International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 670-691
https://doi.org/10.70917/ijcisim-2026-2542
Article

# DIAGNOSIS OF THYROID CANCER USING BOBCAT OPTIMIZATION ALGORITHM (BOA) AND ENHANCED TWO STAGE ATTENTION WITH LONG SHORT-TERM MEMORY (ETSA-LSTM)

**Sruthi. V.S[1], Dr.Kokilamani,M[2]**

[1]Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, svspulappatta@gmail.com
[2]Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Kokilamanikcas1@gmail.com

**Abstract:** The malignant (cancerous) cells develop in the thyroid gland's tissues, is known as thyroid cancer (TC). For these conditions to be effectively treated and for patient care to be provided, an accurate and quick diagnosis is essential. Thyroidectomy is still the principal therapeutic option, despite significant efforts to improve diagnosis. A precise preoperative diagnosis may not always be ensured by the current human evaluation of Thyroid Nodule (TN) malignancy, because this TN maligancy is prone to errors. Medical (DA) Data Analysis (MDA) problems can be easily solved with the use of Data Mining (DM) algorithms. With several techniques for classification, clustering, association, etc., DM offers significant assistance with thyroid datasets. Deep Learning (DL) techniques to predict and identify TC, their development and application present a number of difficulties. The suggested work consists of five primary steps: Data Collection (DC), Data Pre-processing, Feature Selection (FS), Data Classification, and Performance Evaluation. Initially, the Kaggle online repository is used to gather the DC, TC risk prediction dataset. This dataset mimics real-world TC risk factors and includes 212,691 records with 23 features. Then, issues with Data Encoding (DE), Data Resampling (DR), Data Normalization (DN), Data Imputation (DI) and handling missing data are addressed by using data pre-processing techniques. DN is carried out using Min-Max Normalisation (MMN) or Min-Max Scaling (MMS). To extract pertinent feature information, the Z-score- (ZS) based Boruta-SHapley Additive exPlanations (BorutaSHAP) attribute importance technique, is used by the Feature importance (FI). SMOTE-EkNN is a data imbalance problem-solving technique that uses Edited k-Nearest Neighbors (EkNN) and Synthetic Minority Over-sampling Technique (SMOTE) to remove noise. Thirdly Feature Selection (FS), Bobcat Optimization Algorithm (BOA) is used to choose appropriate features. For TC detection, the most pertinent features (an optimal reduced feature subset) are chosen using BOA. The Search Space (SS) can be effectively explored and exploited by BOA, a Nature-Inspired Algorithm (NIA). Next, the data is classified using Enhanced Two Stage Attention with Long Short-Term Memory (ETSA-LSTM) for predicting the TN malignancy. Accuracy (ACC), Specificity (SP), Precision (PR), Recall (R) or Sensitivity (SE), and the Area Under Receiver Operating Characteristic (AUROC) curve are the final metrics used to evaluate efficiency. According to the study of the results, the suggested model outperformed the other methods currently in use in terms of ACC .

**Keywords:** Thyroid Cancer (TC), Deep Learning (DL), Thyroid Nodule (TN), Data Mining (DM), Bobcat Optimization Algorithm (BOA), Synthetic Minority Over-sampling Technique (SMOTE), Edited k-Nearest Neighbors (EkNN), Enhanced Two Stage Attention with Long Short-Term Memory (ETSA-LSTM), Area Under Receiver Operating Characteristic (AUROC), Feature Selection (FS), and Classification.

# 1. INTRODUCTION

A large percentage of people worldwide are suffering from thyroid disease (TD). This TD is a common medical problem. Over the past three decades, TC has become more prevalent. One of the most common endocrine system malignant tumours is TC. Around 1% to 3% of all newly identified malignant tumours globally are TC [1]. According to the NIH's Surveillance, Epidemiology, and End Results Program (SEER), the prevalence of TC increased by 5.5% per year between 2005 and 2015 [2]. An estimated 42 million people in India alone are thought to be afflicted with thyroid-related conditions such TC, thyroid goitre, and iodine deficiency [3]. The prevalence of TC is still increasing in the USA. In 2022, about 2,230 TC-related deaths and 43,800 new cases are anticipated, according to the most recent estimation from the American Cancer Society [4]. Tumor is usually affected by variations or mutations in genes that disrupt the genes that control the function of cells. As a result, unchecked cell division and tissue invasion occur.

The medical community urgently requires some advanced diagnostic methods to identify and prevent the beginning of these illnesses, because TD occurs so frequently. Total blood thyroxine (T4) and total serum triiodothyronine (T3) are two significant hormones produced by the thyroid gland, an endocrine organ presents in vertebrates. Body temperature, heart rate, and the development of bones are all impacted by these hormones. These factors are essential for controlling the body's metabolism. TD may result from thyroid gland dysfunction. Euthyroidism, hyperthyroidism, and hypothyroidism are the three types of thyroid-related disorders that are categorised according to hormone levels.

Normal thyroid gland function is referred to as euthyroidism. Overactivity of the thyroid gland results in hyperthyroidism. Then, hormone production is increased due to hyperthyroidism. On the other side, an underactive thyroid gland that does not generate enough hormones is the characteristic of hypothyroidism. For patients, early detection (ED) and diagnosis of TD is crucial since it can help them get treatment quickly and save money on healthcare costs [5].

Effective treatment and reduced damage can be achieved by treating ED of the malignant TN before the cancerous cells in the thyroid gland spread [6]. Two key methods for detecting TC: (1) Neck palpation as part of a physical examination. (2) Ultrasonography (US) can detect both palpable nodules (PN) and non-PN, especially those with a diameter of less than 1 cm. Papillary TC (PTC), Follicular TC (FTC), Medullary TC (MTC), and Anaplastic TC (ATC) are the four histological types that are frequently included in TC. The primary treatments for Differentiated TC (DTCs) include radioactive iodine remnant ablation, and surgery particularly PTC.

Differentiating among benign and malignant TN is one of the surgical operation's main goals. In order to prolong the survival duration, it also aids in the selection of thorough post-surgery treatment [7]. The following stages are typically included in the clinical protocol for TC. (a) TN presurgical diagnosis. (b) The primary tumour mass is surgically removed. (c) The staging system based on the T (tumor), N (nodule), and M (metastasis) systems, along with other prognostic classifications. (d) Hormonal therapy or suppression of thyroid stimulating hormone (TSH). (e) In order to prevent or treat distant metastases and local/regional recurrences, radioactive iodine is administered after surgery. Chemotherapy, radiotherapy, immunotherapy, or a combination of these may be used less frequently. (f) monitoring by finding tumour indicators (such as thyroglobulin and antithyroglobulin antibodies), whole-body scintigraphy, and neck US [8].

As a result, using thyroid US, blood tests, and other fundamental medical data to create precise treatment and prognosis is essential. Human judgement is frequently laborious and prone to mistakes. To help with medical decisions and cut down on labour, researchers must find Prediction Models (PM) that is accurate and comprehensible. The process of extracting information from difficult data using innovative approaches is known as the DM methodology. In order to build and implement a Decision-Supporting System (DSS), all useful variables are realised through the application of computer skills and DM methodologies.

Artificial Intelligence (AI) has significantly improved in recent years due to algorithmic advancements in Machine Learning (ML). One class of algorithms used in ML is called an Artificial (NN) Neural Networks (ANNs). ANN and Decision Trees (DT) have been used for nearly 20 years to diagnose and identify TC [9]. There are a number of clinical factors that may influence the prognosis of TC. These consist of the patient's age, general health, the kind and location of the disease, and the grade of the malignancy.

Furthermore, in terms of Pattern Recognition (PR) and classification, ANNs perform better than classical statistical modelling. ANN models are built in layers to learn progressively Higher-Dimension (HD) and distant representations of the input data. The uncertainty in the PM was not taken into account by other ML-based models,

which merely offered point estimation of the model performance. ML can sometimes struggle to generalize well to unseen data, meaning they may not perform as well on new data as they do on the training data. DL methods for TC detection and prediction have a lot of promise for clinical use; their development and use are filled with difficulties.

In order to discover critical features, remove irrelevant, redundant, or noisy features, and minimise the size of the feature space, data pre-processing is necessary for DL approaches. Their main features are not entirely independent of one another, but rather have intricate relationships with one another. Optimization is the method of choosing the best response from a variety of possible solutions because of the problems with these approaches. ETSA-LSTM and FS based on BOA were introduced in this work for TC classification.

In this paper, suggested an ETSA-LSTM DL framework to predict nodule malignancy accurately. Firstly, data collection, TC risk prediction dataset is collected from Kaggle online repository. This dataset simulates real-world TC risk factors and includes 212,691 records with 23 features. Secondly data pre-processing, MMN (or) Min-max Scaling algorithm for normalization. FI performed ZS-based BorutaSHAP FI technique to extract relevant feature information. Noise removal using EkNN. SMOTE for addressing data imbalance problems. SMOTE-EkNN to deal with Missing Values (MV). Thirdly FS; suitable features are selected by using BOA. The most pertinent features, an ideal reduced feature subset, were chosen using BOA. Then data classification, ETSA-LSTM to predict the thyroid nodule malignancy to achieve both better estimation and clearer interpretation. ACC, R or SE, PR, AUROC, and SP are the final metrics used to evaluate performance. The publicly available benchmark data from the Kaggle online repository have been taken into consideration to evaluate and contrast the appropriateness of the recommended method with the other approaches that are already in use.

The following is how the paper processes: An overview of relevant research on TC prediction using several classifiers and hybrid techniques is given in Section 2. Section 3 provides an explanation of the DC procedure, data pre-processing, FS, and data classification techniques. The system's implementation and outcomes are detailed in Section 4. Section 5 concludes with a brief discussion and some suggestions for further research.

## 2. LITERATURE REVIEW

This section lists and analyses the various TC detection models that use ML and DL models. According to the study, a novel framework is created to address the core problems in TC detection and classification.

For TC detection, Ahmed et al. [10] suggested a Deep NN (DNN). The two primary forms of TD are as follows. Slower metabolism is a symptom of hypothyroidism. Possible symptoms include weight gain, fatigue, anxiety, difficulty concentrating, memory problems, sluggish speech, and trouble with movements and ideas. The body's metabolism, gets accelerated by hyperthyroidism. An increased heart beat (HB), weight loss, diarrhoea, thirst, shaking and sweating, feeling too hot, and other symptoms are caused by this. Numerous methods, such as ML and DL, can be used to tackle this. The suggetsed work introduced a DNN for TC detection. The dataset collected from online kaggle repository. Compared to other (SOTA) state-of-the-art methods, the suggested DNN's accuracy (ACC) is improved.

In order to forecast TD, Random Forest (RF), DT, KNN, Gradient Boosting (GB), and Multilayer Perceptron (MLP) were developed by Uddin et al. [11]. The dataset's data was first pre-processed to handle MV, data encoding, FS, resampling, and normalisation. Vital attributes from the dataset are then found using FS approaches like XGBoost and SelectKbest. Random Oversampling (ROS) is used to optimise the dataset. Several ML algorithms are then employed to train the models using the training data. To select the best thyroid prediction result, five ML models: RF, DT, KNN, GB, and MLP were used in addition to the Ensemble ML (EML) classifier (hard voting). Performance metrics like F-Measure (or F1-score), ACC, R (or SE), SP, PR, and AUC are used to determine which of these five classifiers is the best ML model. The University of California Irvine (UCI) ML repository provided the dataset. It includes 6 databases with 2800 training and 972 test cases from the Garavan Institute in Sydney, Australia. When employing the XGBoost and SelectKBest FS techniques, the EML classifier performs best for hard voting on RF and DT with higher SE and ACC.

Alshayeji [12] introduced a Synthetic Minority Oversampling Technique (SMOTE), Logistic Regression (LR), NN, DT, KNN, Support Vector Machine (SVM), Bagging, Boosting for early thyroid risk prediction. Researchers were able to replace the manual examination of these parameters with an ML framework that may forecast TD by analysing premature signs by the clinical features. Understanding the function of attributes in task of predicting thyroid is made easier by feature analysis and visualisation. Additionally, 5-fold (CV) cross-validation and data balancing with the SMOTE were used to solve the overfitting problem. LR, NN, DT, KNN, SVM, bagging, and boosting are examples

of ensemble learning (EL) algorithms. Due to the use of many classifiers in the forecast choices, these EL models provide PM reliability. In order to enhance early treatment and work with real-time (RT) Computer-Aided Diagnosis (CAD) systems for ED, the suggested model used the boosting method to obtain higher ACC, SE, and SP. From the UCI ML repository, the dataset is offered. It includes 6 databases with 2800 training and 972 test cases from the Garavan Institute in Sydney, Australia. True (P) Positive (TP) Rate (TPR), True (N) Negative (TN) Rate (TNR), P (PV) Predictive Value (PPV), Negative PV (NPV), ACC, and Misclassification Rate are the performance metrics.

In order to reduce the risk of recurrence and address DTC recurrence, Sibarani and Suharjito [13] developed LR, SVM, kNN, ANN, AdaBoost, RF, Bagging, Stochastic Gradient Descent (SGD), Boosting, k-means, hierarchical, and Louvain Clustering. 383 patients in total with 17 attributes were monitored to create the dataset. To identify the most important factors, an exploratory DA was carried out. Following that, a range of supervised algorithms were used to evaluate the classification accuracy of both single and ensemble models (EM), including LR, SVM, kNN, ANN, AdaBoost, RF, Bagging, SGD, and Boosting. Several unsupervised learning techniques, such as Louvain Clustering, hierarchical, and k-means, were used for cluster determination. With remarkable scores of 0.971, the results indicate that the ensemble stacking algorithm performed better and achieved a higher classification ACC. The dataset may be divided into 2 distinct clusters, according to the study of clustering techniques, particularly k-means and hierarchical clustering. The factors "Response," "Risk," "Adenopathy," and "N" were shown to be the most significant in affecting the recurrence of TC with a high correlation. By identifying vital components and dependable models, the diagnostic model is refined to improve the prediction of DTC recurrence. An excellent resource for examining the recurrence of DTC Recurrence is the dataset obtained from the UCI ML Repository.

LR, K-Neighbors (KN), Support Vector Classifier (SVC), Gaussian Naive Bayes (GNB), DT, RF, Ada Boost classifier (AdaB), and GB were all used by Alghamdi [14] to predict TC. The Prostate, Lung, Colorectal, and Ovarian (PLCO) dataset, which includes over 155,000 patients with TC occurrence and death incidence, can be subjected to various ML techniques. The use of ML models to forecast the death rates of TC patients is investigated. The findings show that out of all the models, the AdaB and GB classifiers perform the best. ACC, PR, R, F-Score, and AUC are some of the metrics that indicate how different PMs perform in comparison to one another. This study demonstrates how classifier performance is influenced by the parameters selected. Therefore, prior to final implementation, it is crucial to investigate and assess them.

Anuhya et al., [15] adopted a SVM, KNN, DT, and RF approaches for the classification of TC using a properly chosen dataset. Important methodologies like feature engineering, training models, data preparation, and detailed assessment are all covered in the research. It offers insights on the merits and downsides of every technique. Through major discussions and detailed evaluations, brings new information to the area of ML-driven TC detection with the objective of increasing clinical decision support systems moving forward and boosting patient outcomes.

Vanitha and Perumal [16] presented a Spatial Convolution Based LSTM Network (SCBLN) and Principal Component Analysis (PCA) for prediction of the TC. The SCBLN method can be used to forecast the following conditions: binding protein (increased binding protein), autoimmune thyroiditis (compensated hypothyroidism), Hashimoto's thyroiditis (primary hypothyroidism), and non-thyroidal syndrome (NTIS) (concurrent non-thyroidal sickness). The normalisation approach can be used to process the initially obtained dataset. The features linked to cancer can then be extracted using the PCA approach. At last, the SCBLN classifier can be used to predict TC. In a Python environment, the entire experiment was conducted. The SCBLN classifier has the greatest ACC and F1 score, based on numerous tests. The results show that, in terms of both ACC and computational complexity, the suggested framework is a better choice for TC detection. The superiority of the SCBLN technique is confirmed by comparing its performance to previous investigations.

For TC therapy, Xi et al. [17] proposed the Gradient Boosting Machine (GBM), LR, Linear Discriminant Analysis (LDA), SVM, and RF. Initially, clinical data from medical records was gathered and preprocessed for 724 patients and With a ten-fold CV, the PM performance was assessed using five metrics: ACC, AUROC, SE, SP, and P. Next, bootstrap analysis (BA) and permutation predictor importance (PPI) were used to further analyse the model uncertainty and key factors. At last, the PM and expert evaluation were contrasted. The suggested approach can be used as supplementary evidence in the preoperative diagnosis of TC because it is precise and interpretable. Based on the gathered new clinical data, this study suggested a ML architecture to forecast TN malignancy. The model performance under uncertainty was estimated and interpreted using the ten-fold CV, BA, and PPI.

Based on the analysis, ML algorithms drawbacks are ANN-high computational costs, SVM-increased computational complexity, LR -lack of interpretability, and security vulnerabilities. A DL-based TC classification model is required to improve the efficiency and AC of TC in order to address the aforementioned problems. In this

study, an ETSA-LSTM based on cancer cell detection and classification is developed to identify TC and classify the disease's severity. Addressing the difficulties of precisely detecting and diagnosing TC early on and enhancing the AC are the main objectives.

## 3. PROPOSED METHODOLOGY

For data classification, the suggested work presented an FS based on BOA and ETSA-LSTM. DC, Data Pre-processing, FS, Data Classification, and Performance Evaluation are the five main steps of the suggested study. Firstly, dataset collection, TC risk prediction dataset is collected from Kaggle online repository. Secondly, data pre-processing methods are normalization is performed by using MMN (or) Min-max Scaling. By utilizing BST, an effective data balancing method, for resolving the data imbalance issue. Feature importance performed Z-score-based BorutaSHAP FI technique for extracting pertinent feature data. Noise removal using EkNN and SMOTE for addressing data imbalance problems referred as SMOTE-EkNN. Then, to identify the TC, the most relevant features (an ideal reduced feature subset) were chosen using FS and BOA. Next data classification, ETSA-LSTM to predict the TN malignancy to achieve both better estimation and clearer interpretation. At last, the PR, R or SE, SP, ACC, and AUROC are used to evaluate performance. The applicability of the suggested ETSA-LSTM technique has been evaluated and compared with publicly available benchmark data from the Kaggle repository.

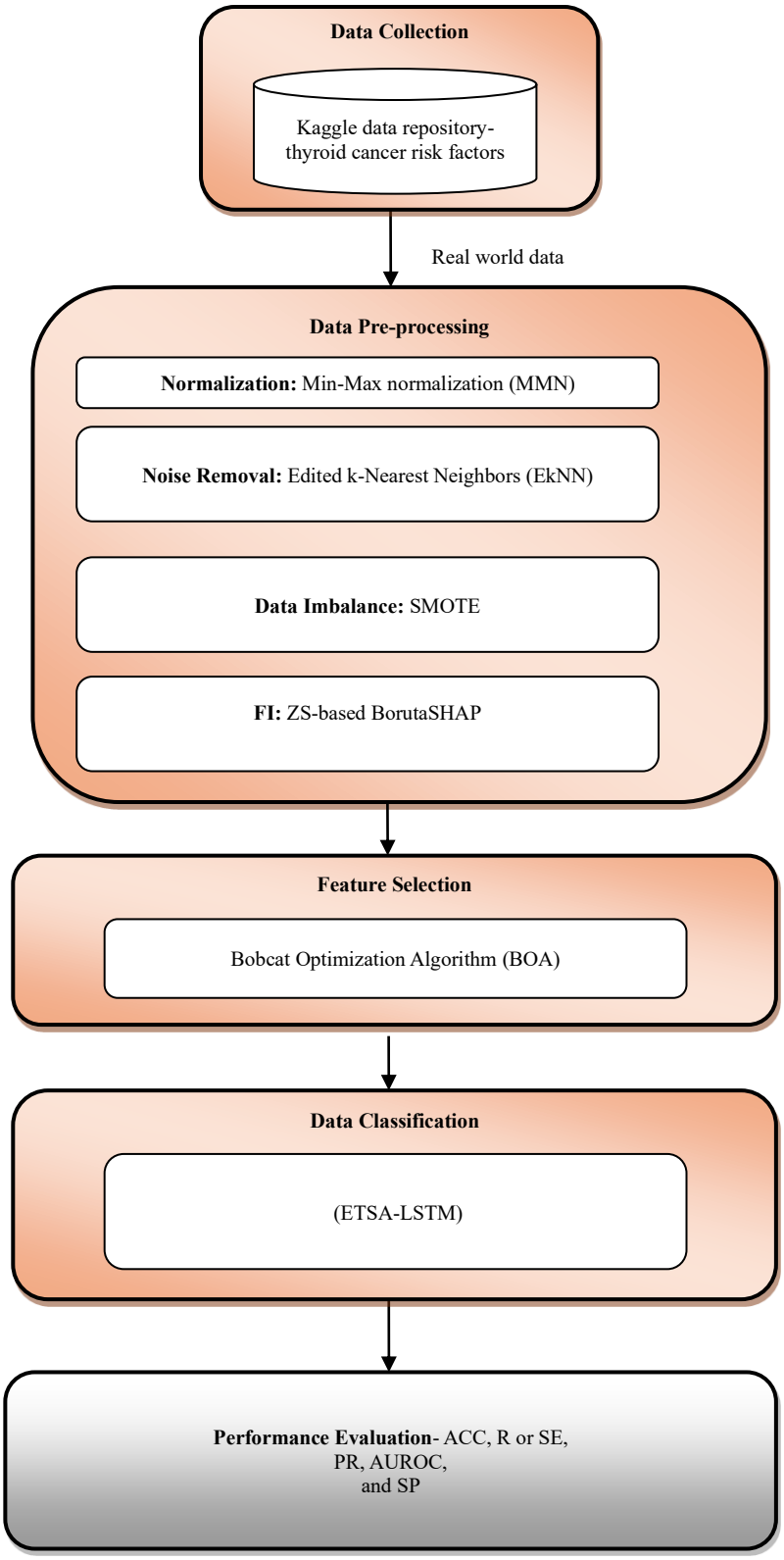


**FIGURE 1. FLOW CHART OF SUGGESTED SCHEME**

## 3.1.*DC*

TC risk prediction dataset is collected from Kaggle online repository [18]. This dataset simulates real-world TC risk factors and includes 212,691 records with 23 features. The collected dataset taken as input data pre-processing to solve missing values, noise, feature importance and class imbalance problems is described in section 3.2.

## 3.2.*DATA PRE-PROCESSING*

For retrieving valued insights from the original dataset, data pre-processing is a vital stage in the training of any DL model. Essentially, these processing techniques serve to convert the original data into a comprehensible and readable format. The gathered dataset had issues with noise removal, MV, data imbalance, so data pre-processing procedures are necessary. As a result, cleaning it up is necessary before using the DL model. To deal with MV and noise removal, some common techniques can be employed, such SMOTE-EkNN. Data must also be standardised or normalised before DL techniques can be used. Relevant feature information was extracted using the ZS-based BorutaSHAP FI approach for FI. The normalisation is done using MMN (or MMS).

### 3.2.1. **Normalization**

Regarding predictive methods, normalisation is the procedure of rescaling the features of numerical data to a predetermined range. A data preprocessing method called normalisation is used to align the values of several attributes on a common scale. By doing this, it is ensured that all features contribute equally and that features with wider ranges do not significantly affect the analysis or the model's learning process [19]. Several methods are used for normalisation, such as MMS, ZS normalisation (ZSN), and resilient scaling. When the risk factors in a dataset have different values, it's referred to as DN. MMN is used to normalise the dataset.

**MMS (or) MMN**

To alter numerical values within a specific range, often from 0 to 1, a data pre-processing technique known as Min-Max scaling, or MMN. MMN rescales each feature to a specific range using  and  value [20]. The MMN formula is represented by equation (1).

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

The scaled value is denoted as .  represents the attribute's initial value. denotes the attribute's  value in the dataset. he feature's  value in the dataset is denoted as .

### 3.2.2. **Noise Removal and Data Imbalance (DIMB)**

A number of algorithms employing various techniques have been developed in order to eliminate the noise by DA. DM uses methods such as clustering, description, estimate, prediction, association rules, and classification. When there are substantially differing numbers of examples for various classes in a dataset, named as DIMB.  Noise removal and DIMB issue using an effective method SMOTE-EkNN.

**SMOTE-EkNN**

To balance the dataset, the data balancing (DB) strategies employed a hybrid DB methodology that combined a clustering strategy with an oversampling (OS) and undersampling (UndS) method. Based on the data characteristics identified by K-means (KM) clustering (KMC), the dataset was split up into multiple clusters. Researchers employed the elbow approach to determine the ideal number of clusters. A range of potential values for k is specified in the KMC procedure. The Sum of Squared Distances (SSD) between each data point (DP) and its assigned cluster centroid was calculated for each k in this range. Using k values on the x-axis and their corresponding SSD values on the y-axis, a graphic representation was created.

Researchers eventually determined the plot's "elbow point "(EP). It indicates the point at which there is a discernible slowdown in the SSD. For this EP, the ideal number of clusters was indicated. The elbow approach was then used to separate the training data into five clusters. SMOTE, an OS method, was used in each cluster. Some noises were added after SMOTE was applied. SMOTE-EkNN is the name given to the first approach, which eliminated the noise by using the EkNN UndS technique. Create the final balanced training dataset by merging all of the cluster data after applying SMOTE-EkNN to the training data in each cluster [21]. Preserved outgoing data was identical to the original data during this operation. Algorithm 1 displays the SMOTE-EkNN algorithm, the suggested DB approach.

**Procedure 1: KM SMOTE-EkNN DB**

1: **Input:** $X$ (dataset), $k_{max}$ (max count of clusters), $k_{neighbors1}$(count of SMOTE neighbors), $K_{neighbors2}$ (count Of ENN Neighbors)

2: **Output:** $X_{balanced}$ (balanced dataset)

3: *distortions*← ∅

4: **for** $k = 1$ to $k_{max}$

5: Cluster the DP in $X$ into $k$ clusters using KM algorithm.. Use the K-Means technique to cluster the DP in $X$ into $k$ clusters.

6: *distortions*[$k$] ← SSD of DP to their nearest cluster centroid

7: **Endfor**

8: Plot the corresponding SSD on the y-axis of a graph with $k$ values on the x-axis.

9: Determine the ideal count of clusters by identifying the EP with the largest SSD loss.

10: Using the elbow approach that matches the EP, get the ideal $k$.

11: To obtain the final clustering results, use the KM approach to cluster the DP in $X$ into the ideal $k$ clusters.

12: **for** $i = 1$ to $k$

13: $X_i$ ←DP in cluster $i$

14: Compute the centroid $C_i$ of cluster $X_i$

15: **Endfor**

16: $X_{smote} \leftarrow \emptyset$

17: **for** $i = 1$ to $k$

18: **for** DP $x$ in $X_i$

19: Find the $k_{neighbors1}$ nearest neighbors of $x$ in $X_i$

20: **for** each nearest neighbor $n$

21: Use SMOTE among $x$ and $n$ to create a synthetic example $s$

22: $X_{smote} \leftarrow X_{smote} \cup \{s\}$

23: **Endfor**

24: **Endfor**

25: **Endfor**

26: $X_{smotebalanced} \leftarrow X \cup X_{smote}$

27: Remove noisy samples from $X_{smotebalanced}$ by using ENN (Edited Nearest Neighbours).

28: Determine the nearest $k_{neighbors2}$ neighbors for each DP in $X_{smotebalanced}$

29: Eliminate DP whose class label differs from the majority class of its $k_{neighbors2}$ neighbors

30: **return** $X_{balanced}$

### 3.2.3. **FI**

Methods that assess the input features according to their capacity to forecast output variables are referred to as FI. The insights into the model attributes are reflected in the FI scores in PM. For HD data, it forms the foundation for attribute selection (AS) and dimensionality reduction (DR), both of which could affect PM efficiency. The most widely utilised techniques include F-score-based FI, LRmodel coefficients, RF and DT-focused features scores, statistical correlation scores, and permutation-based scores. For XGB-based FI, the Select From Model (SFM) class transforms a model into a subset with predetermined features [22]. Boruta AS and SHAP values are used by wrapper-based BorutaSHAP. In terms of time and feature subset reliability, this combination performs better than the traditional permutation importance technique. By offering superior attributes and the most reliable and accurate global feature rankings, this approach enhances model inference.

## 3.3.*FEATURE SELECTION*

The data pre-processing is completed TC dataset taken as input for BOA based FS. Typically, FS is used for detection, which increases accuracy of model. For the best selection of features from the TC dataset, the BOA algorithm is employed, a subset of the original features that are most pertinent to the issue at hand are chosen. It entails picking particular dataset features to be used in a learning algorithm's training procedure. BOA based FS technique are employed to measure the effectiveness of features in classifying data.

**BOA**

For detection purpose, the FS is utilized. The AC of the framework is upgraded by the FS. The application simulates the hunting and behaviour methods of bobcats (BC) and is called a nature-inspired optimisation algorithm (NIOA). The search space (SS) is intended to be efficiently explored and exploited by the algorithm. Following a statement of the design motivation for the BOA technique, it is statistically modelled for use in optimisation tasks.

***Inspiration of BOA***

From southern Canada through the majority of the contiguous United States [US] to Oaxaca in Mexico, the BC is a native of North America. As a versatile predator, the BC inhabits semidesert, urban, forest, and swampland habitats. Season, abundance, and habitat all affect prey selection (PS). Tracking the target is the initial step in the BC's hunting method. After a chase procedure, it attacks the target at the ideal moment and pursues it. The BC's hunting technique has been far the most noticeable of its natural behaviours in the wild [23]. (i) tracking and approaching the prey, and (ii) pursuing and capturing the prey are the two activities that make up this hunting method. The BOA technique is designed using mathematical modelling of these intelligent processes, that will be described as follows.

***Mathematical Model of BOA***

The BOA technique draws inspiration from the hunting tactics of BC in the wild to update the population of the algorithm in the problem solving space (PSS). Tracking and approaching the prey and pursuing and capturing the prey are the two components that comprise the BC's location in its habitat during the hunting phase. In each iteration, the population members (PopMem)' positions are modified in two phases, drawing inspiration from this natural strategy in BC lifestyle. These are (i) exploration, which is based on simulating the BC's position change as it approaches the prey, and (ii) exploitation, which is based on simulating the BC's position change as it pursues and captures the prey. Each of these BOA update steps is explained as follows.

***Initialization***

The BOA is a population-based optimisation technique that finds appropriate solutions through an iterative procedure. In order to investigate the PSS and find the best feature subsets from the TC dataset, BOA makes use of the collective intellect of its "BC" members, each BC stands for a potential solution. The BC's wildlife habitat matches the PSS, and the BC's placement inside this habitat matches the BOA members' position within the PSS, according to the design inspiration of the BOA.

Since a vector may be used to model the problem mathematically, the position of each BC reflects a candidate solution (CS) to the problem. Here, a decision variable (DV) is represented by each member in this vector. Together, BC make up the algorithm's population, which Equation (1) allows to be mathematically represented using a matrix. Equation (2) is used to randomly initialise BC's primary location in the PSS.

$$x = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times M} = \begin{bmatrix} x_{1,1} & \dots & x_{1,d} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,d} & \dots & x_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,d} & \dots & x_{N,m} \end{bmatrix}_{N\times m} \quad (1)$$

$$x_{i,d} = lb_d + r.(ub_d - lb_d) \quad (2)$$

The BOA (Pop Mx)population matrix is denoted by $X$. The i[th] BC (CS) is $X_i$. Its d[th] dimension in the (SS) search space (DV) is $x_{i,d}$. The number of BC is N. The number of DV is $m$. In the interval [0,1], a random number $r$exists. The d[th] DV's lower and upper bounds are denoted by $lb_d$ and $ub_d$.

The objective function (OF) of the problem can be assessed in relation to the position of each BC, which indicates a CS for the problem. Equation (3) can be used to express the set of evaluated values for the OF using a vector.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1} \quad (3)$$

$F$ is the evaluated OF's vector in this case. The evaluated OF based on the i[th] BC is denoted by $F_i$. In the same way, the best BOA member is the one with the best feature solution in the OF $F$, and the worst BOA member is the one with the worst evaluated value for the OF. Every iteration of BOA design updates the BC's position in the PSS, which in turn updates the CS and OF values.

***Stage 1: Exploration Stage***

Based on the simulation of BC tracking and movement towards prey during hunting, the PM positions in the PSS are updated during the first phase of BOA. The exploration ability of BOA to control the global search (GS) is increased by modelling the movement of BC towards the prey, which results in significant changes in the position of the PM in the PSS. The candidate prey (CP) set for every BC is determined by equation (4).

$$CP_i = \{X_k : F_k < F_i and\ k \neq i\}, where\ i = 1,2,\dots,N\ and\ k \in \{1,2,\dots,N\} \quad (4)$$

The set of CP' positions for the i[th] BC is denoted by $CP_i$. The PopMem with a higher OF value than i[th] BC is $X_k$. Its OF value is $F_k$. Equation (5) is used to determine a new position for every BOA member based on the modelling of the BC's position for FS change as it moves near the prey in this technique.

If this novel location raises the OF value in accordance with Equation (6), it takes the place of the corresponding member's prior position.

$$x_{i,j}^{P1} = x_{i,j} + \left(1 - 2r_{i,j}\right).\left(SP_{i,j} - I_{i,j}.x_{i,j}\right) \quad (5)$$

$$X_i = \begin{cases} X_i^{P1}, F_i^{P1} \leq F_i \\ X_i, else, \end{cases} \quad (6)$$

The selected prey (SP) by i[th] BC is denoted as $SP_i$. Its j[th] dimension is $SP_{i,j}$. Based on the BOA's exploration phase, the new position for the i[th] BC is $X_i^{P1}$. Its jth dimension is $x_{i,j}^{P1}$. Its OF value is $F_i^{P1}$. Arbitrary values from the interval [0,1] make up $r_{i,j}$. The integers $I_{i,j}$ are chosen at random to be either 1 or 2.

***Stage 2: Exploitation Stage***

The chase simulation between the BC and the prey during hunting is used to update the PopMem position in the PSS during the 2[nd] stage of BOA. In order for the BC in order to capture the prey, this chasing phase takes place close to the hunting location. The ability of BOA to exploit (LS) local search is enhanced by simulating the movement of BC during the pursuit and capture of prey, which results in slight adjustments to PopMem position in the PSS.

Equation (7) is used in BOA design to determine a new position for each BOA member nearer to the hunting location based on the modelling of BC position varied throughout the chase procedure. According to Eqn (8), the associated member's prior location is then replaced by this novel one if it raises the OF value.

$$x_{i,j}^{P2} = x_{i,j} + \frac{1 - 2r_{i,j}}{1 + t}.x_{i,j} \quad (7)$$

$$X_i = \begin{cases} X_i^{P2}, F_i^{P2} \leq F_i \\ X_i, else, \end{cases} \quad (8)$$

The new location determined for the i[th] BC based on the BOA's exploitation phase is denoted by $X_i^{P2}$. Then its j[th] dimension is $x_{i,j}^{P2}$. The value of its OF is $F_i^{P2}$. Random numbers from the interval [0,1] make up $r_{i,j}$. The iteration counter is denoted by t.

***Pseudocode and Flowchart of BOA***

After changing each BC's location in the PSS according to the exploration and exploitation (E-E) stages, an initial repetition of the suggested BOA technique is performed. The process of updating BC continues until the final iteration of the procedure based on Equations (4)–(8). Following that, the algorithm enters the subsequent iteration with updated values for the position of BC and the OF.

Until the iteration is changed and saved, the optimal solution is found in each iteration. The best CS found during the algorithm's iterations is displayed as the BOA solution for the given problem once the algorithm has been fully implemented. Figure 2 displays a flowchart of the BOA implementation processes, and Algorithm 2 presents the pseudocode for the system.

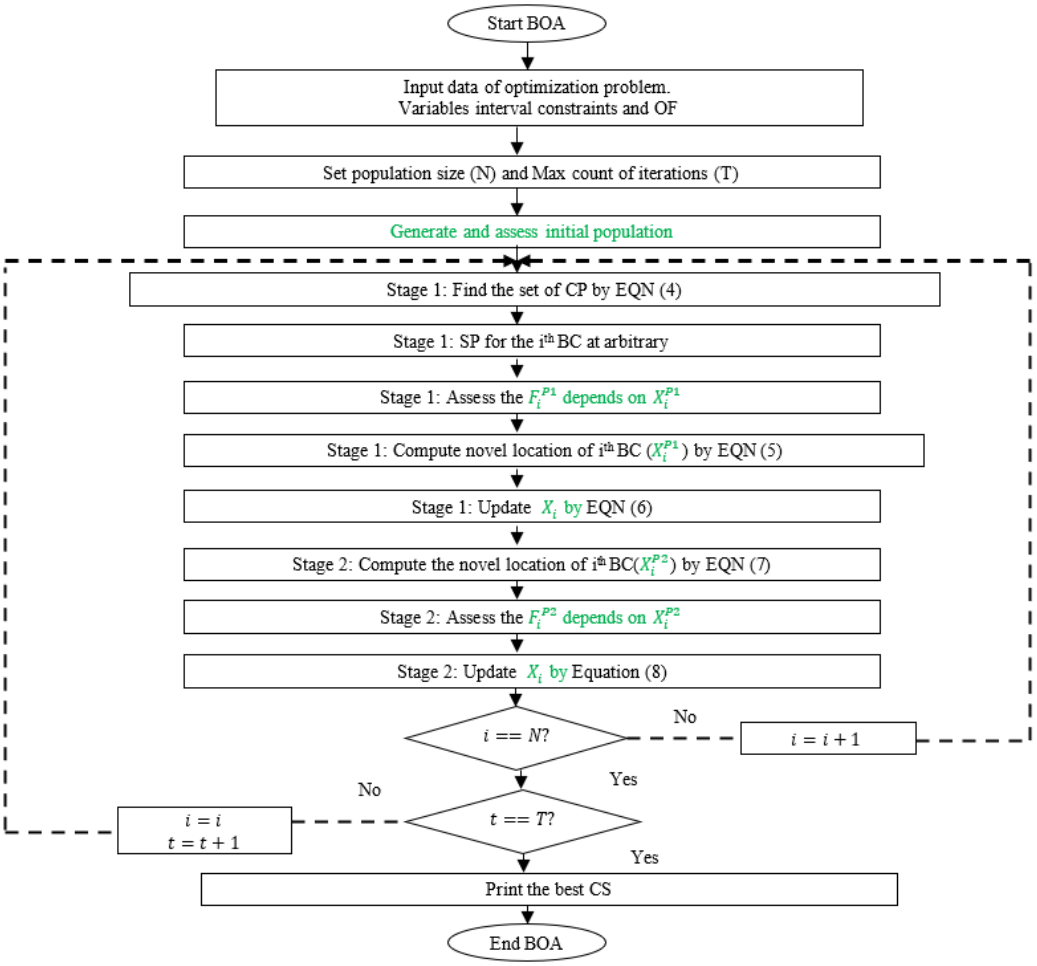


**FIGURE 2. FLOWDIAGRAM OF BOA**

**Procedure 2: Pseudocode of BOA**

Begin BOA

1.Input problem data: variables, OF, and constraints.

2.Set BOA of the N and T.

3.Create the initial Pop Mx at arbitrary by EQN (2). $x_{i,d} \leftarrow lb_d + r.(ub_d - lb_d)$

4.Assess the OF for best selected feature solution

5. For t=1 to T

6. For i=1 to N

7. Stage 1: exploration stage

8. Find the preys set for the i$^{th}$ BOA member by EQN (4). $CP_i \leftarrow \{X_k : F_k < F_i and\ k \neq i\}$

9. Choose the terminate mounds for the i$^{th}$ BOA member at arbitrary.

10. Compute novel location of i$^{th}$ BOA member by EQN (5). $x_{i,j}^{P1} \leftarrow x_{i,j} + \left(1 - 2r_{i,j}\right).\left(SP_{i,j} - I_{i,j}.x_{i,j}\right)$

11. Update i$^{th}$ BOA member by equation (6). $X_i \leftarrow \begin{cases} X_i^{P1}, F_i^{P1} \leq F_i \\ X_i, else, \end{cases}$

12. Stage 2: exploitation stage

13. Compute novel location of i$^{th}$ BOA member by EQN (7). $x_{i,j}^{P2} \leftarrow x_{i,j} + \frac{1-2r_{i,j}}{1+t}.x_{i,j}$

14. Update i$^{th}$ BOA member by EQN (8). $X_i \leftarrow \begin{cases} X_i^{P2}, F_i^{P2} \leq F_i \\ X_i, else, \end{cases}$

15. end

16. store the optimal CS so far.

17. end

18. Output the best quasi-optimal solution attained through the BOA.

End BOA.

***Population Diversity (PD), E-E Analysis***

The way that PopMem are dispersed throughout the problem space is referred to as the PD of BOA. This PD is essential for monitoring the search processes of the algorithm. This metric indicates whether population are more inclined to focus on optimising existing solutions or investigating novel options. The capacity of the algorithm to efficiently explore and exploit as a group may be evaluated and adjusted by taking into account the variety within the BOA population.

| (9) |
|---|
| (10) |

is the number of PopMem. The number of problem dimensions is denoted by . In the dimension *d*, *d*th represents the population mean. Thus, Equations (11), (12), can be used to define the extent of E-E within the population for each iteration.

| (11) |
|---|
| (12) |

## 3.4.*DATA CLASSIFICATION*

The best selected number of features taken as input to ETSA-LSTM algorithm for TC detection. The process of classifying data or objects into predetermined groups or classifications according to their features is known as classification. An algorithm is trained on a labelled dataset to predict the class in DL classification, a kind of Supervised Learning (SL) technique. ETSA-LSTM algorithm is to predict the thyroid nodule malignancy to achieve both better estimation and clearer interpretation.

**LSTM**

A MC and four gates make up the LSTM, a type of RNN. Forget gate , input gate , control gate , and output gate o are those four gates in LSTM. According to LSTM, after executing certain operations on the prior memory content , it produces current memory content and current cell state as final outputs. It accepts bias $b$, current input vector , and previous cell state as inputs [24]. Equation (13), which defines the operation of ,

$$Diversity = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{d=1}^{m}\left(x_{i,d} - \overline{x_d}\right)^2} \tag{9}$$

$$\overline{x_d} = \frac{1}{N}\sum_{i=1}^{N} x_{i,d} \tag{10}$$

$N$ is the number of PopMem. The number of problem dimensions is denoted by $m$. In the dimension $d$, $d^{\text{th}}$ represents the population mean. Thus, Equations (11), (12), can be used to define the extent of E-E within the population for each iteration.

$$\text{Exploration} = \frac{\text{Diversity}}{\text{Diversity}_{\max}} \tag{11}$$

$$\text{exploitation} = \frac{|\text{Diversity} - \text{Diversity}_{\max}|}{\text{Diversity}_{\max}} \tag{12}$$

## 3.5. *DATA CLASSIFICATION*

The best selected number of features taken as input to ETSA-LSTM algorithm for TC detection. The process of classifying data or objects into predetermined groups or classifications according to their features is known as classification. An algorithm is trained on a labelled dataset to predict the class in DL classification, a kind of Supervised Learning (SL) technique. ETSA-LSTM algorithm is to predict the thyroid nodule malignancy to achieve both better estimation and clearer interpretation.

**LSTM**

A MC and four gates make up the LSTM, a type of RNN. Forget gate $f$, input gate $i$, control gate $c$, and output gate o are those four gates in LSTM. According to LSTM, after executing certain operations on the prior memory content $c_{t-1}$, it produces current memory content $c_t$ and current cell state $h_t$ as final outputs. It accepts bias $b$, current input vector $x_t$, and previous cell state $h_{t-1}$as inputs [24]. Equation (13), which defines the operation of $f$,

$$f_t = \sigma_g(w_f x_t + u_f h_{t-1} + b_f) \tag{13}$$

The amount of input from the current timestamp I that is added to the memory cell (MC). Equation (14) express it as follows,

$$i_t = \sigma_g(w_i x_t + u_i h_{t-1} + b_i) \tag{14}$$

By taking into account the output of the $f$ $and$ $i$, which is specified by equation (15), the $c$ oversees the process of updating MC content from $c_{t-1}$ to $c_t$.

$$c_t = f_t \times c_{t-1} + i_t \times \sigma_h(w_c x_t + u_c h_{t-1} + b_c) \tag{15}$$

The $o$ generates the current timestamp's final output. The cell state $h_{t-1}$ is also updated to $h_t$ by $o$. Equations (16-17) define the final output.

$$o_t = \sigma_g(w_o x_t + u_o h_{t-1} + b_o) \quad (16)$$

$$h_t = o_t \times \sigma_h(c_t) \quad (17)$$

The sigmoid function is denoted as $\sigma_g$. $\sigma_h$ indicates hyperbolic tangent functions. The vanishing gradients problem (VGP) is avoided by using the weight values $w$ and $u$. After the inputs are sent to an input layer (IL) .Then it can be processed through two hidden layers (HL). At last, a dense layer (DenL) acts as the output layer (OL), carrying out the last classifications.

An estimated attention value for every input in every layer indicates how crucial each input is to executing the final TC prediction. With the use of an attention vector, the DenL ultimately determines whether a patient has TC. When the o of the LSTM is closed, it is unable to access the contents of its prior MC. ETSA-LSTM addresses this problem by introducing an additional connection. ETSA-LSTM referred to as a peephole connection, across every gate and the prior memory content.

**ETSA-LSTM**

The complexity and timeliness problems caused by the three gates are resolved by LSTM network applications. To ensure that the model encoder (E) converges to particular features of the input sequence when predicting the specific features of the output sequence, the ETSA-LSTM model attention mechanism (AM) first and foremost gives priority to the input characteristics, particularly their significance. This improves the prediction quality and self-learning development of the model. Choosing network features and preserving model performance in real-time (RT) improve prediction quality.

To concentrate on a single aspect of a complicated input until categorisation, the network employs AM. The suggested model performs better in studies than baseline models due to input attention, temporal attention, and repetition. For TC data, the ETSA-LSTM model is employed. Users can better comprehend the model's SE to a variety of scenarios and possible sources of uncertainty by using sensitivity analysis (SA). Decision-making by the user (DMak) is enhanced by this. By altering input features and examining the outcomes, these objectives are accomplished. By offering model-specific documentation on the model's structure, training procedures, input features, and output evaluations, healthcare professionals can develop confidence in the model and comprehend its limitations. By improving its interpretability, these methods enable healthcare professionals without prior knowledge of ML to use the ETSA-LSTM model. For the precise assessment and long-term (LT) prediction of cancer occurrences, the suggested ETSA-LSTM model offers a time series (TS) model [25].

The input sequence is $n$, and the window size length is $\tau$. The equation $X = (x^1, x^2, x^3, \dots, x^T) = (x_1, x_2, x_3, .., x_n) \in \mathbb{R}^{n\times\tau}$ represents the annual incidence movement. Then, d is a vector of $n$ moving input sequence at time t is thus represented as $X = (x_t^1, x_t^2, x_t^3, \dots, x_t^n)^T \in \mathbb{R}^n$. The moving sequence of length $\tau$ is then represented as $X^a = (x_1^a, x_2^a, x_3^a, .., x_\tau^a)^T \in \mathbb{R}^\tau$. Using the target sequence's prior values as $(y_1, y_2, y_3, .., y_{\tau-1})$ and $y_t \in \mathbb{R}$ again. As $(x_1, x_2, x_3, .., x_\tau); x_t \in \mathbb{R}^n$, the present and previous values of the $n$ moving sequence are represented. With $F(\bullet)$ as the non-linear (NL) mapping function to be determined via optimisation, the NL autoregressive exogenous models seek to optimise NL mappings to the current value of a target sequence $y_\tau$ using $\widehat{y_\tau} = F(y_1, y_2, .., y_\tau, x_1, x_2, \dots, x_t)$.

The LSTM units can be formalised as follows, $W$ and $U$ are the weight matrices, given each time *t*, an input vector $x_t$, a memory state vector $c_t$, and $h_t$ as the output of $c_t$. A bias vector is denoted by $b$. Input, forget, and output gating units are denoted by $i_t, f_t, o_t$. The gating units' (AF) activation function is denoted by $\sigma$. To determine the optimal mapping from $x_t$ to $h_t$ using $\mathrm{h_t} = \mathrm{f_1(h_{t-1}, xt)h_t} \in \mathbb{R}^s$, apply the suggested LSTM E to the input sequence $X = (x_1, x_2, x_3, .., x_\tau)|x_t \in \mathbb{R}^n$ for the suggested TS prediction at time t. Here, hidden state (HS) of the E at the time $t$ is denoted as $h_t$. $s$ is the size of the $h_t$, and $f_1$ is LSTM NL AF.

**Two Stage Attention mechanism (TSA)**

For TSA, LSTM E reads input sequences and encodes them into fixed-length vectors (FLV). Models are unable to learn longer sequences when they are encoded as FLV (Equation 18-22). Prediction ACC is increased by

augmenting the encoding unit of the ETSA model. Initially, the convergence of the model's LSTM E on input and output sequence features is investigated. Learning enhanced the prediction of the model. Then, in order to maintain model performance as sequence length increases, build a second AM, temporal attention, to select crucial network features continuously throughout decoding.

By using the LSTM cell state $c_{t-1}$ and the prior HS $h_{t-1}$ in equation (18), an input AM with an MLP is created by assuming the a-th moving sequence input, $x^a = (x_1^a, x_2^a, x_3^a, \dots, x_\tau^a) \in \mathbb{R}^\tau$.

$$e_t^a = v_e^T \tanh(W_e[h_{t-1}, c_{t-1}] + U_e x^k) \,|\alpha_t^a = \exp(e_t^a) \left[\sum_{t=1}^{n} \exp(e_t^a)\right]^{-1} \tag{18}$$

Here, $v_e^T \in \mathbb{R}^\tau, W_e \in \mathbb{R}^{\tau \times 2s}, U_e \in \mathbb{R}^{\tau \times \tau}$ are learnable parameters, The attention weight (AW) that maintains the significance of the $a$-th input feature at time $t$ is denoted by $\alpha_t^a$.

Finally, conduct a soft action to $e_t^a$ so that the sum of all the AW equals 1. The HS of the E is updated as ($\widehat{h_t} = f_1(h_{t-1}, \widehat{x_t})$ when the AM and LSTM E are combined, resulting in the sequence $\hat{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \alpha_t^3 x_t^3, .. \alpha_t^n x_t^n)$. This enables the LSTM E unit of the model to preferentially focus on distinctive features of the input sequence. Next, build an LSTM (D) decoding unit that predicts output $\widehat{y_t}$ by using the internal representation of the model. Build an LSTM D unit that predicts output $a$ by utilising the internal representation of the model.

As the length of the sequence expands, build a temporal AM that consistently selects critical network features and D maintains the model performing well across the network. Since the AW of an E's HS is calculated using equations (19–20) at time $t$,

$$l_t^i = v_d^T \tanh(W_d[d_{t-1}; \hat{c}_{t-1}] + U_d h_t) \,|1 \le i \le \tau \tag{19}$$

$$\beta_t^i = \exp(l_t^i) \left[\sum_{t=1}^{\tau} \exp(l_t^i)\right]^{-1} \tag{20}$$

Here, $\hat{c}_{t-1} \in \mathbb{R}^p$ is the cell state of the LSTM unit and $d_{t-1} \in \mathbb{R}^p$ is the HS of a prior D. Learnable parameters are $v_d^T \in \mathbb{R}^s, W_d \in \mathbb{R}^{s \times 2p}, U_d \in \mathbb{R}^{s \times s}$.

The temporal AM contains the pertinent features of the $i$-th E HS for prediction utilising AW $\beta_t^i$ at the D level, just like the E unit's input AM. The AM evaluates the context vector $c_t$ (unique at each time) as the weighted sum of all the HS $\{h_1, h_2, h_3, \dots, h_\tau\}$ of the E units in equation (21), taking into account that the E's HS $h_i$ is mapped to the input's temporal component.

$$\widehat{c}_t = \sum_{t=1}^{\tau} \beta_t^i h_i \tag{21}$$

Equation (22) is thus attained by combining the target sequence and $c_t$.

$$\widehat{y}_{t-1} = \widehat{W^T}[y_{t-1}; c_{t-1}] + \hat{b} \tag{22}$$

Here, $\widehat{w} \in \mathbb{R}^{m+1}$ and $\hat{b} \in \mathbb{R}$ maps $[y_{t-1}; c_{t-1}]$ to the size of D. The D's HS at time $t$ is then updated using $\widehat{y}_{t-1}$. the TSA E-D model as it outperformed other SOTA models. Figure 3 depicts the ETSA-LSTM design.

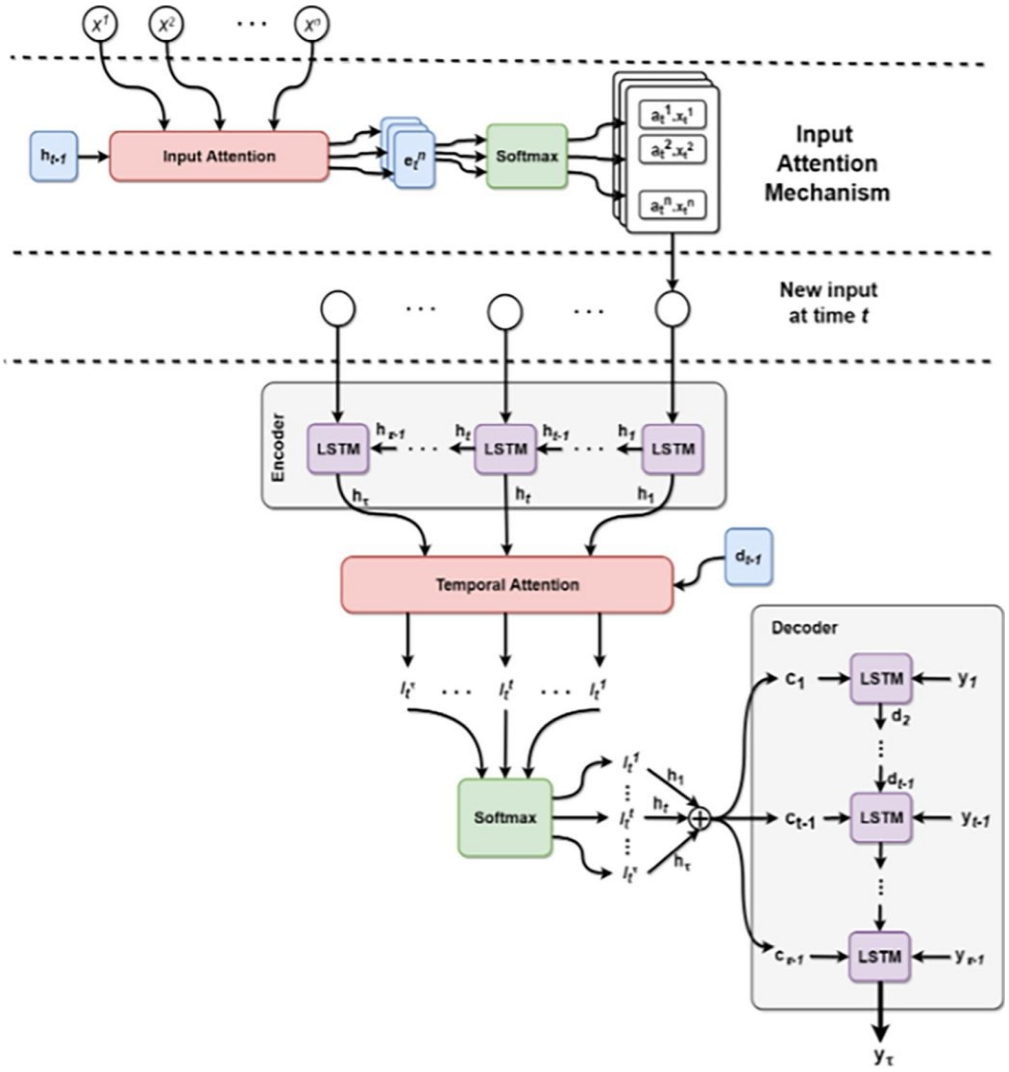


**FIGURE 3. ARCHITECTURE OF ETSA-LSTM**

## 4. RESULTS AND DISCUSSION

This part reveals the outcomes of several classifiers experimented on the TC risk prediction dataset is collected from Kaggle online repository [26]. It simulates real-world TC risk factors and has 212,691 records with 23 features.

### 4.1.*EVALUATION METRICS*

**Performance Measures:** The classifiers in this work were evaluated using the PR, R or SE, SP, AC, and AUROC curves. Table 1 displays the Confusion Matrix's (CM) format in general. Samples for Binary Classification (BC) problems are categorised as true positive (tp), false positive (fp), true negative (tn), and false negative (fn) based on the combinations of actual class (AC) and predicted class (PC).

**TABLE 1. CM IN THIS RESEARCH**

| AC | PC | |
|---|---|---|
| | P | N |
| P | tp | fn |
| N | fp | tn |

The results of all the classifiers are measured via the metrics like PR, R or SE, SP, ACC and AUROC curve as shown by equations (23, 24, 25, 26, & 27). PR is an incorrect positive diagnosis could subject a patient to unnecessary treatments. Similarly, R or SE is paramount to avoid missing a TC disease diagnosis, which can have serious health implications. SP to ensure that those without the disease are not falsely diagnosed, thus preventing unnecessary treatments. ACC, which is assumed to be one of the extensively regarded measures, is used to study the performance of classifiers. AUROC metric is utilized to assess the discriminative efficiency of the frameworks. This metric is crucial in clinical diagnostics as it helps in effectively identifying tp (sensitivity) while minimizing fp (specificity).

PR is a proportion of how many thyroid cases the model correctly identified. Equation (23) provides its definition.

$$\text{Precision(PR)} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{23}$$

The model's capacity to accurately identify thyroid in living individuals is indicated by R or SE. Equation (24) defines a SE maximum percentage 100, which means that the algorithm chosen correctly predicted each instance of TC.

$$\text{Recall(R)/Sensitivity(SE)} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{24}$$

Equation (25) defines SP. SP assesses the model's capacity to identify individuals who will never experience thyroid problems.

$$\text{Specificity(SP)} = \frac{\text{tn}}{\text{tn} + \text{fp}} \tag{25}$$

Equation (26) defines ACC. The percentage of all forecasts that the framework accurately forecast is measured by AC.

$$\text{ACC} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \tag{26}$$

The TPR against FPR trade-off is displayed by AUROC. The discriminating thresholds can be changed to provide a sequence of TPR and FPR. AUROC is defined by Equation (27) illustrates that TPR and FPR,

$$\text{FPR} = \frac{\text{fp}}{\text{fp+fn}}, \text{TPR} = \frac{\text{tp}}{\text{tp+fn}} \tag{27}$$

The term "$tp$" refers to a thyroid person who is appropriately predicted to acquire TC. An unaffected thyroid person who was correctly predicted to be TC-free is referred to as $tn$. The incorrect diagnosis of affected in an unaffected person is known as $fp$. Affected individuals is not impacted if they receive a $fn$ result.

## 4.2.*SIMULATION EXPERIMENT*

Then, various existing techniques are DNN, RF, and Conv-LSTM has been suggested for results comparison of suggested ETSA-LSTM. The various performance indices now in utilisation are obtained through the application of classification and FS. The comparative analysis of the suggested classifier with current methods is discussed in table 2.

**TABLE 2. COMPARATIVE PERFORMANCE OF THE SUGGESTED CLASSIFIERS AND CURRENT METHODS**

| Evaluation Metrics/Classifiers | DNN | RF | Conv-LSTM | ETSA-LSTM |
|---|---|---|---|---|
| Precision (PR) | 81.2 | 84.8 | 86.8 | 88.7 |
| Recall (R) or Sensitivity (SE) | 82.2 | 84.5 | 87.5 | 89.6 |
| Specificity (SP) | 68.6 | 73.4 | 77.6 | 80.9 |

| | | | | |
|---|---|---|---|---|
| Accuracy (ACC) | 82.5 | 86.8 | 90.5 | 92.8 |
| Area Under Receiver Operating Characteristic (AUROC) | 82.5 | 87.9 | 89.7 | 91.8 |

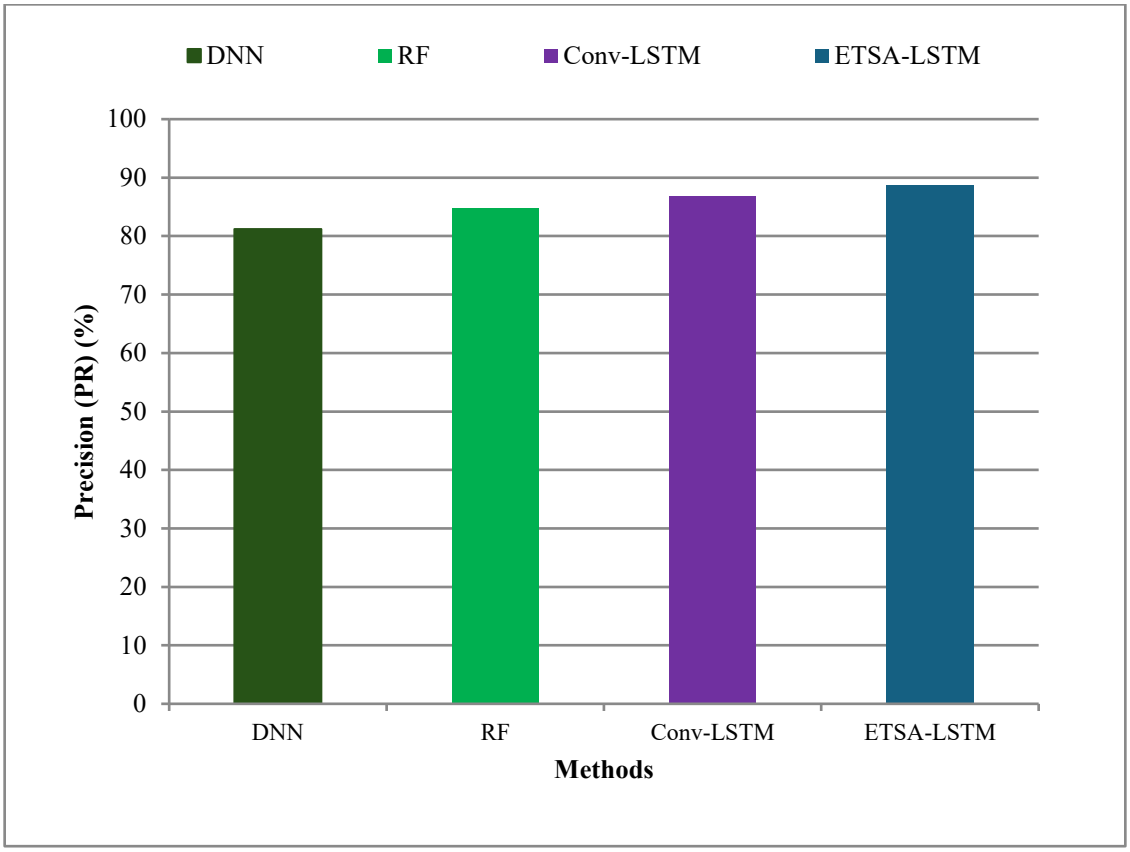


**FIGURE 4. P COMPARISON VS. CLASSIFIERS**

Figure 4 shows the PR performance metric comparison between existing classifiers like DNN, RF, Conv-LSTM, and suggested ETSA-LSTM for TC detection. The classification methods are indicated in the X-axis and the y-axis represents the PR results. Suggested work using SMOTE for data imbalance and it increases the PR outcomes. From the outcomes it is concluded that the suggested ETSA-LSTM classifier delivers highest PR outcomes of 88.7% and the existing DNN, RF, and Conv-LSTM has lowest precision of 81.2%, 84.8%, and 86.8% (Refer Table 2).

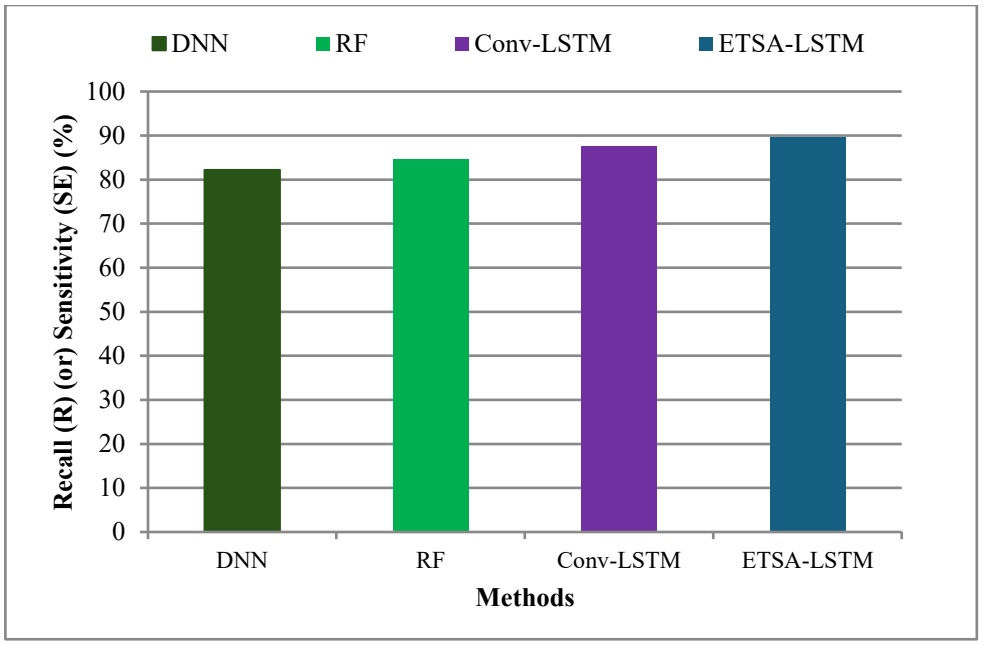


**FIGURE 5. RECALL (R) (or) SENSITIVITY (SE) COMPARISON VS. CLASSIFIERS**

The performance comparison outcomes for the current classifiers like DNN, RF, Conv-LSTM, and proposed ETSA-LSTM for TC detection in terms of Recall (R) (or) sensitivity (SE) are shown in Figure 5. In the above figure, the classification methods are plotted in X-axis and the y-axis denotes the R results. From the outcomes it is concluded that the suggested ETSA-LSTM classifier delivers highest R outcomes of 89.6% and the existing DNN, RF, and Conv-LSTM has lowest R of 82.2%, 84.5% and 87.5% (Refer Table 2).

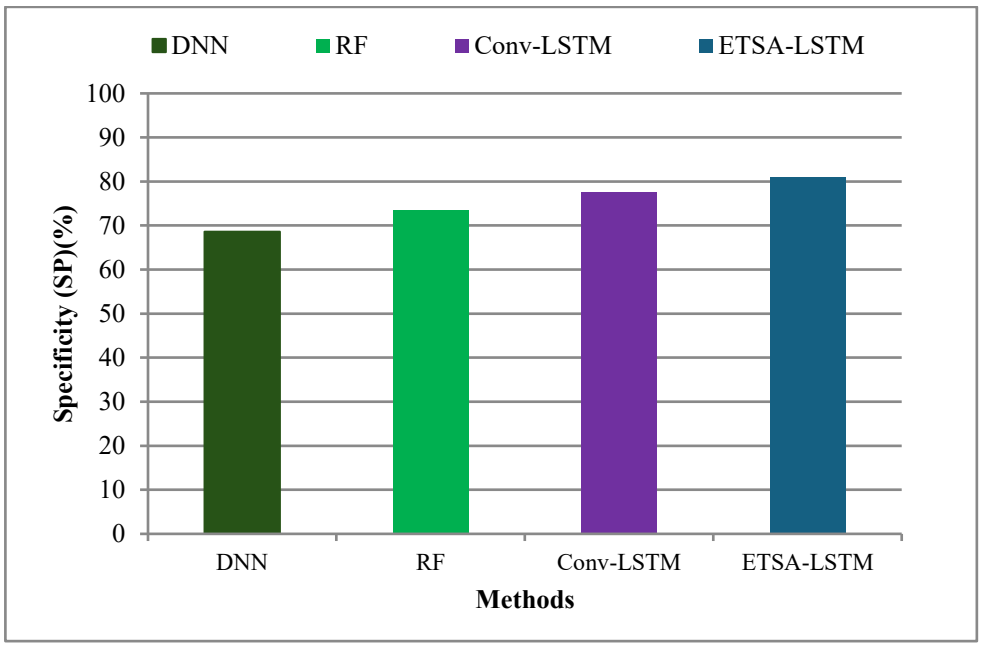


**FIGURE 6. SPECIFICITY (SP) COMPARISON VS. CLASSIFIERS**

Specificity (SP) performance metric comparison between existing classifiers like DNN, RF, Conv-LSTM and proposed ETSA-LSTM for TC detection are shown in figure 6. In the above figure, the classification methods are indicated in x-axis and the y-axis denotes the SP results. From the results it is concluded that the ETSA-LSTM classifier delivers highest specificity results of 80.9% while the existing DNN, RF, and Conv-LSTM has lowest specificity of 68.6%, 73.4% and 77.6% (Refer Table 2).

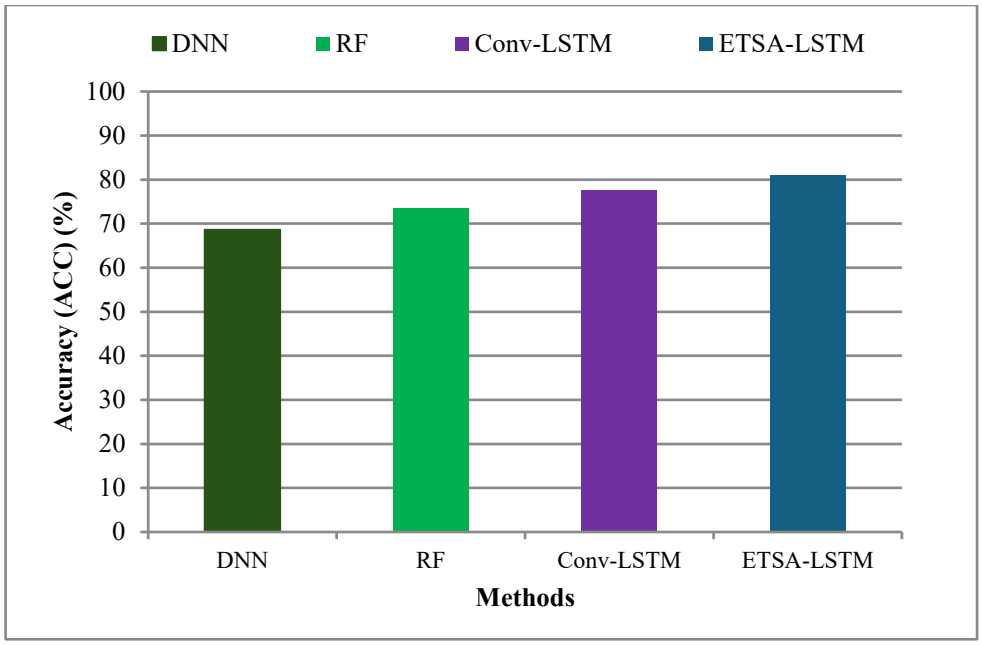


**FIGURE 7. ACCURACY (ACC) COMPARISON VS. CLASSIFIERS**

Figure 7 shows the ACC performance metric comparison between existing classifiers like DNN, RF, Conv-LSTM and proposed ETSA-LSTM for TC classification. The ACC results are showed on the Y-axis in the above image, while the classification methods are displayed on the X-axis. From the results it is concluded that the suggested classifier ETSA-LSTM delivers highest ACC outcomes of 92.8% while the existing DNN, RF, and Conv-LSTM has lowest ACC of 82.5%, 86.8% and 90.5% (Refer Table 2).

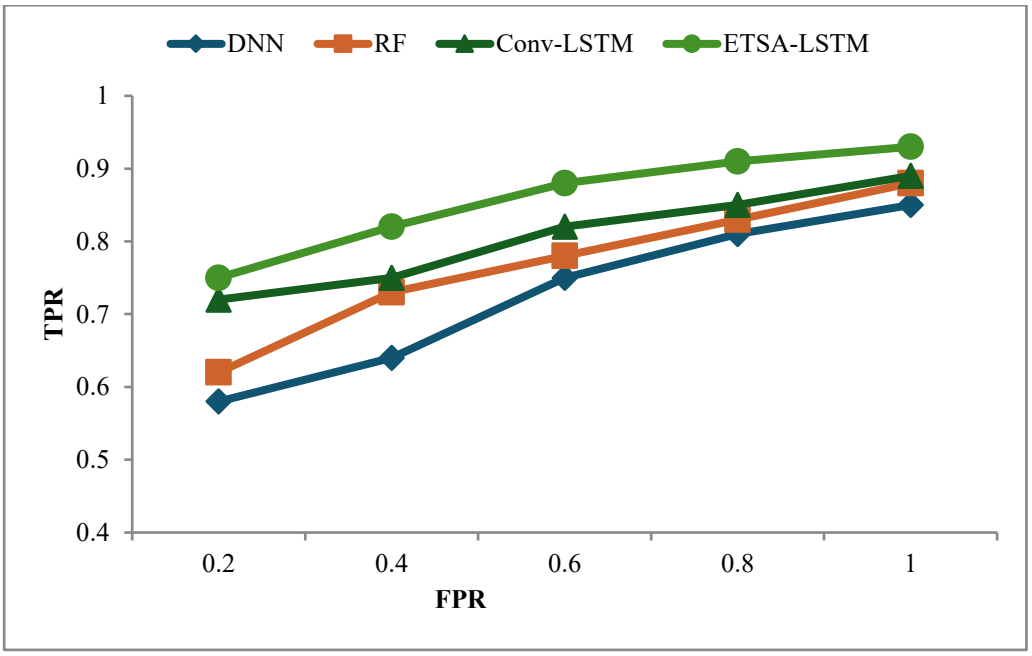


**FIGURE 8. AUROC COMPARISON VS. CLASSIFIERS**

Figure 8 shows the AUROC performance metric comparison between existing classifiers like DNN, RF, Conv-LSTM and proposed ETSA-LSTM for TC classification. The FPR is shown on the X-axis and the TPR results are presented on the y-axis, as presented in the above figure. From the outcomes it is concluded that the suggested classifier ETSA-LSTM delivers highest AUROC results while the existing DNN, RF, and Conv-LSTM has lowest AUROC (Refer Table 2).

## 5. CONCLUSION AND FUTURE WORK

TC disease is a major healthcare concern globally, significantly affecting individual's quality of life and health. TC issues are on the rise, and treatment that reduces mortality and avoids complications depends on ED. Although DL methods for TC detection and prediction have a lot of clinical utility potential, its development and application present a number of difficulties. In order to identify important features and remove irrelevant, redundant, or noisy features, data pre-processing is necessary for DL techniques. This reduces the size of the feature space.

This work aimed to provide a DL model for diagnosis of TC. The ETSA-LSTM framework for simulating real-world TC risk factors, integrating techniques for DC, data pre-processing, FS, data classification and performance evaluation. Firstly data collection, TC risk prediction dataset is collected from Kaggle online repository. Secondly, data pre-processing using MMN (or) MMS, is used to perform DN. Feature importance performed Z-score-based BorutaSHAP FI technique to extract relevant feature information, noise removal using EkNN and SMOTE for addressing data imbalance. Finally, the most pertinent features (an ideal reduced feature subset) were chosen using FS and BOA in order to identify the TC. Then data classification, ETSA-LSTM to predict the TN malignancy to achieve both better estimation and clearer interpretation of the parameters. Finally, performance evaluation, evaluated based on PR, R or SE, SP, ACC, and AUROC curve. Outcomes indicate that the suggested model achieves 92.8% accuracy which is better than other current frameworks. The simulation's results show that the ETSA-LSTM model overtakes other frameworks for TC prediction in most cases. Additionally, (1) A meta classifier algorithm that boosts classification accuracy by combining the new ensemble model with bagging, stacking, voting, and boosting techniques; (2) In several datasets, data quality is calculated by the model prediction accuracy.

### References


1. Rossi, E.D., Pantanowitz, L. and Hornick, J.L., 2021. A worldwide journey of thyroid cancer incidence centred on tumour histology. The lancet Diabetes & endocrinology, 9(4), pp.193-194.
2. Nguyen, Q.T., Lee, E.J., Huang, M.G., Park, Y.I., Khullar, A. and Plodkowski, R.A., 2015. Diagnosis and treatment of patients with thyroid cancer. American health & drug benefits, 8(1), pp.30-40.
3. Chen, L., Li, X., Sheng, Q.Z., Peng, W.C., Bennett, J., Hu, H.Y. and Huang, N., 2016. Mining health examination records-a graph-based approach. IEEE Transactions on Knowledge and Data Engineering, 28(9), pp.2423-2437.
4. Davies, L. and Hoang, J.K., 2021. Thyroid cancer in the USA: current trends and outstanding questions. The lancet Diabetes & endocrinology, 9(1), pp.11-12.
5. Rao, A.R. and Renuka, B.S., 2020. A machine learning approach to predict thyroid disease at early stages of diagnosis. In 2020 IEEE international conference for innovation in technology (INOCON), pp.1-4.
6. Siegel, R.L., Miller, K.D., Wagle, N.S. and Jemal, A., 2023. Cancer statistics, 2023. CA: a cancer journal for clinicians, 73(1), pp.17-48.
7. Abe, I. and Lam, K.Y., 2021. Anaplastic thyroid carcinoma: Updates on WHO classification, clinicopathological features and staging. Histol Histopathol, 36, pp.239-248.
8. Nguyen, M., He, G. and Lam, A.K.Y., 2022. Clinicopathological and molecular features of secondary cancer (metastasis) to the thyroid and advances in management. International Journal of Molecular Sciences (IJMS), 23, pp.1-25.
9. Zhanga, X. and Lee, V.C.S., 20204. Deep Learning Empowered Decision Support Systems for Thyroid Cancer Detection and Management. Procedia Computer Science, 237, pp.945-954.
10. Ahmed, I., Mohiuddin, R., Muqeet, M.A., Kumar, J.A. and Thaniserikaran, A., 2022. Thyroid cancer detection using deep neural network. In 2022 International conference on applied artificial intelligence and computing (ICAAIC), pp.166-169.
11. Uddin, K.M.M., Al Mamun, A., Chakrabarti, A. and Mostafiz, R., 2024. An ensemble machine learning-based approach to predict thyroid disease using hybrid feature selection. Biomedical Analysis, 1(3), pp.229-239.
12. Alshayeji, M.H., 2023. Early thyroid risk prediction by data mining and ensemble classifiers. Machine Learning and Knowledge Extraction, 5(3), pp.1195-1213.
13. Sibarani, I.J.B. and Suharjito, S., 2024. Enhancing predictive accuracy for differentiated thyroid cancer (DTC) recurrence through advanced data mining techniques. TIN: Terapan Informatika Nusantara, 5(1), pp.11-22.
14. Alghamdi, N.S., 2019. Evaluation of classification models for predicting mortality rate using thyroid cancer data. Journal of Computational Science, 15(1), pp.131-142.
15. Anuhya, K., Saie, N.S., Pravinya, G., Hemanth, P. and Pothireddy, A.R., 2024. Enhanced Thyroid Cancer Classification: Leveraging Advanced Machine Learning Techniques with a Focus on Random Forest for Optimal Accuracy. In 2024 2nd World Conference on Communication & Computing (WCONF), pp.1-8.
16. Vanitha, R. and Perumal, K., 2025. Prediction of the Thyroid Cancer using the Deep Learning Based Hybrid Spatial Convolution based LSTM Network (SCBLN). Journal of Information Systems Engineering and Management, 10(3), pp.220-237.
17. Xi, N.M., Wang, L. and Yang, C., 2022. Improving the diagnosis of thyroid cancer by machine learning and clinical data. Scientific reports, 12(1), pp.1-11.

18. https://www.kaggle.com/datasets/ankushpanday1/thyroid-cancer-risk-prediction-dataset
19. Islam, M., Chen, G. and Jin, S., 2019. An overview of neural network. American Journal of Neural Networks and Applications, 5(1), pp.7-11.
20. Shantal, M., Othman, Z. and Bakar, A.A., 2023. A novel approach for data feature weighting using correlation coefficients and min–max normalization. Symmetry, 15(12), pp.1-18.
21. Akter, S. and Mustafa, H.A., 2024. Analysis and interpretability of machine learning models to classify thyroid disease. Plos one, 19(5), pp.1-30.
22. Latif, M.A., Mushtaq, Z., Arif, S., Rehman, S., Qureshi, M.F., Samee, N.A., Alabdulhafith, M. and Al-masni, M.A., 2024. Improving Thyroid Disorder Diagnosis via Ensemble Stacking and Bidirectional Feature Selection. Computers, Materials & Continua, 78(3), pp.1-17.
23. Benmamoun, Z., Khlie, K., Bektemyssova, G., Dehghani, M. and Gherabi, Y., 2024. Bobcat Optimization Algorithm: an effective bio-inspired metaheuristic algorithm for solving supply chain optimization problems. Scientific Reports, 14(1), pp.1-62.
24. Rahman, M.M. and Siddiqui, F.H., 2019. An optimized abstractive text summarization model using peephole convolutional LSTM. Symmetry, 11(10), pp.1-24.
25. Khan, R. and Jie, W., 2025. Using the TSA-LSTM two-stage model to predict cancer incidence and mortality. Plos one, 20(2), pp.1-39.
26. https://www.kaggle.com/datasets/ankushpanday1/thyroid-cancer-risk-prediction-dataset