

## RESEARCH ARTICLE

# Conv1D-LSTM: Autonomous Breast Cancer Detection Using a One-Dimensional Convolutional Neural Network With Long Short-Term Memory

MITANSHI RASTOGI<sup>1</sup>, MEENU VIJARANIA<sup>1</sup>, NEHA GOEL<sup>2</sup>, AKSHAT AGRAWAL<sup>3</sup>,  
CRESANTUS N. BIAMBA<sup>4</sup>, AND CELESTINE IWENDI<sup>5</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, K. R. Mangalam University, Gurugram 122103, India

<sup>2</sup>Department of Computer and Applications, Vivekananda Institute of Professional Studies, Pitampura 110034, India

<sup>3</sup>Department of Computer Science & Engineering, Amity University Haryana, Gurgaon 122413, India

<sup>4</sup>Department of Educational Sciences, University of Gävle, 801 76 Gävle, Sweden

<sup>5</sup>Centre of Intelligence of Things, School of Creative Technologies, University of Bolton, BL3 5AB Bolton, U.K.

Corresponding author: Cresantus N. Biamba (cresantus.biamba@hig.se)

**ABSTRACT** Breast cancer is an increasingly serious problem in contemporary society, with millions of women and men worldwide affected by the disease. While traditional cancer detection strategies are at times effective, they typically require costly and time-intensive methods for implementation. The major drawback of using conventional methods for identifying breast cancer using the available data sets is that a single algorithm is not sufficient for accurate breast cancer diagnosis due to the heterogeneity of tumors, diverse data types, pattern complexity, feature engineering and dataset overfitting. The aim is to surpass the constraints of the conventional models and develop a hybrid model. The idea is to attain higher accuracy and lower computational time than existing models. This paper introduces a new method for detecting breast cancer using a one-dimensional convolutional neural network (1D CNN) and long short-term memory (LSTM). The model combines the strengths of both approaches, extracting sequential features from local data and modeling temporal dependencies and relationships. To detect and classify breast cancer, the 1D CNN and LSTM are used to automatically extract and analyze features from distinguishing features from a real dataset generated from mammography reports. The developed model has been assessed on the extracted feature of the primary available dataset consisting of mammograms from over 760 patients. The developed model achieves 99% accuracy on the test data, demonstrating its potential to provide an automated approach to breast cancer detection. The work emphasizes a significant improvement in feature extraction, accuracy, and robustness. Additionally, the proposed model's versatility allows it to handle diverse data types, achieve better generalization and lower computational time. The model offers a high level of interpretability, which is crucial for medical professionals to understand and trust the decision-making process of the system. The developed hybrid model outperforms various other state-of-the-art techniques like ANN, CNN, CNN-Bi-LSTM-GRU-AM (Convolutional Neural Network-Bidirectional Long Short-Term Memory-Gated Recurrent Unit-Attention Mechanism), and CNN-GRU (Convolutional Neural Network- Gated Recurrent Unit) in terms of accuracy, feature extraction and computational time. This work emphasises the potential of 1D CNN augmented with LSTM to create an automated system for identifying breast cancer. Hence, provides a promising foundation for further development and practical usage of deep learning for automated cancer diagnosis.

**INDEX TERMS** Breast cancer, convolutional neural network, LSTM, deep learning, machine learning, max-pooling layer, RNN.

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy<sup>1</sup>.

## I. INTRODUCTION

The worst disease affecting women globally is likely breast cancer [1]. In the last five decades, the USA has experienced nearly 90% overall survival for BC (breast cancer), compared to 60% in India [2]. The primary driver of cancer growth is infection [3]. Because of the unique growth of breast cells, women may develop tumors. The division of these sizable tumor cells into cancerous or non-cancerous tissues may vary depending on factors such as their location, size, and area [4]. “Benign” refers to the primary tumour region of a non-cancerous tumour, and “malignant” depicts the tumour area of a melanoma tumour. Benign tumors do not endanger the lives of women as they can be treated easily. On the other hand, malignant require medical treatment such as surgery or radiation [5].

Machine learning approaches that offer a better presentation, such as K-NN, LR (logistic Regression), SVM (support Vector Machine), and others, have been used in a number of historical studies. Breast cancer is now being characterized using deep learning to circumvent machine learning [6]. The most popular deep learning techniques include convolutional neural networks, recurrent neural networks, MP, long-short-term memory networks, and many more. It is a process that is frequently employed in data science [7]. As stated by the WHO, breast tumors affect a lot of people all over the globe [8]. Mammography, MRI, and histopathology tests are all used in determining the illness’s current assessment methods.

The utilization of a one-dimensional Convolutional Neural Network (CNN) in tandem with Long Short-Term Memory (LSTM) networks presents a cutting-edge approach to autonomous breast cancer detection. This fusion of architectures enables the extraction of intricate spatial features from breast cancer-related images or sequences, such as mammograms or histopathology data, via CNN’s adeptness at feature extraction from one-dimensional data representations. The CNN layers in this hybrid model proficiently discern essential patterns and features within the input data, allowing for robust feature representation. Following the CNN layers, the integration of LSTM networks becomes instrumental in processing the extracted features sequentially. LSTMs excel at capturing temporal dependencies and prolonged relationships within sequential data, effectively encoding the temporal information embedded in breast cancer-related data. This synergistic coupling of CNN and LSTM networks improves the model’s efficacy in comprehending not only spatial structures but also temporal intricacies, thus significantly bolstering the accuracy and efficacy of autonomous breast carcinoma detection.

By harnessing the capabilities of a one-dimensional CNN combined with LSTM networks, the autonomous detection of breast cancer achieves a novel advancement in medical diagnostics. The one-dimensional CNNs efficiently extract pertinent spatial features from breast cancer-related data, while the subsequent integration of LSTM networks

empowers the methods to comprehend the consecutive arrangement and time-related dependencies within the data. This amalgamation of architectures allows the model to discern intricate patterns and nuances that are crucial in diagnosing breast cancer, thereby the model has its potential to significantly enhance the precision and reliability of breast cancer detection, offers a promising path towards more accurate and effective diagnostic tools. The hybrid model has unique ability to capture both spatial features and temporal behavior inspires a future where breast cancer can be comprehensively understood and categorized, leading to improved healthcare outcomes and motivating further research and development in the field.

## A. MOTIVATION

Traditional methods frequently depend on handcrafted features as they cannot capture the intricate patterns [11]. CNNs automatically learn relevant attributes derived from the input information, making them adept at capturing intricate details that might be missed by manual feature engineering. CNNs are excellent at learning spatial hierarchies and patterns in data. In the context of diagnosing breast cancer from data sequences, a 1D CNN can effectively sequentially capture important features [12]. Combining it with LSTM, renowned for its ability to recognize temporal correlations in sequential data, may improve the precision of identifying subtle patterns that point to the advancement of cancer.

Breast cancer data might involve sequences or time-dependent patterns, such as medical records, sequential imaging (like MRI scans over time), or genetic data. LSTMs are useful for problems involving data order because they can recognize and retain dependencies in sequential data [13]. Automated systems based on deep learning models aim to reduce human bias and subjectivity in interpretation. These algorithms may reliably detect malignant trends through training on various datasets, which may yield more objective and trustworthy results.

The authors in [14] suggest future research directions in improving the system model by enhancing the CNN architecture and utilizing trained neural networks to achieve more accurate measurements. In [15], the authors using Mammographic Images over CNN mention that the investigation is ongoing and new advancements are being made in enhancing the CNN architecture and utilizing trained neural networks, indicating potential future research directions for more accurate measurements. The authors in [16] have shown the importance of an ML-based model using the LSTM algorithm, a variant of recurrent neural network, for predicting breast cancer. This method works well with sequential data and has outperformed other machine learning algorithms in the study. As a result, this study suggests a different method for addressing the issue of discovering breast cancer by proving the effectiveness of a 1D-CNN enhanced with LSTM for the independent identification of cancerous tissue. Early

detection of breast cancer significantly improves treatment results. Leveraging advanced deep learning models like 1D CNNs with LSTM could aid in detecting subtle changes indicative of presence of cancer at an earlier and more responsive phase. Autonomous detection systems have the potential to be scalable and accessible. Once trained, these systems could be deployed in various healthcare settings, especially in areas where access to expert radiologists or healthcare facilities is limited. Instead of replacing human expertise, these tools can facilitate quicker and more accurate analysis, which could lighten the strain and allow specialists to work on more complicated issues. These AI systems can continually learn from new data, potentially improving their accuracy and performance over time as they encounter more diverse cases and patterns. The CNN architecture was improved using the LSTM model to produce more accurate performance outcomes. By evaluating the effectiveness of various algorithms, such as ANN, CNN, CNN-Bi-LSTM-GRU-AM, CNN-GRU, and LSTM, for cancer diagnosis, the suggested model improves on current understanding.

The following highlights the study's primary contribution:

- 1) The work is on employing a 1D-CNN supplemented with LSTM to recognise breast cancer in mammographic data.
- 2) The study employs the primarily available dataset consisting of mammograms, which consists of over 760 patients, to train and test their model.
- 3) The study aims to accurately and quickly identify potentially cancerous areas in mammograms, which can help optimize the time of histologists and improve the efficiency of cancer detection.
- 4) An experimental comparison between the suggested models and the most advanced classifiers that have undergone training as well as evaluation.

By investigating the use of 1D-CNN enhanced with LSTM models on mammogram data, the study advances the field of breast cancer diagnosis overall and aims for more precise forecasts. The integration of a 1D-CNN augmented with LSTM in autonomous breast cancer detection marks a significant advancement. This hybrid model harnesses CNN's proficiency in feature extraction from sequential data and the LSTM's ability to capture temporal dependencies. By leveraging these strengths, the model achieves improved accuracy in identifying intricate patterns within sequential breast cancer-related data, potentially enhancing early detection, reducing false positives, and providing automated decision support in clinical settings. The fusion of 1D CNN with LSTM represents a pivotal step toward more effective and reliable autonomous breast cancer detection systems. The remaining paper was embellished as follows: The history and relevance of finding carcinoma of the breast are covered in sections I and II. The deep learning techniques aided in the diagnosis of breast carcinoma have been put out in Section III. The experimental outcomes of the investigation are depicted in section IV. In Section V, the suggested model is contrasted

with various learning algorithms already in use. Section VI concludes the analysis and outlines its future directions.

## II. LITERATURE REVIEW

this section presents related work done by researchers in this field. Begam et al. [9] in the earlier examination, the authors proposed an ensembled model of CNN with LSTM and random forest for diagnosing breast cancer. The experimental study was implemented on a secondary image WDBC dataset that has 32 attributes. The proposed model shows an accuracy of 98%. Alanazi et al. [10] proposed CNNs for the identification of tumors in breasts in an efficient manner. A huge dataset of images of about 275000 was retrieved from Kaggle. The implemented model offers an accuracy of 87%. Masud et al. [11] state that the female mortality rate is at the surge. CNNs helped in identifying the deadliest breast cancer with better precision. Two datasets of 250 images and 780 images of different dimensions are used for the research purpose. The decision parameters made helped in analysis using 5-fold cross-validation techniques. Zheng et al. [12] proposed an efficient model of AdaBoost with DL\_NN for the prognosis of tumors in the breast as compared to other DL models. The developed model has attained an acc. of 97%. The study [13] offered a deep-learning model implemented on DDSM BUS1 and BUS2 datasets achieved through mammography and image ultrasound of breast cancer patients. The BUS1 dataset consists of 780 images and BUS2 consists of 250 images of breast cancer patients. The implemented model ShuffleNet-RESNET offered a maximum accuracy of 96.52% and 98.13%. The offered technique is also suitable for smaller datasets. Botlagunta et al. [14] in this paper the repository of cancer patients was sourced from the year 2012-21 from BIACH & RI. The dataset consists of 25652 records. The Split () function was used for data pre-processing. ML techniques are employed for the prognosis of illness at an early stage. Ensembled DT outperformed with a maximum accuracy of 83%. Azam et al. [15] worked on mammography reports of 53273 females related to BC. The spots of microcalcifications lead to more than 20% of breast tumors in the body. Age, menopausal status, BMI index, and mammographic density are some of the parameters used for the analysis. Mahesh et al. [16] this paper drew attention to preparing an automated system for the identification of diseases which results in minimizing mortality rate. A hybrid model of SVM, K-NN, LR, RF, and DT was made and their performances were analyzed based on performance metrics. The ensemble model of ML algorithms delivers better results as compared to individual algorithms. Srinivas et al. [17] discussed twenty algorithms out of which 9 models are implemented in Python using Google Collab on the WISCONSIN dataset and 11 are implemented using WEKA. The stochastic gradient descent technique was achieved with 98.5% and 97% accuracy using collab and WEKA respectively. Sharma and Kumar [18] the authors worked on WBC and WDBC datasets. In this paper, textual (1-D) data was converted into image (2-D) data using CNN with

VGGnet-16. The processing cost of CNN with VGGnet-16 is very high. Vaka et al. [19] the authors raised concern about the surge in breast tumor cases. 5 ML and DL algorithms NB classifiers, SVM, Bi-clustering and Adaboost, RCNN, HA-BiRNN, and DNNS were applied on a real-time dataset collected from M.G Cancer Hospital. The proposed classifier DNNS outperforms with a maximum accuracy of 97.2%. Divyavani and Govindaswamy [20] the authors implemented SVM and ANN on the UCI repository dataset with 98% and 99% accuracy respectively. Performance analysis was done on 10-fold cross-validation techniques. Mohammed et al. [21] authors worked on J48, Naive Bayes, and SMO using the WBC dataset. Standard deviation, ROC curve, and accuracy are some of the parameters used for performance evaluation. With the help of the resampling filtering technique, sequential minimal optimization depicts the maximum accuracy in comparison to DT and NB. Dalal et al. [22] highlight that early identification of breast cancer can help to stop the disease from spreading to different parts of the body. RF, XBoosting, LR, and multilayer perceptron were trained and implemented on the WDBC dataset. The proposed model Xboost tree attained an accuracy of 99.6. Elkorany and Elsharkawy [23] mammograms are often employed for the identification of BC. The categorizing of the onography region of interest spots (normal, aggressive, or innocuous) is an extremely important step in this procedure because it assists medical practitioners in identifying BC. Feature selection CNN-based Inception V3, ReseNet 50, and Alexnet techniques applied in comparison with multiclass SVM. The MIAS dataset was used for the research. MSVM attained maximum accuracy concerning other deep learning models. Hooda et al. [24] offer an instance investigation of the Malwa region Belt in India, which has resulted in the escalation of mortality rate owing to mammary tumors. BagBoosting was proposed for the forecasting of disease in the early period which resulted in an accuracy of 98.21%. Lee et al. [25] highlight the importance of the hybrid model over individual approaches for BC prognosis. An image dataset, TCGA was used for the research purpose. The combination of 6 approaches, GB, random forest, naïve\_bayes, k-nn, support\_vector\_machine, and logistic\_regression was used to develop the hybrid model. The comparative analysis was done for ROC for lymphocytes and stroma cancer. Bhise et al. [6] conducted an evaluation of machine learning techniques including CNN, SVM, KNN, logistic regression, naïve Bayes, and random forest using the BreakeHis 400X dataset, with accuracy and precision as decision parameters. Their analysis revealed that CNN outperforms other algorithms in terms of both accuracy and precision. On the other hand, Masood [26] applied preprocessing, feature selection, and extraction methods to reduce dimensionality. The author explored various ML models such as ANN, ELM, SVM, KNN, MLP, NB, and CART, concluding that SVM stands out as the most suitable machine learning model based on the analysis. Islam et al. [27] a comparative study was

conducted on machine learning models including K-NN, ANN, SVM, RF, and LR, utilizing the UCI dataset and considering accuracy, precision, and F1 score as decision parameters. The experimental results demonstrated that ANN achieved a maximum accuracy of 98.57%, precision of 97.82%, and an F1 score of 0.9890. Additionally, Ming et al. [28] analyzed random forest, logistic regression, k-nearest neighbor, and MCMC GLMM alongside existing BCRAT and BOADICEA models, using accuracy and AU-ROC as comparison parameters. The utilization of machine learning models led to an increase in accuracy by up to 30 to 35% compared to the existing models. Dhahri et al. [29] use various techniques of machine learning for predicting and diagnosing breast tumor with the WISCONSIN dataset encompassing sensitivity, accuracy, specificity, precision, and ROC curves as analysis parameters. Genetic programming can identify the most appropriate model using a combination of feature extraction, preprocessing, and classifier algorithms. Abdulrahman Jasim et al. [30] state breast cancer costs many lives and is spreading rapidly across the globe. The paper conducted an analysis wherein twelve machine learning algorithms were employed on the WBC dataset. It then evaluated their performance based on multiple metrics, including accuracy, recall, sensitivity, and F1-score. SVM outperforms other classifiers with the highest accuracy of 90%. Thakur et al. [31] emphasize the growth in the number of cancer patients every year. The METABRIC dataset is used for data analysis. ANOVA technique is used for pre-processing. SVM, DT, RF, AdaBoost, and ANN are the five techniques used. This dataset is available on Kaggle.com with 1904 instances and 31 features. As a result, ANN depicts a maximum accuracy of 87.43%. Khan et al. [4] discussed the escalation in the death rate of breast cancer patients in India, the USA, and China. WBCD dataset retrieved from UCI repository. The algorithms applied to the dataset were DT, RF, K-NN, and Logistic Regression. The classifier that achieved maximum accuracy in logistic regression is 98.6%.

In the aforementioned survey, it has been observed that convolution neural networks are effective in extracting spatial features from the datasets, but mammography reports are spatial as well as temporal data [30]. Hence, only CNN for breast cancer prediction was inefficient and delivered low accuracy rates. With the increased size of a dataset, the computation time also increases. CNN achieved large size data after feature extraction with high computational cost leading to an inefficient system [31]. Therefore, there is a need for an efficient and reduced feature extraction technique. In the proposed Conv1D-LSTM model, CNN is utilized in combination with LSTM which allows for the processing of spatial and temporal information simultaneously. LSTMs are designed for sequential data, making them well-suited for tasks involving sequences, such as time-series data. LSTM further helps to regularize the data by preventing over-fitting by capturing dependencies in the data. In the proposed model, the max pooling layer is used as it helps in reducing the



**TABLE 1. Comparison of machine learning techniques in breast cancer prediction.**

| Author                                          | Methodology                                                      | Dataset                                                        | Results                                                                                                          | Remarks                                                                                                                     |
|-------------------------------------------------|------------------------------------------------------------------|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Mehedi Masud et al. [11]                        | CNN                                                              | 250 and 780 ultrasound images                                  | The testing accuracy of the technique is 100% whereas the training accuracy of ResNet 50 is 98%.                 | Two different datasets are combined to enhance the efficiency of the model                                                  |
| Jing Zheng et al. [12]                          | Deep Learning Assisted Efficient Adaboost Algorithm              | Images of PET, CT, and MRI screenings of breast tumor patients | The attained accuracy is 97.2%                                                                                   | Images collected before, after, and during the medical treatment                                                            |
| Adyasha Sahu et al. [13]                        | Shuffle Net-RESNET                                               | BUS1 and BUS2 image datasets                                   | The attained accuracy is 96.52% and 98.13% on the respective datasets                                            | This model is suitable only for small datasets                                                                              |
| Mahendran Botlagunta et al. [14]                | Ensembled DT                                                     | Image datasets of 25652 records                                | The attained accuracy is 83%                                                                                     | ML techniques implied for the prognosis of breast tumor                                                                     |
| Shadi Azam et al. [15]                          | Cox Regression and Logistic Regression                           | Mammography image dataset of 53273 females                     | 95% microcalcification shows an adverse impact on BC patients.                                                   | Impact of microcalcification on breast tumor patients                                                                       |
| T.R. Mahesh et al. [16]                         | K-NN, LR, RF, DT, and ensemble model                             | WDBC dataset                                                   | A combination of K-NN, LR, RF, and DT gives an accuracy of 98.14%                                                | For the earlier identification of carcinoma of the breast accuracy, recall, precision, and f1-score                         |
| Taarun Srinivas et al. [17]                     | Gradient Descent Technique                                       | WISCONSIN Dataset                                              | The accuracy of the offered technique is 98.5% and 97% using collab and WEKA, respectively.                      | The models were implemented using Collab and WEKA                                                                           |
| Anuraganand Sharma and Dinesh Kumar [18]        | CNN with VGGNet-16                                               | WBC and WDBC datasets                                          | The developed model attained maximum accuracy as compared to other classification algorithms on the WDBC dataset | Converting 1-D tabular data of BC patients into a graphical image                                                           |
| Anji Reddy Vaka et al. [19]                     | SVM, Bi-clustering, Adaboost, RCNN, HA-BiRNN, and DNNS           | Real-time DATASET                                              | DNNS outperforms with maximum accuracy of 97.2%                                                                  | Performance analysis of models was done based on accuracy, precision, and recall                                            |
| M. Divyavani and Kalpana Govindaswamy [20]      | SVM and ANN                                                      | UCI Repository Dataset                                         | ANN attained an accuracy of 99%.                                                                                 | Performance analysis was done on 10-fold cross-validation techniques                                                        |
| Siham A. Mohammad et al. [21]                   | J48, NB, and SMO                                                 | WBC Dataset                                                    | SMO attained 99.56%                                                                                              | Performance comparison is done using Std. deviation, recall, accuracy, Roc-curve                                            |
| Surjeet Dalal et al. [22]                       | Multilayer perceptron, XBoost, RF, and LR                        | WDBC Dataset                                                   | XBoosting attained the highest accuracy of 99.65%                                                                | WDBC image dataset used for the research purpose                                                                            |
| Ahmed S. Elkorany and Zeinab F. Elsharkawy [23] | CNN with Inception V3, ResNet 50, Alexnet, and Multiclass SVM    | MIAS dataset                                                   | The Multiclass SVM model attained maximum accuracy of 97.81%                                                     | Feature selection-based techniques applied to mammographic data                                                             |
| Nishtha Hooda et al. [24]                       | Bagboosting                                                      | Real-time Dataset of the Malwa region of India                 | The accuracy of the proposed model is 98.21%                                                                     | The research was done using K-fold cross-validation techniques                                                              |
| Sanghoon Lee et al. [25]                        | GB, RF, NB, KNN, SVM, LR and their hybrid model                  | TCGA image dataset                                             | The hybrid model has maximum accuracy                                                                            | The comparative analysis was done to ROC for lymphocytes and stroma cancer                                                  |
| Sweta Bhise et al. [6]                          | CNN, SVM, KNN, LR, NB and RF                                     | BreaKHis 400X Dataset                                          | CNN gives the highest accuracy and precision as compared to other models                                         | Probabilistic predictions of the results have been calculated using activation coefficients like ReLu.                      |
| Hiba Masood [26]                                | ANN, ELM, SVM, KNN, MLP, NB, and CART                            | WISCONSIN Dataset                                              | SVM has achieved maximum accuracy                                                                                | Preprocessing, feature selection, and extraction were applied to minimize the total number of dimensions                    |
| MD. M. Islam et al. [27]                        | SVM, KNN, RF, ANN and LR                                         | WISCONSIN Dataset                                              | ANN has the highest accuracy of 98.57%                                                                           | Accuracy, sensitivity, specificity, precision, f1-score, and Matthews Correlation Coefficient used for performance analysis |
| Chang Ming et al. [28]                          | RF, LR, KNN, MCMC GLMM                                           | Real-time dataset                                              | RF attained a maximum accuracy of 98%                                                                            | Accuracy and Roc curve used as decision parameters for performance analysis                                                 |
| Habib Dhahri et al. [29]                        | KNN, RF, GNB, LR, AdaBoost, GP, and Linear Discriminant Analysis | WISCONSIN Dataset                                              | Genetic Programming has the best accuracy of 98%                                                                 | Accuracy, precision, Roc curve, and specificity are the parameters used for performance analysis                            |

dimensionality and lowers the computational time. SoftMax activation is used with Conv1D-LSTM allowing to merging of the predictions achieved from CNN and LSTM. Hence, providing a coherent and interpretable prediction as SoftMax ensures that the output is the probability distribution.

### III. MATERIALS AND METHOD

This section entails a detailed explanation of the self-gathered dataset. Furthermore, a novel hybrid Conv1D-LSTM along with a max pooling layer and SoftMax function have been proposed. In this section, the framework, and algorithm have also been discussed. A mammogram database of 760 patients related to breast cancer has been self-gathered from a healthcare hospital. The individual reports of patients do not allow for efficient decision-making for disease diagnosis in the body. By utilizing the application of Python script distinguished features have been extracted from individual reports in tabular format. Train and test data are two further classifications for the retrieved dataset. The results for breast tumor diagnosis are then obtained by feeding the training and testing data onto the suggested Conv-1D with the LSTM model. The diagrammatic representation of research methodology is shown in Figure 1

The dataset comprises mammography reports from over 760 patients, collected from a healthcare institution with informed consent obtained from each participant. Processing the PDF mammography reports presented many challenges. 1) The inconsistent structure of reports. An algorithm that reads PDF files and uses regular expressions to identify and extract the relevant information. This was crucial for ensuring that key data points were accurately collected from reports, which may contain a mix of text, tables, and other content. 2) The statistics showed a class imbalance, with most reports indicating non-cancerous diagnosis and fewer incidences of breast cancer. The Synthetic Minority Oversampling Technique (SMOTE) was used to balance the dataset by oversampling the minority class. This allowed the model to learn equally from positive and negative diagnoses. The reports of patients do not allow for efficient decision-making directly for disease diagnosis. 3) Feature Extraction and Scaling: The features extracted from the reports varied widely in scale and significance. Some features, such as age or tumor size, were continuous, while others were binary or categorical, such as the presence of certain symptoms or treatments. Label encoding and one-hot encoding convert categorised input to numerical representations that the model can handle. This was crucial for Conv1D, which is sensitive to the scale of input data. 4) Data Overfitting: Overfitting was a problem given the dataset's modest size. Utilizing the 5-fold cross-validation technique is essential for enhancing the accuracy and reliability of your model. This approach effectively divides your dataset into five distinct subsets, allowing for thorough training and validation. This method allowed us to evaluate the model's efficacy over multiple sets of data, ensuring that it was generalisable beyond a single training set. A reliable assessment model is generated

to measure the performance by averaging the data from each fold. By addressing these challenges through targeted preprocessing techniques, it has been ensured that the data was cleaned, normalized, and ready for sequential analysis using the Conv1D-LSTM architecture. These steps were crucial in handling the unique complexities of our dataset and ensuring that the model could accurately capture both local and long-range dependencies.

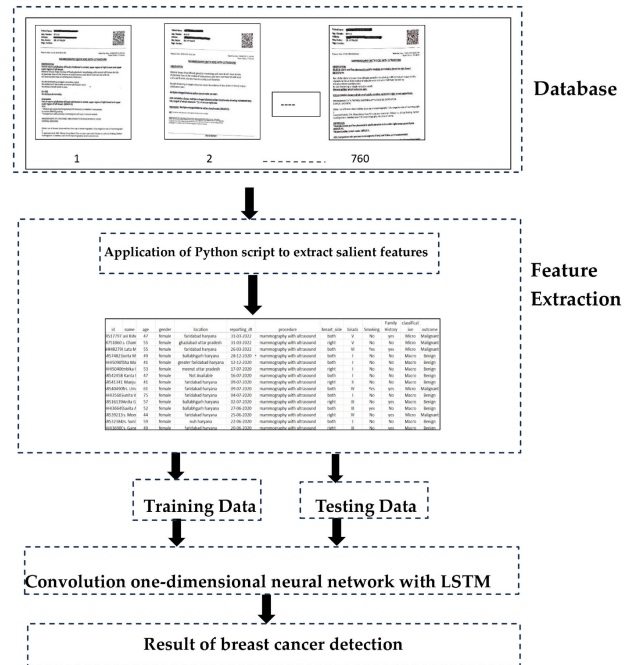


FIGURE 1. Research methodology for proposed work.

#### Pilot and Pre-Pilot Study

In the preliminary study, mammography scans were evaluated by an oncologist. A sample of data from a hundred patients was collected for a pilot study. Factor analysis was conducted on the patients' reports, revealing that some of the data was not significant and needed to be excluded. Subsequently, the relevant scan reports were chosen for further data analysis.

#### Construct Validity

A thorough KMO (Kaiser-Meyer-Olkin) test was conducted to ensure the suitability of our dataset for exploratory factor analysis. With a KMO score of 0.86 well exceeding the minimum threshold of 0.80, our data proved highly appropriate for factor analysis. Moreover, our assessment of internal consistency and validity using Cronbach's alpha coefficient resulted in an impressive value of 0.83, affirming the robustness and reliability of the relationships among the variables and bolstering the credibility of our research findings.

#### Correlation analysis

Correlation analysis is crucial in identifying potential risk factors associated with breast cancer. When dealing with non-linear datasets, Pearson correlation analysis emerges as

the most suitable technique for gaining insights into the relationships between demographic characteristics, lifestyle choices, and breast cancer. Furthermore, it aids in identifying variables correlated with early-stage cancer, thus contributing to developing more effective screening programs. In our investigation as shown in Figure 2, a great emphasis on examining fundamental attributes that play pivotal roles in understanding the disease. To examine the potential correlation between age, an intrinsic factor, and Breast Imaging Reporting and Data System (BI-RADS) scores, offering important insights into age-related trends in breast health. The comprehensive exploration involves identifying and characterizing calcifications, a crucial radiological feature, to elucidate patterns and associations. Additionally, we thoroughly investigate whether the affected breast side contributes to distinctive observations, thus providing a strong and comprehensive foundation for breast cancer analysis. The thorough process of data validation and correlation analysis ensures that the most reliable and significant factors are including in training, which ultimately improves the accuracy and relevance of breast cancer detection.

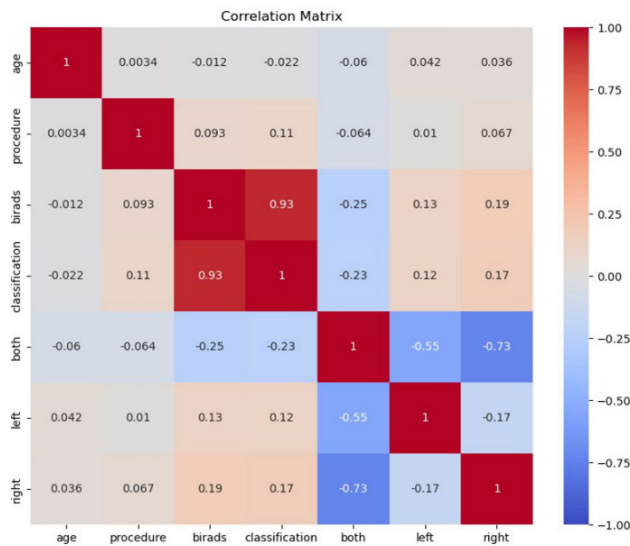


FIGURE 2. Pearson correlation analysis.

## A. DATASET DESCRIPTION

To eliminate the restrictions mentioned in the literature study, non-linear real-time data in the form of mammography reports of patients related to breast cancer has been self-gathered from October 2019 to October 2022 from a healthcare hospital. The collection of data contains 760 occurrences, 523 of which are innocuous and 237 of which are carcinogenic and have Bi-Rads levels of IV or above.

In Table 2, the list of features is listed which is extracted from the mammography reports of the patients. The location represents patients that belong to a geographical area, ID is the distinctive identification of the patient, name, and age

TABLE 2. List of attributes.

| S.No | Attributes        | S.No | Attributes     |
|------|-------------------|------|----------------|
| 1    | Location          | 8    | Procedure      |
| 2    | Patient ID        | 9    | Breast Side    |
| 3    | Name              | 10   | Reporting_dt   |
| 4    | Age               | 11   | Bi-Rad         |
| 5    | Gender            | 12   | Calcification  |
| 6    | Breast Density    | 13   | Smoking        |
| 7    | Menopausal status | 14   | Family History |
| 15   | Outcome           |      |                |

of the patient, the outcome indicates whether the tumor is benign or malignant, gender describes the sexuality of patients, breast side indicates left, right, or bilateral, procedure indicates the process followed for the diagnosis of a disease like mammography with or without ultrasound, BI-RADS groups the outcomes in 0 to 6 categories [32], reporting\_dt (reporting date) mentions the timeline of reporting, microcalcification signifies early chances of having cancer in breast and macro-calcifications are calcareous encrustations that develop in breast tissue [33]. Table 3 presents the data distribution of breast cancer patients based on distinguished parameters. The attributes that impacted the health conditions related to breast cancer are age, breast density, calcification, smoking, menopausal status, family history, and Bi-rad level.

- 1) Age: The experimental analysis and study show breast cancer majorly happens between 45 to 60 years of age [34].
- 2) Breast Density: Higher breast density in women tends to have a higher risk for breast cancer [35].
- 3) Calcification: Females show the presence of microcalcification likely to have breast cancer [15].
- 4) Smoking: Women who were contacted by the nicotine from cigarettes as children or who have been exposed to smokers are highly likely to get breast carcinoma in their lifetime [34].
- 5) Family History: A family history of breast cancer is strongly associated with an elevated chance of getting the illness. The biological relative afflicted by the same ailment or at risk for breast cancer is reported by less than 20% of patients who receive a breast cancer diagnosis [35].
- 6) Bi-Rad Grading: The grading level ranges from 0 to 6. The Bi-Rad levels 4,5, and 6 are prone to breast cancer [36].

The patients' records were retrieved in PDF format as shown in Figure 2, which prevents the ML or deep learning models from being used to process the data directly. It is necessary to convert the summary report of individual patients into a tabular format for efficient decision-making [37].

Steps for data extraction in tabular format have been given below:

- 1) Input the reports of patients.
- 2) Split the data and extract the features.

**TABLE 3.** Data distribution of breast cancer patients.

| Factors                        | Frequency |
|--------------------------------|-----------|
| Age (min-max)                  | 30-84     |
| Calcification                  |           |
| Micro                          | 187       |
| Macro                          | 573       |
| Bi-Rads                        |           |
| I, II, III                     | 523       |
| IV, V, VI                      | 237       |
| Smoke                          |           |
| Yes                            | 182       |
| No                             | 578       |
| Breast Side                    |           |
| Left                           | 170       |
| Right                          | 214       |
| Both                           | 376       |
| Procedure                      |           |
| Mammography with ultrasound    | 453       |
| Mammography without ultrasound | 307       |
| Menopausal Status              |           |
| Yes                            | 133       |
| No                             | 627       |
| Breast Density                 |           |
| Fatty                          | 105       |
| Fibro glandular dense          | 310       |
| Heterogeneously dense          | 257       |
| Extremely dense                | 88        |
| Family history                 |           |
| Yes                            | 139       |
| No                             | 621       |
| Outcome                        |           |
| Benign                         | 523       |
| Malignant                      | 237       |

- 3) Read until <eof> is encountered.
- 4) Create a data frame and convert the data frame into a tabular format.
- 5) Display result.

### B. PROPOSED MODEL- CONV1D-LSTM

Convolutional neural networks are used to examine data patterns. Convoluting data and recognition of trends are utilized to do this. Even with minimal CNN layers, the system can recognize edges and corners [38]. Such patterns may be transferred via our neural network and start to recognize more complicated traits. This characteristic makes CNNs particularly good at spotting patterns [39]. Conv1D are often employed to evaluate pictures of breast tissues to diagnose breast cancer.

Convolution is a computational technique that reduces a tensor, matrix, or vector to a minuscule one to “summarize”

**FIGURE 3.** Sample of the mammography report collected from the hospital.

it. The data can be summarized across one axis of your input matrix if it has one, and you can do the same for all  $n$  levels of a tensor if it does. Conv1D and Conv2D combine information in one or more dimensions. For illustration, concatenate a vector into a shorter vector as seen below. Convolve a “long” vector  $A$  with a total of  $n$  elements into a “short” vector  $B$  with  $n-m+1$  elements by a weight vector  $W$  with  $m$  elements:

$$B_i = \sum_{j=0}^{m-1} a_{i+j} \quad (1)$$

where  $i = [1, n-m+1]$  Therefore, if the weight matrix and vector both have lengths of  $n$ ,

$$w_i = \frac{1}{n} \quad (2)$$

then transformation will create a scalar or vector of one dimension that represents the median value of all the parameters in the provided input matrix. A shifting average of two dimensions and beyond is produced when the identity matrix with weights is one lower than the source matrix.

$$\begin{bmatrix} a : a_1 & a_2 & a_3 \\ w : \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} b : \frac{a_1 + a_2}{2} & \frac{a_2 + a_3}{2} \end{bmatrix} \quad (3)$$

This one-dimensional convolution is more efficient and operates identically as a multidimensional convolution, but it only considers a one-dimensional array with components. To visualize it, a matrix with either rows, columns, or a single dimension can be considered. After multiplying it, we obtain an array of numbers with the same structure but greater or



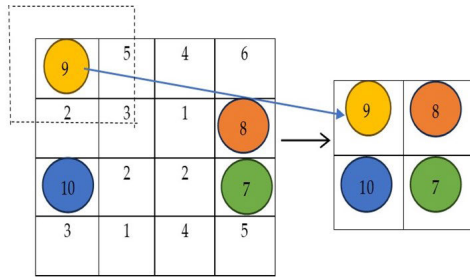


FIGURE 4. Dimensionality reduction using max pooling layer.

lesser numbers, which lets both maximize or minimize the magnitude of the data.

This study employed structured characteristics taken from PDF files in CSV format to describe each patient's breast cancer diagnostic values. Conv1D is applied to capture local patterns within these feature sequences. Although Conv1D is typically used in image-based data, it is equally effective in detecting correlations and trends across adjacent data points in sequential data. Conv1D helps identify localized patterns within the patient's feature sequence, which is crucial for understanding short-range dependencies. However, while Conv1D excels at extracting these localized patterns, it may fail to capture the longer-range dependencies essential in sequential data, where relationships between distant data points carry significant information. At this point LSTM layer complements Conv1D. LSTM is designed to model temporal dependencies and track relationships across the entire sequence of features. This ability is crucial in the dataset, where understanding how certain features evolve across time or relate to earlier patterns can improve the performance of classification tasks. Therefore, combining Conv1D with LSTM allows the architecture to leverage Conv1D's strength in capturing localized trends while LSTM models the long-term dependencies and relationships across the feature sequence. This synergy is useful for structured feature sequences in breast cancer diagnosis, ensuring that both short—and long-range patterns are captured effectively.

**MAX Pooling Layer:** This layer works as a filter with CNNs [40]. The purpose of the layer is to reduce the aspect ratio while preserving features in it [41]. For performing the max pooling function, the filter is required to be fixed with any particular window size, for instance [2 X 2], [3 X 3], [4 X 4] etc. The selected window size calculates the maximum value from the given window and converts it into a matrix of the same window size. An example is illustrated in Figure 3, taking window size as [2 X 2] and a reduced dimensionality matrix is generated by selecting maximum values as 9,8,10 and 7 from each window. The feature map generated by the synthesis of Conv-1D layer was further reduced dimensionally by the max pool layer. This feature reduction is done because the features that were gained from the PDF reports dataset hold irrelevant data in the form of noise. Max pooling helps mitigate this by selecting the

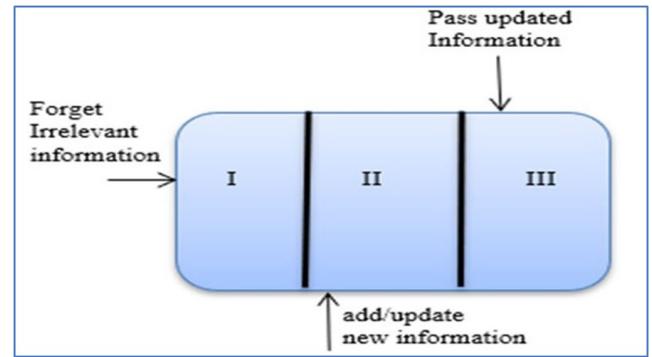


FIGURE 5. Architecture of LSTM.

most prominent values in small regions and focusing on the most important patterns. Max pooling, by downsampling the output of the Conv1D layer, minimizes computational burden and prevents overfitting. This is achieved by eliminating less significant features, ensuring the model identifies significant trends in data. These patterns are critical for identifying subtle indicators of breast cancer and contribute to the model's generalization ability. After max pooling, the LSTM layer can operate on a more distilled feature set version. This allows the LSTM to focus on learning long-term dependencies and relationships between the most critical features, leading to more accurate predictions. As a result, the max pooling serves as an intermediary, simplifying the representation of features so that the following LSTM layer may train more efficiently. LSTM: A branch of RNN, used for the identification of extraction of attributes [42]. In contrast to conventional RNN, LSTM is an enhanced RNN that uses memory segments to address gradient and vanishing-point issues. A network using LSTM can recall and link data from previous periods with data from today, which is another significant distinction between it and RNNs [43]. Long-term states of cells are generated by cumulative states of cells which include their three states. It includes a mix of various layers like 1ConvD Layer, 1 layer of LSTM, multiple convolution layers, pooling layers and 1 output layer which includes SoftMax Function for the result [44]. LSTM networks are specifically designed to handle time-series data or sequences where patterns across time must be recognized. The initial phase is to choose whether to continue or not to retain the data from the step that preceded it.

$$f_t = \sigma(x_t * U_f + H(t-1) * W_f) \quad (4)$$

xt= enter the present date and time

Uf= value corresponding to the input

Ht-1= the prior timestamp's concealed state

Wf= coefficient matrix related to the concealed state

The above equation (4) allows the LSTM to decide which past information is no longer relevant to the current task of cancer detection (such as discarding outdated medical data or irrelevant features). In the next phase, the new/updated information has been added through the input gate. For new

information:

$$N_t = \tanh(x_t * U_c + H_{t-1} * W_c) \quad (5)$$

Equation (5) introduces the potential new information from the input, where the non-linear activation function ‘tanh’ normalizes the input. For updated information:

$$C_t = f_t * C_{t-1} + i_t * N_t \quad (6)$$

$C_t$  is the updated cell state, containing both the retained information from the past and the new information introduced at the current step. In the last phase, updated information has been passed through the output gate.

$$O_t = \sigma(x_t * U_o + H_{t-1} * W_o) \quad (7)$$

This step ensures that only the most relevant information is passed forward, which is particularly useful in detecting cancerous changes over time in medical data.

The LSTM tracks how these spatial features change over time. A non-cancerous cyst is monitored over multiple screenings. The LSTM can help detect long-term patterns, such as irregular growth, increased density, or other risk factors indicating a shift toward a cancerous state. By analyzing temporal dependencies, LSTMs can help predict whether a cyst’s development follows a pattern more commonly associated with malignancy. **Batch Normalization & Dropout Layer:** In deep learning models, overfitting poses a serious concern, especially when the predictive algorithm is complex and the dataset is limited or changeable. Generally due to the problem of overfitting model fails to perform on an unknown external dataset. To mitigate this, batch normalization and dropout were incorporated into the model architecture. The batch normalisation approach stabilises and accelerates the training process by normalising the inputs to each layer, ensuring that the input distribution stays stable throughout the training. In breast cancer classification, where the dataset consists of extracted features from PDF reports, the data can vary in scale or distribution. By normalizing the activations, batch normalization reduces the sensitivity to initial weights and helps prevent overfitting by reducing the internal covariate shift, which can cause the model to adapt too closely to the training data. It also allows for higher learning rates, speeding up convergence. Another technique to reduce overfitting is the dropout method, which suggests randomly dropping out a few neurons so that the model can perform well over the unknown and naïve dataset rather than depending purely on the attributes of the primary provided dataset. By introducing randomness, dropout encourages the model to learn distributed representations of the data, making it less likely to overfit [45]. Together, batch normalization and dropout work in tandem to ensure that the Conv1D-LSTM model doesn’t overfit the breast cancer dataset, allowing it to generalise well across different patient cases and make accurate predictions on new data. The magnitude of every input can be equalized by adjusting the data being provided. The initially activated neurons in the initial segment of the NN

must be calculated. Pre-activation is understood to be nothing more than the skew added to the balanced total of inputs. The formula for every  $i$ th layer of the pre-activation function is:

$$a1 = W1 * x + b1 \quad (8)$$

Then, apply activation fn. to the outcomes of every  $i$ th layer of pre-activation fn.

$$hi = g(ai(x)) \quad (9)$$

The architecture of the proposed Conv1D-LSTM model was meticulously designed based on the dataset’s characteristics and the nature of the classification task (breast cancer detection) shown in Figure 6. Each layer and its parameters were chosen to balance complexity, efficiency, and the need for robust feature extraction and pattern recognition. The number of units in the LSTM layer was chosen to ensure that the model could discover long-term associations in the data sequence without overfitting. Too few units limit the model’s capacity to identify intricate designs in the feature sequences, while an excessive number of units may lead to overfitting or unnecessary complexity, a risk that should be carefully considered. The chosen number strikes a balance by providing sufficient capacity to capture meaningful relationships between features while keeping the model efficient for training on the dataset, including dense layers after the LSTM, which is meant to provide a final abstraction layer before classification. Dense layers are fully connected, which allows the model to integrate the information learned by the Conv1D and LSTM layers. The high-dimensional outcome from the LSTM was progressively reduced into a more compact form appropriate for binary classification by varying the count of dense layers and the neurones in every single layer. Avoiding the addition of too many dense layers, which might lead to the algorithm memorising the training data instead of generalising, also helps avoid overfitting. Additionally, hyperparameters were adjusted through experimentation, using validation data to determine the optimal design. These factors included the amount of LSTM units and neurones in the dense layers. The model’s effective generalisation of unknown data was guaranteed using cross-validation approaches. The parameters were heavily influenced by the need for the model to effectively capture both local patterns (via Conv1D) and long-term dependencies (via LSTM). This balance ensures the model’s efficiency and accuracy in classifying breast cancer. First, we use Conv1D from TensorFlow to pass our input where we are using 16 filters and kernel size 3. In addition, we add Dropout and Batch Normalization layers to avoid overfitting problems. Next, MaxPool1D was combined for down sampling to make a hybrid model, afterwards, the data was passed into the LSTM layer for the extraction of features. At last, fully connected multiple dense layers are generated to show how neurones are coupled to neurones in the layer above them. Conv-1D produces a tensor of responses by constructing a kernel for convolution that is twisted with an associated input layer over a single structural

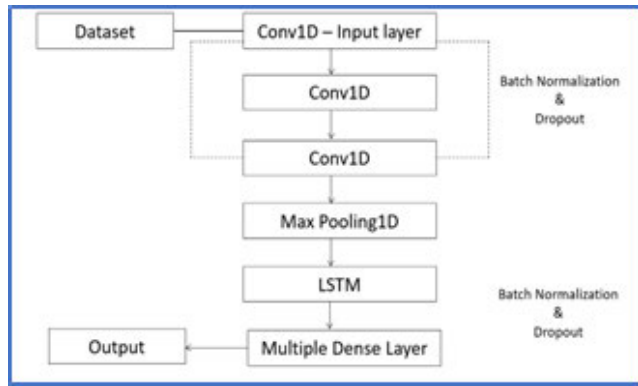


FIGURE 6. The proposed framework of CNN & LSTM with max pooling layer.

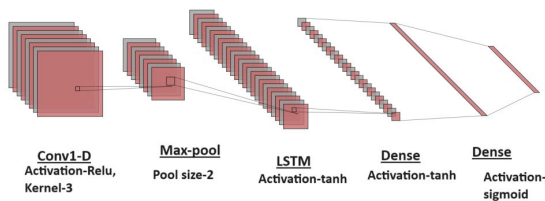


FIGURE 7. The architecture of the proposed system.

dimension. A neural network will frequently use extremely particular arrangements of weights, or a “conspiracy” of weights, the dropout layer helps to recognize these erroneous patterns. To dispel these theories, arbitrarily remove a portion of a layer’s unit of input at each round of training. This makes it far more difficult for the network to recognize such fictitious tendencies in the simulated data. Then, each batch that is received by a batch normalization layer is examined, and once each batch has been normalized with its average and standard deviation, the data is also scaled to a new scale using two trainable rescaling parameters. Later, passing the outcome of phase 2 into the LSTM with 32 units and using the tanh activation function. After adding one LSTM layer, 3 hidden layers with 128, 64, and 32 neurons have been added with the tanh activation function. Now finally, with two neurons and a SoftMax stimulation, we used the resulting layer.

Read all PDF files and extract relevant information using regular expressions: The algorithm starts by reading PDF files and extracting relevant information from them using regular expressions. This step is likely part of data collection and preprocessing.

#### 1) DATA PRE-PROCESSING

Employ data preprocessing techniques to perform feature extraction and cleaning of data: After extracting information from the PDF files, data preprocessing techniques are applied to clean the data and extract relevant features. Feature extraction involves selecting or deriving relevant attributes from the data. For data preprocessing, label encoding, one

#### Algorithm 1 Hybrid CNN-LSTM Model for Image Classification

**Input:** Input image set  $\{(x_i, y_i)\}$ ,  $1 \leq i \leq N$ ; Iterations (number of epochs)

**Output:** Weights  $w$

- 1) Read all pdf files and extract relevant information using regular expressions
- 2) Use data preprocessing techniques to perform feature extraction and data cleaning.
- 3) Split data into training and testing sets.
- 4) **for**  $epoch = 0$  **to**  $iter$  **do**

##### Feature extraction:

Using CNN branch to extract features:  $F = T(f(x_i))$ ;

##### 5) Hybrid Model Implementation:

$F$  is combined with LSTM for Hybrid Model with tanh activation and applied dropout and batch normalization:  $H = T(F(x))$ ;

$H$  is combined with Fully Connected Network and output to obtain  $V$ ;

##### 6) Loss and Perform Optimization:

Calculate Loss  $L_H$  using  $V$  according to Sparse Categorical Cross Entropy formula;

Check for early stopping if loss is constant for 3 iterations;

- 7) Update parameter  $w$  by minimizing total loss;

- 8) **return**  $w$ ;

hot encoding, feature scaling, k-fold cross-validation, and SMOTE (Synthetic Minority Oversampling Technique) for imbalanced classification have been used. K-fold cross-validation is utilized as a method for assessing prediction models. This strategy divides the input data into k separate folds, each as a validation set for one of the model’s k training and evaluation cycles. The model’s generalization performance is then calculated by aggregating the effectiveness metrics from each fold. Feature scaling helps in building precise and efficient models and is made simpler by the ability to convert the dataset’s characteristics onto a more uniform scale.

#### 2) TRAIN AND TEST DATASET

The dataset is split into a train and a test dataset, following common practice in machine learning to evaluate model performance. for  $epoch = 0, \dots, iter$  do: This loop iterates through a specified number of epochs (iterations) to train the model. Feature extraction: Within each epoch, a Convolutional Neural Network (CNN) branch is used to extract features from the data. These features are represented as ‘ $F = T(f(x_i))$ ’, where ‘ $f(x_i)$ ’ represents the features extracted from input image ‘ $x_i$ ’.

#### 3) HYBRID MODEL IMPLEMENTATION

The extracted features ‘ $F$ ’ is combined with a Long Short-Term Memory (LSTM) model with tanh activation. Dropout and batch normalization techniques implemented to

bolster the model's resilience, the outcome of this procedure is represented as  $H = T(F(x_i))$ .  $H$  is combined with a Fully Connected Network and output to obtain  $V$ : The ' $H$ ' is further processed by a fully connected neural network to obtain an output ' $V$ '.

#### 4) LOSS AND PERFORM OPTIMIZATION

In this step, the algorithm calculates a loss ' $LH$ ' using the output ' $V$ ' based on a formula involving Sparse Categorical Cross Entropy. This loss is a measure of how well the model is performing.

Check for early stopping if the loss is constant till 3 iterations: The algorithm monitors the loss over several iterations, and if the loss remains constant for three consecutive iterations, it checks for early termination. In order to avoid overfitting, early stopping involves terminating the training procedure if the model's efficacy remains unchanged. If the validation loss fails to improve substantially, the call back procedure stops training. The minimal change in validation loss necessary to carry on training is indicated by the ' $\text{min\_delta}$ ' variable. It is set to 0.0001 in this instance, which indicates that training will cease if the loss in validation does not decrease by a minimum of this number. When the validation effectiveness of the model reaches a plateau, this callback lowers the learning rate. By adjusting the processing rate progressively as required, it helps optimise the training process. The loss function plays a crucial role in evaluating the performance of binary classification models by measuring the difference between predicted outcomes and the actual desired outcomes. Adam is an optimization algorithm that combines the advantages of both the AdaGrad and RMSprop optimizers utilized for efficiently updating the model's parameters during training. Furthermore, the techniques trained for 5 epochs mean the model will see the entire dataset five times. These hyperparameters are crucial for training a deep learning model. The combination of callbacks, loss function, optimizer, and number of epochs helps in achieving better model performance, preventing overfitting, and efficiently adjusting the learning rate during training. Update parameter  $w$  by minimizing total loss: The algorithm updates the model's parameters represented as ' $w$ ' by minimizing the total loss. This typically involves using optimization techniques like gradient descent to adjust the model's weights. return  $w$ : Finally, the process returns the trained model weights, " $w$ " which may be used to forecast future data. This algorithm explains how a learning model is trained that combines CNN and LSTM architectures for some tasks, likely related to processing and extracting information from PDF files. The goal is to optimize the model's performance by updating its parameters based on the loss.

#### IV. EXPERIMENTAL RESULTS OF THE PROPOSED MODEL

The key concept underlying this strategy is to employ the CNN part as a feature extractor, to feed the most discriminatory aspects of the input to the LSTM part, resulting

in concurrent decreases in dimensionality. Furthermore, the suggested model employs a focal-loss (FL) function to reduce errors in forecasting and handle data inequalities.

The test outcomes indicate that the proposed architecture improves overall forecast accuracy, demonstrating notable stability and resilience in the model. Performance metrics were utilized to assess the proposed Conv1D-LSTM model, which incorporates a max-pooling layer. These metrics are frequently used to evaluate classification results, which have the goal of identifying an ordered label for every input occurrence. The matrix shows how many real positives (RP), real negatives (RN), fraudulent positives (FP), and fraudulent negatives (FN) the simulation generated on the experimental data.

As per the confusion matrix in Figure 6(a), the performance evaluation has been done with cross-validation techniques. A DL approach's accuracy statistic specifies the number of times the technique properly identifies an item of data. Accuracy can be defined as a percentage of accurately anticipated observations divided by the entire number of observations.

$$\text{Accuracy} = \frac{\text{Count of favorable observations}}{\text{Total count of observations}} \quad (10)$$

The inaccuracy caused by the framework is quantified as loss highlighted by the loss graph in Figure 7(b). However, an elevated loss the quantity typically implies that the framework delivers incorrect results, whereas a modest loss value suggests the approach contains much fewer flaws. After visualizing the accuracy graph in Figure 7(c), the attained accuracy of the proposed Conv1D-LSTM model is 99%.

An overall estimate of effectiveness along all the potential categories is provided by AUC. AUC in Figure 7(d) can be seen as the likelihood that the algorithm values a randomly chosen favourable instance higher than a randomly chosen adverse instance. The graph depicts a trend of decreasing loss over epochs, which is a positive sign for a learning model. The element on the X-axis (Epochs) represent show many times the model was trained using the complete training dataset. Every epoch denotes a single training cycle. Each epoch represents one complete training cycle. The element on the Y-axis (Loss) represents the loss value, a metric that indicates the degree to which the model's forecasts differ from ground truth values. The graph likely shows two lines:

- Train Loss: This line shows the loss estimated from the experimental data. The line decreases over time as the framework improves its predictions based on experimental data.
- Test Loss (Possible): The loss was calculated using a distinct test dataset from the one used to train the model. This constitutes an important measure for assessing the model's capacity to generalize to current information.

The graph shows an increase in accuracy on each epoch for both the training and test set. The x-axis in Figure 7 likely represents epochs but the graph only shows a small range (around 4 epochs). Both the train accuracy and test accuracy



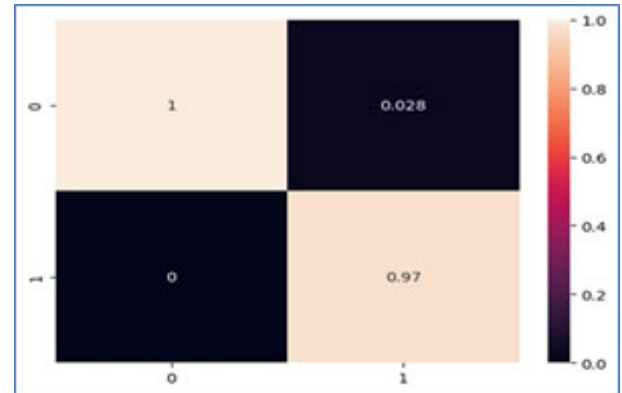
lines start at a relatively high value (around 0.9) and fluctuate slightly. While there's a slight upward trend for train accuracy, it's not definitive across all epochs shown. The test accuracy shows some fluctuation within the limited range. The graph's X-axis (False-Positive Rate - FPR) shows the fraction of all adverse cases mistakenly categorized as favorable by the model. The Y-axis number (True-Positive Rate - TPR) shows the ratio of favorable outcomes categorized adequately by the model. The value on the Y-axis (True Positive Rate - TPR) represents the proportion of positive instances that are correctly classified as positive by the model. The diagonal line reflects a random classifier. A random classifier has an equal chance of mistaking positive and negative instances, resulting in an ROC curve that follows the diagonal line. The curved line in the graph represents the performance of your model. The model performs better when the ROC curve expands in the upper-left corner. In Figure 7(d), the curve leans significantly towards the top-left corner, indicating good performance. The AUC value of 0.99 indicates AU\_ROC curve. This summarizes the framework's capacity to discriminate across classes.

In this section, the efficiency of a developed model Conv1D-LSTM with max-pooling layer and Softmax function has been compared with other machine and deep learning models: Logistic Regression, K-NN, ANN, and SVM on a collection of experimental data which is evaluated with a confusion matrix. As per the confusion matrix of the models, the performance evaluation has been done with 5 fold cross-validation techniques. The classification report in Table 4 shows a detailed analysis of the performance parameter. The graphical representation of confusion matrices and AU-ROC is depicted in Figures 8 and 9 respectively.

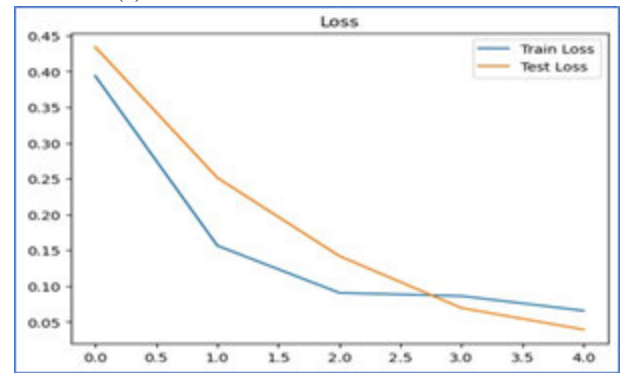
The analysis of the performances of classification models is done as per the cross-validation techniques. In Table 4, logistic regression gives the overfitting problem with the dataset, it is not capable of handling complex machine-learning problems because it simply uses a distance formula to classify objects. Other than that ANN and SVM give comparative results of accuracy of 88% and 89% respectively along with the same precision value, f1-score, and AU value of the ROC curve. The proposed model Conv1D-LSTM with a max-pooling layer outperforms with a maximum accuracy of 99%. The output of the study depicts that the suggested methodology (Conv1D-LSTM with max-pooling layer) surpasses both the DL and ML models.

## V. DISCUSSIONS, LIMITATIONS, AND FUTURE DIRECTIONS

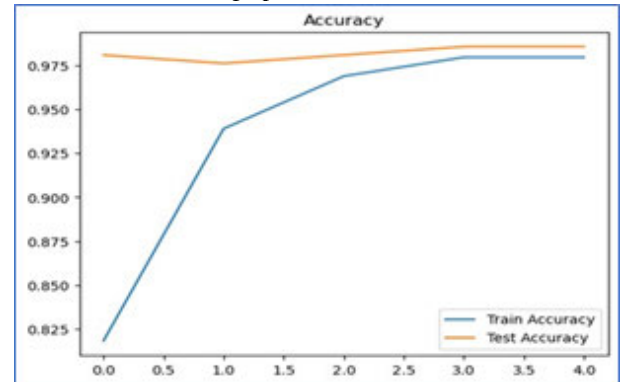
This section highlights the performance comparison of existing deep learning techniques with our proposed model. The accuracy for different algorithms is compared using image and non-image datasets such as WISCONSIN, WBC, MIAS, and CBIS-DDSM datasets. The same comparative study is shown in Table 5.



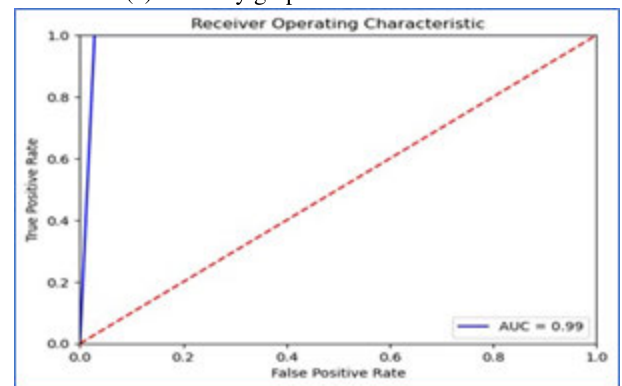
(a) Confusion Matrix of Conv1D-LSTM.



(b) Loss graph of Conv1D-LSTM.



(c) Accuracy graph of Conv1D-LSTM.



(d) AU-ROC of Conv1D-LSTM.

FIGURE 8. (a),(b),(c),(d) describes Conv-1D-LSTM.

Abdullah-Al Nahid et al. [46] BreakHis dataset for the experimental analysis. CNN-based architecture achieves an accuracy of 91% which is comparatively lower than the accuracy attained by the proposed Conv1D-LSTM model. The proposed model, despite using 80% training data achieved 99% accuracy and in the paper [27] ANN technique achieves an accuracy of 98.57% with 90% training data. The author in [48] measured XGBoost combined with the K-CGAN method and achieved an accuracy of 99.1% when applied to the WISCONSIN dataset. The Conv1D-LSTM model was confidently trained for just 5 epochs, optimizing computational resources while still achieving model convergence. Remarkably, this approach resulted in an accuracy that matches that of the 10,000-epoch model. The author [49] proposed a CNN-Bi-LSTM-GRU-AM model employed by the CBIS-DDSM dataset.

**TABLE 4.** Confusion matrix values.

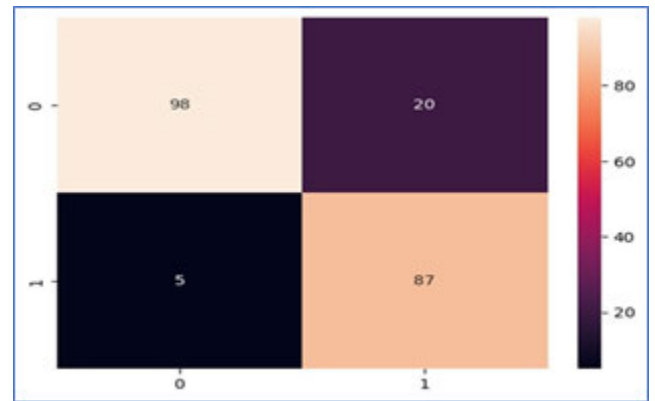
| Models      | True Positives | True Negatives | False Positives | False Negatives |
|-------------|----------------|----------------|-----------------|-----------------|
| Conv1D-LSTM | 1              | 0.97           | 0.028           | 0               |
| ANN         | 98             | 87             | 20              | 5               |
| K-NN        | 81             | 77             | 37              | 15              |
| LR          | 75             | 43             | 76              | 16              |
| SVM         | 92             | 26             | 35              | 57              |

The confusion matrix helps in calculating accuracy, recall and f1-score value. Table 4 is plotted through confusion matrix graphs.

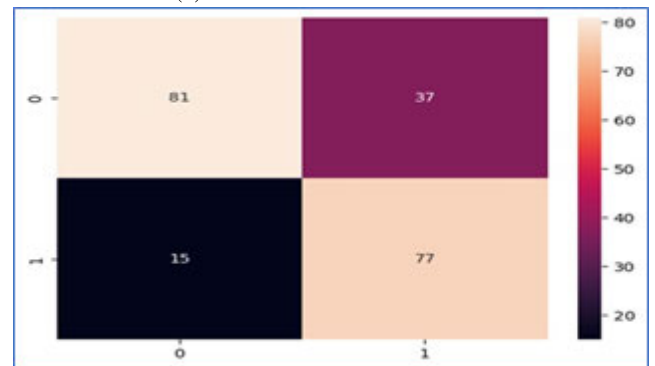
**TABLE 5.** Comparative analysis classification report of ML & DL models.

| Model    | Accuracy | Precision | F1-score | Specificity | Recall | AUC-ROC |
|----------|----------|-----------|----------|-------------|--------|---------|
| CNN-LSTM | 0.99     | 0.99      | 0.99     | 0.99        | 0.99   | 0.99    |
| ANN      | 0.88     | 0.83      | 0.88     | .83         | 0.95   | .89     |
| K-NN     | 0.75     | 0.68      | 0.75     | 0.68        | 0.84   | 0.76    |
| LR       | 1.00     | 0.56      | 0.71     | 1.00        | 1.00   | 1.00    |
| SVM      | 0.89     | 0.83      | 0.88     | 0.90        | 0.95   | 0.89    |

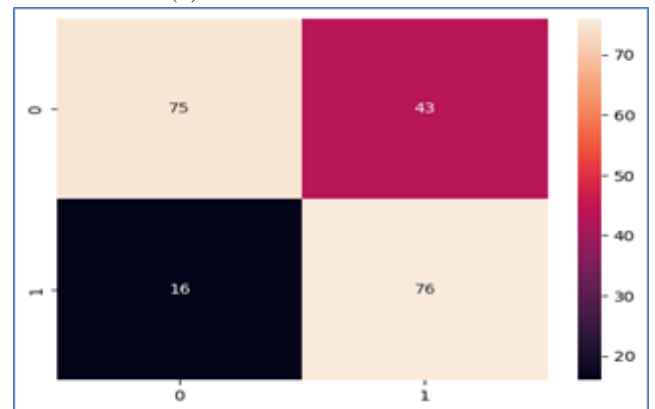
In this paper, principal component analysis is utilized to reduce dimension. Wang et al. [50] proposed an ensemble CNN-GRU model. In this paper, the dataset of female patients suffering from IDC was taken from Kaggle.com. Table 5 demonstrates that the combined approach can achieve an optimal accuracy of 86.21% while the designed model performs at 99%. The analysis of the performances of classification models is done as per the cross-validation techniques. In Table 5, logistic regression gives the overfitting problem with the dataset, it is not capable of handling complex machine-learning problems because it simply uses a distance formula to classify objects. Other than that ANN and SVM give comparative results of accuracy of 88% and 89% respectively along with the same precision value, f1-score, and AU value of the ROC curve. The suggested Conv1D-LSTM model achieves better, with an optimal accuracy of 99%. The outcome depicts that the suggested methodology (Conv1D-LSTM with max-pooling layer) surpasses both the DL and ML models. CNN has



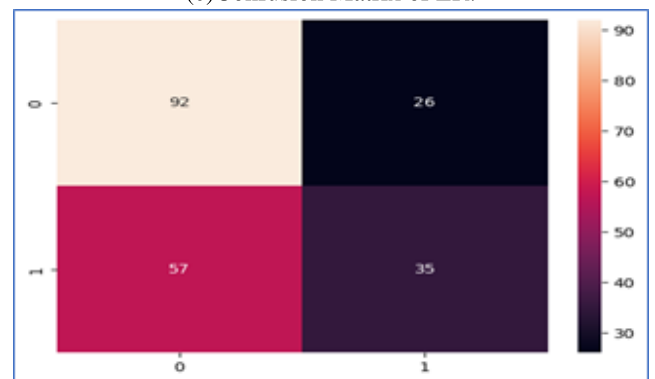
(a) Confusion Matrix of ANN.



(b) Confusion Matrix of K-NN.

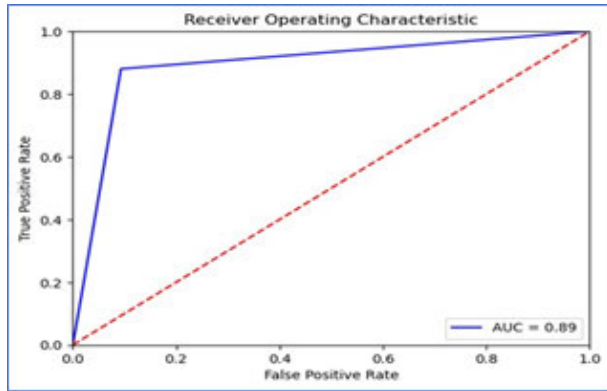


(c) Confusion Matrix of LR.

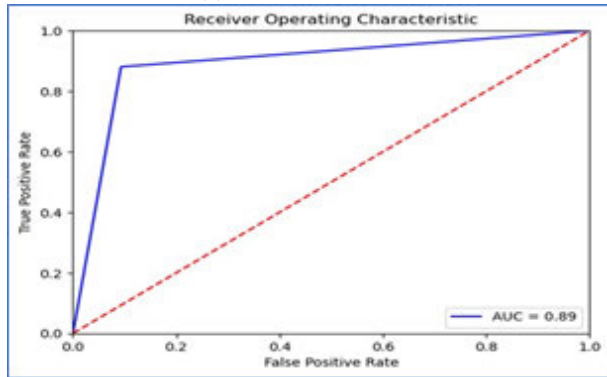


(d) Confusion Matrix of SVM.

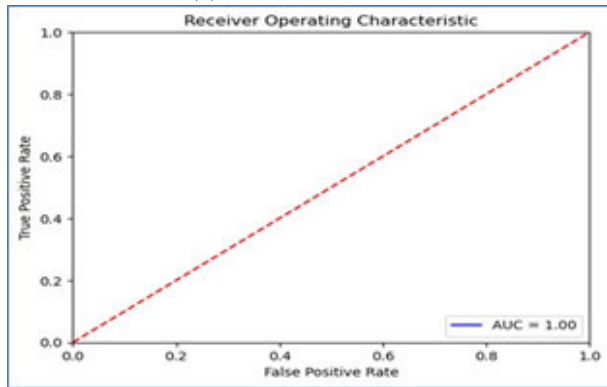
**FIGURE 9.** (a),(b),(c),(d) describes confusion matrix.



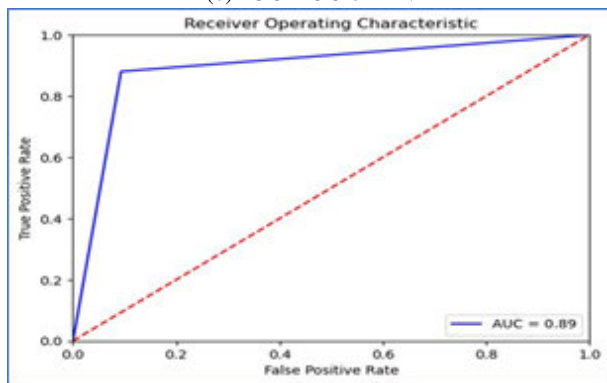
(a)AUC-ROC of ANN



(b)AUC-ROC of K-NN.



(c)AUC-ROC of LR.



(d)AUC-ROC of SVM.

**FIGURE 10.** (a),(b),(c),(d) describes AUC-ROC.**TABLE 6.** Analysis of Conv1D-LSTM with alternative learning methods.

| Model               | Exhibit Feature Learning | Handle Spatial Features | Handle Temporal Data |
|---------------------|--------------------------|-------------------------|----------------------|
| Conv1D-LSTM         | YES                      | YES                     | YES                  |
| ANN                 | YES                      | NO                      | NO                   |
| K-NN                | NO                       | NO                      | NO                   |
| Logistic Regression | NO                       | YES                     | NO                   |
| SVM                 | YES                      | NO                      | NO                   |

a feature extraction and learning mechanism that helps in distinguishing minority class instances. Additionally, CNN provides regularisation methods that assist in avoiding over-fitting to the dominant class. In this model, the max-pooling layer has been utilized for dimensionality reduction which minimizes the computation time. As the dataset used is spatial and temporal, CNN1D automatically learn spatial hierarchies of features. For 1D, data, this means detecting patterns over the sequence. In addition, LSTM layers are capable of learning long-term dependencies. Conv1D-LSTM model learn hierarchical features and capture temporal relationships, integrating the distinct benefits of 1D-CNNs with LSTMs. Whereas other learning models have certain limitations ANN cannot capture spatial hierarchies and dependencies as effectively as convolutional layers in CNNs. K-NN does not perform feature learning and handle temporal dependencies as it treats all features equally and independently. Logistic Regression is a linear model and does not perform feature learning. Also, it is not suitable for temporal dependencies or complex patterns in sequential data. SVM is not designed for sequential data and cannot capture long-term dependencies effectively [47]. In conclusion, a CNN1D-LSTM model excels over ANN, KNN, SVM, and Logistic Regression in handling sequential data due to its combined ability to learn hierarchical features with convolutional layers and capture long-term dependencies with LSTM units. This combination allows CNN1D-LSTM models to effectively recognize patterns, trends, and temporal relationships in data, making them superior for tasks involving time series, text sequences, and other sequential datasets. Other models either lack the feature learning capacity (ANN, SVM, Logistic Regression) or are not designed to handle temporal dependencies effectively (KNN, SVM, Logistic Regression), making them less suitable for such tasks.

**TABLE 7.** Evaluation of the Conv1d-LSTM accuracy with existing models.

| Model                                                         | Dataset                       | Accuracy (%) |
|---------------------------------------------------------------|-------------------------------|--------------|
| CNN                                                           | BreakHis dataset              | 91.00        |
| ANN                                                           | UCI repository dataset        | 98.57        |
| XGBoost with K-CGNA                                           | WISCONSIN dataset             | 99.10        |
| CNN-Bi-LSTM-GRU-AM                                            | CBIS-DDSM dataset             | 98.65        |
| CNN-GRU                                                       | PCam Kaggle dataset           | 86.21        |
| Proposed:CNN-LSTM with max-pooling layer and SoftMax function | Real data mammography dataset | 99.00        |

### A. THE PROPOSED MODEL LIMITATIONS AND FUTURE DIRECTIONS

Using a Conv1D enhanced alongside LSTM (long-short-term memory) for autonomous breast cancer detection may have the following drawbacks:

- **Interpretability and Explainability:** Using complex neural network models like 1D CNN with LSTM can make it hard to understand how the model makes its predictions. This lack of interpretability means that clinicians and patients might not easily see why the model came to a certain conclusion. Improving interpretability is important because it helps build trust in the model's decisions, showing that its predictions are reliable and based on understandable reasoning.

- **Data Heterogeneity:** It refers to the diversity and variations present in medical datasets. These variations can arise from multiple sources:

- 1) **Data Protocols:** Different hospitals might use various equipment, have different ways of taking data, or treat patients with different demographics (like age, ethnicity, or health conditions), the model might not work as well when it encounters data that's different from what it was trained on.
- 2) **Data Quality:** Variations in data quality can stem from factors such as equipment differences or noise levels. These factors can influence the interoperability and consistency of data utilized for training neural network algorithms.
- 3) **Patient Demographics:** Patient populations vary in age, gender, ethnicity, and underlying health conditions. These demographic differences can introduce biases or variations in data distribution, affecting the generalization capacity of the model across various patient populations.

- **Generalization:** For machine learning models to be helpful in the real world, they must successfully generalize across different demographics and healthcare systems. A generalization is the capacity to function correctly on untested data that can be different from the training set. The possibility of performance differences between various demographic groups in models trained on unbalanced or biased datasets raises concerns regarding justice and equity in healthcare applications.

- **Limited Availability of Annotated Data:** Annotated datasets with extensive labels for breast cancer data (such as lesion boundaries, and tumor characteristics). Since larger datasets are frequently required to capture the wide variety of cancer symptoms and changes, this shortage hinders the capacity to train models using machine learning.

In conclusion, the Conv1D+LSTM model holds promise for advancing breast cancer detection through its ability to analyze sequential and temporal patterns in medical data. While offering advantages such as improved feature extraction, multi-modal integration, and interpretability, its

real-world deployment necessitates addressing challenges related to data availability, model complexity, validation, and ethical considerations. By considering these variables, the Conv1D+LSTM model has the potential to significantly enhance the care of cancer patients by identifying the disease early and creating customized treatment regimens.

Implications of the proposed model in the real world:

- **Early Detection:** By leveraging both spatial and temporal features, the model can potentially detect breast cancer at earlier stages by identifying changes over time that single models might miss.
- **Robustness to Noise:** CNN layers can act as a noise filter in medical data. LSTM layers can then focus on the cleaned, high-level features for better prediction.
- **Applicability to various types of data:** This approach is versatile and can be applied to various types of data, including mammograms, ultrasound, MRI sequences, and even genomic data where sequential patterns are important.

**Future Research Directions:** Automated tumor screening the proposed model could explore several avenues to advance this field:

- **Enhancing Model Interpretability for Conv1D with LSTM:** In Conv1D layers, the model identifies patterns within the input data through learning. Also, the model helps in highlighting the areas that indicate potential abnormalities or lesions. With LSTM layers, the model can analyze sequences of data over time. This is crucial in medical contexts where changes in conditions (like tumour growth or treatment responses) matter. The model can help provide user-friendly interfaces that display outputs alongside visual explanations that can make it easier for clinicians to interpret results quickly and effectively.
- **Integration of Multi-Modal Data:** The combination of multimodal data with a Conv1D-LSTM framework for breast cancer diagnosis represents an innovative technique that utilizes various data sources to enhance diagnostic accuracy. The use of multi-modal data can help predict how patients will respond to specific treatments based on their integrated data profile and optimise their treatment plans.
- **Addressing Class Imbalance and Uncertainty:** The future scope is broad and impactful, more reliable and robust models can be developed with the usage of advanced data augmentation, ensemble methods and active learning. Coupling these improvements with real-world validations and ethical considerations can contribute to a great extent to the advancement of diagnosing breast cancer. Finally leading to accuracy in the diagnosis and improving the treatments.

Addressing these limitations and pursuing these research directions will contribute to the development of more interpretable, robust, and widely applicable autonomous breast cancer detection systems using various DL-based models, thus enhancing their clinical utility and impact. By working on these research directions, advancements in Autonomous Breast Cancer Detection using 1D CNN with LSTM will lead to more accurate, reliable, and interpretable systems that positively impact breast cancer diagnosis and patient care.



## VI. CONCLUSION AND FUTURE SCOPE

Automating cancer screening for breasts is a difficult task that must be overcome to enhance the treatment of patients. This paper introduces an advanced automated method for accurately identifying breast cancer utilizing a Conv1D-LSTM architecture with a max-pooling layer. Convolutional, max-pooling, and wholly completely connected layers were employed to extract essential characteristics from the data during the pretraining stage. An efficient classifying layer was then added to discern between innocuous and malignant samples, significantly improving the detection process's accuracy and reliability. After running fivefold cross-validation tests, accuracy is determined, and performance metrics are computed. As demonstrated, the processing of mass circularity in the dataset above will likely increase its overall size. This statistical assessment makes our decision-making process more transparent, and according to the trustworthy inquiry, it yielded 99% correctness. The suggested framework can eventually be examined for other datasets with more readily available records. Future model enhancements may include attention mechanisms, residual connections, and transfer learning approaches. These will focus on specific input sequences, reduce gradient problems, and enhance performance. Additionally, the use of pre-trained models from extensive datasets can accelerate convergence. Further research is needed to advance histological examination and to gather details for follow-up examinations.

## DATA AVAILABILITY

The dataset will be provided upon an appropriate request.

## CONTRIBUTIONS

The authors confirm their contribution to the article as follows: Study conception and design: Mitanshi Rastogi, Meenu Vijarania, and Neha Goel; Data collection: Neha Goel, Akshat Agrawal, and Cresantus N. Biamba; Analysis and interpretation of results: Meenu Vijarania, Neha Goel, and Celestine Iwendi; Draft manuscript preparation: Meenu Vijarania and Akshat Agrawal; Manuscript review and revision: Celestine Iwendi, Cresantus N. Biamba, Neha Goel, and Akshat Agrawal. All authors reviewed the results and approved the final version of the manuscript.

## ETHICS DECLARATIONS

Ethical guidelines were followed in the study. Throughout the study procedure, patient anonymity was upheld. All relevant preparatory steps have been completed to collect data.

## CONSENT FOR PUBLICATION

Not applicable.

## COMPETING INTERESTS

The authors declare no competing interests.

## ACKNOWLEDGMENT

The hospital management and patients provided support in the research process.

## REFERENCES

- [1] R. Mehrotra and K. Yadav, "Breast cancer in India: Present scenario and the challenges ahead," *World J. Clin. Oncol.*, vol. 13, no. 3, pp. 209–218, Mar. 2022.
- [2] A. Rasool, C. Bunternegchit, L. Tiejian, M. R. Islam, Q. Qu, and Q. Jiang, "Improved machine learning-based predictive models for breast cancer diagnosis," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, p. 3211, Mar. 2022.
- [3] E. Y. Kalafi, N. A. M. Nor, N. A. Taib, M. D. Ganggayah, C. Town, and S. K. Dhillon, "Machine learning and deep learning approaches in breast cancer survival prediction using clinical data," *Folia Biologica*, vol. 65, nos. 5–6, pp. 212–220, 2019.
- [4] M. M. Khan, S. Islam, S. Sarkar, F. Ayaz, M. K. Ananda, and T. Tazin, "Machine learning based comparative analysis for breast cancer prediction," *J. Healthc Eng.*, vol. 2022, no. 1, 2022, Art. no. 4365855.
- [5] V. P. T. Vy, M. M.-S. Yao, N. Q. Khanh Le, and W. P. Chan, "Machine learning algorithm for distinguishing ductal carcinoma in situ from invasive breast cancer," *Cancers*, vol. 14, no. 10, p. 2437, May 2022.
- [6] S. Bhise, S. Bepari, and S. D. K. Gaddekar, "Breast cancer detection using machine learning," *Int. J. Eng. Res. Technol.*, vol. 10, no. 7, p. 9103, 2021.
- [7] K. S. Priyanka, "A review paper on breast cancer detection using deep learning," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1022, no. 1, Jan. 2021, Art. no. 012071.
- [8] M. Hosni, I. Abnane, A. Idri, J. M. Carrillo de Gea, and J. L. F. Alemán, "Reviewing ensemble classification methods in breast cancer," *Comput. Methods Programs Biomed.*, vol. 177, pp. 89–112, Aug. 2019.
- [9] A. Begum, V. D. Kumar, J. Asghar, D. Hemalatha, and G. Arulkumar, "A combined deep CNN: LSTM with a random forest approach for breast cancer diagnosis," *Complexity*, vol. 2022, no. 1, pp. 1–9, Jan. 2022.
- [10] S. A. Alanazi, M. M. Kamruzzaman, M. N. I. Sarker, M. Alruwaili, Y. Alhwaiti, N. Alshammari, and M. H. Siddiqi, "Boosting breast cancer detection using convolutional neural network," *J. Healthcare Eng.*, vol. 2021, pp. 1–11, Apr. 2021.
- [11] M. Masud, A. E. E. Rashed, and M. S. Hossain, "Convolutional neural network-based models for diagnosis of breast cancer," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 11383–11394, Jul. 2022, doi: [10.1007/s00521-020-05394-5](https://doi.org/10.1007/s00521-020-05394-5).
- [12] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis," *IEEE Access*, vol. 8, pp. 96946–96954, 2020. [Online]. Available: <https://wiki.cancerimagingarchive.net/>
- [13] A. Sahu, P. K. Das, and S. Meher, "High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104292, doi: [10.1016/j.bspc.2022.104292](https://doi.org/10.1016/j.bspc.2022.104292).
- [14] M. Botlagunta, M. D. Botlagunta, M. B. Myneni, D. Lakshmi, A. Nayyar, J. S. Gullapalli, and M. A. Shah, "Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms," *Sci. Rep.*, vol. 13, no. 1, pp. 1–17, Jan. 2023, doi: [10.1038/s41598-023-27548-w](https://doi.org/10.1038/s41598-023-27548-w).
- [15] S. Azam, M. Eriksson, A. Sjölander, M. Gabrielson, R. Hellgren, K. Czene, and P. Hall, "Mammographic microcalcifications and risk of breast cancer," *Brit. J. Cancer*, vol. 125, no. 5, pp. 759–765, Aug. 2021.
- [16] T. R. Mahesh, V. V. Kumar, V. Vivek, K. M. K. Raghunath, and G. S. Madhuri, "Early predictive model for breast cancer classification using blended ensemble learning," *Int. J. Syst. Assurance Eng. Manage.*, vol. 15, no. 1, pp. 188–197, Jan. 2024, doi: [10.1007/s13198-022-01696-0](https://doi.org/10.1007/s13198-022-01696-0).
- [17] T. Srinivas, A. K. K. Madhusudhan, J. A. Dhanraj, R. C. Sekaran, N. Mostafaeipour, N. Mostafaeipour, and A. Mostafaeipour, "Novel based ensemble machine learning classifiers for detecting breast cancer," *Math. Problems Eng.*, vol. 2022, pp. 1–16, May 2022.
- [18] A. Sharma and D. Kumar, "Classification with 2-D convolutional neural networks for breast cancer diagnosis," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, Dec. 2022, doi: [10.1038/s41598-022-26378-6](https://doi.org/10.1038/s41598-022-26378-6).
- [19] A. R. Vaka, B. Soni, and S. Reddy, "Breast cancer detection by leveraging machine learning," *ICT Exp.*, vol. 6, no. 4, pp. 320–324, Dec. 2020, doi: [10.1016/j.ict.2020.04.009](https://doi.org/10.1016/j.ict.2020.04.009).

- [20] M. Divyavani, G. Kalpana, and P. D. R. Scholar, "An analysis on SVM & ANN using breast cancer dataset," *Aegaeum J.*, vol. 8, no. 1, 2020, Art. no. 36979.
- [21] S. A. Mohammed, S. Darrab, S. A. Noaman, and G. Saake, "Analysis of breast cancer detection using different machine learning techniques," in *Proc. Int. Conf. Data Mining Big Data*, vol. 1234. Singapore: Springer, pp. 108–117, doi: [10.1007/978-981-15-7205-0\\_10](https://doi.org/10.1007/978-981-15-7205-0_10).
- [22] S. Dalal, E. M. Onyema, P. Kumar, D. C. Maryann, A. O. Roselyn, and M. I. Obichili, "A hybrid machine learning model for timely prediction of breast cancer," *Int. J. Model., Simul., Sci. Comput.*, vol. 14, no. 4, pp. 1–21, Aug. 2023.
- [23] A. S. Elkorany and Z. F. Elsharkawy, "Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance," *Sci. Rep.*, vol. 13, no. 1, pp. 1–12, Feb. 2023, doi: [10.1038/s41598-023-29875-4](https://doi.org/10.1038/s41598-023-29875-4).
- [24] N. Hooda, R. Gupta, and N. R. Gupta, "Prediction of malignant breast cancer cases using ensemble machine learning: A case study of pesticides prone area," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 2, pp. 1096–1104, Mar. 2022.
- [25] S. Lee, M. Amgad, M. Masoud, R. Subramanian, D. Gutman, and L. Cooper, "An ensemble-based active learning for breast cancer classification," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 2549–2553.
- [26] F. Masood, "Breast cancer detection using machine learning," *Int. Res. J. Eng. Technol.*, vol. 8, no. 2, pp. 738–747, 2021, Art. no. 73847.
- [27] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, "Breast cancer prediction: A comparative study using machine learning techniques," *Social Netw. Comput. Sci.*, vol. 1, no. 5, pp. 1–14, Sep. 2020, doi: [10.1007/s42979-020-00305-w](https://doi.org/10.1007/s42979-020-00305-w).
- [28] C. Ming, V. Viassolo, N. Probst-Hensch, P. O. Chappuis, I. D. Dinov, and M. C. Katapodi, "Machine learning techniques for personalized breast cancer risk prediction: Comparison with the BCRAT and BOADICEA models," *Breast Cancer Res.*, vol. 21, no. 1, Dec. 2019.
- [29] S. Choudhary, P. Singh, M. Mittal, and G. Singh, "Automated breast cancer diagnosis based on machine learning algorithms," *Int. J. Adv. Netw. Appl.*, vol. 15, no. 6, pp. 6229–6238, 2024.
- [30] A. A. Jasim, A. A. Jalal, N. M. Abdulateef, and N. A. Talib, "Effectiveness evaluation of machine learning algorithms for breast cancer prediction," *Bull. Electr. Eng. Informat.*, vol. 11, no. 3, pp. 1516–1525, Jun. 2022.
- [31] B. Thakur, N. Kumar, and G. Gupta, "Machine learning techniques with ANOVA for the prediction of breast cancer," *Int. J. Adv. Technol. Eng. Explor.*, vol. 9, no. 87, 2022, Art. no. 23245.
- [32] A. A. N. Gunawan and W. Supardi, "Determination of breast tumor BIRADS variant using physical parameter on the mammography X-ray film," *J. Innov. Technol. Educ.*, vol. 3, pp. 241–250, Mar. 2016.
- [33] R. Baker, K. D. Rogers, N. Shepherd, and N. Stone, "New relationships between breast microcalcifications and cancer," *Brit. J. Cancer*, vol. 103, no. 7, pp. 1034–1039, Sep. 2010, doi: [10.1038/sj.bjc.6605873](https://doi.org/10.1038/sj.bjc.6605873).
- [34] Y.-S. Sun, Z. Zhao, Z.-N. Yang, F. Xu, H.-J. Lu, Z.-Y. Zhu, W. Shi, J. Jiang, P.-P. Yao, and H.-P. Zhu, "Risk factors and preventions of breast cancer," *Int. J. Biol. Sci.*, vol. 13, no. 11, pp. 1387–1397, 2017.
- [35] S. Lukasiewicz, M. Czezelewski, A. Forma, J. Baj, R. Sitarz, and A. Stanislawek, "Breast cancer—Epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review," *Cancers*, vol. 13, no. 17, pp. 1–30, Aug. 2021.
- [36] I. Domingues, P. H. Abreu, and S. J. BiRads, "Classification of breast cancer: A new PreProcessing pipeline for deep models training BI-RADS CLASSIFICATION OF BREAST CANCER," in *Proc. 25th IEEE Int. Conf. Image Process.*, Oct. 2018.
- [37] M. Rastogi, M. Vijarana, and N. Goel, "Extraction of primary data of breast cancer patients and implementation with ML techniques," *Unpubl Work.*, Oct. 2023.
- [38] A. B. Nassif, M. A. Talib, Q. Nasir, Y. Afadar, and O. Elgendy, "Breast cancer detection using artificial intelligence techniques: A systematic literature review," *Artif. Intell. Med.*, vol. 127, pp. 1–21, May 2022, Art. no. 102276.
- [39] Z. Guo, L. Xu, and N. Ali Asgharzadeholiaee, "A homogeneous ensemble classifier for breast cancer detection using parameters tuning of MLP neural network," *Appl. Artif. Intell.*, vol. 36, no. 1, pp. 1855–1875, Dec. 2022, doi: [10.1080/08839514.2022.2031820](https://doi.org/10.1080/08839514.2022.2031820).
- [40] S. Albawi, T. Mohammed, and S. Alzawi, "Layers of a convolutional neural network," in *Proc. Ict.*, 2017.
- [41] J. M. Vaz and S. Balaji, "Convolutional neural networks (CNNs): Concepts and applications in pharmacogenomics," *Mol. Diversity*, vol. 25, no. 3, pp. 1569–1584, Aug. 2021, doi: [10.1007/s11030-021-10225-3](https://doi.org/10.1007/s11030-021-10225-3).
- [42] D. Zhao, R. Jiang, M. Feng, J. Yang, Y. Wang, X. Hou, and X. Wang, "A deep learning algorithm based on 1D CNN-LSTM for automatic sleep staging," *Technol. Health Care*, vol. 30, no. 2, pp. 323–336, Mar. 2022.
- [43] I. Priyadarshini and C. Cotton, "A novel LSTM-CNN-grid search-based deep neural network for sentiment analysis," *J. Supercomput.*, vol. 77, no. 12, pp. 13911–13932, Dec. 2021, doi: [10.1007/s11227-021-03838-w](https://doi.org/10.1007/s11227-021-03838-w).
- [44] Z. Islam, M. Islam, and A. Asraf, "Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19," in *The COVID-19 Resource Centre is Hosted on Elsevier Connect, the Company's Public News and Information*, Jan. 2020, pp. 1–7.
- [45] W. Alsobhi, T. Alafif, W. Zong, and A. E. Abdel-Hakim, "Adaptive batch normalization for training data with heterogeneous features," in *Proc. Int. Conf. Smart Comput. Appl.*, 2023.
- [46] S. R. Gupta, "Prediction time of breast cancer tumor recurrence using machine learning," *Cancer Treatment Res. Commun.*, vol. 32, Jan. 2022, Art. no. 100602.
- [47] A.-A. Nahid, M. A. Mehrabi, and Y. Kong, "Histopathological breast cancer image classification by deep neural network techniques guided by local clustering," *BioMed Res. Int.*, vol. 2018, pp. 1–20, Jan. 2018.
- [48] E. Strelcenia and S. Prakoonwit, "Improving cancer detection classification performance using GANs in breast cancer data," *IEEE Access*, vol. 11, 2023, Art. no. 71594615.
- [49] M. Sarathkumar and K. S. Dhanalakshmi, *CBGAT: An Efficient Breast Cancer Prediction Model Using Deep Learning Methods [Internet]. Multimedia Tools and Applications*. Cham, Switzerland: Springer, 2023, doi: [10.1007/s11042-023-16640-y](https://doi.org/10.1007/s11042-023-16640-y).
- [50] X. Wang, I. Ahmad, D. Javeed, S. Zaidi, F. Alotaibi, M. Ghoneim, Y. Daradkeh, J. Asghar, and E. Eldin, "Intelligent hybrid deep learning model for breast cancer detection," *Electronics*, vol. 11, no. 17, p. 2767, Sep. 2022.



**MITANSHI RASTOGI** received the B.Tech. and M.Tech. degrees in computer science from MDU. She is a Research Scholar with the Department of Computer Science and Engineering, K. R. Mangalam University. She has published many papers in different Scopus-indexed conferences and scientific journals. Her research interests include AI, machine learning, and the IoT.



**MEENU VIJARANA** received the B.Tech. degree in information technology from MDU, Haryana, in 2005, the M.Tech. (IT) degree from GGSIPU, New Delhi, in 2007, and the Ph.D. degree from Amity University Haryana, in 2019. She has 16 years of experience in teaching. Currently, she is with K. R. Mangalam University Gurugram, Haryana, as an Associate Professor with the Department of Computer Science, Centre of Excellence, School of Engineering and Technology. She has published more than 45 research papers in international journals/conferences and book chapters of high repute. Her research interests include wireless networks and genetic algorithm, machine learning, and the Internet of Things.



**NEHA GOEL** is currently an Associate Professor with the Vivekananda Institute of Professional Studies. She was a topper in the NPTEL exam organized by IIT Kharagpur. She is actively involved in the NAAC and NBA Accreditation Process of College. She has published several research papers in national and international conferences/journals of repute. She is a member of review and editorial committees.



**AKSHAT AGRAWAL** received the B.Tech. degree in computer science and engineering from UPTU and the M.Tech. degree from USICT and GGSIPU, Delhi. He is currently pursuing the Ph.D. from GGSIPU. He is an Assistant Professor with Amity School of Engineering and Technology, Amity University Haryana. He has a total of 14 years of teaching and research experience. He has published a total of 47 research papers in SCIE/Scopus indexed journals and conferences.

He has guided 30 M.Tech. thesis and 52 B.Tech. projects. In June 2019, he was a Visiting Faculty with the Technical University of Kosice, Slovakia, and also visited the University of Nova, Portugal, through CABICIN Project (ERASMUS Funded). He actively participated in the peer review of research papers and book chapters. He served on the program committees for several conferences. He has edited one special issue on sustainable technological solutions for next-generation intelligent buildings in smart cities for EAI endorsed transactions on smart cities—EUDL (ISSN: 2518-3893) and two edited books for Wiley publication and IGI Global. His primary research interests include artificial intelligence, deep learning, artificial neural networks, speech processing, and image processing.

**CRESANTUS N. BIAMBA** joined the Department of Education, University of Gavle, Sweden, in January 2014, as a Senior Lecturer in education. He has been involved in high profile projects at the university and international level. As an academic, he has conducted basic and applied research in education, teaching and learning, curriculum development, educational leadership, and education for sustainable development. His research interests cover a wide area of educational and didactical questions. His specific interests include teacher education and professional development of school leaders, school governance and higher education, curriculum development, and early childhood education. He is interested in promoting endogenous and international research in leadership and management, adult education, and education for sustainable development.

**CELESTINE IWENDI** (Senior Member, IEEE) received the Ph.D. degree in electronics engineering. He is the Head of the Centre of Intelligence of Things, University of Bolton, U.K. He was a past ACM Distinguished Speaker, a Seasoned Lecturer, and a Chartered Engineer. He is a highly motivated Researcher and a Teacher with an emphasis on communication, hands-on experience, willing-to-learn and a 23 years technical expertise. He has developed operational, maintenance, and testing procedures for electronic products, components, equipment, and systems; provided technical support and instruction to staff and customers regarding equipment standards, assisting with specific, difficult in-service engineering; inspected electronic and communication equipment, instruments, products, and systems to ensure conformance to specifications, safety standards, and regulations. He is a wireless sensor network Chief of Evangelist, AI, and a ML and IoT Expert and Designer. He is a Reader (Professor) with the University of Bolton. He is also the Counselor of IEEE University of Bolton, Student Branch; a former Board Member of IEEE Sweden Section; and a fellow of The Higher Education Academy, U.K., and the Institute of Management Consultants, to add to his teaching, managerial and professional experiences. He is an Ambassador in the prestigious Manchester Conference Ambassador Program, a Visiting Professor of five universities; and an IEEE Humanitarian Philanthropist. He has received the prestigious recognition of the Royal Academy of Engineering through the Exceptional Talent Scheme, acknowledging his substantial contributions to artificial intelligence and its medical applications. Additionally, he takes pride in his three-year inclusion in Elsevier's publication, featuring the World's Top 2% Influential Scientists. He is the Chair of the Election Committee of IEEE Computer Society Worldwide 2024. He is an IEEE Brand Ambassador.

...