Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 19s (2025)
https://www.jneonatalsurg.com

Modified Grasshopper Optimization Algorithm (Mgoa) Based Feature Selection And Convolutional Long Short-Term Memory (Conv-Lstm) For Diagnosis Of Thyroid Cancer

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

[1]Research Scholar, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Bharathiar University, Coimbatore, Tamil Nadu, India.

[2]Assistant Professor, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Bharathiar University, Coimbatore, Tamil Nadu, India.

Email ID: Kokilamanikcas1@gmail.com

.Cite this paper as: Sruthi. V.S, Dr. Kokilamani, M, (2025) Modified Grasshopper Optimization Algorithm (Mgoa) Based Feature Selection And Convolutional Long Short-Term Memory (Conv-Lstm) For Diagnosis Of Thyroid Cancer. *Journal of Neonatal Surgery*, 14 (19s), 448-462.

ABSTRACT

One prevalent endocrine carcinoma that affects the thyroid gland is Thyroid Cancer (TC). Thyroidectomy is still the principal therapeutic option, despite significant efforts to improve diagnosis. An effective procedure with no unnecessary side effects depends on an accurate preoperative 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 Data Analysis (MDA) problems can be easily solved with the use of Data Mining (DM) algorithms. To find every problem in the dataset, DM helps with complex DA. With several techniques for classification, clustering, association, etc., DM offers significant assistance with thyroid datasets. Though there is enormous scope for clinical utility in applying Deep Learning (DL) techniques to predict and identify TC, their development and application present a number of difficulties. In order to find critical features and remove irrelevant, redundant, or noisy features, data pre-processing is necessary for DL approaches. This reduces the size of the feature space. Data collection (DC), data pre-processing, feature selection (FS), data classification, and performance evaluation are the five main stages of the suggested task. 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, data normalisation (DN), and handling missing data are addressed by using data pre-processing techniques. Min-Max normalisation (MMN), also known as Min-max scaling, is used to perform normalisation. Boosting, the Synthetic Minority Over-sampling Technique (SMOTE), and Tomek links (BST) are effective data balancing techniques for addressing data imbalance problems. Finally, the Modified Grasshopper Optimisation Algorithm (MGOA) is used to choose appropriate features. For TC detection, the most pertinent features (an optimal reduced feature subset) are chosen using MGOA. The search space (SS) can be effectively explored and exploited by MGOA, a nature-inspired algorithm (NIA). The TN malignancy is then predicted by classifying the data using Convolutional- (LSTM) Long Short-Term Memory (Conv-LSTM). This will enhance estimation and makes the parameters easier to understand. The final metrics utilised for assessing efficiency are Accuracy (AC), Specificity (SP), Precision (PR), Recall (R) or Sensitivity (SE), and the Area Under Receiver Operating Characteristic (AUROC) curve. According to the study of the results, the suggested model outperformed the other methods currently in use in terms of AC.

Keywords: *Thyroid Cancer (TC), Deep Learning (DL), Modified Grasshopper Optimization Algorithm (MGOA), Convolutional Long Short-Term Memory (Conv-LSTM), Synthetic Minority Over-sampling Technique (SMOTE), Data Mining (DM), Boosting, SMOTE, and Tomek links (BST), Area Under Receiver Operating Characteristic (AUROC), Feature Selection (FS), and Classification.*

---

1. INTRODUCTION

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]. From 2005 to 2015, the prevalence of TC increased by 5.5% yearly, according to the NIH's Surveillance, Epidemiology, and End Results Program (SEER) [2]. In USA, the prevalence of TC is still rising. According to the American Cancer Society's most current estimate, there will be roughly 2,230 TC-related fatalities and 43,800 new cases in 2022 [3]. The initial indication of TC, a solid tumour, is usually a nodule or mass in the thyroid gland at the front base of the throat. TC results from rogue cells multiplying too quickly for the immune system to handle. Cancer is usually caused by changes or mutations in genes that disrupt the genes that control the function of cells. As a result, unchecked cell division and tissue invasion occur.

In most cases, tumour metastasis (TM) significantly deteriorate the diagnosis of the individuals and may even be the primary cause of the patient's eventual death. Effective treatment and reduced damage can be achieved by treating ED (Early Detection) of the malignant TN before the cancerous cells in the thyroid gland spread [4]. One method for the ED of malignant TN is TC screening. 2 key methods for detecting TC: (1) Neck palpation as part of a physical examination. (2) Both palpable and nonpalpable nodules, particularly those with a diameter of less than 1 cm, can be found using ultrasonography (US). The most important therapeutic procedure for identifying the features of TN is US. These features aid in the nodules classification as either benign or malignant.

Papillary TC (PTC), follicular TC (FTC), medullary TC (MTC), and anaplastic TC (ATC) are the four histological forms that are typically included in TC. FTC and PTC make up 95% of TC, despite the fact that there are numerous additional types of TC [5]. As a result, there is a great deal of variation in the prognosis of TC patients. Of the TC, 99% are DTC, and 96% come from follicular cells. The primary treatments for DTCs include radioactive iodine remnant ablation, surgery, and TSH suppression therapy (levothyroxine), particularly PTC. Currently, surgery is the main therapeutic option, though customised care is depending on the type of lesion [6].

Differentiating among benign and malignant TN is one of the surgical operation's main goals. A seamless procedure, fewer unwanted side effects, and a lower chance of post-operative recurrence are all made possible by an appropriate preoperative diagnosis. In order to prolong the survival duration, it also aids in the selection of thorough post-surgery treatment [7]. As a result, using thyroid US, blood tests, and other fundamental medical data to create precise treatment and prognosis is essential. As of right now, radiologists' and surgeons' clinical expertise plays a major role in diagnosing malignant nodules (MN). 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 are accurate and comprehensible.

Over the past several years, TC research has advanced. Further analysis is required in light of the treatment's intricacy and difficulties. Methods that characterise clinical functions, such the DM approach, could be used for these analyses. The process of extracting information from difficult data using innovative approaches is known as the DM methodology. Medical management and health quality could be enhanced by the new science. 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 [8]. For the accountable health services, survival prediction is crucial because TC treatment is costly and taking a long time.

Tasks like learning, environment feature detection (FD) for self-driving, and clinical applications, by the utilization of 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 network (ANN). Without the use of preset rules, the meaningful relationships are generated by ANN, as it has the potential to detect patterns and learn from inputs and outputs. ANN are compared with classical statistical modelling used in some of the more recent cancer prognostication applications. Furthermore, in terms of pattern recognition (PR) and classification, ANNs perform better than classical statistical modelling. Additionally, there is a strong correlation between the amount of training data and the ANNs' performance. To generate relevant results to feed the subsequent layer, ANN models are built in layers to learn progressively higher-dimension (HD) and distant representations of the input data [9]. Previous studies have created statistical frameworks for predicting the prevalence of MN based on a range of datasets. The complex, nonlinear (NL) relationship between clinical and demographic factors was disregarded by those models, which mostly used descriptive statistics or Logistic Regression (LR) [10–11]. The uncertainty in the PM was not taken into account by other ML-based models, which merely offered point estimation of the model performance.

Traditional ML drawbacks, especially simpler ones, are generally computationally less expensive to train. 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. In HD DA, FS has become a typical method to overcome the high input dimensionality and relatively small sample sizes. Although deep learning (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 processing is necessary for DL approaches. Their main features are not entirely independent of one another, but rather have intricate relationships with one another. Optimisation is the method of selecting the best response from a variety of possible solutions because of the problems with these approaches. Determining the optimal subset while maintaining classification accuracy is the primary objective of FS technique, which ideally searches among feature subsets. It is essential to give the health sector more attention and in developing novel strategies for ED of cancer to lower the mortality rate from this disease. Conv-LSTM and FS based on MGOA were introduced in this work for TC classification.

In this paper, suggested a Conv-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 preprocessing, Min-Max normalization (or) Min-max Scaling algorithm for normalization. BST to deal with missing values (MV). Thirdly FS; suitable features are selected by using MGOA. In order to identify the TC, the most pertinent features, an ideal reduced feature subset, were chosen using MGOA.

Then data classification, Conv-LSTM to predict the thyroid nodule malignancy to achieve both better estimation and clearer interpretation. AC, 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. The suggested system and some performance metrics are then described in detail in section 3. 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. Based on the study, a novel framework is designed to address the core problems in TC detection and classification.

To predict survival in TC, ANN, LR, and Multi Layer Prediction (MLP)were utilized by Jajroudi et al. [12] . These outcomes were compared and survival in TC was predicted using ANN and LR. Data from the SEER site was gathered. Based on evidence and radiation oncologist supervision, effective features of TC have been chosen. Following data reduction (DR), 16 features were examined in 7706 samples. Survival was estimated for one, three, and five years using MLP as the selected NN. The model was evaluated using the parameters AC, SE, and SP. The results were comparable to those of statistical models, suggesting that ANN might effectively serve as an appropriate approach of survival prediction in TC patients.

A Support Vector Machine (SVM) for early TC detection and treatment was created by Jain et al. [13]. An SVM approach based on ML has been used to classify the cancer kinds in thyroid disease. One of the most effective methods in this field is the multiclass (MC) SVM. Using the SVM approach and the ML datasets supplied by the University of California Irvine (UCI), the created AI model has been assessed (trained, tested, and validated). The SVM-based final AI model exhibits higher PR, R, f1-score, and AC. One of the most common forms of cancer is TC. ED of it is quite challenging. Due to its extensive support for Python and its libraries, Google Collaboratory has been used for implementation.

An eXtreme Gradient Boosting (XGboost) and Association Rule Digging with Dynamic Thresholding and Weight Optimization (ARDdtwo), Association Rule Digging by constructing a dynamic threshold model (ARDcdt) and Self-optimizing Component Importance Measurement model (SoCIM) for prediction method for TC was suggested by Jia et al [14]. ED of the TC using a bi-dimensional substratum information mining model based on ARDdtwo. SoCIM for quantitative analysis and ARDcdt for qualitative analysis make up this integrated assessment framework. To solve the distributional bias issue, the temporal and spatial features of sparse data are considered by ARDcdt.

To further identify High-Risk Low-Frequency (HRLF), new importance diagnostic computations were created. The SoCIM can determine the relative weight of each component by assessing its level of risk in the overall system using the Risk Enhancement Level (REL) and Risk Reduction Level (RRL). It recognises the weight setting's self-adjustment and optimisation. At last, an empirical investigation was used to validate the model. The assessment of the research project demonstrates that better outcomes, including PR, f1-score, and AC, were attained. The model's total Area Under the Curve (AUC) was 0.882. This demonstrates the suggested model's suitability for real-world uses. It can lower the cost of tests and make the pathological procedure simpler for patients. The Third Xiangya Hospital submitted medical examination data for 11 provinces in 2021 for this study.

Using a large dataset of several clinical factors, a ML model for the detection of TC has been suggested by Vu et al. [15]. To predict the existence of cancer, several ML algorithms were trained and assessed. The suggested results showed a noteworthy accomplishment, with the produced system attaining a roughly greater AC. This is a significant advancement over current diagnostic techniques and highlights how ML can help physicians detect TC more accurately and effectively. To confirm the suggested model's clinical usefulness and generalisability, more improvements and prospective research are required. The suggested study demonstrates how ML techniques have great promise for enhancing patient outcomes and TC detection.

To find TC mutations early, Shah et al. [16] offer Deep (EL) Ensemble Learning (DEL) techniques such LSTM, Gated Recurrent Units (GRUs), and Bi-directional LSTM (Bi-LSTM). The model is trained with 633 samples collected from individuals of various demographics, containing 969 mutations across 41 genes. The dataset was obtained from asia.ensembl.org and IntOGen.org. Techniques such as Hahn moments, core moments, raw moments, and other matrix-based approaches are all included in feature extraction (FE). Three testing techniques are used in evaluation: the 10-Fold Cross Validation (FCV) Test (10-FCVT), the Independent Set Test (IST), and the Self-Consistency Test (SCT). Higher AC in the IST indicates that the suggested EL framework executes better. For a thorough assessment, statistical metrics such training AC, testing AC, R, SE, SP, Mathew's Correlation Coefficient (MCC), loss, training AC, F1 Score, and Cohen's kappa are used.

For TC detection, a Bi-Directional Recurrent NN (Bi-RNN) was suggested by Begum et al. [17] . Utilising data gathered from the University of California, Irvine ML repository (UCI) store, the investigation and classification models are being used in the thyroid disease. In order to understand difficult learning tasks, like therapeutic conclusion and prognosis assignments, it is essential to ensure a traditional database that may be explored and used as a cross breed demonstration. To predict the assessed risk of a patient developing thyroid disease, ML Calculations Bi-RNN was used. The True Positive (TP) Rate (TPR), False Positive (FP) Rate (TPR), AC, PR, and R are the performance metrics.

A method for ED of the TC diagnosis called Bagged Classification and Regression Trees (Bagged CART) was developed by Çiçek and Küçükakçalı [18]. The study's data set consisted of 724 patients who applied to China Median University Shengjing Hospital between 2010 and 2012. For every patient who had a thyroidectomy, the collection includes data on blood test results, US characteristics, demographics, and nodule malignancy. The modelling technique known as Bagged CART was used with this publicly available data set. F1-score, Accuracy (ACC), Balanced Accuracy (BACC), Positive Predictive Value (PPV), Negative Predictive Value (NPV), SE, and SP were the performance metrics used to evaluate the model's predictive performance. Additionally, the model's validity was assessed using the 10-FCV approach. The degree to which the input factors affect the output variable was also determined by establishing variable importance. From the outcomes, is clear that the Bagged CART model was successful in detecting TC. These findings will enable the condition's decision-making process to proceed more quickly, which will improve the effectiveness of therapeutic procedures.

For the diagnosis of TC, the SMOTE, Principal Component Analysis (PCA), StratifiedKFold (SKF), and GridSearchCV (GSCV) were utilized by Reddy et al. [19]. The class imbalance and high-dimensionality, the system incorporates SMOTE and PCA. Additionally, SKF and GSCV are utilized for optimization and validation, ensuring robust model performance. The integration of domain-specific features leveraging expert knowledge further improves model efficacy. Model performance is improved through feature engineering, comprehensive data pre-processing, and hyperparameter (HP) tuning. The ensemble approach outperforms individual classifiers, achieving a higher accuracy, F1-Score. These outcomes highlight the efficiency of advanced ensemble techniques in improving thyroid cancer diagnosis and offer a promising direction for future advancements in medical diagnostics.

To predict Differentiated TC (DTC), the SMOTE for Nominal and Continuous (SMOTE-NC), Regularised Class Association Rules (RCAR), Classification Based on Association Rules (CBAR), and Binomial LR (BLR) were suggested by Firat Atay et al. [20] . Pre-processing, classification, and DC are the suggested work steps. The first publicly available dataset includes 13 clinicopathological features and 550 DTC samples gathered over a 15-year period. Next, the class imbalance in recurrence status was addressed using SMOTE-NC. A hybrid model was created by fusing association rule mining with classification algorithms. RCAR and CBAR, two relational classification techniques, were used. Independent predictors of recurrence were examined by BLR. The model's performance was evaluated using the F1 score, PPV, NPV, AC, SE, and SP. By obtaining higher AC, SE, and F1 scores, the suggested RCAR model outperformed the CBAR model. In DTC management, the model efficiently identifies high-risk people, enabling customised follow-up tactics that may enhance patient outcomes and maximise (RA) resource allocation.

To examine a SEER database that contained clinical data from de-identified TC patients, a Fisher's discriminant ratio, Kruskal-Wallis' analysis, and Relief-F was suggested by Mourad et al. [21] . 34 distinct clinical factors, including the patient's age at diagnosis and lymph node information, were covered by the data. These clinical features were used to develop a number of novel classifiers that distinguished between patients who survived at least five years after diagnosis and those who did not.It has been shown that unspecialised and current clinical record can be consistently transformed into power of prediction to support clinicians make informed and optimised treatment decisions by properly optimising supervised NN, specifically MLP, employing information from sizable TC groups of patients (ranging from 6,756 to 20,344 for various approaches). This is because the highest AC has been achieved in distinguishing patients based on prognosis. Next, consider how the suggested ML approach might be used for various illnesses and objectives, such creating clinical trials to reveal the greatest advantages and reduce the risks of novel medication candidates for particular groups. PPV, NPV, MCC, AC, F1 score (or F-measure), and F1 score are the performance metrics for assessment.

Based on the analysis, ML algorithms drawbacks are ANN-high computational costs, Bi-RNN-increased computational complexity, Binomial LR (BLR)-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, a Conv-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 MGOA and Conv-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 Min-Max normalization (or) Min-max Scaling. By utilizing BST, an effective data balancing method, for resolving the data imbalance issue. Then, to identify the TC, the most relevant features (an ideal reduced feature subset) were chosen using FS and MGOA.Next data classification, Conv-LSTM to predict the thyroid nodule malignancy to achieve both better estimation and clearer interpretation. At last, the PR, R or SE, SP, AC, and AUROC are used to evaluate performance. The applicability of the suggested Conv-LSTM technique has been evaluated and compared with publicly available benchmark data from the Kaggle repository.

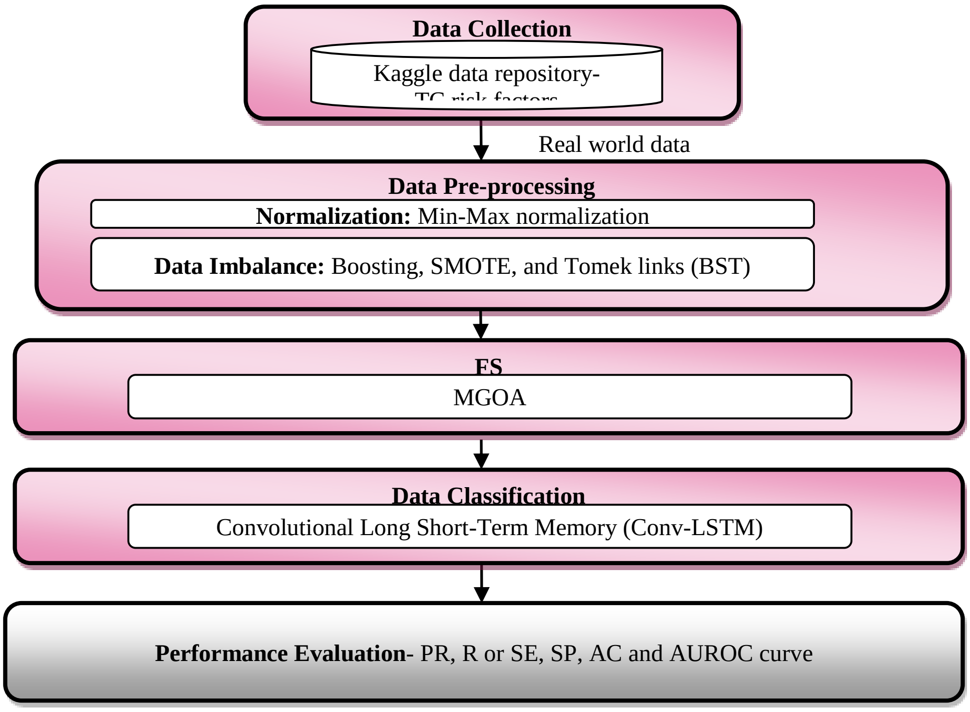

FIGURE 1. FLOW CHART OF SUGGESTED SCHEME

3.1 DC

TC risk prediction dataset is collected from Kaggle online repository [22]. 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 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 step 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, MV, and class 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, some common techniques can be employed, such BST. Data must also be standardised or normalised before DL techniques can be used. MMN (or Min-max Scaling) is used to perform the normalisation.

3.3 Normalization

In the context of statistical approaches, normalisation is the process of rescaling the features of numerical data to a predetermined range. Normalisation attempts to place the values of different variables onto a comparable scale in order to give reliable comparisons and prevent features with larger scales from predominating those with lower scales during analysis [23]. Numerous methods, such as robust scaling, Z-score normalisation (ZSN), and Min-Max scaling, are used to conduct normalisation. DN stands for when a dataset's risk factors have varying values. The dataset is normalised using MMN.

Min-Max Scaling (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, is employed. Each feature is rescaled to a range with a minimum and maximum value using this method [24]. 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 X_n. The original value of the feature is denoted as X. The minimum value of the feature in the dataset is denoted by $X_{\min}$. The maximum value of the feature in the dataset is denoted by $X_{\max}$.

3.3.1 Data Imbalance

When there are substantially differing numbers of examples for various classes in a dataset, this is known as data imbalance. A significant difference between the number of occurrences in one class and the other class is known as data imbalance. Data imbalance issue using an effective data balancing method Boosting, SMOTE, and Tomek links (BST).

BST

Because minority classes are under-represented and do not have enough examples to fully capture complex patterns, they frequently face misclassification problems. Therefore, the BS was first applied across I iterations to the imbalanced dataset, represented by D. In this phase, D is trained using samples that have identical weights $(1/\mathrm{N})$, and the Learning Rate (LR) is calculated. and number of samples are denoted as N. The LR is then utilized to increase the weights for minority class samples, ensuring that in subsequent stages, greater emphasis is placed on minority instances. This deliberate weighting helps to produce more varied synthetic instances and improves the minority class's representation. After adjusting the weights, applied the SMOTE technique to D.

This technique computes the imbalance ratio as M/S. Here, the number of samples are denoted as S and the number of minority classes are denoted as M. Then, it classifies the K-Nearest Neighbors (KNN) from M and randomly selects these neighbors. The imbalance ratio M/S is computed using this method. Here, the number of minority samples are denoted as M, and the samples count are denoted as S. It then selects these neighbours at random after determining which K-Nearest Neighbours (KNN) are present in M. The newly generated synthetic instances are then incorporated into the Augmented Dataset (AG).

However, there is a potential drawback in this process, as the inclusion of noisy and irrelevant synthetic instances may introduce high complexity and hinder result reproducibility. Therefore, in the final stages, address these concerns by applying TL to the AG. In this stage, once again identify KNN from both majority and minority samples in AG [25]. This entails figuring out the Euclidean Distance (ED) between each instance in AG and the feature vector, then selecting the samples with the shortest distances from both classes.

After that, find and eliminate the majority class data samples that are most similar to the minority class data. Due to the removal of irrelevant and noisy samples, these approaches greatly minimize the complexity of AG.

3.4 FEATURE SELECTION

The data pre-processing is completed TC dataset taken as input for MGOA based FS.With FS, a subset of the original features that are most pertinent to the issue at hand are chosen. Reducing the dataset's dimensionality while keeping its most significant attributes is the aim. It entails picking particular dataset features to be used in a learning algorithm's training procedure. MGOA based FS technique are employed to measure the effectiveness of features in classifying data.

MGOA

For detection purpose, the FS is utilized. The AC of the framework is upgraded by the FS. The natural foraging and swarming behaviors of the grasshoppers (GH) are simulated by the GOA algorithm, and this method is employed for the optimal selection of features from the TC dataset [26]. The crop production was damaged by these GH, and it is not suitable for processing or consumption.

In nature, nymphs, or immature forms, and adults, or fully grown forms are the 2 different forms of GH. The distinctive feature of the GH swarm is that it exhibits swarming behaviour in both nymphs and adults. The TC dataset's features are selected using GH swarming. The following equation (2) describes the X_i, which is the i^{th} GH location for FS.

$$X_i = SO_i + GR_i + AD_i \qquad (2)$$

Here, social interaction is denoted by SO_i. The gravitational force that i^{th} GH experiences are denoted as GR_i. Wind advection is AD_i. The random behaviour of GH is given by equation (2). The primary element that the GH itself has developed is SO_i. Equation (3) is the mathematical expression for it.

$$SO_i = \sum_{j=1, j\neq 1}^{N} s(d_{ij})\acute{d}_{ij} \qquad (3)$$

Here, $d_{ij} = |xj - xi|$ denotes the distance between the i^{th} and j^{th} GH. A unit vector from the i^{th} GH to the j^{th} GH is indicated by ($\acute{d_{ij}} = \frac{x_j - x_i}{d_{ij}}$. Equation (4) represents the function s, that can be employed for indicating the strength of social pressures.

$$s(r) = \frac{fe - r}{l - e - r} \tag{4}$$

Here, the attractive length scale is denoted as l, and f is the intensity of attraction.

Equation (5) below illustrates the calculation of the GR_i component described in Equation (2).

$$GR_i = -geg \tag{5}$$

Here, the gravitational constant is denoted as g. A unit vector directed towards the globe centre is denoted as eg.

Equation (6) illustrates the calculation of the AD component in Equation (2).

$$AD_i = uew \tag{6}$$

Here, a constant drift is denoted by u. A unit vector with the similar direction as that of the wind is denoted as ew. Due to their lack of wings, the nymphs' motions will inevitably be closely tied to the wind's direction.

Equation (7) is obtained by substituting the elements $SO, GR\ and\ AD$ in Equation (2).

$$Xi = j = 1, jiN\ s\left(d_{ij}\right)\left(d_{ij}\right) - geg + uew \tag{7}$$

Here, $s(r) = fe - r/l - e - r$. The number of GH is N. Only swarms may use this concept in the free space. One can use Equation (6) to simulate the interaction among GH in a swarm. GOA has certain disadvantages. (1) The outcomes of the study show that local optimal stagnation and insufficient global exploration remain its shortcomings. (2) Because it prevents the algorithm from investigating and taking advantage of the SS close to the solution, equation (7) is inapplicable to swarm simulation and optimisation algorithms. An optimisation algorithm must conduct constructive exploration and exploitation in order to determine the precise approximation of the global optimum. Equation (8) below illustrates the improved version of MGOA [27].

$$Xid = cj = 1, jiNcubd - lbd2s(|xjd - xid|)xj - xjdij + Td \tag{8}$$

Here $s(r) = fe - r/l - e - r$, the dth dimension's upper bound is denoted as ubd. In the dth dimension, lbd is the lower bound. The value of the target's dth dimension (best Feature solution) is Td. To lessen the comfort, attraction, and repulsion regions, use c = decreasing coefficient. The SO component indicated in Equation (2) is equal to s.

As the iteration counter increases, the outer c is in charge of decreasing the search coverage. The decrease of the attraction or repulsive forces between the GH is the responsibility of the inner c. The exploration and exploitation (E-E) space of the GH is linearly reduced by $cubd - lbd2$. If the GH should explore is shown by $s(|xjd - xid|)$. Habitual attraction of the GH to the food source is denoted by Td. To balance E-E, the parameter c is decreased proportionately to the iteration count. Equation (9), which expresses how the coefficient c reduces the comfort region proportionately with the iteration count,

$$C = cmax - lcmax - cmin * Maxitr \tag{9}$$

Here, the maximum value is denoted as $cmax$. The minimal value is $cmin$. current iteration is denoted as l. The maximum number of iterations is denoted by $Maxitr$. The following is a description of the suggested MGOA algorithm:

```
Intialize $c_{\max,}c_{\min}$ and $Maxitr$
Set population of swarms $X_i(i = 1,2,..,n)$ randomly
Calculate fitness of every solution by Conv-LSTM classifier
Fix T as the optimal solution
While $(t < Maxitr)$
    Update c with Eqn (9)
    for every solution
        Normalize the separation among GH in [1,4]
        Update the location of present solution by Eqn (8)
        Reset the current solution if it violates the boundary of SS
```

```
  end for
  Update T if there is an optimal solution in the population
  $t = t + 1$
end while
To create $c$ classifiers
for $i = 1\ to\ c$
    Randomly sample the training data $D$ with replacement to produce $D_i$
    Generate a root node, $N_i$ containing $D_i$
    Call BulidTree($N_i$)
end for
BulidTree($N$):
if N contains instances of 1 class then
    return
else
Randomly choose number of probable splitting features in N
Choose the feature F
Create f child nodes of $N, N_1, \ldots, N_f$ where F has f probable values $(F_1, \ldots\ldots, F_f)$
for $i = 1\ to\ f$
        Fix the contents of $N_i$ to $D_i$ is all the samples in $N$ that match
        $F_i$
        Call BulidTree($N_i$)
end for
end if
Find the best feature
End
```

3.5 DATA CLASSIFICATION

The selected number of features taken as input to Conv-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.Conv-LSTM algorithm is to predict the thyroid nodule malignancy to achieve both better estimation and clearer interpretation.

LSTM

A memory cell 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. The fundamental data pattern that fixes the long-term data dependency among data that arises in traditional RNN algorithms can be extracted and retained. According to LSTM, after executing certain operations on the prior memory content c_{t-1}, it produces current memory content ct and current cell state ht as final outputs. It accepts bias b, current input vector x_t, and previous cell state h_{t-1}as inputs.

The amount of data from the previous memory cell that should be ignored is determined by the value that the f generates in the range of 0 to 1 [28]. A value near 0 means that the majority of the information from the prior timestamp will be forgotten at the present timestamp, whereas a value near 1 suggests the reverse. Equation (10), which defines the operation of f,

$$f_t = \sigma_g(w_f x_t + u_f h_{t-1} + b_f) \tag{10}$$

The amount of input from the current timestamp that is added to the memory cell is controlled by this i. Eqn 11 express it as follows,

$$i_t = \sigma_g(w_i x_t + u_i h_{t-1} + b_i) \tag{11}$$

By taking into account the output of the f and i, which is specified by equation (12), the c oversees the process of updating memory cell content from c_{t-1} to c_t.

$$c_t = f_t \times c_{t-1} + i_t \times \sigma_h(w_c x_t + u_c h_{t-1} + b_c) \tag{12}$$

The o generates the current timestamp's final output. The cell state h_{t-1} is also updated to h_t by this o. Equations (13–14) define the final output.

$$o_t = \sigma_g(w_o x_t + u_o h_{t-1} + b_o) \tag{13}$$

$$h_t = o_t \times \sigma_h(c_t) \tag{14}$$

In this case, the sigmoid function is denoted as σ_g and hyperbolic tangent function are indicated by σ_h. The vanishing gradients problem (VGP) is avoided by using the weight values w and u. A dense layer serves as the output layer (OL), performing the final classifications after the inputs are sent to an input layer (IL) and then passed through two hidden layers (HL). There are 100 LSTM units in each layer of the model. These 100 LSTM units store the data pattern as the inputs, that are passed through them.

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 dense layer ultimately determines whether a patient has TC. When the output gate of the LSTM is closed, it is unable to access the contents of its prior memory cell. Conv-LSTM addresses this problem by introducing an additional connection. Conv-LSTM referred to as a peephole connection, across every gate and the prior memory content.

Conv-LSTM

Conv-LSTM is an LSTM variation with a peephole connection. Even when the output gate is closed, Conv-LSTM permits all of the gates to use the contents of the prior memory cell. Figure 2 displays the Conv-LSTM diagram. In this case, each gate accepts an extra parameter (c_(t-1)) as input to guarantee that previous memory cells continue to have an impact even when the o_t is closed [29]. Even with a lengthy input sequence, this relationship ensures that the earliest input will have an impact. Equations (15–19) describe how the Conv-LSTM's four gates operate.

$$f_t = \sigma_g(w_f * x_t + u_f * h_{t-1} + v_f \times c_{t-1} + b_f \tag{15}$$

$$i_t = \sigma_g(w_i * x_t + u_i * h_{t-1} + v_i \times c_{t-1} + b_i \tag{16}$$

$$c_t = f_t \times c_{t-1} + i_t \times \sigma_h(w_c x_t + u_c * h_{t-1} + b_c) \tag{17}$$

$$o_t = \sigma_g(w_o * x_t + u_o * h_{t-1} + v_o \times c_{t-1} + b_o \tag{18}$$

$$h_t = o_t \times \sigma_h(c_t) \tag{19}$$

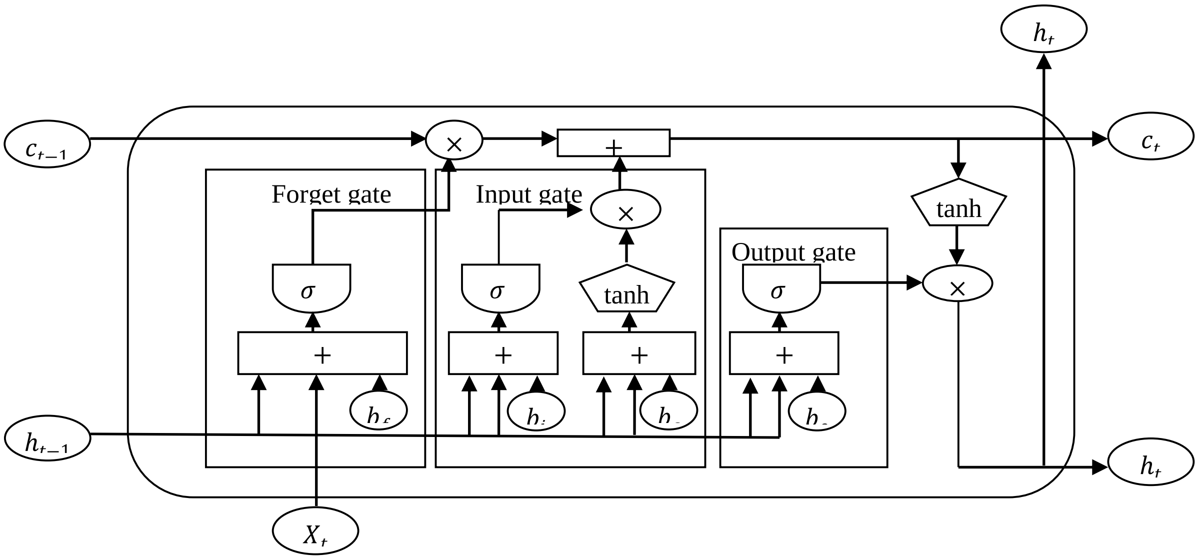

FIGURE 2. CONVOLUTIONAL LSTM (CONV-LSTM)

4. RESULTS AND DISCUSSION

This part discusses the outcomes of several classifiers experimented on the TC risk prediction dataset is collected from Kaggle online repository [22]. It simulates real-world TC risk factors and has 212,691 records with 23 features.

5. 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. Based on the combinations of actual class and predicted class, samples for binary classification (BC) problems is classified into true positive (tp), false positive (fp), true negative (tn), and false negative (fn).

TABLE 1. CM IN THIS RESEARCH

Actual Class	**Prediction Class**	
	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, AC and AUROC curve as shown by equations (20, 21, 22, 23, & 24). PR is an incorrect positive diagnosis could subject a patient to unnecessary treatments. Similarly, R or SE is paramount to avoid missing a thyroid cancer disease diagnosis, which can have serious health implications. SP to ensure that those without the disease are not falsely diagnosed, thus preventing unnecessary treatments. AC, 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 (20) provides its definition.

$$\text{Precision(PR)} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{20}$$

The model's capacity to accurately identify thyroid in living individuals is indicated by R or SE. Equation (21) defines a SE of 100%, which means that the algorithm chosen correctly predicted each instance of TC.

$$\text{Recall(R)/Sensitivity(SE)} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{21}$$

Equation (22) defines SP. SP assesses the model's capacity to identify individuals who will never experience thyroid problems.

$$\text{Specificity(SP)} = \frac{\text{tn}}{\text{tn} + \text{fp}} \tag{22}$$

Equation (23) defines AC. The percentage of all predictions that the model accurately predicted is measured by AC.

$$(\text{AC}) = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \tag{23}$$

The True Positive Rate (TPR) against False Positive Rate (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 (24) illustrates that TPR and FPR,

$$\text{FPR} = \frac{\text{fp}}{\text{fp+fn}}, \text{TPR} = \frac{\text{tp}}{\text{tp+fn}} \tag{24}$$

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

5.1 SIMULATION EXPERIMENT

Then, various existing techniques are Bi-RNN, and Random Forest (RF), have been suggested for results comparison of suggested Conv-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	**Bi-RNN**	**RF**	**Conv-LSTM**
Precision (PR)	80.5	84.6	85.8
Recall (R) or Sensitivity (SE)	81.6	83.4	86.2
Specificity (SP)	65.1	70.7	73.2
Accuracy (AC)	80.2	84.3	88.5
Area Under Receiver Operating Characteristic (AUROC)	81.3	84.5	86.2

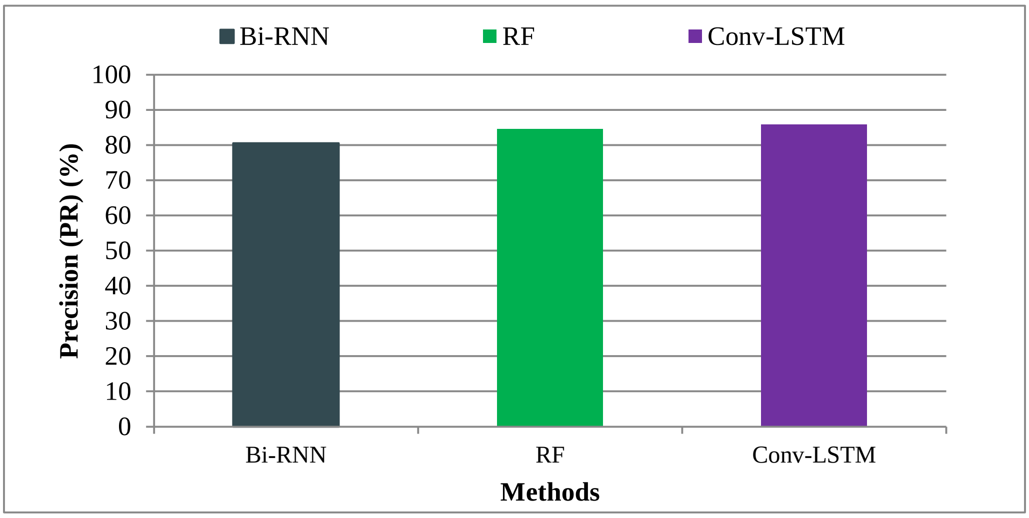

FIGURE 3. PR COMPARISON VS. CLASSIFIERS

Figure 3 shows the PR performance metric comparison between existing classifiers like Bi-RNN, RF and proposed Conv-LSTM for thyroid cancer detection. The classification methods are indicated in the X-axis and the y-axis represents the PR results. Suggested work using BST for data imbalance and it increases the precision outcomes. From the outcomes it is concluded that the suggested Conv-LSTM classifier delivers highest PR outcomes of 85.8% and the existing Bi-RNN, and RF has lowest precision of 80.5%, and 84.6% (Refer Table 2).

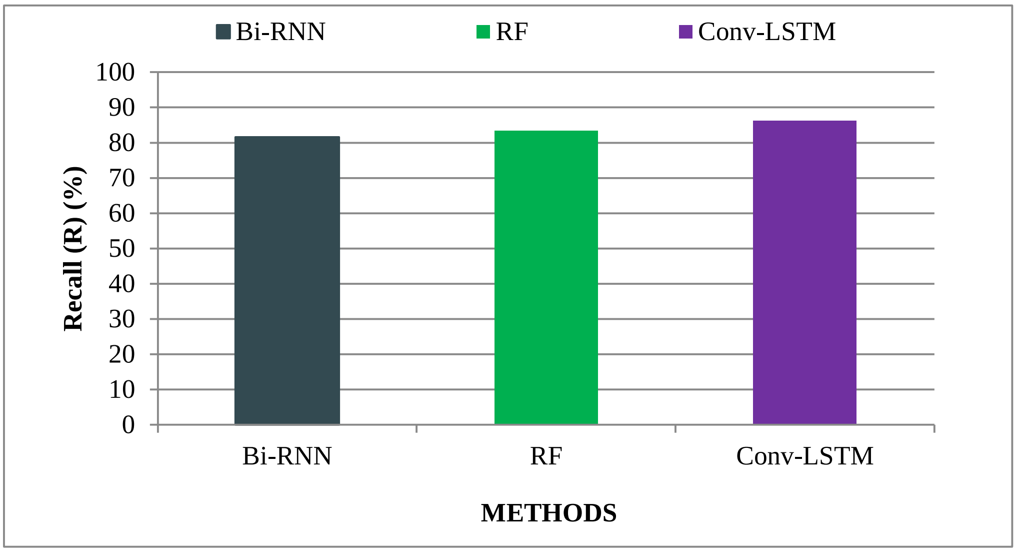

FIGURE 4. RECALL (R) COMPARISON VS. CLASSIFIERS

The performance comparison outcomes for the current classifiers like Bi-RNN, RF and proposed Conv-LSTM for thyroid cancer detection in terms of recall (R) are shown in Figure 4. 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 Conv-LSTM classifier delivers highest R outcomes of 86.2% and the existing Bi-RNN, and RF has lowest recall of 81.6%, and 83.4% (Refer Table 2).

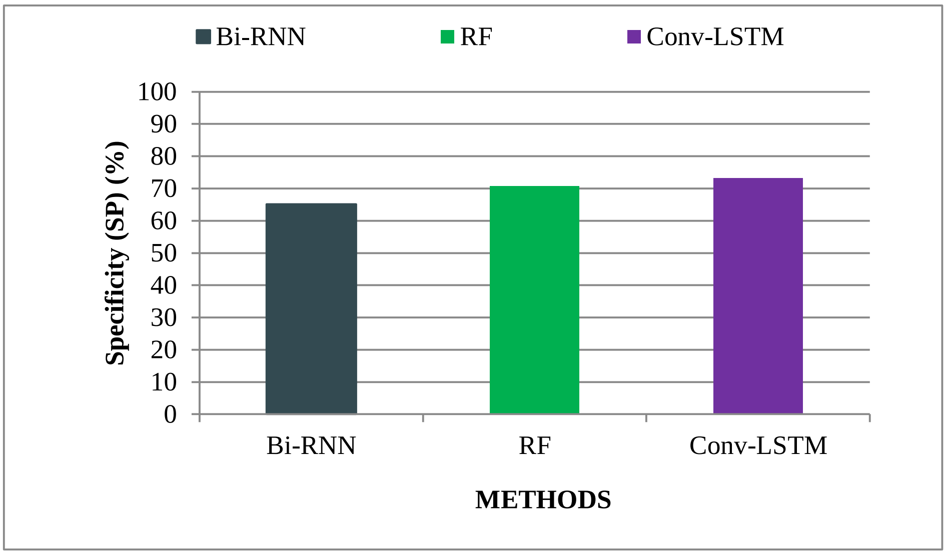

FIGURE 5. SPECIFICITY (SP) COMPARISON VS. CLASSIFIERS

Specificity (SP) performance metric comparison between existing classifiers like Bi-RNN, RF and proposed Conv-LSTM for thyroid cancer detection are shown in figure 5. 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 Conv-LSTM classifier delivers highest specificity results of 73.2% while the existing Bi-RNN, and RF has lowest specificity of 65.1%, and 70.7% (Refer Table 2).

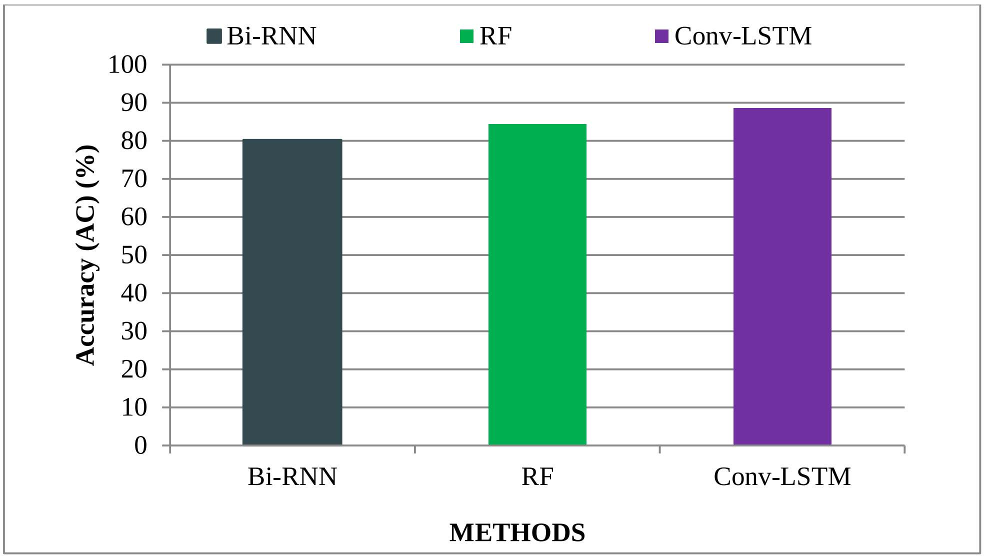

FIGURE 6. ACCURACY (AC) COMPARISON VS. CLASSIFIERS

Figure 6 shows the Accuracy (AC) performance metric comparison between existing classifiers like Bi-RNN, RF and proposed Conv-LSTM for thyroid cancer classification. In the above figure, the classification techniques are shown on the X-axis, and the AC results are shown on the Y-axis. From the results it is concluded that the suggested classifier Conv-LSTM delivers highest AC outcomes of 88.5% while the existing Bi-RNN, and RF has lowest accuracy of 80.2%, and 84.3% (Refer Table 2).

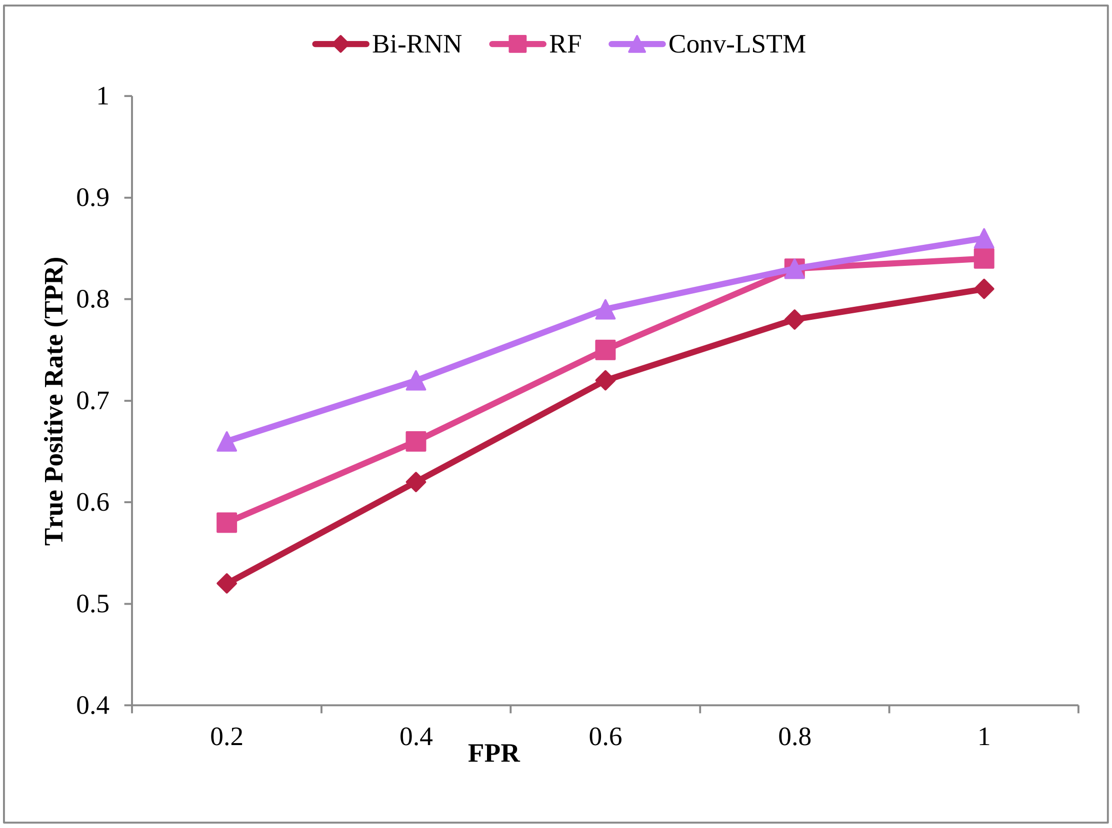

FIGURE 7. AUROC COMPARISON VS. CLASSIFIERS

Figure 7 shows the AUROC performance metric comparison between existing classifiers like Bi-RNN, RF and proposed Conv-LSTM for thyroid cancer 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 Conv-LSTM delivers highest AUROC results while the existing Bi-RNN, and RF has lowest AUROC (Refer Table 2).

6. 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. These factors have made it more crucial than ever for the medical field to identify TC abnormalities. For clinical treatment and therapy, it is vital to accurately predict how a disease will progress and to understand the interactions between clinical features. 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 Conv-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) Min-max Scaling for normalization, and then data imbalance using BST. Finally, the most pertinent features (an ideal reduced feature subset) were chosen using FS and MGOA in order to identify the TC. Then data classification, Conv-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, Specificity SP, AC, and AUROC curve. Outcomes indicate that the suggested model achieves 88.5% accuracy which is better than other current frameworks. The simulation's results show that the Conv-LSTM model overtakes other frameworks for TC prediction in most cases. Additionally, (1) this research can be expanded using a self-developed dataset. (2) A meta classifier algorithm that boosts classification accuracy by combining the new ensemble model with bagging, stacking, voting, and boosting techniques; (3) 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] 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.

[4] 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.

[5] Chen, D.W., Lang, B.H., McLeod, D.S., Newbold, K. and Haymart, M.R., 2023. Thyroid cancer. *The Lancet*, *401*(10387), pp.1531-1544.

[6] Lamartina, L., Grani, G., Durante, C., Filetti, S. and Cooper, D.S., 2020. Screening for differentiated thyroid cancer in selected populations. *The lancet Diabetes & endocrinology*, *8*(1), pp.81-88.

[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] Banu, G.R., 2016. A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease. *International Journal of Computer Sciences and Engineering*, *4*(11), pp.64-70.

[9] Hirsch, D., Levy, S., Tsvetov, G., Gorshtein, A., Slutzky-Shraga, I., Akirov, A., Robenshtok, E., Shimon, I. and Benbassat, C.A., 2017. Long-term outcomes and prognostic factors in patients with differentiated thyroid cancer and distant metastases. *Endocrine Practice*, *23*(10), pp.1193-1200.

[10] Hou, C.J., Wei, R., Tang, J.L., Hu, Q.H., He, H.F. and Fan, X.M., 2018. Diagnostic value of ultrasound features and sex of fetuses in female patients with papillary thyroid microcarcinoma. *Scientific reports*, *8*(1), pp.1-6.

[11] Saito, D., Nakajima, R. and Yasuda, S., 2020. Examination of malignant findings of thyroid nodules using thyroid ultrasonography. *Journal of Clinical Medicine Research*, *12*(8), pp.499-507.

[12] Jajroudi, M., Baniasadi, T., Kamkar, L., Arbabi, F., Sanei, M. and Ahmadzade, M., 2014. Prediction of survival in thyroid cancer using data mining technique. *Technology in cancer research & treatment*, *13*(4), pp.353-359.

[13] Jain, M., Singh, M.N., Somani, V. and Kumar, A., 2022. Technique to Identify & Classify Thyroid Cancer Using Supervise & Supervise Learning. *Journal of Algebraic Statistics*, *13*(3), pp.4512-4520.

[14] Jia, Z., Huang, Y., Lin, Y., Fu, M. and Sun, C., 2023. Multidimensional Prediction Method for Thyroid Cancer Based on Spatiotemporally Imbalanced Distribution Data. *IEEE Access*, *12*, pp.4674-4686.

[15] Vu, T.A., Huyen, N.A., Huy, H.Q. and Huong, P.T.V., 2023. Enhancing Thyroid Cancer Detection Through Machine Learning Approach. In *2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp.188-193.

[16] Shah, A.A., Daud, A., Bukhari, A., Alshemaimri, B., Ahsan, M. and Younis, R., 2024. DEL-Thyroid: deep ensemble learning framework for detection of thyroid cancer progression through genomic mutation. *BMC Medical Informatics and Decision Making*, *24*(1), pp.1-15.

[17] Begum, M.A., Tresa, I.M., Sandhya, S., Vidhya, S. and Vinodhini, G., 2021. Machine learning based dysfunction thyroid cancer detection with optimal analysis. *Turkish Journal of Computer and Mathematics Education*, *12*(7), pp.818-823.

[18] Çiçek, İ.B. and Küçükakçalı, Z., 2023. Machine learning approach for thyroid cancer diagnosis using clinical data. *Middle Black Sea Journal of Health Science*, *9*(3), pp.440-452.

[19] Reddy, N.N., Maddula, S., Deepika, P., Sanjana, R., Nipun, L. and Shilpa, T., 2025. Enhancing thyroid cancer diagnosis with ensemble machine learning techniques. In *Recent Trends in VLSI and Semiconductor Packaging*, pp.421-430.

[20] Firat Atay, F., Yagin, F.H., Colak, C., Elkiran, E.T., Mansuri, N., Ahmad, F. and Ardigò, L.P., 2024. A hybrid machine learning model combining association rule mining and classification algorithms to predict differentiated thyroid cancer recurrence. *Frontiers in Medicine*, *11*, pp.01-10.

[21] Mourad, M., Moubayed, S., Dezube, A., Mourad, Y., Park, K., Torreblanca-Zanca, A., Torrecilla, J.S., Cancilla, J.C. and Wang, J., 2020. Machine learning and feature selection applied to SEER data to reliably assess thyroid cancer prognosis. *Scientific reports*, *10*(1), pp.1-11.

[22] https://www.kaggle.com/datasets/ankushpanday1/thyroid-cancer-risk-prediction-dataset

[23] 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.

[24] 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.

[25] Wang, Z.H.E., Wu, C., Zheng, K., Niu, X. and Wang, X., 2019. SMOTETomek-based resampling for personality recognition. *IEEE access*, *7*, pp.129678-129689.

[26] Saremi, S., Mirjalili, S. and Lewis, A., 2017. Grasshopper optimisation algorithm: theory and application. *Advances in engineering software*, *105*, pp.30-47.

[27] Goel, N., Grover, B., Gupta, D., Khanna, A. and Sharma, M., 2020. Modified grasshopper optimization algorithm for detection of autism spectrum disorder. *Physical Communication*, *41*, pp.1-11.

[28] Rahman, M.M. and Siddiqui, F.H., 2019. An optimized abstractive text summarization model using peephole convolutional LSTM. *Symmetry*, *11*(10), pp.1-24.

[29] Rahman, M., Islam, D., Mukti, R.J. and Saha, I., 2020. A deep learning approach based on convolutional LSTM for detecting diabetes. *Computational biology and chemistry*, *88*, pp.1-10.