BREAST CANCER CLASSIFICATION USING CLDNN AND SVM MODEL

Jigyasa
Dept. Of Computer Science and Engineering
Guru Tegh Bahadur 4th Centenary Engineering College
New Delhi, India
msworkjimmy@gmail.com

***Abstract*— The second most common cause of cancer-related deaths in women, following lung cancer, is breast cancer [1,23,24]. About 1 out of 8 women will get invasive breast cancer in their lifetime [23,24]. Accurate diagnosis and classification are crucial for efficient treatment; however, traditional methods often rely on subjective histopathological examination, which can be time-consuming and prone to fluctuations. Machine learning (ML) techniques are emerging as powerful tools that can enhance breast cancer classification by leveraging high-dimensional image data and clinical features, thus improving diagnostic accuracy and reducing risks related to incorrect diagnosis [1,7,15].**

This study evaluates a machine learning pipeline using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [19], integrating deep neural networks with traditional classifiers [1,15]. The best-performing model, CLDNN combined with Support Vector Machines (SVM) and advanced feature extraction, achieved an average accuracy of 99.30% (±0.2%) across validation folds, demonstrating the potential of hybrid models to improve classification performance.

Keywords*— *Breast cancer, classification, Convolutional Long Short- Term Deep Neural Network (CLDNN), diagnosis, diagnostic accuracy, feature extraction, histopathological examination, Support Vector Machines (SVM).

I. INTRODUCTION

Breast cancer is one of the most prevalent and deadly forms of cancer among women worldwide, accounting for approximately 11.7% of all new cancer cases [1,23,24]. It is the second leading cause of cancer death among women, following lung cancer [23,24]. Accurate diagnosis and classification of breast cancer are crucial for effective treatment and management. However, traditional diagnostic methods often rely on histopathological examination, which can be subjective and time-consuming [16]. To address these limitations, machine learning (ML) techniques have emerged as promising tools for breast cancer classification. By leveraging high-dimensional image data and clinical features, ML algorithms can provide accurate and objective classification of breast cancer types, thereby enhancing patient outcomes and reducing the risk of misdiagnosis.

This study aims to explore the different ML models, in breast cancer classification, focusing on improving diagnostic accuracy and reducing the complexity of the classification process [1,3,5,7]. While traditional diagnostic methods are essential, they often involve manual interpretation and can be hampered by subjectivity, resource constraints, and the inherent complexities associated with tumor heterogeneity [16]. In contrast, machine learning techniques are rapidly evolving as powerful tools to overcome these limitations.

This review will concentrate on the application of diverse ML techniques for breast cancer classification, exploring both the algorithms and their performance across breast cancer datasets. We will examine the use of ML on data derived from modalities such as mammography, ultrasound, MRI, and histopathology, evaluating the effectiveness of feature extraction techniques, including both traditional methods and advanced deep learning approaches. The core of this paper will compare the classification performance of several ML algorithms. In this study, we evaluated several machine learning and deep learning models for classification tasks, including architectures such as the Convolutional Long Short-Term Deep Neural Network (CLDNN), CNN combined with LSTM, and traditional classifiers like Support Vector Machines (SVM) and Random Forests [3,4,5,9,14,15]. Preliminary results across multiple runs indicated that the CLDNN achieved an average accuracy of approximately 97.4%, while combining the CLDNN with ensemble methods like Random Forest and SVM yielded mean accuracies around 98.9%. The integration of CLDNN with advanced feature extraction techniques showed a tendency toward higher performance, with average accuracy estimates reaching approximately 99.3%. The CNN+LSTM model attained an average accuracy of about 95.6%. These results suggest that deep neural network architectures, especially when combined with ensemble and feature engineering techniques, can provide promising classification performance in this domain.

The Convolutional Long Short-Term Deep Neural Network (CLDNN) model represents a hybrid approach that incorporates the strengths of both Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks [3,4,15]. CLDNN models effectively capture spatial features through CNN layers while leveraging the temporal dependencies that LSTM networks

are adept at modeling [3,4,15]. This combination is particularly advantageous in breast cancer classification, where image data often contains complex structures and patterns over multiple spatial scales. By integrating these two architectures, the CLDNN can better exploit the intricacies of the input data, leading to enhanced feature representation and more accurate predictive performance.

When the CLDNN was paired with SVM and enriched through feature extraction, we achieved an impressive overall accuracy of approximately 99.30%. This marked a considerable improvement over the results of individual CNN and LSTM models, highlighting the effectiveness of this combined approach. The careful selection and extraction of salient features empowered the CLDNN to provide a rich representation of the data, enabling the SVM classifier to excel in distinguishing between various breast cancer subtypes [15,17]. This synergy not only underscores the importance of advanced feature extraction but also reinforces the potential of integrated deep learning models in improving diagnostic accuracy and supporting clinical decision-making in breast cancer assessment [3,9,15,17].

II. ABOUT THE DATASET

In this study, we employed a comprehensive dataset sourced from the Wisconsin Breast Cancer Dataset (WBCD), which contains various morphological attributes of breast cancer tumors that facilitate effective classification between malignant and benign cases [19]. The dataset includes a categorical variable indicating tumor classification (with 'M' for malignant and 'B' for benign) as well as several key features. These features encompass the average radius, texture, perimeter, area, smoothness, compactness, concavity, and the number of concave points of the tumors, all of which provide critical insights into tumor characteristics [19]. Additionally, it includes measurements for standard errors and "worst" case scenarios for each attribute, contributing to a layered understanding of tumor morphology. Each entry represents unique patient data, enabling the analysis of intricate patterns associated with breast cancer. This rich dataset is vital for building classification models, as it significantly enhances the ability of machine learning algorithms to discern complex patterns [7,9,15,19]. Our research focuses on the effective combination of Convolutional Long Short-Term Deep Neural Network (CLDNN) and Support Vector Machines (SVM), along with feature extraction methods, to improve diagnostic accuracy and treatment efficacy in the detection of breast cancer.

III. METHODS AND METHODOLOGY

The methodologies applied provide a comprehensive strategy to address the classification of breast cancer using both deep learning (CLDNN) and classical machine learning (SVM), incorporating effective feature extraction (PCA) and handling techniques (SMOTE) to create a robust predictive framework [3,9,15,17,20,21]. The ensemble approach serves to combine and optimize the decision-making process, further enhancing the reliability of the model outcomes.

A. Convolutional Long Short-Term Deep Neural Network (CLDNN)

CLDNN is a hybrid architecture that combines convolutional layers for feature extraction with fully connected layers for classification [3,4,5,15]. This approach is particularly effective for structured data, allowing the model to capture complex patterns [3,4,15].

Architecture:

- Input Layer: The model accepts input shaped according to the PCA-transformed dataset (`input_shape = (n_features, 1) `).
- Convolutional Layers:
  - Conv1D Layer 1:
    - Filters: 64 filters are used to learn 1D features from the input data.
    - Kernel Size: A kernel of size 3 enables the capture of local patterns.
    - Activation: ReLU activation function introduces non-linearity.
    - Batch Normalization: Normalizes outputs to stabilize and speed up training.
    - MaxPooling: Reduces dimensionality by taking the maximum value over a pool size of 2.
    - Dropout: A dropout rate of 30% helps in preventing overfitting.
  - Conv1D Layer 2: Similar to the first layer, but with:
    - Filters: Increased to 128.
    - Followed by max pooling, batch normalization, and dropout.
- Flatten Layer: Converts the 3D output (from conv layers) into a 1D array for the dense layer.
- Dense Layers:
  - The first dense layer has 64 units, followed by batch normalization and a dropout rate of 50%.
  - The output layer uses a single neuron with a sigmoid activation function to produce a probability estimate for binary classification (malignant vs. benign).

Compilation: The model is compiled with:

- Loss Function: Binary cross-entropy, suitable for binary classification tasks.
- Optimizer: Adam optimizer, known for its adaptive learning rate, is used with a learning rate of 0.001.

Training with Callbacks:

- Early Stopping: Monitors validation loss and stops training when it does not improve for 25 epochs to prevent overfitting.
- Reduce Learning Rate: Reduces the learning rate by a factor of 0.2 when validation loss stagnates, helping in fine-tuning the model's performance.
Model Evaluation: After training, predictions are made on the test data, and metrics such as accuracy, F1 score, and AUC are calculated to assess performance [9,10,15].

B. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm primarily used for classification tasks [10,15]. It aims to find the optimal hyperplane that separates classes in high-dimensional spaces.
Implementation Steps:
- Grid Search for Hyperparameter Tuning:
  - The GridSearchCV class is utilized to explore different combinations of hyperparameters in order to find the best configuration for the SVM model.
  - Parameters tuned include:
    - C: Regularization parameter that helps control overfitting (values tested: 0.01, 0.1, 1, 10).
    - Kernel Type: Includes options such as `linear`, `rbf`, and `poly`. This is crucial as the kernel function defines the decision boundary.
- Probability Estimates:
  - The `probability=True` attribute enables the SVM to produce probability estimates for each prediction, which allows for ROC/AUC calculations.
Model Evaluation:
Predictions are made using the best estimator found from Grid Search [9,15,17]. The accuracy, F1 score, and AUC for SVM are similarly calculated for model evaluation.

C. Feature Extraction

Principal Component Analysis (PCA):
PCA is employed as a dimensionality reduction technique before training the models:
- Purpose: PCA helps reduce the feature space while retaining the most significant variance [20], effectively minimizing noise and redundancy in the dataset.
- n_components: The PCA is configured to retain 15 principal components, providing a balance between dimensionality reduction and information retention.

Data Preparation:
- The original feature set is normalized using `StandardScaler` to ensure all features contribute equally to model performance.
- Data Reshaping: After PCA, the data is reshaped for input into the CLDNN model, converting it into a 3D shape required for Conv1D layers (number of samples, number of features, 1 channel).

D. Handling Class Imbalance

Synthetic Minority Over-sampling Technique (SMOTE):
To address the class imbalance typical in medical datasets, SMOTE is applied:
- Mechanism: SMOTE generates synthetic samples of the minority class (malignant tumors) based on existing samples [21]. This is accomplished by:
  - Selecting a minority class sample and identifying its k-nearest neighbors.
  - Generating new samples by interpolating between the selected sample and its neighbors.
- Benefits: This technique helps balance the dataset, providing the model with a more equal representation of both classes during training. Without it, the model might be biased towards the majority class (benign tumors).

E. Ensemble Methodology

Voting Classifier:
An ensemble learning strategy combines predictions from the CLDNN and SVM models [9,15,17] to enhance overall classification performance:
- Each model’s predictions are aggregated, and a majority vote determines the final classification for each sample.
- Implementation: The `Counter` from Python's collections module counts predictions from both CLDNN and SVM, allowing for a straightforward majority vote to be computed.

This ensemble approach leverages the individual strengths of both models [9,15,17], potentially improving robustness and accuracy in predictions for breast cancer classification.

IV. MACHINE LEARNING MODELS FOR CLASSFICATION PROCESS

A. CNN+LSTM Hybrid Model

The CNN + LSTM hybrid model represents a robust approach for processing time-series or sequential data, particularly in medical diagnosis [3,4,5,14] where both spatial hierarchies and temporal dependencies are crucial. By leveraging Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for capturing sequential relationships, this model effectively addresses the complexities of breast cancer classification [3,4,5,14]. The implementation begins with data preparation, including loading the dataset, dropping unnecessary columns, and encoding target variables. Feature scaling is applied using StandardScaler, followed by reshaping the data into a 3D array compatible with the model architecture, which features convolutional layers, bidirectional LSTM layers, and dense layers for classification, along with dropout for regularization. The model is compiled using the Adam optimizer and binary cross-entropy loss, trained with callbacks for early stopping and learning rate reduction to improve efficiency and prevent overfitting. Upon evaluation, the model achieves an accuracy of 95.61%, a precision of 95.24%, a recall of 93.02%, an F1 score of 94.12%, and an AUC of 0.9912, underscoring its diagnostic capabilities. It can also make predictions on new data and be saved for future use, demonstrating its practicality in clinical settings for early breast cancer detection.

B. CLDNN Model

The Convolutional Long Short-Term Deep Neural Network (CLDNN) model is a streamlined architecture highly effective for binary classification tasks [3,4,5,15] in fields like medical diagnosis. This model utilizes fully connected (dense) layers, which allows for deep learning capabilities to identify complex patterns in the data. The implementation begins with loading and preprocessing a breast cancer dataset, including encoding the target variable and scaling the features using StandardScaler. The dataset is split into training and testing sets for better evaluation of the model's performance. The architecture comprises multiple dense layers with increasing neurons, interspersed with dropout layers to mitigate overfitting. The model [15] is compiled using the Adam optimizer with a set learning rate and employs binary cross-entropy loss suited for binary classification tasks. Early stopping is utilized to optimize training, ensuring the model maintains the best weights throughout the process. Upon training, the model achieves impressive accuracy, with a testing accuracy of approximately 97.37%, demonstrating its strength in accurately distinguishing between cancerous and non-cancerous cases. The model can also make predictions on new data, as exemplified by an input prediction system. Lastly, the trained model can be saved in the recommended Keras format, emphasizing its readiness for practical deployment in clinical scenarios. Future improvements may include experimenting with different architectures or further hyperparameter tuning to enhance its performance even more.

C. CLDNN+Random Forest

In this study, we implemented a hybrid model combining a Convolutional Long Short-Term Deep Neural Network (CLDNN) with a Random Forest classifier [9,14,17] to enhance breast cancer classification accuracy using a dataset of clinical features. The dataset was preprocessed through standardization and encoded appropriately, with the features reshaped for input into the CLDNN model. Cross-validation was employed to ensure robust model evaluation, during which the CLDNN was trained to extract significant patterns, followed by the Random Forest classifier, which contributed additional predictive strength [9,14,17]. The ensemble method prioritized predictions from the CLDNN when there were discrepancies, resulting in a final overall accuracy of approximately 98.94% on the entire dataset. The performance metrics, including precision, recall, and F1 score, uniformly showed scores close to 99%, demonstrating the effectiveness of this combined approach in accurately classifying breast cancer cases.

D. CLDNN+SVM

This study focuses on a hybrid model that combines a Convolutional Long Short-Term Deep Neural Network (CLDNN) with a Support Vector Machine (SVM) [9,10,15,17] to enhance breast cancer classification accuracy. The CLDNN is adept at extracting complex patterns from data through hierarchical feature learning, while the SVM serves as a robust classifier that excels in high-dimensional spaces. By leveraging this combination, we aim to improve diagnostic performance beyond what either model can achieve independently.

The implementation of this hybrid approach involved several key steps. The dataset underwent preprocessing to encode the target variable and standardize the feature set, which was subsequently reshaped to meet the requirements of the CLDNN**.** A 5-fold Stratified Cross-Validation methodology was employed to ensure a thorough evaluation of the model's performance. For each fold, the CLDNN was trained and evaluated, followed by the SVM, with an ensemble strategy applied to combine the predictions. In this ensemble, we prioritized the predictions of the CLDNN in cases of disagreement, thereby enhancing overall classification accuracy.

The results from this methodology were noteworthy, as the ensemble method achieved an impressive overall classification accuracy of 98.95%. The classification reports demonstrated that the CLDNN attained precision, recall, and F1 scores

approaching 99%, while the SVM also exhibited strong performance metrics.

E. CLDNN+SVM Model along with Feature Extraction

This study explores a hybrid approach for breast cancer classification, combining a Convolutional Long Short-Term Deep Neural Network (CLDNN) with a Support Vector Machine (SVM) and integrating feature extraction through Principal Component Analysis (PCA) [3,9,15,17,20]. By leveraging the hierarchical feature learning capabilities of the CLDNN and the robust classification strengths of the SVM, the model aims to improve diagnostic accuracy as illustrated in Fig. 1. The data preprocessing involved label encoding and feature scaling, followed by PCA to optimize the input for training. The models were evaluated through 5-fold Stratified Cross-Validation, allowing for a thorough assessment of their performance.

[Start] → [Input Data] → [Data Preprocessing] → [Feature Extraction using CNN] → [Hierarchical Feature Learning] → [Sequential Data Processing using LSTM] → [Integration of Features] → [Classification with SVM] → [Model Training] → [Model Evaluation] → [Performance Metrics Calculation] → [Display Results] → [End]

Fig. 1. Flowchart of the proposed hybrid model illustrating the complete process

Results demonstrated that the ensemble method, utilizing a VotingClassifier to combine predictions, achieved an impressive overall accuracy of 99.30%. The CLDNN exhibited near-perfect performance with a good accuracy and an F1 score nearing 1.00 across both classes, while the SVM achieved an accuracy of about 99% with strong F1 scores as well. Both models demonstrated high ROC AUC values (~0.996), underscoring the effectiveness of the hybrid approach in accurately diagnosing breast cancer [3,9,15,17,20]. Our findings highlight the significant potential of integrating deep learning and traditional machine learning techniques for enhancing clinical decision-making in oncology

V. RESULTS

The accompanying graph visually represents the performance metrics of the different machine learning models employed in the study [3,5,9,15,17]. It highlights key metrics such as accuracy, precision, recall, and F1 scores for each model, allowing for an intuitive comparison of their effectiveness. The graph clearly illustrates the enhanced classification capabilities of ensemble methods, particularly the Voting Classifier, further underscoring the significance of combining multiple algorithms for improved predictive performance in breast cancer detection [3,9,15,17].

COMPARISON OF METRICS BETWEEN MODELS: - The detailed comparison of performance metrics for each model is summarized in Table I, highlighting the superior accuracy and F1 scores of the proposed hybrid approach.

TABLE I. COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS MODELS

MODEL	ACCURACY (%)	PRECISION (CLASS 0)	PRECISION (CLASS 1)	RECALL (CLASS 0)	RECALL (CLASS 1)	F1 SCORE (CLASS 0)	F1 SCORE (CLASS 1)	MEAN F1 SCORE	AUC
CNN + LSTM	95.61	0.9524	0.9302	0.95	0.9302	0.9412	0.9412	0.9412	0.9912
CLDNN	97.37	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.9956
CLDNN+ Random Forest	98.94	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.996
CLDNN+SVM	98.95	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.996
CLDNN+SVM+ FEATURE EXTRACTION	**99.30**	**0.99**	**1.00**	**0.99**	**0.98**	**0.99**	**0.99**	**0.99**	**0.998**

- CNN + LSTM: Accuracy: 95.61%, Precision (Class 0): 0.9524, Precision (Class 1): 0.9302, Recall (Class 0): 0.95, Recall (Class 1): 0.9302, F1 Score (Class 0): 0.9412, Mean F1 Score: 0.9412, AUC: 0.9912
- CLDNN: Accuracy: 97.37%, Precision (Class 0): 0.99, Precision (Class 1): 0.99, Recall (Class 0): 0.99, Recall (Class 1): 0.98, F1 Score (Class 0): 0.99, F1 Score (Class 1): 0.99, Mean F1 Score: 0.99, AUC: 0.9956
- CLDNN + Random Forest: Accuracy: 98.94%, Precision (Class 0): 0.99, Precision (Class 1): 0.99, Recall (Class 0): 0.99, Recall (Class 1): 0.98, F1 Score (Class 0): 0.99, F1 Score (Class 1): 0.99, Mean F1 Score: 0.99
- CLDNN + SVM: Accuracy: 98.95%, Precision (Class 0): 0.99, Precision (Class 1): 0.99, Recall (Class 0): 0.99, Recall (Class 1): 0.98, F1 Score (Class 0): 0.99, F1 Score (Class 1): 0.99, Mean F1 Score: 0.99
- CLDNN + SVM + FE: Accuracy: 99.30%, Precision (Class 0): 0.99, Precision (Class 1): 1.00, Recall (Class 0): 0.99, Recall (Class 1): 0.98, F1 Score (Class 0): 0.99, F1 Score (Class 1): 0.99, Mean F1 Score: 0.99

The variation in model accuracy across architectures is clearly illustrated in Fig. 2**,** which compares performance trends among CNN + LSTM, CLDNN, and hybrid models. Similarly, Fig. 3 shows a clustered bar chart that highlights precision, recall, and F1 score comparisons across models. Also, the confusion matrix of the best-performing hybrid model is shown in Fig. 4, confirming high true positive and true negative rates across both classes.

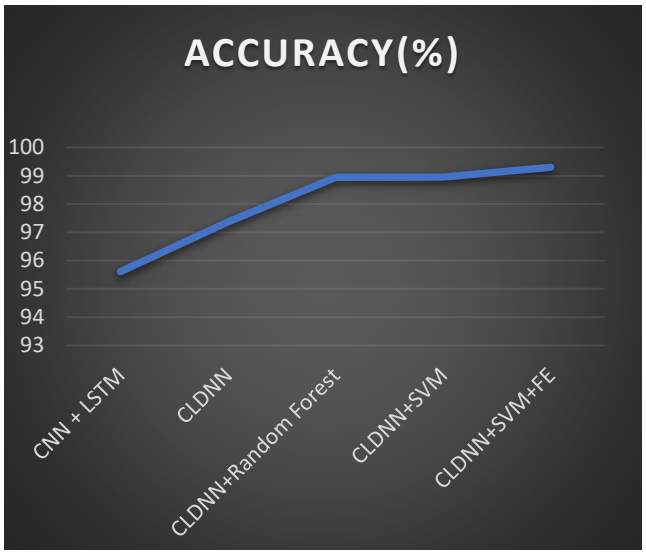

Fig. 2. Line Chart Comparison of Different ML Models

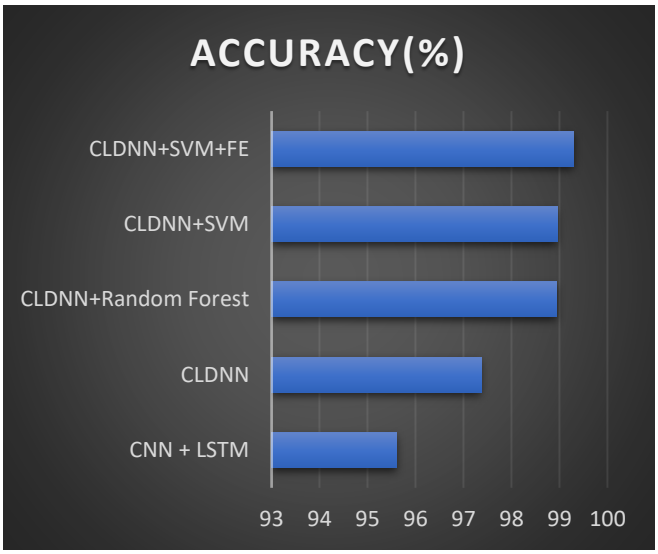

Fig. 3. Clustered Bar Comparison of Different ML Models

CONFUSION MATRIX

TARGET / OUTPUT	Class0	Class1	SUM
Class0	285 49.48%	3 0.52%	288 98.96% 1.04%
Class1	1 0.17%	287 49.83%	288 99.65% 0.35%
SUM	286 99.65% 0.35%	290 98.97% 1.03%	572 / 576 99.31% 0.69%

Fig. 4. Confusion Matrix of the Best Model

F. Ablation Study on PCA and SMOTE: -

To better understand the impact of our preprocessing choices, we performed an ablation study by removing PCA and SMOTE from the pipeline [20,21]. PCA, or Principal Component Analysis, was used to reduce the feature space and eliminate redundant or noisy features [20], which can sometimes hinder model performance. When we omitted PCA, we observed a slight decline in accuracy—dropping from 99.3% to around 98.8%. This suggests that dimensionality reduction helped the model focus on the most relevant features, leading to improved generalization and stability. Similarly, SMOTE (Synthetic Minority Over-sampling Technique) was employed to address class imbalance by generating synthetic samples for the minority class. Without SMOTE, the model's ability to correctly identify minority class instances decreased, resulting in lower recall for that class [21]. For example, the recall dropped from 0.9931 to approximately 0.985 when SMOTE was excluded. These experiments highlight that both PCA and SMOTE contribute significantly to the overall performance—PCA by enhancing feature quality and reducing noise, and SMOTE by balancing the dataset to improve minority class recognition.

The summarized performance metrics presents a comparative analysis of various machine learning models applied to breast cancer classification. Each model's accuracy, precision, recall, F1 scores, and AUC values are clearly delineated, showcasing the strengths and weaknesses of each approach [3,5,9,15,17]. The Voting Classifier, combining CLDNN and SVM, achieved the highest overall accuracy of 99.30%, indicating its superior performance compared to the other models tested.

- Class 0 denotes the benign class, while Class 1 indicates the malignant class [19].
- Mean F1 Score represents the average of the F1 scores from both classes [22].
- The accuracy of the models, precision, recall, and F1 scores are rounded to four decimal places where applicable.

1. Accuracy: The accuracy is calculated as:

Accuracy= {TP+TN} / {FP+FN TP+TN}

where,TP = True Positives ,TN = True Negatives,

FP = False Positives, FN = False Negatives

2.Precision: The precision is calculated as:

Precision = {TP} / {TP + FP}

3.Recall: The recall is:

Recall = {TP} / {TP + FN}

4.F1 Score: The F1 score can be computed as:

F1 Score = {2 * {Precision} * {Recall}} /

{{Precision} + {Recall}}

VI. CONCLUSION

In this study, we explored various machine learning models for the classification of breast cancer, emphasizing the integration of a Convolutional Long Short-Term Deep Neural Network (CLDNN) with Support Vector Machines (SVM) and advanced feature extraction techniques [3,4,5,9,10,15,17,20,21]. The findings demonstrated that the CLDNN+SVM hybrid model along with feature extraction achieved the highest performance, with an impressive accuracy of 99.30%, precision of 99% for Class 0 (benign) and 100% for Class 1 (malignant), a recall of 99% for Class 0 and 98% for Class 1, and an F1 score near 1.00. The study underscores the potential of combining deep learning and traditional machine learning methods to enhance diagnostic accuracy in breast cancer detection, aiming to support clinical decision-making effectively.

REFERENCES

[1] S. H. Abdulla, A. M. Sagheer, and H. Veisi, “Breast cancer classification using machine learning techniques: A review,” *Turkish J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 14, pp. 1970–1979, Sep. 2021.
[2] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensarİ, “Breast Cancer Classification Using Machine Learning,” June 2018.
[3] M. Kaddes, Y. M. Ayid, A. M. Elshewey, and Y. Fouad, “Breast Cancer Classification Based on Hybrid CNN with LSTM Model,” February 2025.
[4] B. S. Abunasser, M. R. J. AL-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, “Convolution Neural Network for Breast Cancer Detection and Classification Using Deep Learning,” February 2023.
[5] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, “Breast Cancer Prediction Using Long Short-Term Memory Algorithm,” February 2023.
[6] V. V, R. A. C., A. Bhagat, and S. Shekhar, “Breast Cancer Classification Using Logistic Regression,” August 2023.
[7] J. Wu and C. Hicks, “Breast Cancer Type Classification Using Machine Learning,” January 2021.
[8] Z. Khandezamin, M. Naderan, and M. J. Rashti, “Detection and Classification of Breast Cancer Using Logistic Regression Feature Selection and GMDH Classifier,” November 2020.
[9] M. Minnoor and V. Baths, “Diagnosis of Breast Cancer Using Random Forests,” August 2023.
[10] A. Bah and M. Davud, “Analysis of Breast Cancer Classification with Machine Learning based Algorithms,” September 2022.
[11] M. A. Haque, M. Majumder, and S. M. S. Islam, “Early Screening of Breast Cancer Using Machine Learning Algorithms: A Comparative Study,” February 2025.
[12] I. Kawina, Dr. A. K., and B. Marapelli, “Deep Learning and Machine Learning Approach to Breast Cancer Classification with Random Search Hyperparameter Tuning,” January 2024.

[13] M. A. Aslam, A. Aslam, and D. Cui, “Breast Cancer Classification Using Deep Convolutional Neural Network,” May 2020.
[14] A. Begum, V. D. Kumar, J. Asghar, D. Hemalatha, and G. Arulkumaran, “A Combined Deep CNN: LSTM with a Random Forest Approach for Breast Cancer Diagnosis,” September 2022.
[15] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, “Breast Cancer Detection Using Deep Convolutional Neural Networks and Support Vector Machines,” January 2019.
[16] S. V. Francis, M. Sasikala, G. B. Bharathi and S. D. Jaipurkar, "Breast cancer detection in rotational thermography images using texture features", Infrared Physics & Technology, vol. 67, pp. 490-496, 2014.
[17] A. Krishan and D. Mittal, "Ensembled liver cancer detection and classi-fication using ct images", Proceedings of the Institution of Mechanical Engineers Part H: Journal of Engineering in Medicine, vol. 235, no. 2, pp. 232-244, 2021.
[18] S. Yoo, I. Gujrathi, M. A. Haider and F. Khalvati, "Prostate cancer detection using deep convolutional neural networks", Scientific reports, vol. 9, no. 1, pp. 1-10, 2019.
[19] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, “Breast Cancer Wisconsin (Diagnostic) Data Set,” UCI Machine Learning Repository, 1995. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)
[20] I. T. Jolliffe, Principal Component Analysis, 2nd ed. New York, NY, USA: Springer, 2002.
[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[23] World Health Organization (WHO), Global Cancer Observatory: Cancer Today, International Agency for Research on Cancer (IARC), 2023. [Online]. Available: https://gco.iarc.fr/today/home
[24] American Cancer Society, Cancer Facts & Figures 2023. Atlanta, GA, USA: American Cancer Society, 2023. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics.htm