histories. The dataset must be large and diverse enough to capture a wide range of thyroid conditions, from benign conditions like goiters to more serious disorders such as thyroid cancer. Medical imaging data is often preprocessed to extract relevant features, such as the size, shape, and texture of thyroid nodules in ultrasound images, which can be analyzed using machine learning algorithms like Convolutional Neural Networks (CNNs) or directly with clustering methods like K-Means and FCM.
Algorithm Evaluation Criteria

To evaluate the performance of K-Means and Fuzzy C-Means algorithms, several key criteria need to be considered:

1. Clustering Accuracy: This refers to how well the algorithm can classify patients into meaningful groups that align with clinically significant thyroid conditions. Evaluation metrics like Silhouette Score, Rand Index, or Adjusted Rand Index can be used to measure the quality of clustering. For example, a high Silhouette Score indicates that the clusters are well-separated and meaningful. In clinical applications, the goal is to ensure that each cluster corresponds to a distinct thyroid disease or health condition, aiding in accurate diagnosis and treatment planning.
2. Convergence Rate: The rate at which the algorithm converges to a stable solution is crucial, particularly in large medical datasets where computational resources can be a limiting factor. The K-Means algorithm generally converges faster than FCM due to its simpler membership function (hard assignment of data points), but FCM may take longer to converge as it calculates the degree of membership for each data point in each cluster. The number of iterations required for convergence, as well as the computational time for clustering, should be evaluated for efficiency.
3. Cluster Interpretability: It is important that the clusters generated by the algorithms are interpretable and actionable from a clinical perspective. This means that the clusters should reflect distinct and recognizable patterns of thyroid disease, such as different stages of thyroid dysfunction or the presence of thyroid nodules. The Fuzzy C-Means algorithm, with its soft clustering approach, may provide more insight into cases where patients exhibit mixed symptoms or conditions, such as subclinical hypothyroidism. The interpretability of the clusters should be evaluated by clinical experts who can assess whether the groups make sense in terms of real-world thyroid disease diagnoses.
4. Clinical Application: The ultimate goal of using clustering algorithms is to assist in improving patient diagnosis and treatment. The system's performance should be evaluated based on its ability to assist healthcare providers in identifying thyroid conditions early, suggesting treatment pathways, and reducing diagnostic errors. This can be assessed through retrospective studies, where the algorithm is applied to a set of clinical cases, and its performance is compared to traditional diagnostic methods or expert opinions. A confusion matrix (showing true positives, false positives, true negatives, and false negatives) is useful for evaluating how well the system performs in terms of correctly diagnosing different thyroid conditions.

III. LITERATURE SURVEY

- The paper "An Improved Framework for Detecting Thyroid Disease Using Filter-Based Feature Selection and Stacking Ensemble" [1] presents methodological strategies for thyroid disease classification and prediction have been provided. The performance of five distinct machine-learning base learners and their integration into a stacked ensemble were explored. This approach sets our study apart from prior thyroid disease classification research using ML.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14666

The classifiers were applied to a thyroid disease dataset, where the combined predictive power of the base classifiers through the stacking method, together with the filter-based method, consistently surpassed individual model predictions. Our findings highlight the stacking ensemble model's effectiveness in improving thyroid disease

- The paper " Analysis of Thyroid Disease Using K Means and Fuzzy C Means Algorithm"[2] presents is designed to help medical professionals who work in diagnosing thyroid disease. The application requires a Lab user and a medical professional to upload a thyroid image of the patient. The application pre-processes the thyroid image and infers the image to the predictive model. The output of the model is then displayed to the medical professional.The system should be able to give information that medical professional can appropriately understand and gain insight from it. This project contains many aspects of research that support deep learning's ability to find thyroid disease within thyroid image. The data used to train the model was gathered from a deep learning competition. The radiologists determined whether a nodule is thyroid disease or not and this location has been specified in coordinates. The project integrates different topics in Computer Science to try and solve a real world problem in the medical domain. The project consists of data mining and software development to deliver a proof of concept. The application is a thyroid disease detection system to help doctors make better and informed decisions when diagnosing thyroid disease. In this study, thyroid disease data set is clustered based on fuzzy algorithm. In this fuzzy k-means algorithm used for classification of thyroid disease.

- The paper "Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach" [3] signifies machine learning and data mining techniques to benefit the medical field and healthcare system. According to the regular protocol, this study will help the doctors use this as a supplementary system. We have evaluated the dataset based on precision and recall. Random forest was performed to be 94.8 percent accurate on average. Random forest is the most efficient in classification, and KNN is the least efficient. On the other hand, ANN and naïve Bayes performed a level above the average of the KNN. With more training and a more extensive dataset, as expected, there will be better results from the artificial neural network. Our proposed method may also be helpful in creating a medical-related application or use it with neuro-fuzzy interference. The efficient and accurate diagnosis of thyroid disease will benefit the whole medical community. The healthcare system can be further enhanced, and better medical decisions can be taken.

- The paper " Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach " [4] signifies machine learning and data mining techniques to benefit the medical field and healthcare system. According to the regular protocol, this study will help the doctors use this as a supplementary system. We have evaluated the dataset based on precision and recall. Random forest was performed to be 94.8 percent accurate on average. Random forest is the most efficient in classification, and KNN is the least efficient. On the other hand, ANN and naïve Bayes performed a level above the average of the KNN. With more training and a more extensive dataset, as expected, there will be better results from the artificial neural network. Our proposed method may also be helpful in creating a medical-related application or use it with neuro-fuzzy interference. The efficient and accurate diagnosis of thyroid disease will benefit the whole medical community. The healthcare system can be further enhanced, and better medical decisions can be taken.

- The paper " Exploring the Challenges of Diagnosing Thyroid Disease with Imbalanced Data and Machine Learning:"[5] presents the uses imbalanced data to discover the most recent ML-based and data-driven developments and strategies in diagnosing thyroid disease. When developing ML-based systems for predicting thyroid disease in the real world, including real-patient data and using interpretable machine learning methods to explain the final predictions is essential accurately. A comprehensive review of 41 papers suggests that more research is needed to prove reliable performance in healthcare settings. Although Deep Learning has come to dominate the area, SMOTE is still widely used as an Over-Sampling technique for handling unbalanced data by many academics and practitioners. Many researchers have noticed the development of an RF-based model for predicting thyroid disease since it is easier to train and can handle many features. Another big attraction is that they resist overfitting, making them useful in various machine-learning applications. The limits of ML that are discussed in the discussion sections may guide the direction of future research. Regardless, ML-based thyroid disease detection utilizing imbalanced data and innovative techniques is expected to uncover numerous undiscovered opportunities in the future.

- The paper " Thyroid Disease Prediction Using Machine Learning Approaches " [6] Rafikhan et al. has used a clinical data of Kashmir of 807 patients and UCI thyroid repository of ''new thyroid'' has only 215 instances. Proposed method has not taken this data set for thyroid prediction; it will consider in future work and measure accuracy using decision tree and kNN. Hence, according to the data set which is used in this work, the accuracy obtained is satisfactory. The current scenario is of the developing of the models that help in the various sectors of life using the machine

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14666

learning. The availability of data and its generation day by day increased a chance for the computer scientists to make prediction and analysis on such data sets that make the human life better and comfort. This study is concern with this motivation. The prediction and classification of any data depends on the data set itself and the various algorithms that are used. If anyone organizes a better data set of real time and applies various other machine leaning and deep learning algorithms such as SVM, Naive Bayes, auto encoders, ANNs and CNNs then further better results may be achieved.

- The paper " Hypothyroidism Prediction and Detection Using Machine Learning" [7] signifies Computer Aided diagnosis system becomes the vital goal in different countries, due to its effectiveness in reducing the human mistakes that comes from less experience and rare of specialists in rural countries. Machine learning which is formed a major part of artificial intelligence (AI) behaves a vital role in CAD system. different research exhibits the effectiveness of machine learning in diagnosis and detection of abnormalities cases. On top of that, it is facilities the rule of the physician and save time. The goal of our work is to use or embody machine learning to benefit the different fields, especially the health field, to reduce errors and incorrect diagnosis. This paper aimed to analyze a large database to build a classifier can diagnose hypothyroidism cases. It turns out that the learner Decision Tree is the most suitable learner, due to the nature of his sequential and simple work. In the next stage, we seek to expand the work of this learner to include other disease categories central and secondary hypothyroidism, and we may include other diseases like hyperthyroidism and cancer

- The paper " Improving the diagnosis of thyroid cancer by machine learning and clinical data" [8] signifies utilized machine learning methods to improve the diagnosis of malignant thyroid nodules. We collected a real dataset of 724 patients' demographic and clinical information. Overall, the machine models exhibit satisfactory prediction performance. The average accuracy and AUROC of the six models are 0.78 and 0.85, respectively. In practice, an AUROC greater than 0.8 indicates excellent discrimination between binary outcomes19. One encouraging result of our study is the superior model performance over the expert assessment. The best-performed model, random forest, beat the expert assessment by 11% on accuracy, and 12% on F1 score, the two general measurements. One interpretation of better prediction by machine learning models is that they are able to capture the complex nonlinear relationships among different variables. Such relationships are implicitly contained in the dataset and are challenging for humans to identify. Te models are also more aggressive in predicting nodules as malignant. As a result, the machine learning model is valuable for diagnosing thyroid cancer.

- The paper " Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution" [9] signifies y detection of thyroid disorders is critical to avoid such complications. This study employs a differential evolution-based optimization algorithm to find optimal parameters for machine learning models to obtain higher performance for thyroid disease detection. It is further aided by data augmentation using the CTGAN model. Experimental results suggest that an accuracy of 0.998 can be obtained using the optimized AdaBoost model by differential evolution. These results are further validated by k-fold cross-validation and performance appraisal with state-of-the-art approaches. Results indicate that contrary to linear models, ensemble models tend to show better performance. Machine learning models show better results using augmented datasets than deep learning models. This study provides two major contributions to enhancing thyroid detection. Using a differential evolution algorithm for hyperparameter optimization provides improved performance by the machine learning models compared to existing studies where conventional hyperparameter optimization is carried out. Secondly, CTGAN helps to balance the number of samples of each class which mitigates the probability of model bias and overfitting. Therefore, the models show robust performance and are generalizable compared to existing models. We intend to increase the dataset size to further analyze the performance of deep learning models in the future.

- The paper " Thyroid Detection using Machine learning by Savita Anil Adhav " [10] presents the problem of domain is classification of thyroid patient.so our algorithms for that are varies classification algorithms. K-means, random forest, decision tree, svm and some deep learning algorithms have given best results to classify thyroid patient. K-means is a clustering algorithm in which each observation is partitioned into a single cluster with no information about how confident we are in this assignment. When a new patient's data is provided to model it check for nearest cluster and then accordingly classify the patient. Decision Tree- it in machine learning algorithm in which at each layer classification is done on the basis of data.in training phase, a tree is built which classify the new instance. Random Forest - A random forest consists of multiple random decision trees. Two types of randomness are built into the trees. First, at each tree node, a subset of features is randomly selected to generate the best split. Second, each tree is built on a random sample from the original data. it is generally used to increase the accuracy. Svm-support vector machine classifies the data using support vectors. it can also work for nonlinear data.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14666

SL NO.	NAME OF THE PAPER	METHOD	ACCURACY	ADVANTAGES	DISADVANTAGES
1	An Improved Framework for Detecting Thyroid Disease Using Filter- Based Feature Selection and Stacking Ensemble	Filter-Based Feature Selection, Stacking Ensemble Learning	95%	Improved generalization, reduces bias and variance	High computational cost due to multiple base models and meta- model training
2	Analysis of Thyroid Disease Using K-Means and Fuzzy C-Means Algorithm	K-Means, Fuzzy C-Means (FCM)	95%	Effective for well-separated clusters	Requires pre- specifying the number of clusters, less effective for complex datasets
3	Retracted: Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach	SVM, GBM, XGBoost, LightGBM	93.7%	Automates diagnostic processes, reduces review time for doctors	Heavily dependent on high-quality labeled data, potential credibility issues due to retraction
4	Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach	SVM, ANN, Naïve Bayes	85% - 95%	Detects complex patterns in large datasets	Risk of overfitting, requires careful tuning
5	Exploring the Challenges of Diagnosing Thyroid Disease with Imbalanced Data	Logistic Regression, Random Forest, XGBoost, LightGBM	92% - 95%	Fast processing for large datasets	Sensitiveto noisy/missing data, requiring extensive preprocessing
6	Thyroid Disease Prediction Using Machine Learning Approaches	Decision Tree, Random Forest, SVM	81.25%	Can detect early-stage thyroid disease	Risk of overfitting if not properly regularized
7	Hypothyroidism Prediction and Detection Using Machine Learning	Decision Tree, Random Forest, ANN, Naïve Bayes	80% - 90%	High accuracy in difficult cases	Model overfitting, tuning complexity
8	Improving the Diagnosis of Thyroid Cancer by Machine Learning and Clinical Data	XGBoost, LightGBM, CatBoost, CNN, SVM	90% - 95%	Adaptability across different populations	Overfitting risk with too many features

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

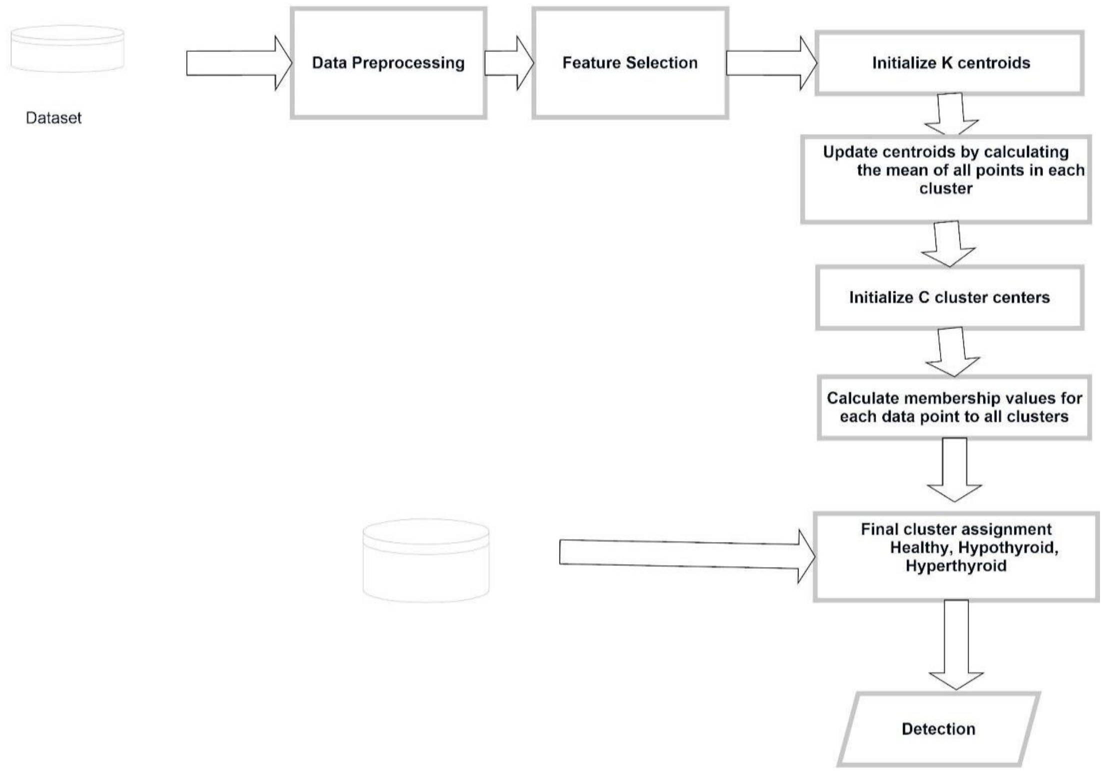
Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14666

9	Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution	Differential Evolution (DE)	85% - 95%	No need for gradient information, works well with complex spaces	Requires parameter optimization, increasing complexity
10	Thyroid Detection Using Machine Learning by Savita Anil Adhav	Naïve Bayes, Logistic Regression, KNN	85% - 95%	Reduces manual diagnosis costs	Lack of interpretability in complex models

Table 2.2.3.1: Comparison Of Different Papers Review.

IV. SYSTEM FLOW DIAGRAM



V. PSEUDOCODE

Pseudocode for K-Means Algorithm

1. Load the dataset (thyroid disease dataset with clinical features)
2. Preprocess the data: a. Handle missing values (remove or impute missing data) b. Normalize the data (scale features to a common range) c. Perform feature selection (choose relevant features)
3. Define the number of clusters (k) based on the dataset or using a method like the Elbow method
4. Initialize k cluster centroids randomly
5. Repeat until convergence: a. Assign each data point to the nearest centroid b. Update the centroid of each cluster based on the mean of the data points assigned to it
6. Once convergence is reached, output the final clusters
7. Evaluate the clustering results using metrics such as accuracy, precision, recall, and silhouette score
8. Visualize the results (optional)
9. End

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14666

Pseudocode for Fuzzy C-Means Algorithm

1. Load the dataset (thyroid disease dataset with clinical features)
2. Preprocess the data:
   a. Handle missing values (remove or impute missing data)
   b. Normalize the data (scale features to a common range)
   c. Perform feature selection (choose relevant features)
3. Define the number of clusters (c) and the fuzziness parameter (m, typically m = 2)
4. Initialize the fuzzy membership matrix randomly (membership values between 0 and 1)
5. Repeat until convergence (or a maximum number of iterations):
   a. Compute the cluster centroids based on the weighted average of the data points, weighted by their membership
   b. Update the membership matrix based on the distance between data points and centroids, adjusting membership values
   c. Check for convergence by comparing the old and new membership matrices (if the change is below a threshold, stop)
6. Once convergence is reached, output the final clusters with membership values for each data point
7. Evaluate the clustering results using metrics such as accuracy, precision, recall, and fuzzy partition coefficient
8. Visualize the results (optional)
9. End

VI. CONCLUSION

The research on thyroid disease detection using machine learning, particularly through the application of Convolutional Neural Networks (CNNs), demonstrates significant potential in enhancing diagnostic accuracy and efficiency. By leveraging advanced algorithms and deep learning techniques, such as K-Means, Fuzzy C-Means, and CNN, we can move toward more automated, reliable, and faster methods for identifying thyroid-related disorders. The study also proposes an integrated hospital system that streamlines the diagnostic process, making it more accessible and effective for healthcare providers, technicians, and patients alike.

Through the use of CNN, this research has shown that medical image data (such as thyroid scans) or clinical data (hormone levels) can be effectively processed to identify patterns and classify thyroid diseases with high accuracy. While traditional methods of diagnosis are still widely used, they often involve subjective judgment and lengthy test results. By employing deep learning models like CNN, we have the ability to automate these processes, reducing human error and improving the overall speed of diagnosis.

The integration of this machine learning approach into a hospital management system brings additional benefits by automating administrative tasks, enabling doctors to access diagnostic results and patient histories efficiently, and allowing patients to receive timely feedback on their health status. Furthermore, by allowing technicians to input data easily into the system and enhancing their workflow, the overall operational efficiency of the hospital is improved.

In conclusion, this research illustrates how machine learning can be leveraged not just as a diagnostic tool but as part of a holistic healthcare solution that brings together data management, diagnosis, and patient care. Despite the promising results, challenges like data quality, class imbalance, and the need for model interpretability remain areas of active research. Future work could focus on further improving model performance, integrating larger and more diverse datasets, and incorporating more advanced hybrid models to better handle medical data's inherent complexities.
Ultimately, the combination of CNN-based models for thyroid disease detection and systematic hospital management integration will lead to faster, more accurate diagnoses, better patient outcomes, and more efficient healthcare systems.

REFERENCES

[1]. L. Aversano, M. L. Bernardi, M. Cimitile, A. Maiellaro, and R. Pecori, ‘‘A systematic review on artificial intelligence techniques for detecting thyroid diseases,’’ PeerJ Comput. Sci., vol. 9, p. e1394, Jun. 2023.
[2]. S. Keestra, V. H. Tabor, and A. Alvergne, ‘‘Reinterpreting patterns of variation in human thyroid function: An evolutionary ecology perspective,’’ Evol., Med., Public Health, vol. 9, no. 1, pp. 93–112, 2021.
[3]. S. Prathibha, D. Dahiya, C. R. Rene Robin, C. Venkata Nishkala, and S. Swedha, ‘‘A novel technique for detecting various thyroid diseases using deep learning,’’ Intell. Autom. Soft Comput., vol. 35, no. 1, pp. 199–214, 2023.
[4]. G. Mariani, M. Tonacchera, M. Grosso, F. Orsolini, P. Vitti, and H. W. Strauss, ‘‘The role of nuclear medicine in the clinical management of benign thyroid disorders—Part 1: Hyperthyroidism,’’ J. Nucl. Med., vol. 62, no. 3, pp. 304–312, Mar. 2021.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14666

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

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

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

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

I. Iakovou, E. Giannoula, and C. Sachpekidis, ''Imaging and imagingbased management of pediatric thyroid nodules,'' J. Clin. Med., vol. 9, no. 2, p. 384, Feb. 2020.

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

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

[10]. S. S. Islam, M. S. Haque, M. S. U. Miah, T. B. Sarwar, and R. Nugraha, ''Application of machine learning algorithms to predict the thyroid disease risk: An experimental comparative study,'' PeerJ Comput. Sci., vol. 8, p. e898, Mar. 2022.

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

[12]. T. Alyas, M. Hamid, K. Alissa, T. Faiz, N. Tabassum, and A. Ahmad, ''Empirical method for thyroid disease classification using a machine learning approach,'' BioMed Res. Int., vol. 2022, pp. 1–10, Jun. 2022.

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

[14]. D. C. Yadav and S. Pal, ''Prediction of thyroid disease using decision tree ensemble method,'' Hum.Intelligent Syst. Integr., vol. 2, nos. 1–4, pp. 89–95, Dec. 2020.G. Ciaburro, ''An ensemble classifier approach for thyroid disease diagnosis using the adaboostm algorithm,'' in Machine Learning, Big Data, and IoT for Medical Informatics. Amsterdam, The Netherlands: Elsevier, 2021, pp. 365–387.

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

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

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

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

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

[19]. T. Akhtar, S. Arif, Z. Mushtaq, S. O. Gilani, M. Jamil, Y. Ayaz, and S. I. Butt, ''Ensemble-based effective diagnosis of thyroid disorder with various feature selection techniques,'' in Proc. 2nd Int. Conf. Smart Syst. Emerg. Technol. (SMARTTECH), May 2022, pp. 14–19.

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

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

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

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

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

DOI: 10.17148/IJARCCE.2025.14666

[24]. Idarraga, A.J., Luong, G., Hsiao, V., Schneider, D.F.: False negative rates in benign thyroid nodule diagnosis: Machine learning for detecting malignancy. J. Surg. Res. 268, 562– 569 (2021)
[25]. Aversano, L., Bernardi, M.L., Cimitile, M., Iammarino, M., Macchia, P.E., Nettore, I.C., Verdone, C.: Thyroid disease treatment prediction with machine learning approaches Procedia Computer Science 192, 1031–1040 (2021)
[26]. Kwon, M.-R., Shin, J., Park, H., Cho, H., Hahn, S., Park, K.: Radiomics study of thyroid ultrasound for predicting braf mutation in papillary thyroid carcinoma: preliminary results. Am. J. Neuroradiol. 41(4), 700–705 (2020)
[27]. Chen, D., Hu, J., Zhu, M., Tang, N., Yang, Y., Feng, Y.: Diagnosis of thyroid nodules for ultrasonographic characteristics indicative of malignancy using random forest. BioData mining 13(1), 1–21 (2020)
[28]. Razia, S., Siva Kumar, P., Rao, A.S.: Machine learning techniques for thyroid disease diagnosis: a systematic review. Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough, 203–212 (2020)
[29]. 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. 34(2), 396–411(2020)
[30]. Chaganti, R., Rustam, F., De La Torre Díez, I., Mazón, J.L.V., Rodríguez, C.L., Ashraf, I.: Thyroid disease prediction using selective features and machine learning techniques. Cancers 14(16),3914 (2022)
[31]. Shankar, K., Lakshmanaprabu, S., Gupta, D., Maseleno, A., De Albuquerque, V.H.C.: Optimal feature-based multi-kernel svm approach for thyroid disease classification. J. Supercomput. 76(2), 1128–1143 (2020)
[32]. Riajuliislam, M., Rahim, K.Z., Mahmud, A.: Prediction of thyroid disease (hypothyroid)in early stage using feature selection and classification techniques. In: 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), pp. 60–64 (2021). IEEE
[33]. Hosseinzadeh, M., Ahmed, O.H., Ghafour, M.Y., Safara, F., Ali, S., Vo, B., Chiang, H.- S., et al.: A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things. J. Supercomput. 77(4), 3616– 3637 (2021)
[34]. Sankar, S., Potti, A., Chandrika, G.N., Ramasubbareddy, S.: Thyroid disease prediction using xgboost algorithms. Journal of Mobile Multimedia 18(3), 1–18 (2022)
[35]. Jha, R., Bhattacharjee, V., Mustafi, A.: Increasing the prediction accuracy for thyroid disease: A step towards better health for society. Wireless Pers. Commun. 122(2), 1921–1938 (2022)
[36]. 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 Computer Science 8, 898 (2022)
[37]. Prathibha, S., Dahiya, D., Robin, C., Nishkala, C.V., Swedha, S.: A novel technique for detecting various thyroid diseases using deep learning. Intell. Autom. Soft Comput 35(1), 199–214 (2023)
[38]. Tang, F., Ding, J., Wang, L., Ning, C.: A novel distant domain transfer learning framework for thyroid image classification. Neural Process. Lett. 55(3), 2175–2191 (2023)
[39]. Sureshkumar, V., Balasubramaniam, S., Ravi, V., Arunachalam, A.: A hybrid optimization algorithm-based feature selection for thyroid disease classifier with rough type-2 fuzzy support vector machine. Expert. Syst. 39(1), 12811 (2022)
[40]. Kaggle, E.F.W.: Thyroid diseases dataset. https://www.kaggle. com/datasets/emmanuelfwerr/thyroid-disease-data (2022)
[41]. Salerno, M., Improda, N., Capalbo, D.: Management of endocrine disease subclinical hypothyroidism in children. Eur. J. Endocrinol. 183(2), 13–28 (2020)