INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Early Prediction of Hypothyroidism using Chi-Square Feature Selection and SMOTE-ENN with Randomized Search CV Compared to Grid-Based Search

K Chitra[1]*, **Dr. G Arockia Sahaya Sheela**[2]*

1 Research Scholar, Department of Computer Science, St. Joseph's college (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, Tamil Nadu, India
2 Assistant Professor, Department of Computer Science, St. Joseph's college (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, Tamil Nadu, India

OPEN ACCESS

Received: 03/12/2025
Accepted: 29/12/2025
Published: 18/01/2026

Citation: Chitra K, Dr. G Arockia Sahaya Sheela (2026) Early Prediction of Hypothyroidism using Chi-Square Feature Selection and SMOTE-ENN with Randomized Search CV Compared to Grid-Based Search. Indian Journal of Science and Technology 19(1): 8-18. https://doi.org/10.17485/IJST/v19i1.1928

* **Corresponding authors**.

chitravijayalakshmi75@gmail.com

arockiasahayasheela_cs1@mail.sjctni.edu

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: To predict hypothyroid disorder at an earlier stage using Machine Learning Algorithms. **Method:** An early diagnosis of hypothyroidism was addressed through a machine-learning algorithm-based framework incorporating data pre-processing, handling imbalance dataset, selecting relevant features, data splitting, and model optimization. An unbalanced dataset was converted to a balanced dataset using SMOTEEN. To address the effect of class imbalance and boost the learning of minority class data elements. Feature selection is performed using a filter-based Chi-square test to discover the most appropriate features, thus enhancing the model performance and minimizing overfitting. Random Search Cross-Validation (Randomized CV) was used to train and optimize a subset of important machine learning classifiers, including Decision tree (DT), Random Forest (RF), and K closest neighbor (KNN), to determine the optimal hyperparameters. **Findings:** An ensemble learning strategy using voting techniques was implemented to improve predictive performance. The effectiveness of each model was measured using performance metrics such as precision, recall, F1-score, and accuracy of 98.81%. The results of the experiments show that the proposed method greatly enhances categorization and can be a reliable instrument for the early detection of hypothyroidism in clinical decision support systems. **Novelty:** Random Search is a well-defined structure in hyperparameter tuning for the prediction of hypothyroidism earlier and faster, ensuring computational efficacy and robust performance, and assisting as a stepping stone for advanced optimization strategies.

Keywords: Hypothyroidism; SMOTE-ENN; Pre-processing; Feature selection; Performance Metrics

1 Introduction

In the early days, the basic resampling technique, SMOTE, was frequently used for thyroid disease prediction. However, this approach is not suitable for avoiding noise and leads to overfitting, particularly in highly imbalanced datasets. In this study, a hybrid resampling strategy with SMOTE-ENN (Edited Nearest Neighbor) was introduced to overcome this limitation. This method not only creates synthetic samples for minority classes but also uses Edited Nearest Neighbor filtering to eliminate noisy and wrongly categorized samples. This guarantees a more balanced and clearer dataset. Additionally, we employed randomized search cross-validation (CV) instead of conventional random search techniques to improve model resilience and reduce bias caused by random data distributions. Compared with current methods, this unified strategy produces a more stable and dependable classification model by enhancing the data quality and generalization performance.

Hypothyroidism occurs when the thyroid gland secretes a small quantity of thyroid hormones to meet the body's requirements. The lack of these hormones, which regulate metabolism, can lead to a number of symptoms and health concerns, such as sadness, headaches, fatigue, unexpected weight gain, and cold sensitivity. Because these symptoms develop gradually, it may be difficult to obtain an early diagnosis using routine clinical examinations. Hypothyroidism is a serious problem; if it is not detected at the right time, it can result in serious health issues, such as cardiovascular disease, infertility, and cognitive difficulties. Therefore, commencing early medical action and improving patient outcomes depend on early and accurate condition prediction. Recent developments in machine learning, which can recognize intricate patterns in medical data that might not be apparent using traditional techniques, have made powerful tools for early disease identification.

However, issues such as class imbalance and superfluous or unnecessary features can significantly impair the performance of prediction models. This study proposes an inclusive architecture that integrates several machine learning models with advanced data preprocessing techniques to address this problem. The class imbalance problem was handled by SMOTEEN (Synthetic Minority Over-Sampling Technique with Edited Nearest Neighbor), which efficiently balanced a dataset through over-sampling minority classes and noise removal of samples. To improve the model efficiency and accuracy, the chi-square statistical test was used. Cross-validation was used to verify the robustness and generalizability of the model across various data splits. In addition, an ensemble learning approach based on voting was used to enhance the predictive performance and mitigate individual model biases. This is a combination of the predictions of several classifiers, taking advantage of each classifier's strength to generate a more accurate and consistent result. The proposed system was tested using common classification metrics, such as precision, recall, F1-score, and accuracy. As a promising tool for clinical decision support and preventive medicine, the results illustrate the efficiency of the combined method for the early prediction of hypothyroidism.

Recently, research on hypothyroidism has significantly expanded. It uses machine learning to improve prediction, diagnosis, and treatment with various algorithms, such as decision tree, random forest, and K-nearest neighbor. The objective was to create a range of predictive models that would assist healthcare professionals in diagnosing and treating hypothyroidism. To start working on the early diagnosis of hypothyroidism, Tirumala AK, et al.[(1)] applied a range of machine learning approaches, like Decision Tree (DT), Support Vector Machine (SVM), and Logistic Regression (LR) and concluded Decision Tree makes higher accuracy compared with other algorithms. By removing irrelevant, unnecessary, or noisy features, the model's accuracy and efficiency can be increased by selecting the most crucial features to predict hypothyroid disease.

Krishnasamy L, et al.[(2)] combined and tested the accuracy of three congruent ma-chine learning algorithms: Random Forest, Support Vector Machine, and Logistic Regression. The author stated that the random forest accuracy level is greater than that of the other two methods. To investigate thyroid disease, Aversano et al.[(3)] used machine learning techniques. For the categorization of thyroid disorders, popular machine learning methods include Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), Nave Bayes (NB), K-nearest neighbor (KNN), Discriminate Function analysis (DFA), and Multilayer Perceptron (MLP). This study concludes that the Naïve Bayes and MLP algorithms performed well compared to the other algorithms. Haripriya E[(4)] Four machine learning algorithms— decision tree classifier, support vector machine, random forest, and K Neighbour classifier —were tested to detect hypothyroidism based on performance criteria such as accuracy, recall, F1 score, and precision. The authors concluded that the Decision Tree Classifier approach performed better than the other models. To forecast thyroid disease, Almahshi HM, et al.[(5)] designed and analysed classification models that considered many parameters. Many machine learning techniques, such as support vector machines (SVM), naïve Bayes (NB), decision trees (DT), and ensemble approaches, have been utilized to obtain the best prediction.

The authors claimed that the decision tree algorithm ultimately attained the highest level of accuracy. Using a range of machine learning techniques, including decision trees, random forests, naive Bayes, multiclass classifiers, and an artificial neural network (ANN) based on deep learning (DL), Guleria K, et al.[(6)] created prediction models. After analysing performance indicators such as accuracy, precision, recall, and F1-score, it was discovered that decision trees and random forests had the best accuracy.

Chaganti R, et al.[7] study feature engineering models for machine learning and deep learning. The methods included forward feature selection, backward feature removal, bidirectional feature elimination, and machine learning-based feature selection with additional tree classifiers. Numerous experiments have shown that using a random forest classifier with highly randomized tree classifier-based selected features produces the best results. These results suggest that machine learning techniques are a solid option for predicting thyroid disease. Jha R, Bhattacharjee V, Mustafi A[8] applied dimension reduction techniques, data augmentation and classifier prediction to find accuracy in disease prediction. Finally, dimension reduction and data augmentation can be used effectively to achieve high accuracy in hypothyroid disease prediction. Gupta P, Rustam F, et al.[9] examined a differential evolution (DE)-based optimization algorithm with hyperparameter tuning in machine learning models. The authors concluded that an accuracy score was obtained using AdaBoost with DE optimization.

Thabit QQ, Ibrahim A.[10] Based on performance criteria, including accuracy, recall, F1 score, and precision, correlated the efficiency of machine learning algorithms such as Decision Tree Classifier, Support Vector Machine, Random Forest, Naïve Bayes, and Logistic Regression classifiers in predicting the disease. The results demonstrate that the Random Forest and Decision Tree classifier algorithms are more accurate than the other models. Sutradhar A, et al.[11] unveiled two hybrid Machine Learning classifiers particularly, RDKVT and RDKST. The recommended classifiers were validated using two thyroid datasets acquired from two separate repositories. The experimental results showed that the RDKST classifier delivered a more robust accuracy than the RDKVT. Jain A, Yadav S.[12] used several machine learning approaches to forecast thyroid disorders. Among these, K closest neighbor (KNN), Random Forest (RF), and Decision Tree (DT) are frequently utilized owing to their strong performance and interpretability. Random forest has attained the highest accuracy of 98.4%, followed by KNN and DT.

The authors of Kumari P, et al.[13] compared three fundamental classifications: k nearest neighbor (KNN), random forest (RF), and decision tree (DT). With an accuracy of 99.45% following feature selection, they proved that Random Forest (RF) performed better than other classifiers. Decision Tree (DT) and Random Forest (RF) were combined to explain ensemble approaches by Kumar D, Rani A.[14]. This approach improved the prediction accuracy for thyroid disorders. Kumar SP, et al.[15] explained that SMOTEEN is used to balance unbalanced datasets; this method combines the Synthetic Minority Over-Sampling Technique (SMOTE) with edited nearest neighbors (ENN). Medical datasets are frequently unbalanced. Imbalanced datasets can bias the model towards the majority class. The SMOTEEN technique showed significant improvements in recall and F1 scores in the minority class. Khan MA, Hussain A, Majid A.[16] used SMOTEEN in conjunction with ensemble classifiers. These classifiers achieved enhanced performance in the early prediction of thyroid cancer. Peya ZJ, Islam MS, Chumki MKN.[17] used the chi-square test to choose features for predicting thyroid disease. By removing unnecessary or superfluous features, feature selection is essential for enhancing the model performance. A test of statistics for the independence between observed and expected values was evaluated using the chi-square test.This test resulted in a higher accuracy.

Challenges and Motivation

1. For the past three decades, a hypothyroid disorder has been increasing gradually. To avoid severity of this issue, early prediction is significant.
2. Laboratory testing (T3, T4, and TSH) are a major component of traditional diagnostic techniques, which may not detect the thyroid dysfunction at an earlier stage. Besides, the clinical datasets are imbalanced with fewer hypothyroid cases compared with normal or hyperthyroid.
3. The imbalance reduces the effectiveness of standard machine learning models, and it results in low recall when identifying examples of minority classes.
4. Even though there are various machine-learning algorithms that have been available to predict thyroid disease, integration of the SMOTE data-balancing technology has received little attention.
5. Still, medical professionals face challenges like obstructing proper biased learning, noisy and uninformative samples, poor generalization and over fitting.

To address these challenges, SMOTE oversampling technique was integrated with Edited Nearest Neighbor (ENN) such as SMOTEEN was introduced with Chi square feature selection method and hyper parameter tuning using Randomized CV (cross validation) technique to enhance the Recall and accuracy of prediction model. Concentrating Recall is very important in disease prediction because false negative is very dangerous.

2 Methodology

In this study, a detailed explanation of methods and materials was discussed. Here, the working methodology was divided into six parts contains – 1) data collection, 2) data pre-processing, 3) feature selection and data splitting, 4) ML classifiers, 5) hyper parameter tuning and 6) performance Evaluation metrics.

2.1 Data collection

The hypothyroid dataset taken from Kaggle. The dataset contains different data types namely Integer, Float, Character, and Binaries. This dataset contains 3771 records with 30 different features.

Table 1. Features of Dataset

S. No	Column	Description
1	Age	Patient's age
2	Sex	Gender (M = Male, F = Female)
3	on thyroxine	Whether on thyroxine treatment (t/f)
4	query on thyroxine	Whether there is a question regarding thyroxine treatment
5	on antithyroid medication	Whether on antithyroid medication
6	Sick	Whether the patient is ill
7	Pregnant	Whether the patient is pregnant
8	thyroid surgery	Did the patient have thyroid surgery?
9	I131 treatment	Whether radioactive iodine treatment was administered to the patient
10	query hypothyroid	Question if patient has hypothyroidism
11	query hyperthyroid	Question if patient has hyperthyroidism
12	Lithium	Whether patient is on lithium treatment
13	Goitre	Whether patient has goitre
14	Tumor	Whether patient has a tumor
15	Hypopituitary	Whether patient has hypopituitarism
16	Psych	Whether patient has psychiatric disorder
17	TSH measured	If TSH (thyroid-stimulating hormone) level was tested
18	TSH	Tested TSH value
19	T3 measured	If T3 hormone was tested
20	T3	Tested T3 value
21	TT4 measured	If total thyroxine (TT4) was tested
22	TT4	Tested TT4 value
23	T4U measured	If thyroxine uptake (T4U) was tested
24	T4U	Tested T4U value
25	FTI measured	If Free Thyroxine Index was tested
26	FTI	Tested FTI value
27	TBG measured	If thyroxine-binding globulin was tested
28	TBG	Tested TBG value
29	referral source	Where the patient was referred from (e.g., SVHC, other)
30	binaryClass	Target variable: P = Positive (hypothyroid), N = Negative

2.2 Data Pre-processing

Data Pre-processing plays an important role in ML models to effectively transform unprocessed data. Nonnumeric categorical attributes are also an additional barrier; the classifier could struggle to interpret them. Here, label-encoder approaches are used to transform nonnumerical values into numerical values. At the outset the attribute TBG with missing values greater than 50% was eliminated. To balance the unbalanced data, the novel balancing method SMOTEEN was applied.

2.3 Feature selection and Data Splitting

Usually, a dataset has large features yet all features will not participate in prediction. We can decrease feature space while increasing a model's performance by choosing pertinent characteristics from the original dataset. There were more features in a dataset, some of which were noisy and some of which were irrelevant. In this circumstance accuracy will be compromised

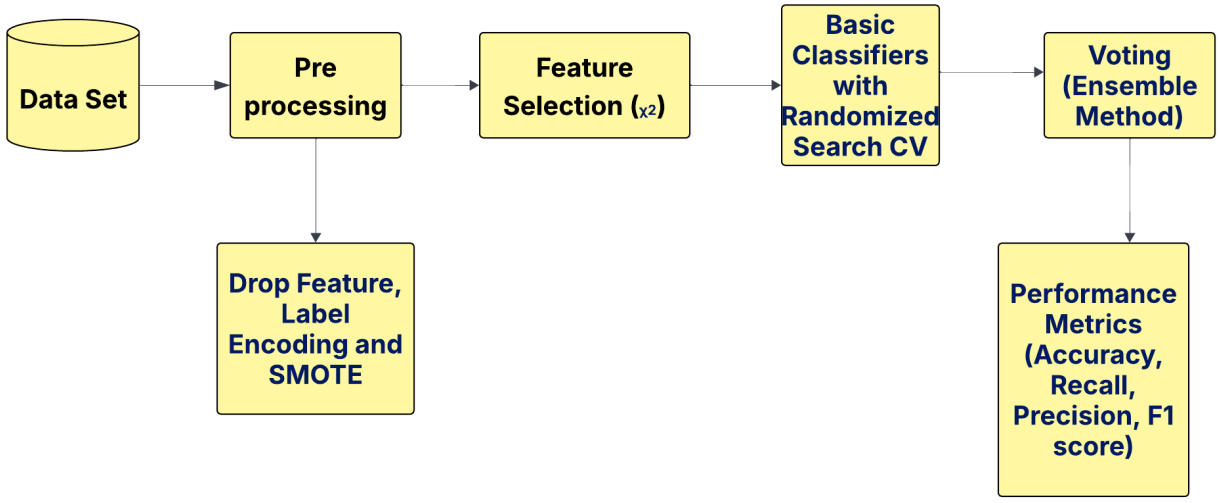

Fig 1. Proposed Workflow of this study

by slowing down the training process. The feature selection strategy was applied to limit overfitting and increase accuracy in order to prevent this kind of issue.

The following is a list of the three primary feature selection algorithms:

- Filter method
- Wrapper methods
- Embedded methods

This study focused on the filter method of feature selection. This strategy examines each and every aspect individually with the target variable. Features with a high relationship with the target variable are selected because this helps to make predictions easier.

Filter method Implementation

2.3.1. Chi-Square Feature Selection

The Chi-Square statistical test was applied to evaluate the dependency between each input feature and the target class. Features with low Chi-Square scores, indicating weak relevance to the class label, were removed, thereby reducing the dimensionality of the dataset. This dimensionality reduction step removed irrelevant and redundant features, reducing computational complexity, minimizing overfitting, and improving model interpretability. The selected feature subset retained only the most informative attributes, leading to more efficient and stable model training. Consequently, the classifiers achieved better generalization on unseen data.

Overall, the inclusion of Chi-Square based dimensionality reduction strengthened the robustness, scalability, and predictive capability of the proposed machine learning framework. When working with categorical input and output variables, Chi-Square feature selection was employed.

Chi-Square feature selection was used as part of the suggested methodology to rank input features according to how well they matched the target class. Only features exhibiting strong statistical relevance were chosen for additional analysis after each feature's Chi-Square score was calculated independently. A smaller and more significant feature set was produced by eliminating less informative features. All of the study's classifiers were trained and evaluated consistently thanks to this dimensionality reduction step, which also improved learning efficiency and reduced noise.

It examines each feature's connection with the target variable using Chi-Square ($\chi 2$) test.

$$\chi^2 \leftarrow \sum(\mathrm{Oi} - \mathrm{Ei})/\mathrm{Ei} \tag{1}$$

- Oi – Observed Value
- Ei – Expected Value

2.3.2. Data Splitting

In order to assess machine learning models that are used to generalize new and unknown data, splitting datasets is crucial. Here, a dataset was partitioned into 80% for training and 20% for testing separately. By splitting 80:20 train–test, the hold-out validation method, which provides an effective balance between model training and evaluation accuracy. This approach is widely recognized in machine learning research and ensures a reliable and unbiased assessment of model performance.

2.4 Basic Classifiers of Machine Learning

While multiple machine learning algorithms such as Decision Tree, Random Forest, and KNN were discussed, the final algorithm selection was performed using a Voting Classifier framework. The ensemble evaluation identified Random Forest as the best-performing classifier, and it was consequently adopted for the proposed study. Subsequently, we present one hybrid classifier BCRVT by combining these basic classifiers with the help of voting ensemble approach.

2.4.1. Decision Tree

In machine learning, the Decision Tree (DT) algorithm is a type of supervised learning. A tree-like structure is used to represent the Decision Tree. Internal nodes are attributes, and each branch describes the result of the test. The leaf node represents the predicted value. The main goal of DT is to shape a training model to predict target variables by understanding decision tree rules. This decision tree algorithm selects the best feature to split the data at each node to minimize the impurity. This process continues until the condition is met.

2.4.2. Random Forest

A supervised machine learning algorithm called Random Forest (RF) can be applied to classification problems. RF is built on using multiple decision trees, is a significant and adaptable method used as ensemble learning to make predictions with increased accuracy

2.4.3. K-Nearest Neighbour

One type of supervised machine learning technique is the K-Nearest Neighbour (KNN) algorithm. It operates by predicting the K-nearest data points to a new data point and making predictions based on their characteristics. Because KNN is a non-parametric method that makes no assumptions about the original data distribution, it can be applied to a number of circumstances.

2.4.4. Proposed Algorithm (BCRVT)

In this study, we performed classification on the hypothyroidism dataset utilizing an inclusive machine learning pipeline to increase prediction accuracy and decrease class imbalance. To evaluate the models' dependability, the dataset was divided into training and testing sets of 80% and 20% proportions.

Three basic classifiers were applied independently:

- Decision Tree (DT)
- Random Forest (RF)
- K-Nearest Neighbour (KNN)

To increase the significance of the input features, we employed feature selection based on Chi-Square test, which assesses the statistical significance of each feature and the target variable. This helps reduce dimensionality and keep the greatest useful characteristics.

A hybrid data balancing method called SMOTE-ENN is used to address imbalanced datasets in classification problems. To enhance minority class representation while eliminating noisy samples, it combines SMOTE (Synthetic Minority Over-sampling Technique) and ENN (Edited Nearest Neighbors).

By interpolating between current minority instances and their closest neighbors, SMOTE creates synthetic samples for the minority class in the first stage. This reduces overfitting by increasing the number of minority class samples without merely duplicating data. ENN cleans the data in the second stage by eliminating samples from both majority and minority classes that have been incorrectly classified by their closest neighbors.

By doing this, noise and overlapping instances close to class boundaries are reduced. Overall, SMOTE-ENN produces a balanced and cleaner dataset, enhances class separability, and improves the performance of machine learning models. To guarantee generalizability and prevent overfitting, Randomized Search cross-validation was implemented. This systematic

method of hyper parameter tuning and model validation helped to enhance performance metrics. The primary objective of this pipeline technique was to maximize performance metrics such as accuracy, precision, recall, and F1-score in the prediction of hypothyroidism.

BCRVT Algorithm:
Input: Dataset DS consisting of M instances, DS = { (Xi, Yi) | i = 1 to M }
Output: Binary prediction indicating whether a person is affected by hypothyroidism

1. **DSa** ← DS.drop{TBG}
2. **DSb** ← LabelEncoder(DSa)
3. **DSc** ← SMOTE-EN(DSb)
4. $\chi^2 \leftarrow \Sigma\ (Oi - Ei)^2\ /\ Ei$
   (Applying Chi-Square Feature Selection Method)
5. **Xi, Yi** ← Input(DSd)
   (DSd is an $N \times M$ matrix)
6. **Xtr, Ytr, Xte, Yte** ← TrainTestSplit(Xi, Yi, 0.2)
7. **While** (executing training datasets) **do**
   7.1 **BC1** ← DT(Xtr_i, Ytr_i, Yte_i)
   7.2 **BC2** ← RF(Xtr_i, Ytr_i, Yte_i)
   7.3 **BC3** ← KNN(Xtr_i, Ytr_i, Yte_i)
   End While
8. **Apply Voting Classifier**
9. **For each classifier** BC ∈ {BC1, BC2, BC3} **do**
   9.1 Optimize BC using **RandomizedSearchCV**
   9.2 $h^* = \text{argmax}\ \Sigma\ (Xi, Yj) \in CV\ M\ (f(Xi; h), Yj)$
   $h \in H$, where h P(H)
   9.3 Save best model performance $\mathbf{M}^*$
   End For
10. **Apply BCRVT** to combine or select the best model
11. **BCRVT** ← argmax { M*(BC1), M*(BC2), M*(BC3) }
12. **Result** ← BCRVT. Predict(New Data)
13. **Print Result**

2.5 Hyper parameter tuning

In the hyperparameter tuning process, machine learning is employed to identify the hyper hyperparameter values. These factors impact the model's learning process and are established earlier in the training phase. A model's recall, accuracy, and ability to generalize to new data can all be significantly impacted by the hyperparameters used. There are several methods available for tuning hyperparameters, such as

- Random Search
- Bayesian Optimization
- Gradient based Optimization
- Grid based CV search

Grid Search CV is also called Grid based cross validation, which is a hyper parameter tuning technique. It determines pre-defined hyper parameter values for a machine learning model. It assesses each combination of parameters through cross validation and identifies the best combination that gives best performance. The drawbacks of Grid based CV search are

- Computationally expensive and time consuming
- Lack of adaptability
- Inefficiency with Large Search Spaces
- Possible for Suboptimal solutions

Here are the performance metrics of Grid based CV:

Table 2. Performance Metrics using Grid based search CV (Existing Method)

Feature sets	Basic Classifiers	Accuracy	Precision	Recall	F1 score
Chi square FS with Grid Based CV HP search	DT	0.9622	0.9292	0.9946	0.9611
	RF	0.9662	0.9292	0.9946	0.9611
	KNN	0.8947	0.8797	0.8776	0.8836
	RDKVT	0.9741	0.9436	0.9986	0.9703

Randomized Search Vs Randomized Search CV

Random Search is widely used technique for hyperparameter tuning. In this case, samples are selected at random from predetermined distributions of hyperparameter values, and they are subsequently assessed. Although random search is a straightforward optimization technique, it has a number of drawbacks. Because it samples points totally at random and may entirely overlook potential areas of the search space, it cannot ensure that the optimal solution will be found. The curse of dimensionality, which decreases the likelihood of finding a favourable combination, makes random search more ineffective as the problem's dimensionality rises.

But Randomized Search CV provides crucial aspects that improve the dependability and efficiency of hyperparameter tuning, we choose it over simple random search. Randomized Search CV avoids overfitting and provides a more accurate assessment of model performance by evaluating each option using cross-validation, whereas random search merely selects random combinations of parameters. All things considered, Randomized Search CV is the superior option in the majority of machine learning workflows since it provides an organized, statistically sound, and effective method of doing random hyper parameter search.

$$h* = argmax \quad \sum(Xi, Yj) \in CVM(f(Xi; h), Yj) \\ h \in H, whereh \sim P(H) \tag{2}$$

- h* - Best hyperparameter found
- argmax - highest total performance across all cross-validation folding
- Σ (Xi, Yj) ∈ CV -summation of cross validation pair
- H - full hyperparameter space
- M(f(Xi; h), Yj) - evaluation metrics of h
- h P(H) – h is a sample distributed over hypothesis space

The table below explains the types of hyperparameters used for different basic classifiers:

Table 3. Hyper parameter used in Basic Classifiers (DT, RF, KNN)

Basic Classifiers	Hyperparameters Utilized
Decision tree (DT)	n_estimators=10, max_depth=16, min_samples_split=50, min_samples_leaf=20, random_state=42, Selection criterion: gini
Random Forest (RF)	'clf__max_depth': 1, 'clf__min_samples_leaf': 20, 'clf__min_samples_split': 50, Selection criterion: gini
K nearest Neighbor (KNN)	'knn__metric': 'manhattan', 'knn__n_neighbors': 3, 'knn__weights': 'distance', Selection criterion: Euclidean

2.6 Performance evaluation Metrics

Performance metrics provide numerical measures of a model's precision, recall, accuracy, and F1-score, making them crucial for assessing how effectively a machine-learning model performs. Prediction detects number of predicted positive cases are actually correct in the model. Recall shows the percentage of real positive cases that the model successfully detects. Accuracy

indicates the number of total predictions made by the model are correct. The F1 score is calculated by averaging the accuracy and memory measurements when their respective contributions are equal.

The Performance Metrics were calculated by using the following formula

- Precision = True Positive/ (True Positive + False Positive)
- Recall = True Positive / (True Positive + False Negative)
- F1 score = 2(Precision* Recall) / (Precision + Recall)
- Accuracy = (True Positive + True Negative) /
- (True Positive + True Negative + False Positive + False Negative)

Table 4. Performance Metrics using Random search (Proposed Method)

Feature sets	Basic Classifiers	Accuracy	Precision	Recall	F1 score
Chi square FS with Randomized HP search	DT	0.9748	0.9985	0.9741	0.9862
	RF	0.9814	1.0000	0.9799	0.9898
	KNN	0.9549	0.9985	0.9526	0.9750
	BCRVT	0.9881	1.0000	0.9871	0.9935

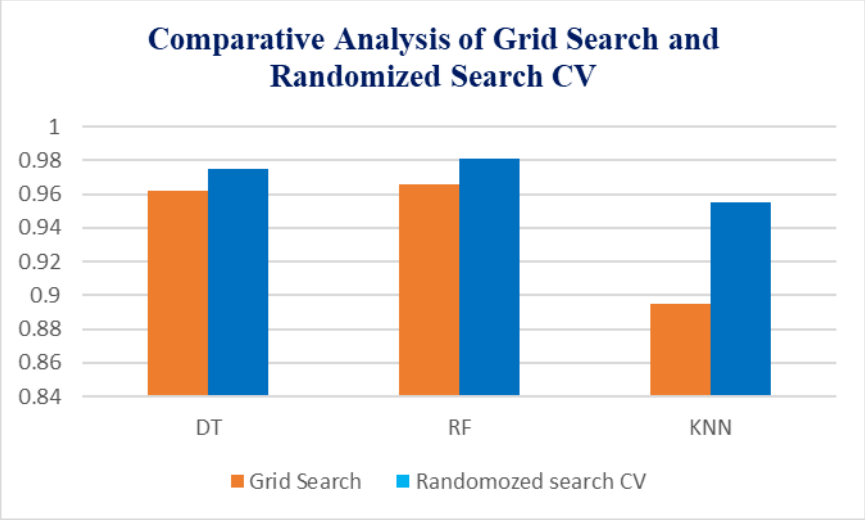

Fig 2. Accuracy of Grid Search Vs Randomized Search CV

3 Results and Discussions

In a recent study, Sutradhar, Akter et al.[11] proposed an interpretable hybrid AI-based thyroid diagnostic system with Grid based search. However, Grid-based searches have several limitations, primarily because of their exhaustive nature. It evaluates all possible combinations of predefined hyperparameter values, which leads to a high computational cost and increased training time, especially for large datasets and complex models. Grid search also explores the hyperparameter space inefficiently by allocating equal effort to all parameters, including those with a small impact on the model performance. Its rigid and discrete search space may miss better parameter values that lie between the specified grid points, and it does not scale well as the number of hyperparameters increases, making it impractical for high-dimensional optimization problems.

Unlike the grid search, which inefficiently spends resources on less influential hyperparameters and is restricted to fixed value sets, the random search explores the hyperparameter space more flexibly and effectively, increasing the chance of identifying near-optimal configurations. It also scales better to high-dimensional and complex models, samples continuous ranges of values, and allows early stopping while still achieving strong generalization performance. Random search Cross Validation overcomes several limitations of grid-based search by significantly reducing the computational cost and time, as it evaluates only a limited number of randomly selected hyperparameter combinations instead of exhaustively testing all possibilities.

Table 5. Performance Analysis of Model A (RDKVT) and Model B (BCRVT)

Metrics	Model A(RDKVT)	Model B(BCRVT)
Accuracy	0.9741	0.9881
Precision	0.9436	0.9990
Recall	0.9986	0.9871
F1-score	0.9703	0.9935

The novelty of the random search CV with SMOTEENN is that it efficiently identifies near-optimal hyperparameters on a noise-free, balanced dataset while reducing the computational cost and improving generalization for imbalanced medical classification problems.

Overall, the results confirm that the proposed combination provides a more reliable and effective diagnostic model for thyroid disorder prediction. The voting method used in the fundamental classifiers (DT, RF, and KNN) for both Grid and Randomized search CV is described in Table 5.

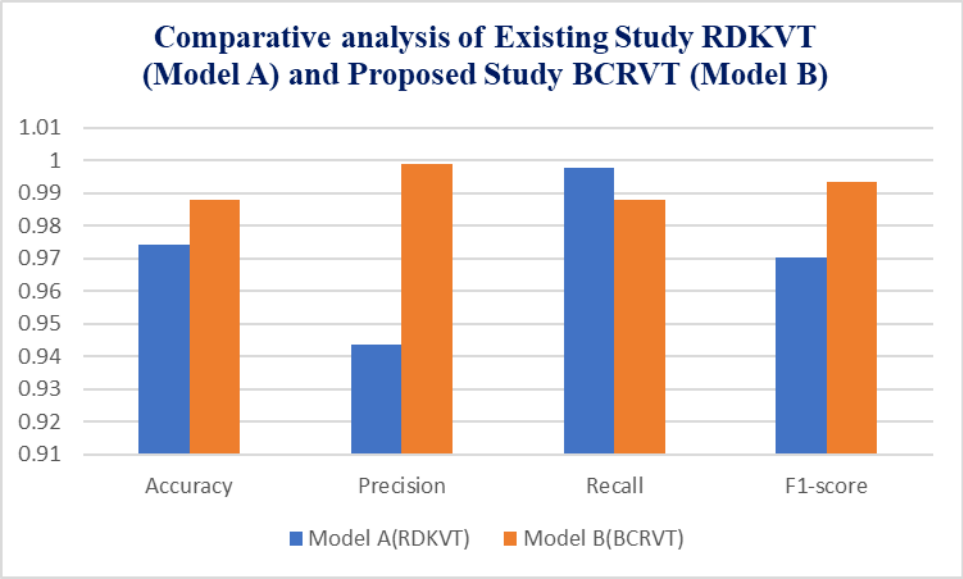

Fig 3. Comparative Analysis of RDKVT and BCRVT

The aforementioned results indicate that Model B (BCRVT) outperforms Model A (RDKVT). While randomized search CV can be utilized for quicker preliminary examination, it is recommended for high-stakes medical applications.

4 Conclusion

A more flexible and resource-efficient optimization approach is represented by the switch to Randomized Search CV from Grid Search. Randomized Search concentrates on cleverly exploring larger parameter spaces in fewer iterations rather than thoroughly investigating parameter combinations, allowing for quicker convergence toward ideal values. Because of this, it is particularly appropriate for models that were trained on complicated or imbalanced medical datasets, where performance stability and computational efficiency are crucial. Randomized Search CV is a practical option for contemporary machine learning pipelines since it prioritizes flexibility and scalability, supporting more reliable tuning results while minimizing needless processing overhead. Future research can incorporate **cost-sensitive classification techniques** in hypothyroid prediction to account for unequal misclassification costs, particularly by reducing false negatives where undetected hypothyroid cases can lead to serious health complications. Integrating cost-sensitive learning will enhance clinical reliability, improve patient safety, and make predictive models more effective for real-world diagnostic use.

5 Acknowledgements

The authors thank, DST-FIST, Government of India for funding towards infrastructure facilities at St. Joseph's College (Autonomous), Tiruchirappalli – 62002.

References

1) Tirumala AK, et al. Predicting thyroid dysfunction using machine learning techniques. IEEE. 2023. 10.1109/ICoAC59537.2023.10249516.
2) Krishnasamy L, et al. Predicting the thyroid disease using machine learning techniques. Springer. 2023. 10.1007/978-981-99-3932-9_2.
3) Aversano L, et al. Thyroid disease treatment prediction with machine learning approaches. *Procedia Computer Science*. 2021. 10.1016/j.procs.2021.08.106.
4) Haripriya E. Performance analysis of machine learning algorithms in thyroid disease prediction. Springer. 2023. Available from: https://link.springer.com/chapter/10.1007/978-3-031-64813-7_42.
5) Almahshi HM, et al. Hypothyroidism prediction and detection using machine learning. IEEE. 2022. 10.1109/IICETA54559.2022.9888736.
6) Guleria K, et al. Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning. *Measurement: Sensors*. 2022;24:100482. 10.1016/j.measen.2022.100482.
7) Chaganti R, et al. Thyroid disease prediction using selective features and machine learning techniques. *Cancers*. 2022;14(16):3914. 10.3390/cancers14163914.
8) Jha R, Bhattacharjee V, Mustafi A. Increasing the prediction accuracy for thyroid disease: a step towards better health for society. *Wireless Personal Communications*. 2022;122(2):1921–1938. 10.1007/s11277-021-08974-3.
9) Gupta P, Rustam F, et al. Detecting thyroid disease using optimized machine learning model based on differential evolution. *International Journal of Computational Intelligence Systems*. 2024. 10.1007/s44196-023-00388-2.
10) Thabit QQ, Ibrahim A. Thyroid disease prediction with machine learning algorithms. *ResearchGate*. 2023. Available from: https://geniusjournals.org/index.php/erb/article/view/3777.
11) Sutradhar A, et al. Advanced thyroid care: an accurate trustworthy diagnostics system with interpretable AI and hybrid machine learning techniques. *Heliyon*. 2024;10. 10.1016/j.heliyon.2024.e36556.
12) Jain A, Yadav S. Machine learning based predictive modelling for classification of thyroid disease. *Materials Today: Proceedings*. 2021;47:2306–2311. 10.1016/j.matpr.2021.05.343.
13) Kumari P, et al. Explainable artificial intelligence and machine learning algorithms. *Discover Applied Sciences*. 2024. 10.1007/s42452-024-06068-w.
14) Kumar D, Rani A. Thyroid disease prediction using hybrid SMOTE-ENN and ensemble learning techniques. *Journal of King Saud University–Computer and Information Sciences*. 2022;34(8):4946–4955. 10.1016/j.jksuci.2022.01.004.
15) Kumar SP, et al. Thyroid disease prediction using machine learning techniques. IEEE. 2024. 10.1109/ICESIC61777.2024.10846626.
16) Khan MA, Hussain A, Majid A. Early thyroid risk prediction by data mining and ensemble classifiers. *Information*. 2021;5(3):61–70. 10.3390/info5030061.
17) Peya ZJ, Islam MS, Chumki MKN. Thyroid disease prediction based on feature selection and machine learning. IEEE. 2022. 10.1109/ICCIT57492.2022.10054746.