Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Prediction of the Thyroid Cancer using the Deep Learning Based Hybrid Spatial Convolution based LSTM Network (SCBLN)

[1] R.Vanitha, [2] Dr.K. Perumal
[1]*Research Scholar, Department of Computer Applications, Madurai Kamaraj University, Madurai.*
[1] *Email: vanithachezian2004@gmail.com*
[2]*Professor, Department of Computer Applications Madurai Kamaraj University, Madurai.*
[2] *Email: perumalmkucs@gmail.com*

ARTICLE INFO

Received: 10 Nov 2024

Revised: 25 Dec 2024

Accepted: 22 Jan 2025

ABSTRACT

The prediction of thyroid cancer has become a significant task in recent years. Binary classification is frequently the objective, despite the existence of extant diagnostic methods; however, the datasets employed are of limited size, and the results are not validated. Model optimization is the primary focus of current methodologies, while the feature engineering component is not as extensively investigated. In order to circumvent these constraints, this study introduces a method that examines feature engineering for deep learning models. Hashimoto's thyroiditis (primary hypothyroidism), autoimmune thyroiditis (compensated hypothyroidism), binding protein (increased binding protein), and non-thyroidal syndrome (NTIS) (concurrent non-thyroidal illness) can be predicted using the SCBLN Hybrid Spatial Convolution Based LSTM Network(SCBLN) approach. The dataset was initially obtained and can be processed using the normalization strategy. After that, the principal component analysis(PCA) method can be employed to extract the cancer-related features. And finally, the thyroid cancer can be predicted by employing the SCBLN classifier. The overall experimentation was carried out under python environment. According to extensive experiments, the SCBLN classifier achieves the highest accuracy and F1 score, with a 0.99 score. Results indicate that the SCBLN model is a superior option for the detection of thyroid cancer in terms of both computational complexity and accuracy. A comparison of the SCBLN approach's performance to existing studies confirms its superiority.

Keywords: Deeplearning, Thyroid Prediction, Hybrid Spatial Convolution based LSTM Network .

I. INTRODUCTION

In recent years, there has been an increase in the number of cases of thyroid cancer. One of the most critical functions of the thyroid gland is to regulate metabolism. Different abnormalities can result from thyroid gland irregularities; hyperthyroidism and hypothyroidism are among the most prevalent. A significant number of individuals are diagnosed with thyroid cancers each year, including hyperthyroidism and hypothyroidism [1]. Hypothyroidism and hyperthyroidism may result from insufficient thyroid hormones, which are produced by the thyroid gland: levothyroxine (T4) and triiodothyronine (T3) [2]. In the literature, numerous methods are suggested for the diagnosis of thyroid cancer. It is imperative to forecast thyroid cancer in advance in order to administer the appropriate treatment to the patient at the appropriate time, thereby reducing medical expenses and saving human lives. As a result of technological advancements in data processing and computation, machine learning and deep learning techniques are employed to predict thyroid diagnoses in the early phases and classify thyroid cancer categories, including hypothyroidism and hyperthyroidism. The healthcare domain has benefited from the utilisation of technology in numerous healthcare areas to promote human well-being, as a result of the advancement in technologies such as data mining, big data, image and video processing, and parallel computing [3]. Data mining-based health care applications may encompass the early detection of maladies and diagnosis, prediction of virus outbreaks, drug discovery and testing, health care data administration, and patient personalised medicine recommendations, among others [4]. Health care professionals endeavour to detect

cancers in their early stages to ensure that patients receive appropriate treatment and the cancer is cured in a timely manner with minimal expenditure. Thyroid cancer is one of the cancers that affects a significant portion of the global human population. The American Thyroid Association, the world's foremost professional organisation, estimates that 20 million Americans are affected by some form of thyroid cancer [5]. At least once in their lifespan, twelve percent of the United States population is diagnosed with a thyroid condition. These statistics indicate that thyroid-based cancers should not be dismissed carelessly. It is highly desirable to enhance the health care practices to detect and prevent thyroid cancers through the use of advanced technologies. While there are only a handful of multiclass-based detection works, the majority of existing research is focused on binary classification problems, in which subjects are classified as either thyroid patients or healthy individuals. Nevertheless, the emphasis is on three categories: normal, hypothyroidism, and hyperthyroidism. The optimisation of machine learning and deep learning models is the primary focus, while the feature selection aspect of thyroid cancer is either completely ignored or incompletely investigated. Although these reporting methods are highly accurate, they are only evaluated on samples that are less than 1000 in number, and the results are not validated. In order to effectively forecast the patient's thyroid condition and proactively manage the patient, it is desirable to categorise the patient based on their status, such as their health condition, treatment condition, and general health issues. The objective of this investigation is to address these concerns and provides the subsequent contributions.

- A novel deep learning-based approach to the prediction of thyroid cancer is SCBLN, with a particular emphasis on the multi-class problem. In contrast to prior research that concentrates on the binary or three-class problem, this study examines a five-class cancer prediction problem.
- The efficacy of the PCA feature elimination approach for the problem at hand is the subject of this study.

The subsequent sections of this article are structured as follows. Section 2 delves into the most recent research on the detection and classification of thyroid cancers. The SCBLN methodology to resolve the issue of thyroid cancer prediction is detailed in Section 3. The experimental results of our study and their comparison with prior art studies are detailed in Section 4. Our contributions conclude the article in Section 5.

II. RELATED WORKS

Recent technical developments in data processing and computing have facilitated the use of machine learning and deep learning methods in several research papers focused on predicting thyroid illness. Early detection and accurate categorization of this illness as either cancer, Hypothyroidism, or Hyperthyroidism are beneficial for prompt medical intervention and recuperation. The literature survey is conducted by using peer-reviewed article databases such as Google Scholar and Scopus. The searches were conducted within the timeframe of the last five years to discover the most current research relevant to our topic. The relevant articles were selected using combinations of the keywords "Thyroid cancer", "Thyroid cancer", "machine learning", and "deep learning". Due to the large amount of search results, we have refined the search queries and implemented a stringent keyword search to discover the most relevant articles. In all, we discovered over 100 relevant publications throughout our first screening process. After doing a more thorough analysis, we have identified and selected some papers that are very pertinent to our research. Machine learning and deep learning techniques are used for the diagnosis of both thyroid illness and thyroid cancer. Due to the distinct nature of the application procedure for each activity, they are addressed individually. The research [6] used the least absolute shrinkage and selection operator (LASSO) and LR model to identify the ultrasonic properties linked with cancerous thyroid nodules. Next, a Random Forest (RF) algorithm is used in conjunction with a score system to categorise the cancerous thyroid nodules. The logistic lasso regression (LLR) with random forest (RF) had the highest performance, with an accuracy of 82%. A separate investigation [7] used machine learning to predict the presence of the BRAF mutation in verified cancerous thyroid nodules. For this investigation, the researchers chose 96 ultrasonic pictures of thyroid nodules. A total of 86 radiomic characteristics were identified from the pictures. Subsequently, three models, namely Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), were used to predict the existence of the BRAF mutation. The classification accuracy for all three models is given as 64.3%. Idarraga et al. [8] used machine learning techniques to predict the malignancy of thyroid nodules. They used ultrasonic and fine-needle aspiration (FNA) features to prevent incorrect negative diagnoses in the first phases of thyroid cancer. The RF approach outperformed other techniques such as decision tree (DT) and gradient descent (GD). The performance of all the aforementioned studies is suboptimal for

predicting thyroid cancer diagnosis and may yet be improved. Various methods for detecting and categorising thyroid cancers have been described in the literature. Garcia et al. [9] used machine learning methods such as RF, LR, GBM, SVM, and deep neural networks (DNN) to identify the compounds that are likely to initiate thyroid hormone homeostasis. Early molecular prediction is advantageous for further testing during the first phases of thyroid illness. The molecular events were derived from ToxCast databases for conducting the studies. According to the research, Thyroid Peroxidase (TPO) and Thyroid Hormone receptor (TR) had the highest level of predictive accuracy, with F1 scores of 0.83 and 0.81, respectively. The authors in [10] used image processing techniques and feature selection approaches to identify the crucial elements from the dataset and get optimal performance in predicting thyroid illness. The categorization of thyroid cancers is a prominent issue that has to be addressed in the healthcare sector. Razia et al. [11] conducted a comparative analysis of several machine learning algorithms to categorise Thyroid illness into normal, Hypothyroidism, or hyperthyroidism groups based on their performance. The authors acquired the datasets from the University of California Irvine (UCI) machine learning library. The dataset has a total of 7200 samples, with each sample possessing 21 distinct properties. The authors found that DT achieved superior performance compared to SVM, NB, and multilinear regression (MLR) with an accuracy of 99.23%. Nevertheless, multi-classification is limited to just three categories, and there is insufficient information about data preparation to evaluate the suitability of the findings for real-time datasets. The work [12] introduces a multi-kernel Support Vector Machine (SVM) for the classification of thyroid disorders. According to the authors, the multi-kernel SVM attained a performance accuracy of 97.49% on the UCI thyroid datasets. The enhanced grey wolf optimisation algorithm conducts feature selection and improves efficiency. A research [13] conducted a multiclass analysis of hypothyroidism utilising specific characteristics and machine learning methods. Hypothyroidism is categorised into four distinct classifications. The findings indicate that the RF algorithm achieved a high level of performance with an accuracy of 99.81%, surpassing the SVM, KNN, and DT algorithms. Nevertheless, the authors failed to address the performance of their suggested technique in classifying thyroid cancers. A separate investigation [14] conducted an experiment to evaluate the effectiveness of three feature selection techniques in combination with Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Logistic Regression (LR), and Naive Bayes (NB) for the purpose of early detection of hypothyroidism. Three feature selection approaches, namely recursive feature selection (RFE), univariate feature selection (UFS), and principal component analysis (PCA), are evaluated in conjunction with machine learning (ML) algorithms. The combination of Recursive Feature Elimination (RFE) and Machine Learning (ML) algorithms outperformed previous approaches of feature selection. When the RFE feature selection method was used in combination with five different machine learning algorithms, all of them achieved an accuracy of 99.35%. Nevertheless, the dataset is very limited, including just 519 entries. They need a comprehensive dataset to assess the efficacy of their approach. The authors [15] assessed the efficacy of several machine learning algorithms in classifying thyroid cancers. SVM, RF, DT, NB, LR, KNN, and MLP are used for illness prognosis. A dataset of 1250 samples is collected from hospitals and labs in Iraq. The Multilayer Perceptron (MLP) had a 96.4% accuracy in predicting the categorization of the thyroid. Nevertheless, there is still potential for enhancing performance. Hosseinzadeh et al. [16] introduced an approach called multiple multi-layer perception (MMLP) to diagnose thyroid disorders. When the Multilayer Perceptron (MLP) is combined with a set of six networks using the Multimodal Learning Platform (MMLP), the accuracy is enhanced by 0.7% in comparison to using a single MLP alone. Despite achieving a 99% classification accuracy on huge dataset samples, training deep learning methods such as MMLP is expensive and requires substantial processing resources to expedite the training process. In [17], the KNN algorithm is used to evaluate the effectiveness of several distance functions in detecting thyroid illness. Prior to using the KNN algorithm with Euclidean and Cosine distances, the chi-square and L1-based feature selection approaches were used to choose the most suitable features. The authors stated that KNN achieved favourable outcomes. Nevertheless, the sample size that was analysed is somewhat limited, consisting of a total of 590 samples. Mishra et al. [18] used machine learning approaches, namely sequential minimal optimisation (SMO), decision tree (DT), random forest (RF), and K-star classifier, for the prediction of hypothyroid illness. This research considers a sample size of 3772 unique records. According to the authors, RF and DT outperformed the other two approaches, with accuracy ratings of 99.44% and 98.97% respectively. However, the authors failed to take into account the prediction of hyperthyroidism. Alyas et al. [19] conducted a comparative examination of the machine learning methods Decision Tree (DT), Random Forest (RF), K-Nearest Neighbours (KNN), and Artificial Neural Network (ANN) for the purpose of detecting thyroid illness. The experiments were performed on the most extensive dataset, taking into account

both sampled and unsampled data with the purpose of predicting thyroid illness. RF achieved the highest level of forecast accuracy, with a precision of 94.8%. Nevertheless, the authors did not conduct the tests for predicting the kind of thyroid illness. The researchers used deep learning algorithms to forecast the categorization of thyroid illness. The authors [20] used a deep neural network (DNN) to forecast the categorization of thyroid illness. The performance assessment is conducted on the UCI dataset, which consists of 3152 distinct samples. The researchers documented a classification accuracy of 99.95% while using Deep Neural Networks (DNN) to categorise thyroid illness. Nevertheless, a substantial dataset is necessary to adequately train the model for accurate performance assessment. Furthermore, an increased allocation of computer resources is required to facilitate the training of deep learning models. Table 1 presents a comparative analysis of the previous studies addressed in this section. Diverse datasets are used in scholarly works to assess the efficacy of thyroid illness diagnosis. However, the majority of the datasets shown in Table 1 are not considered standard datasets for the purpose of evaluating performance and comparing with previous research. Consequently, we used a well recognised UCI dataset for our research. While significant progress has been made in the aforementioned studies, yielding very accurate findings in detecting and classifying thyroid illness, there is a lack of comprehensive study on feature selection for thyroid cancer classification difficulties. In addition, the performance findings presented in the context of accurately classifying thyroid illness are inadequate, leaving room for further improvement. In addition, all previous studies categorise thyroid issues into three distinct groups: normal, hypothyroidism, or hyperthyroidism. It is desirable to categorise patients based on their treatment condition, health condition, and general health difficulties in order to efficiently and proactively forecast and manage their thyroid condition. Furthermore, the comprehensive assessment of machine learning and deep learning methods for classifying thyroid cancers and comparing their effectiveness is lacking in the current literature. We offer a solution for classifying thyroid illness that is based on feature selection, extremely accurate, and supports many classes. This solution aims to address the limits of existing methods and includes a full performance comparison of machine learning and deep learning-based approaches.

Table 1. Summary of the systematic analysis of the state-of-the-art thyroid cancer studies.

"Authors	Year	Sample Size	Dataset Source	Model	Classes	Evaluation Metrics	Results
[9]	2020	-	ToxCast	LR RF SVM XGB ANN	2	F1-score	(TPO) XGB-83% and (TR) RF-81%
[11]	2018	7200 samples, 21 attributes	UCI	SVM, Multiple Linear Re-gression(MLR), NB and DT	2	Accuracy	MLR 91.59% SVM 96.04% Naive Bayes 6.31% Decision Trees 99.23%
[12]	2020	7547,30 features	UCI	multi-kernel SVM	3	Accuracy, Sensitiv-ity, and Specificity	Accuracy (97.49%), Sensitivity (99.05%), and Specificity (94.5%)
[13]	2021	3771 samples, 30 attributes	UCI	DT, KNN, RF, and SVM	4	Accuracy	KNN 98.3% SVM 96.1%DT 99.5% RF 99.81%
[14]	2021	519 samples	diagnostic center Dhaka, Bangladesh	SVM, DT, RF, LR, and NB. Recursive Feature Selection (RFE), Univariate Feature	4	Accuracy	RFE, SVM, DT, RF, LR accuracy-99.35%

				Selection (UFS) and PCA			
[15]	2021	1250 with 17 attributes	external hospitals and laboratories	SVM,RF, DT, NB, LR, KNN, MLP, linear discriminant analysis (LDA) and DT	3	Accuracy	DT 90.13, SVM 92.53 RF 91.2 NB 90.67 LR 91.73 LDA 83.2 KNN 91.47 MLP 96.4
[16]	2021	7200 patients, with 21 features	UCI	multiple MLP	3	Accuracy	multiple MLP 99%
[17]	2021	690 samples, 13 features	datasets from KEEL repo and District Headquarters teaching hospital, Pakistan	KNN without feature selec-tion, KNN using L1-based feature selection, and KNN using chi-square-based fea-ture selection	3	Accuracy	KNN 98%
[18]	2021	3772 and 30 attributes	UCI	RF, sequential minimal opti-mization (SMO), DT, and K-star classifier	2	Accuracy	K = 6, RF 99.44%, DT 98.97%, K-star 94.67%, and SMO 93.67%
[19]	2022	3163	UCI	DT, RF, KNN, and ANN	2	Accuracy	Best performance Accuracy RF 94.8%
[21]	2022	215 with 5 features	UCI	KNN, XGB, LR, DT	3	Accuracy	KNN 81.25 XGBoost 87.5 LR 96.875 DT 98.59
[20]	2022	3152,23 features	UCI	DNN	2	Accuracy	Accuracy 99.95%”

III. SCBLN WORK

In this paper, a novel frame work approach is presented for effective identification of the thyroid. The overall flow of the suggested methodology was illustrated in figure 1.

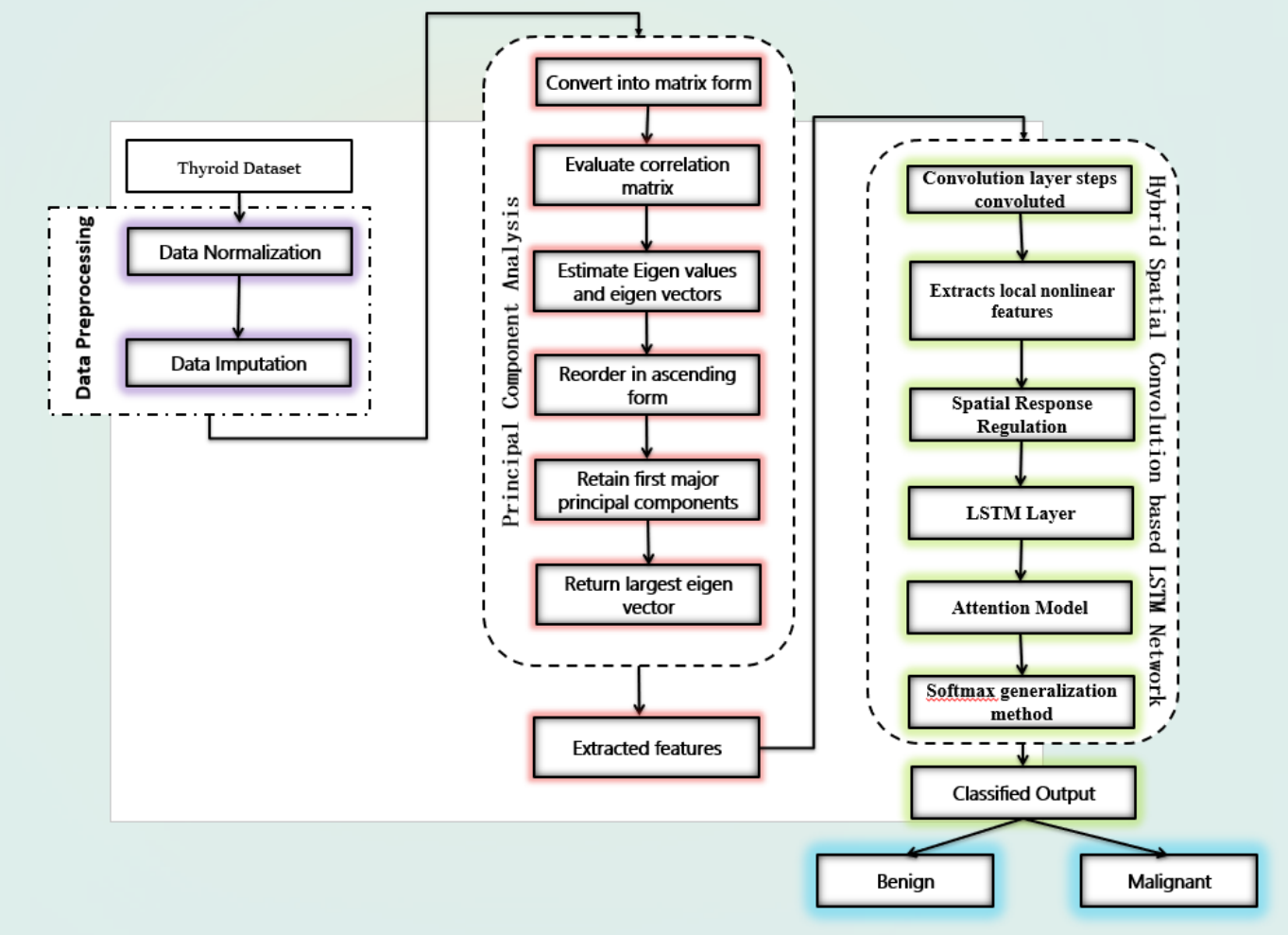


Figure 1 Schematic representation of the suggested methodology

a. Dataset

The dataset was obtained from https://www.kaggle.com/datasets/yasserhessein/thyroid-cancer-data-set. They derived it from Garavan Institute as given by Ross Quinlan. 6 databases are from the Garavan Institute inSydney,AustraliaApproximately the following for each database has 2800 training (data) instances and 972 test instances

b. Data normalization

The distributions of missing data dictate how to manage them in the preprocessing stage to guarantee adequate accuracy. This research used the Missing Completely At Random (MCAR) test to verify that the missing data were really randomly generated. The process that led to the data being missing determines the likelihood of bias resulting from missing data. Analytical strategies for filling in missing data and performing the necessary inferences and computations are examined using the chi-square test for multi-variate quantitative data. The imputer approach was used to fill in the missing variables in this investigation.

In order to enhance prediction, many classification algorithms aim to learn from pure instances and make the border between each class as obvious as feasible. Near the limit, learning to classify synthetic examples is substantially more challenging than far from it, according to most classifiers. There are three steps to the recommended procedure, and they are detailed here.

Step A: Equation (1) is used to build a synthetic instance.

$$N = 2 * (r - z) + z \tag{1}$$

where $r \rightarrow$ majority class samples, z $\rightarrow$ minority class samples and $N \rightarrow$ newly created synthetic instance.

StepB:Togetridoftheoutlierornoise,usethesesteps:If Step A receives a new instance$\hat{S} = \{\hat{S}_1, \hat{S}_2, \hat{S}_3, \ldots \hat{S}_n\}$, we will use Equation (2) to determine the distance between $\hat{S}_i$and the original minority S_m, Min $n_{\text{Rap}}\left(\hat{S}_i, \hat{S}_m\right)$.

$$\text{Min}_{\text{Rap}}\ \left(\hat{S}_i, \hat{S}_m\right) = \sum_{k=1}^{z} \sum_{j=1}^{M} \sqrt{\left(\hat{S}_i^{(j)} - S_{mk}(j)\right)^2} \tag{2}$$

where

$\text{Min}_{\text{Rap}}\ \left(\hat{S}_i, \hat{S}_m\right) \rightarrow$L is determined by using Equation (3) and the samples' rapprochement in accordance with Equation (2).

$$L = \sum_{i=1}^{n} \left(\text{Min}_{\text{Rap}}\ \left(\hat{S}_i, S_m\right)\right) \tag{3}$$

Step C: The distance between each initial majorityS_a, Maj Rap $\left(\hat{S}_i, S_a\right)s$, as specified by Equation (4), and all $\hat{S}_i$, must be calculated.

$$\text{Maj}_{\text{Rap}}\ \left(\hat{S}_i, S_a\right) = \sum_{i=1}^{r} \sum_{j=1}^{M} \sqrt{\left(\hat{S}_i^{(j)} - S_{al}^{(j)}\right)^2} \tag{4}$$

$\text{Maj}_{\text{Rap}}\ \left(\hat{S}_i, S_a\right) \rightarrow$H is calculated using Equation (5) in accordance with the rapprochement of samples and Equation (4).

$$H = \sum_{i=1}^{n} \left(\text{Maj}_{\text{Rap}}\ \left(\hat{S}_i, S_a\right)\right) \tag{5}$$

As a measure of disorder, entropy is all that is required. The connection between heterogeneity and probability was established by Claude E. Shannon using Equations (6) and (7)..

$$H(X) = -\sum \quad (p_i * \log_2\ p_i) \tag{6}$$

$$\text{Entropy (p)} = -\sum_{i=1}^{N} p_i \log_2\ p_i \tag{7}$$

Depend upon the probability the error instances can be removed.

c. Feature extraction

An essential aspect of the principal component analysis cycle is the act of separating connected variables into their respective metrics. In order to facilitate more precise classification, fusion feature (FF) vectors are reduced in dimensionality using principal component analysis. The characteristics that successfully categorise the data are considered throughout the model's construction. A FF has n distinct dimensions.We need to reduce the multi-dimensional F=$z_1, z, \ldots, z_n$to kn in the first place. In the end, the PCA uses these steps to produce a collection of reduced fused features (RFFs):

1. Data scaling:

$$z_j^i = \frac{z_j^i - \overline{z_j}}{\sigma_j} \tag{8}$$

2. Co-variance matrix computation:

$$\sum \quad = \frac{1}{m} \sum_i^m\ (z_i)(z_i)^T, \sum \quad \in R^{n \times n} (9)$$

3. Third, we calculate the eigenvector and eigenvalue.:

$$w^T \sum = \lambda\mu U = \begin{bmatrix} | & | & | \\ w_1 & w_2 \ldots & w_n \\ | & | & | \end{bmatrix}, w_i \in R^n \quad (10)$$

4. Eigenvalue selection. To choose the 100 most important Eigenvalues in K-space for our classification algorithm, we used the following formula

$$z_i^{\text{new}} = \begin{bmatrix} w_1^T z^i \\ w_2^T z^i \\ \cdots \ldots \\ \ldots \\ w_k^T z^i \end{bmatrix} \in R^k (11)$$

5. Finally, the feature matrix was built. At last, the feature significance score that has been acquired is normalised. The formula for this is shown in the following formula.
6. At last the feature matrix was constructed.

Lastly, the normalisation of the resulting feature significance score is shown in the following formula:

$$VIM_j = \frac{\text{VIM}_j}{\sum_{i=1}^{c} VIM_i} \quad (12)$$

Feature name	Correlation score
T3	0.3827
T4U	0.2203
sick	0.1936
age	0.1924
T3_measured	0.1693
TT4	0.1272
TSH_measured	0.1266
TBG_measured	0.0955
TT4_measured	0.0933
T4U_measured	0.0931
FTI_measured	0.0929
on_thyroxine	0.0853
TSH	0.0565
sex	0.0559
query_hyperthyroid	0.0551
pregnant	0.0455
goitre	0.0385
on_antithyroid_medication	0.0275
query_hypothyroid	0.0258
FTI	0.0252
thyroid_surgery	0.0098
query_on_thyroxine	0.0090
lithium	0.0080
tumor	0.0068

Figure 2 thyroid features

d. Prediction

In order to store and control information, the LSTM model adds input gates, output gates, and forget gates, and increases the cell state compared to the regular convolutional neural network model. To create a hybrid classifier for thyroid prediction, we are merging CNN with LSTM in this case. Depending on the current state of the neural network, the following gates are responsible for determining the output information: the forget gate for retaining the previous state, the input gate for determining the retention of the current information, and the cell state for judging the information's usefulness. At any one time, the model contains both the hidden state h_t and the cell state C_t. The forget gate, input gate, and output gate stand in for f_t, i_t, and o_t, respectively, at the t-th time step.

The sigmoid activation function and multiplier make up the cell state, which is important to the SCBLN model. The amount of information that may travel through the cell state is indicated by a number between zero and one that the sigmoid function produces. The SCBLN model primarily comprises the following four stages in its prediction process.

Step 1. The forget gate regulates the amount of information that is forgotten and determines, via the sigmoid function's output probability, whether the hidden cell state from the previous instant is forgotten by the present cell state. In order to determine the degree of information retention, the forget gate takes the state output h_{t-1} from the hidden layer at the last moment and the current input x_t as inputs. Then, it produces an output vector f_t, which in turn determines the amount of state C_{t-1}retained by the previous neuron. Here is the formula expression:

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \tag{13}$$

Here, f_t is the output value of the forget gate and σ is the sigmoid function.

Step 2.

The input gate takes into account which new features to add to the cell state and is responsible for inputting the existing feature information. To begin, $h_{t-1}$and x_t work together to establish the present input data. The next step is for the Tanh network layer to provide the new candidate cell state C_t to $h_{t-1}$and . The input gate then uses weights, typically between 0 and 1, to regulate the amount of new information given to the network by assigning them to the components in C_t . Presented below is the formula:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{14}$$

$$\tilde{C}_t = \text{tach}\ (W_c \cdot [h_{t-1}, x_t] + b_c) \tag{15}$$

Step 3. Cell state update

In order to get fresh cell state information C_t, update the attention cell state information C_{t-1}. The following equation shows the spatial response update formula:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t. \tag{16}$$

Here, $\odot$ is the Hadamard spatial product.

Step 4. Output gate

To get the cell output h_t of the current model, we use and x_t andh_{t-1}) to get the softmax current state o_t. Then, we modify the new cell state C_t using the Tanh activation function and multiply it by this vector. Following follows is the formula expression:

$$so_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{17}$$

When training the SCBLN model, eight sets of parameters must be learnt by the model. These include four sets of weightsw_f, w_i, w_o, and w_c, and four sets of bias itemsb_f, b_i, b_o, and b_c, and b_c, which correspond to f_t, i_t, o_t, and C_t, respectively. The five primary procedures are:

Step 1: Set the weight settings to zero
Step 2: compute each neuron's output value using the algorithm above. Step 3: calculate each neuron's error valuebackwards.
Step 4: Determine each weight parameter's gradient using the associated error value.
Step 5: Update the weights using the recommended method, iterate again until the error approaches the designated threshold, and then cease training.

IV. PERFORMANCE ANALYSIS

At first, we partitioned the datasets into train and test subsets. This technique was used to evaluate the performance of The suggested methodology. The overall experimentation was carried out under python environment. The pytorch framework was used to train the experiment in an environment consisting of an Intel (R) Xeon (R) Gold 5220 CPU and a GeForce RTX 2080 Ti GPU.. This study uses formatted datas from an open- source experimental data set made available on an official public website.

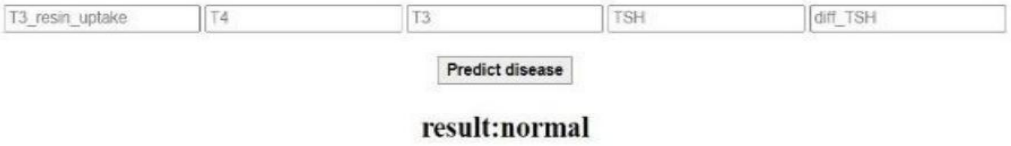


Figure 3 Simulated output

The overall simulated output for thyroid cancer prediction was illustrated in figure 3

Figure 4 Correlation scores for every feature of the thyroid dataset is plotted in a bar chart.

Each column represents the score for relevant feature. The scores are sorted in descending order.

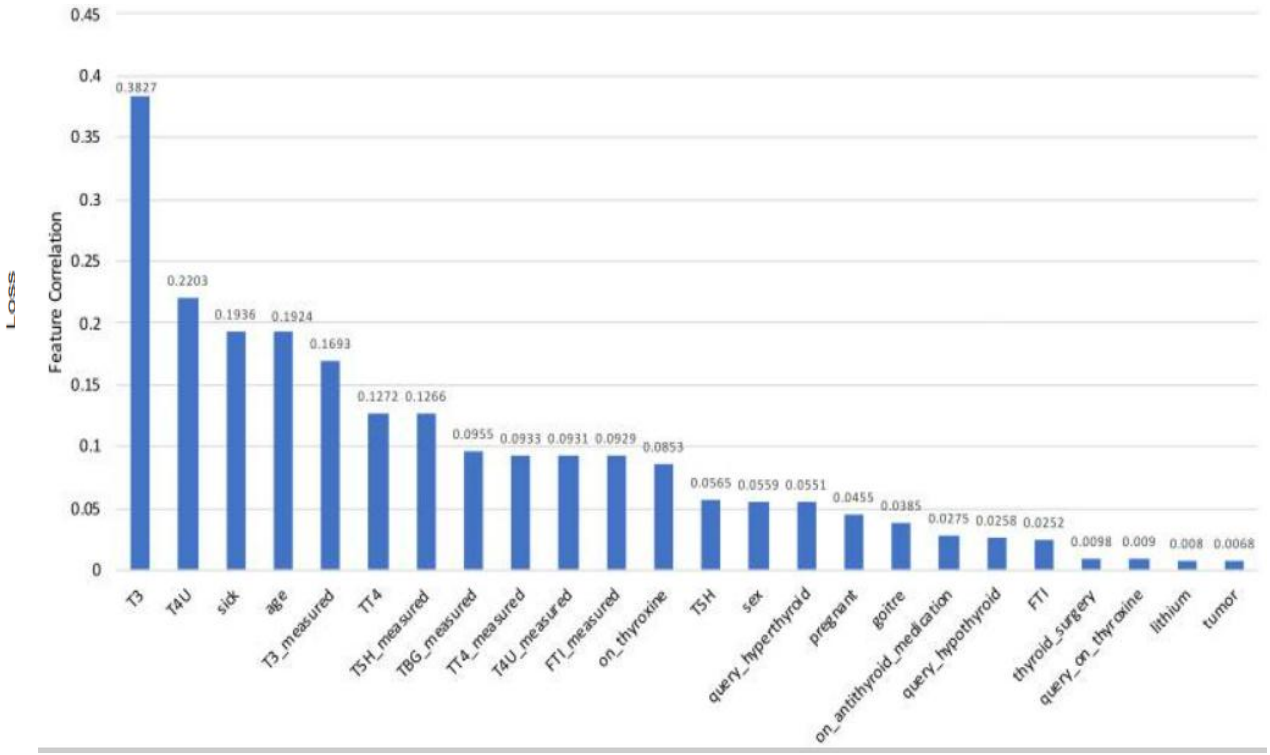


Figure 5 Loss calculation

Figure 5 compares the loss of the SCBLN model with the loss of the base model constructed using the original data. When optimising neural networks in machine learning, the cross-entropy—a popular loss function—corresponds to the loss function. During 100 epochs, isolated features are used to train the base model. From Figure 5, it is evident that over-fitting occurs at about epoch 20, at which point the loss in the training set (unseen data) rises. An over-fitting model loses its capacity to generalise its predictions to fresh inputs from unseen data because it becomes too focused on learning particular information from the training set. When compared to the trained base model, the test set loss does not suggest over-fitting. Up to step 30, it is feasible to recognise continuous learning in the best model's loss curve. From that point on, the error begins to rise, indicating that training should be stopped in the early learning epochs. Therefore, it may be said that the model fits the original dataset's thyroid prediction.

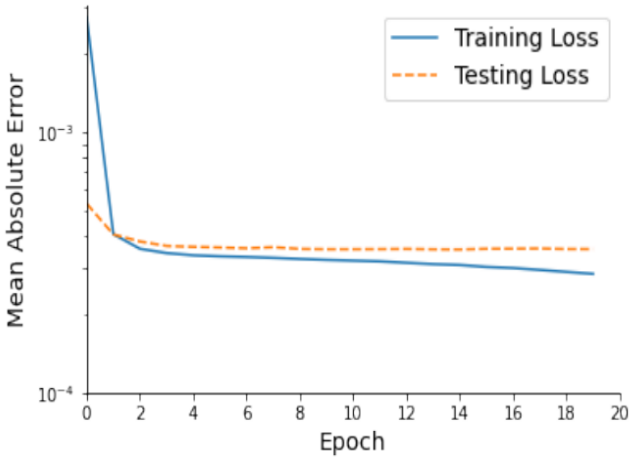


Figure 6 MAE calculation

As of from figure 6 the MAE was seem very less shows the efficiency of the suggested classifier.

We can utilize performance measures for evaluate the performance of the process of the categorization such as Accuracy, Precision, Recall, and F1-score.

Accuracy: In classification problems, accuracy is the proportion of correct predictions that the model makes out of all possible predictions. The number of true predictions is multiplied by the total number of predictions, which is then multiplied by 100. .

$$\text{Accuracy} = \frac{T_P+T_N}{T_P+F_P+F_N+T_N}. \quad (18)$$

In this context, T_P represents True Positive, which refers to a situation in which an individual is diagnosed with thyroid ill syndrome(1) and the model classifies their case as sick-thyroid(1). T_N is a True Negative, which refers to a case in which the model classifies the individual as Negative(0) despite the absence of thyroid ill syndrome. False Positives are classified as F_P when the model classifies a person's case as ill-thyroid(1) despite the fact that they do not have thyroid sick syndrome(0). F_N is False Negative, which refers to a situation in which a person has thyroid ill syndrome(1) and the model classifies their case as Negative(0).

Precision: Precision is the ratio of true positives and total positives predicted.

$$\text{Precision} = \frac{T_P}{T_P+F_P}. \quad (19)$$

Precision is a metric that indicates the percentage of patients who were diagnosed with thyroid ill syndrome and subsequently exhibited the condition. The predicted positives (People predicted as sick-thyroid are T_P and F_P) and the people actually having a euthyroid sick syndrome are T_P.

Recall: A recall is the ratio of true positives to all positives in the ground truth.

$$\text{Recall} = \frac{T_P}{T_P+F_n}. \quad (20)$$

In this context, recall is a metric that denotes the percentage of individuals who were diagnosed as ill-thyroid by the algorithm despite having actually had thyroid sick syndrome. The genuine positives (those with thyroid sick syndrome are TP and FN) and the patients diagnosed with thyroid sick syndrome by the model are TP. FN is included because the Person did, in fact, have a thyroid ill syndrome, despite the model's predictive accuracy.

F1-score: The F1-score metric is a combination of precision and recall. The F1 score is, in fact, the average of the two harmonics. Excellent precision and recall are both indicated by a high F1-score. It exhibits an exceptional combination of recall and precision and is particularly effective in addressing imbalanced classification issues.

$$F_1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{21}$$

V. PERFORMANCE EVALUATION

The evaluation and comparison of the performance with traditional procedures has been conducted. The accuracy is measured and illustrated in figure 7 to validate the projected technique with 99.9% accuracy. Similarly, the ANN,KNN,SVC,Decision Tree and Random forest are achieved95,86,89 and 94 . The analysis showed that the planned technique produced effective results.

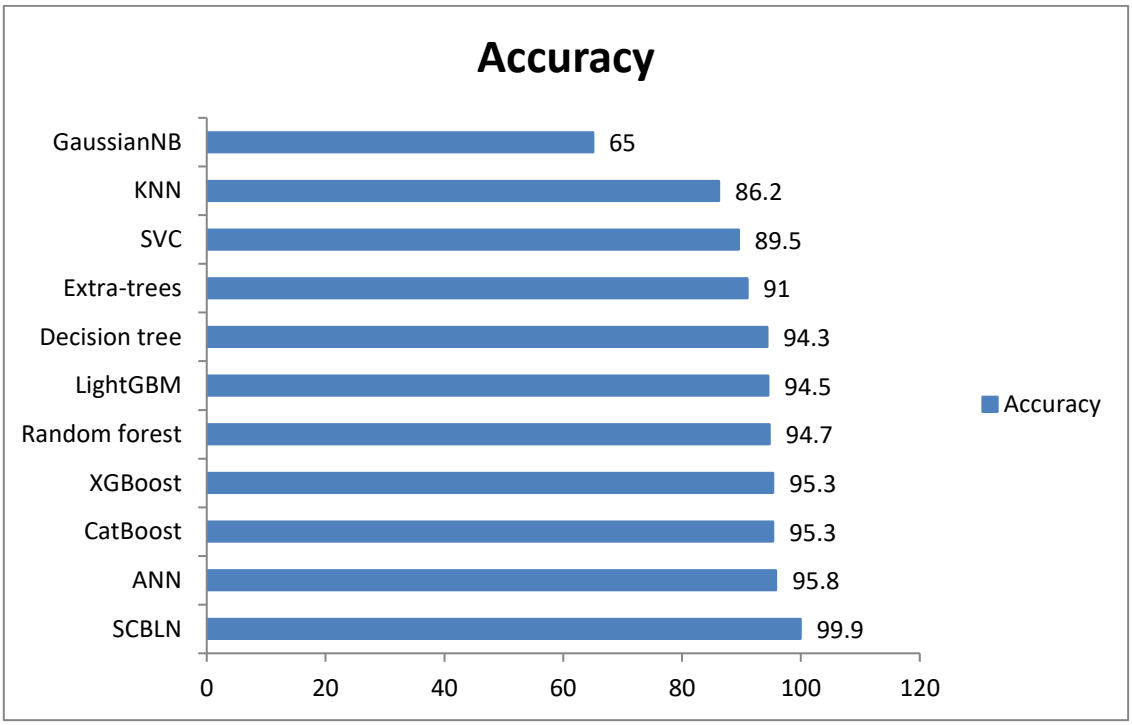


Fig.7 Analysis of Accuracy

To validate the projected technique, Accuracy is measured and illustrated in Table 2.

Method name	**Accuracy**
SCBLN	**99.9**
ANN	95
CatBoost	95
XGBoost	95
Random forest	94
LightGBM	94
Decision tree	94
Extra-trees	91
SVC	89
KNN	86
GaussianNB	65

Table 2. Analysis of Accuracy

The evaluation and comparison of the performance with traditional procedures has been conducted. The accuracy is measured and illustrated in figure 7 to validate the projected technique with 100% Precision. Similarly,

the ANN,KNN,SVC,Decision Tree and Random forest are achieved 95,87,94.and 94. The analysis showed that the planned technique produced effective results.

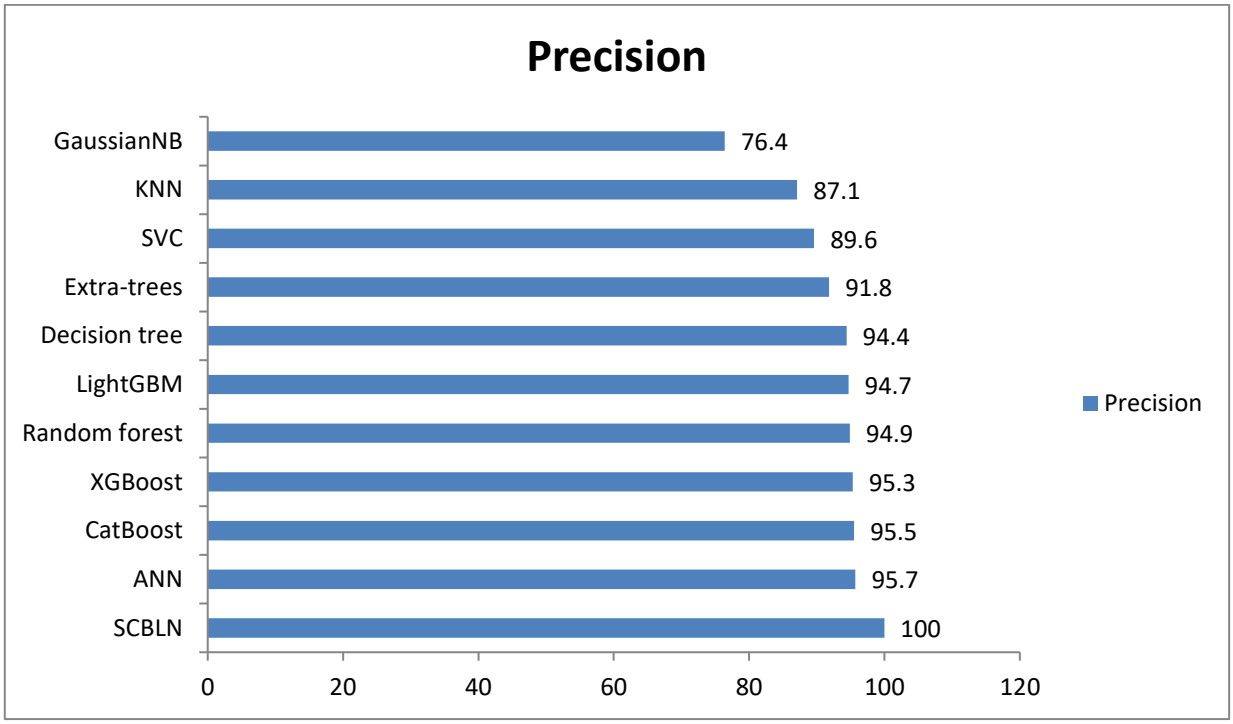


Fig 8.Analysis of Precision

To validate the projected technique, the Precision is measured and illustrated in Table 3.

Method name	**Precision**
SCBLN	**100**
ANN	95
CatBoost	95
XGBoost	95
Random forest	94
LightGBM	94
Decision tree	94
Extra-trees	91
SVC	89
KNN	87
GaussianNB	76

Table 3.Analysis of Precision

The evaluation and comparison of the performance with traditional procedures has been conducted. The accuracy is measured and illustrated in figure 7 to validate the projected technique with 99.9% Recall. Similarly, the ANN,KNN,SVC,Decision Tree and Random forest are achieved 95,86,89,94 and 94 . The analysis showed that the planned technique produced effective results.

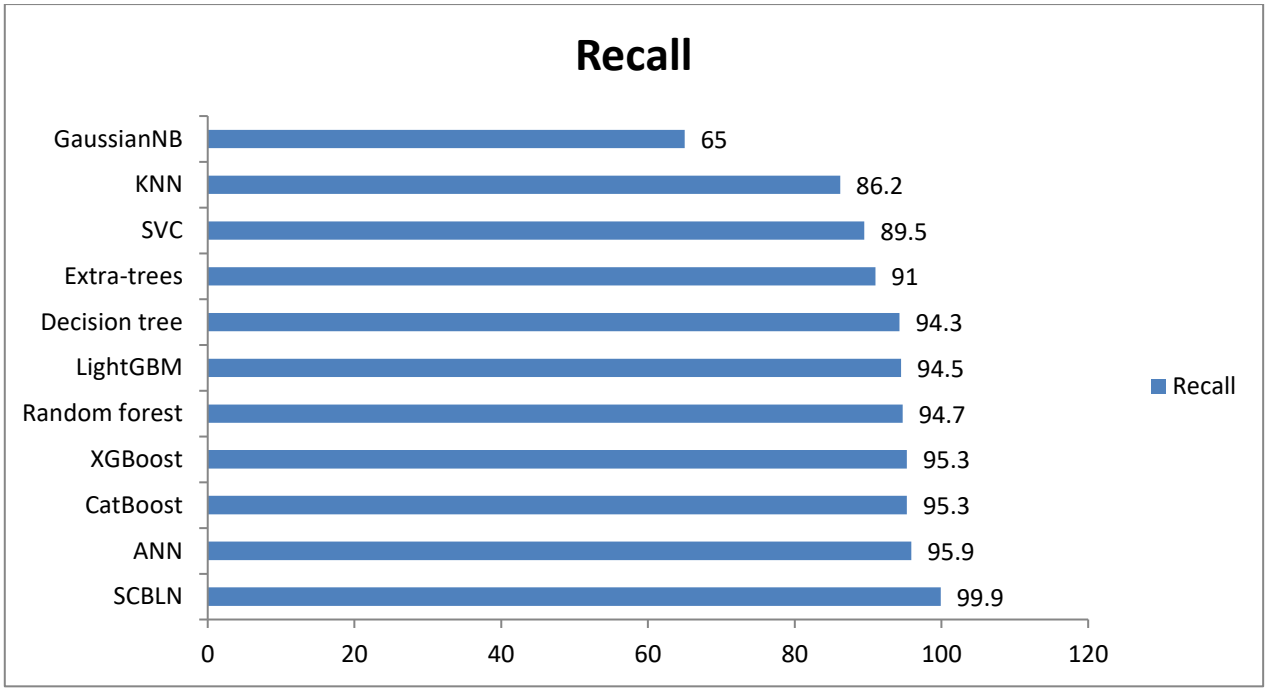


Fig.8. Analysis of Recall

To validate the projected technique, the F1_Measure is measured and illustrated in Table 4.

Method name	**Recall**
SCBLN	**99.9**
ANN	95
CatBoost	95
XGBoost	95
Random forest	94
LightGBM	94
Decision tree	94
Extra-trees	91
SVC	89
KNN	86
GaussianNB	65

Table 4. Analysis of Recall

The evaluation and comparison of the performance with traditional procedures has been conducted. The accuracy is measured and illustrated in figure 7 to validate the projected technique with 99.9% F-Measure. Similarly, the ANN,KNN,SVC,Decision Tree and Random forest are achieved 95,86,89,94 and 94 . The analysis showed that the planned technique produced effective results.

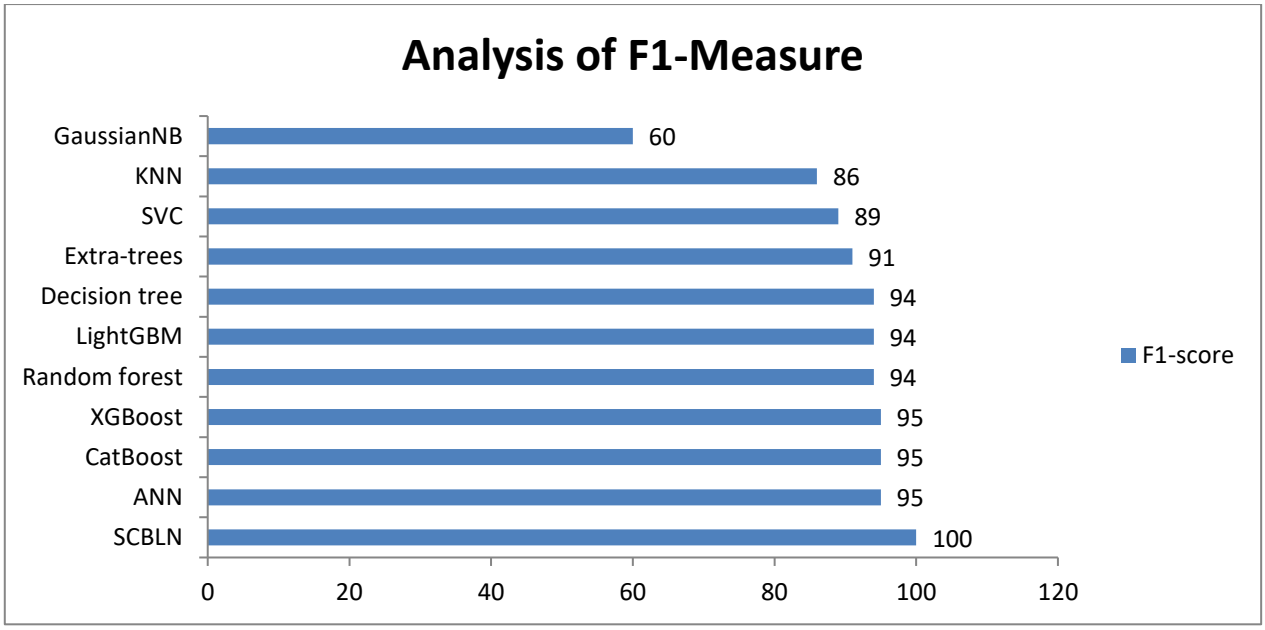


Fig.9.Analysis of F1-Measure

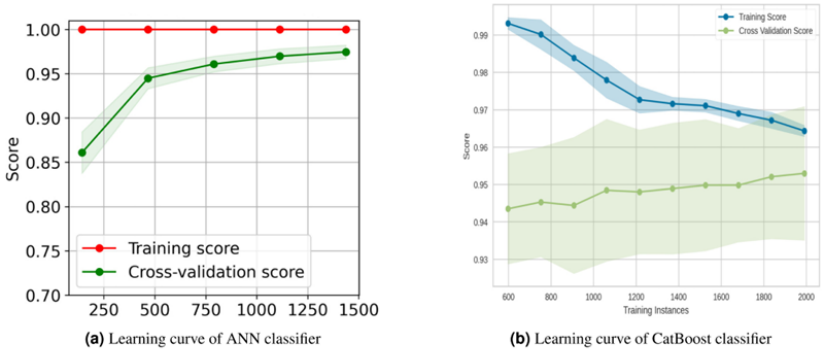
To validate the projected technique, the F1_Measure is measured and illustrated in Table 5.

Method name	**F1-score**
SCBLN	**100**
ANN	95
CatBoost	95
XGBoost	95
Random forest	94
LightGBM	94
Decision tree	94
Extra-trees	91
SVC	89
KNN	86
GaussianNB	60

Table.5.Analysis of F1-Measure

As of from the table 2 to table 5 and figure 6 to fig 9, it was revealed that the suggested methodology outperforms well when compared to other existing mechanisms in use. Hence , it is recommended to consider a suggested classifier when using this dataset to predict thyroid risk for obtaining satisfied performance.The suggested methodology obtained satisfied performance over thyroid cancer prediction by obtaining high range of accuracy(99.9%), precision (100%), recall(99.9%) and f1-score(100%)

In this experiment, we employed a visualization tool called the learning curve to ascertain the benefit of feeding our model with more training data. It demonstrated how test results and training for a machine learning and deep learning model with a variable number of training samples relate to each other. Generally, the cross-validation method was applied for generating the learning curve. We plotted the learning curve using the Python Yellowbrick module. The accuracy score for the train set was shown by the term "Training Score" in the aforementioned graphs (Figs. 10–12), whereas the accuracy score for the test set was indicated by the phrase "Cross-Validation Score." To prove the efficiency of the suggested methodology it can be compared with the existing methods [20],



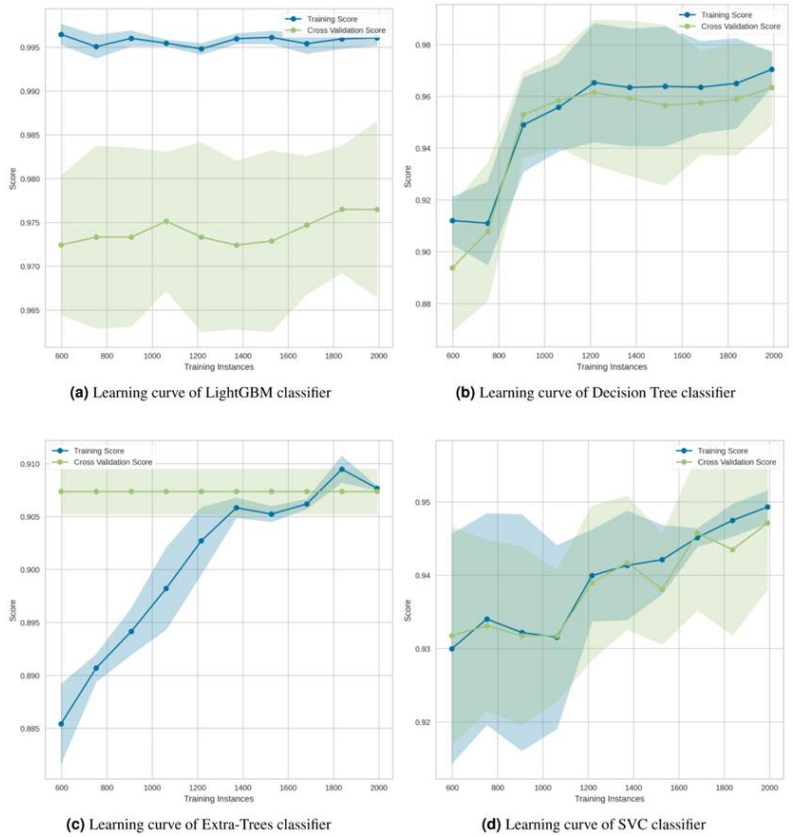
(a) Learning curve of ANN classifier

(b) Learning curve of CatBoost classifier

(c) Learning curve of XGBoost classifier

(d) Learning curve of Random Forest classifier

Fig 10. Learning curves for the top four classification algorithm



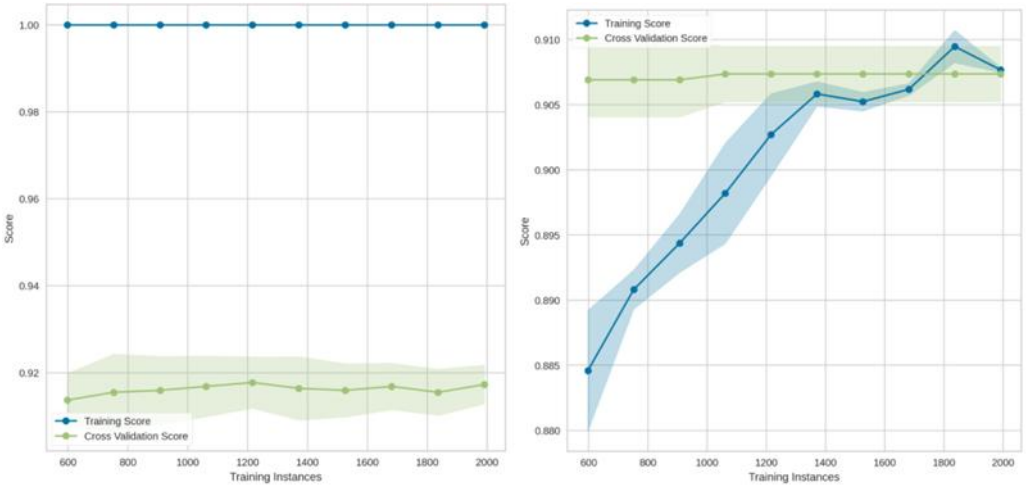
(a) Learning curve of LightGBM classifier

(b) Learning curve of Decision Tree classifier

(c) Learning curve of Extra-Trees classifier

(d) Learning curve of SVC classifier

Fig.11.Learning curves for most important classifier algorithms (Light GBM,DecisionTree,Extra tree, SVC)



(a) Learning curve of K-nearest Neighbour classifier **(b)** Learning curve of Gaussian Naive Bayes classifier

Fig 12.Learning curves for the lowest classifier algorithms (K-Nearest Neighbour,Naïve Bayes)

VI. CONCLUSION

This study provides an experimental investigation of the SCBLN Spatial Convolution-Based LSTM Network(SCBLN) algorithms for the purpose of predicting thyroid risk. In order to forecast thyroid risk, these algorithms are implemented on the Kaggle thyroid dataset. Precision, recall, F1 score, and accuracy are computed to assess the algorithms that are implemented. The SCBLN classifier surpasses all other algorithms with an accuracy of 99.9%. Extensive experiments and analyses have demonstrated that the SCBLN classification algorithm achieves superior results in terms of accuracy, precision, recall, and F1 measure when used to predict thyroid risk on the thyroid dataset. Consequently, it is advisable to priorities the SCBLN classifier over other existing techniques when employing this thyroid dataset to predict thyroid risk. In future the suggested methodology was applied on different cancer prediction dataset for improving its detection efficiency.

REFERENCES

[1] Chaubey, G.; Bisen, D.; Arjaria, S.; Yadav, V. Thyroid cancer prediction using machine learning approaches. Natl. Acad. Sci. Lett. 2021, 44, 233–238.

[2] Ioni¸t˘a, I.; Ioni¸t˘a, L. Prediction of thyroid cancer using data mining techniques. BRAIN Broad Res. Artif. Intell. Neurosci. 2016, 7, 115–124.

[3] Webster, A.; Wyatt, S. Health, Technology and Society; Springer: Berlin/Heidelberg, Germany, 2020.

[4] Hong, L.; Luo, M.; Wang, R.; Lu, P.; Lu, W.; Lu, L. Big data in health care: Applications and challenges. Data Inf. Manag. 2018, 2, 175–197.

[5] Association, A.T. General Information/Press Room|American Thyroid Association. Available online: https://www.thyroid.org/ media-main/press-room/ (accessed on 7 April 2022).

[6] Chen, D.; Hu, J.; Zhu, M.; Tang, N.; Yang, Y.; Feng, Y. Diagnosis of thyroid nodules for ultrasonographic characteristics indicative of malignancy using random forest. BioData Min. 2020, 13, 14.

[7] Kwon, M.R.; Shin, J.; Park, H.; Cho, H.; Hahn, S.; Park, K. Radiomics study of thyroid ultrasound for predicting BRAF mutation in papillary thyroid carcinoma: Preliminary results. Am. J. Neuroradiol. 2020, 41, 700–705.

[8] Idarraga, A.J.; Luong, G.; Hsiao, V.; Schneider, D.F. False Negative Rates in Benign Thyroid Nodule Diagnosis: Machine Learning for Detecting Malignancy. J. Surg. Res. 2021, 268, 562–569.

[9] Garcia de Lomana, M.; Weber, A.G.; Birk, B.; Landsiedel, R.; Achenbach, J.; Schleifer, K.J.; Mathea, M.; Kirchmair, J. In silico models to predict the perturbation of molecular initiating events related to thyroid hormone homeostasis. Chem. Res. Toxicol. 2020, 34, 396–411.

[10] Leng, L.; Li, M.; Kim, C.; Bi, X. Dual-source discrimination power analysis for multi-instance contactless palmprint recognition. Multimed. Tools Appl. 2017, 76, 333–354.

[11] Razia, S.; SwathiPrathyusha, P.; Krishna, N.V.; Sumana, N.S. A Comparative study of machine learning algorithms on thyroid cancer prediction. Int. J. Eng. Technol. 2018, 7, 315.

[12] Shankar, K.; Lakshmanaprabu, S.; Gupta, D.; Maseleno, A.; De Albuquerque, V.H.C. Optimal feature-based multi-kernel SVM approach for thyroid cancer classification. J. Supercomput. 2020, 76, 1128–1143.
[13] Das, R.; Saraswat, S.; Chandel, D.; Karan, S.; Kirar, J.S. An AI Driven Approach for Multiclass Hypothyroidism Classification. In Proceedings of the International Conference on Advanced Network Technologies and Intelligent Computing, Varanasi, India, 17–18 December 2021; pp. 319–327.
[14] Riajuliislam, M.; Rahim, K.Z.; Mahmud, A. Prediction of Thyroid Cancer (Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques. In Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 27–28 February 2021; pp. 60–64.
[15] Salman, K.; Sonuç, E. Thyroid Cancer Classification Using Machine Learning Algorithms. J. Phys. Conf. Ser. IOP Publ. 2021, 1963, 012140.
[16] Hosseinzadeh, M.; Ahmed, O.H.; Ghafour, M.Y.; Safara, F.; Hama, H.; Ali, S.; Vo, B.; Chiang, H.S. A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid cancer diagnosis in the internet of medical things. J. Supercomput. 2021, 77, 3616–3637.
[17] Abbad Ur Rehman, H.; Lin, C.Y.; Mushtaq, Z. Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Cancer. J. Chin. Inst. Eng. 2021, 44, 77–87
[18] Mishra, S.; Tadesse, Y.; Dash, A.; Jena, L.; Ranjan, P. Thyroid disorder analysis using random forest classifier. In Intelligent and Cloud Computing; Springer: Berlin/Heidelberg, Germany, 2021; pp. 385–390.
[19] \19. Alyas, T.; Hamid, M.; Alissa, K.; Faiz, T.; Tabassum, N.; Ahmad, A. Empirical Method for Thyroid Cancer Classification Using a Machine Learning Approach. BioMed Res. Int. 2022, 2022, 9809932. 20. Jha, R.; Bhattacharjee, V.; Mustafi, A. Increasing the Prediction Accuracy for Thyroid Cancer: A Step towards Better Health for Society. Wirel. Pers. Commun. 2022, 122, 1921–1938.
[20] Islam, S. S., Haque, M. S., Miah, M. S. U., Sarwar, T. B., & Nugraha, R. (2022). Application of machine learning algorithms to predict the thyroid cancer risk: an experimental comparative study. PeerJ Computer Science, 8, e898.