Vol. 06, No. 4 (2024) 1731-1740, doi: 10.24874/PES.SI.24.03.006

Proceedings on Engineering Sciences

www.pesjournal.net

SMART HYBRID MODELS FOR IMPROVED BREAST CANCER DETECTION

Nageswara Rao Gali
Panduranga Vital Terlapu[1]
Yasaswini Mandavakuriti
Sai Manoj Somu
Madhavi Varanasi
Vijay Telugu
Maheswara Rao V V R

Received 16.10.2023.
Received in revised form 23.11.2023.
Accepted 28.12.2023.
UDC –004.85:616-091.8

Keywords:

Breast Cancer, Deep Learning, CNN, SVM, Random Forest, VGG-16, XGBOOST

A B S T R A C T

Breast cancer (BC) ranks the second most prevalent cancer among women globally and is the leading cause of female mortality. The conventional method for BC detection primarily relies on biopsy; this might be time-consuming and error prone. The substantial lives lost due to BC underscores its significant threat. Mitigating this threat focuses on early detection and prevention by adopting novel techniques. Many researchers have turned to Machine Learning algorithms to develop prognosis systems. We employ a combination of deep learning (DL) and machine learning (ML) algorithms for BC identification. Our approach is a hybrid Convolutional Neural Network (CNN) model, which performs better than other experimental and existing models. This model effectively categorizes histopathological images into either benign or malignant classes. We explored various methodologies, including CNN, CNN in conjunction with Support Vector Machine (SVM), CNN with Random Forest, and VGG-16 combined with XGBOOST. This research seeks to enhance the accuracy and efficiency of BC diagnosis. It contributes to more effective early detection and improved patient outcomes.

© 2024 Published by Faculty of Engineering

1. INTRODUCTION

Breast Cancer (BC) is a medical condition characterized by the uncontrolled growth of cells within the breast. There are various types of breast cancer, with the specific subtype determined by the type of cells that have undergone malignant transformation. BC can originate in the epithelial cells of the breast's lobules (15%) or ducts (85%), which are part of the glandular tissue (Lukong, 2017; Kim & Villadsen, 2018). This cancer typically remains confined to the lobule or duct, often exhibiting no noticeable symptoms and minimal potential for metastasis (spreading to other body parts). BC will be diagnosed in approximately twenty-three billion women globally in 2020,

[1] Panduranga Vital Terlapu
Email: vital2927@gmail.com

resulting in 685,000 deaths (Khosasi, et al., 2023). By the end of 2020, about 78 million women would have been cancer-free for more than five years. BC may afflict women at any point of their lives and in any nation. Mortality rates for BC experienced fluctuations throughout the 1930s within the 1970s but began making improvements in the 1980s. These improvements can be attributed to early detection programs and a diverse range of treatment strategies aimed at eradicating invasive diseases (PanduRanga et al., 2019).

Table 1. Stages, Symptoms and Treatments of Breast Cancer

Stages	Symptoms	Treatment
Stage 0: the growth is tiny and just present in the organs where they have framed and have not yet developed into neighboring tissues.	There are, by and large, no side effects except that it can occasionally cause breast protuberance.	Hormone Therapy
Stage 1: the growth size is under 2 cm and may spread to different tissues in more modest regions.	Nipple discharge, dimpling of the skin, enlarging or redness of the breast (Vital et al. (2014)).	Radiation Therapy - 4 to 6 weeks
Stage 2: the growth develops to 20-50 mm in size, and some lymph hubs get impacted by disease.	Irregularity in the breast or armpit.	Hormone therapy - for patients above 70 yrs., Radiation Therapy, Chemotherapy, Surgery.
Stage 3: the cancer is more significant than 50 mm with more lymph hubs included. The infection might have migrated to the chest wall or the skin (Vital et al. (2015)).	Same as stage 1 and stage 2	Most commonly surgery, Combination Therapy (Radiation Therapy + Chemotherapy + Hormone Therapy).
Stage 4: The disease spreads to many regions of the human organism.	Weakness or numbness, dry cough, chest pain.	chemotherapy is controlled even before the medical procedure and Radiation Treatment alongside Hormone Therapy.

Survival rates for BC vary significantly, with high-income countries having over 90% survival rates, while countries like South Africa and India have 40% and 66% respectively. Implementing treatment strategies and early detection in resource-limited regions can improve global treatment. The World Health Organization's Global BC Initiative aims to reduce BC mortality by 2.5% annually, preventing 2.5 million fatalities by 2030.

The research gap in breast cancer detection is significant, with current methods lacking sensitivity across different subtypes and facing challenges in resource-constrained settings. The study introduces smart hybrid models, combining the strengths of various detection modalities to refine these mechanisms. The goal is to improve early detection rates and patient outcomes on a global scale, acknowledging the progress made in BC research and treatment. The integration of smart hybrid models aims to contribute to the evolution of BC detection, providing a more robust and efficient framework for clinicians and healthcare professionals.

2. LITERATURE REVIEW

The Literature Review section investigates the current state of information on BC detection. It investigates previous research, methodology, and technical breakthroughs in the area, laying the groundwork for understanding the present status of BC detection approaches and the gaps this study seeks to fill. Chiao et al. (2019) used deep learning, especially Mask R-CNN, to detect and segment breast lesions on UIs with a mean average accuracy of 0.75. Furthermore, the model had an overall accuracy of 85% in categorizing lesions as benign or malignant, indicating that it offers a viable non-invasive approach for complete breast lesion identification and classification. Xie et al. (2022) introduced two CNN models, one for DE speckling ultrasound images and another for classifying them as benign or malignant. When evaluated on the Mendeley Breast Ultrasound dataset, the models achieved an exceptional classification accuracy of 99.89%, outperforming recent methods in the field. Balasubramaniam et al. (2023) Utilized corrected ReLU in LeNet to address the "dying ReLU" issue, enhancing feature discriminability and improving BC diagnosis, detection, and, eventually, better outcomes for patients. Incorporating batch normalization mitigates internal covariate shift, reducing overfitting runtime and outperforming benchmark deep learning models, resulting in breast image identification accuracy of 89.91% is notable. This approach improves performance in recognizing features, segmentation, classification, and identifying BC tumors. Karthik et al. (2022) introduced a novel Stacking Ensemble comprising custom CNN architectures for classifying breast tumours into 'Normal,' 'Benign,' and 'Malignant' categories using ultrasound images. Their ensemble achieved impressive metrics through extensive experimentation with an accuracy of 92.15%, f1-score of 92.21%, precision of 92.26%, and recall of 92.17%. Table 2 shows some existing research works on Breast cancer datasets with various experimental models.

Table 2. Some of Research on BC using UI Images

Author	Description	Dataset Used	Result
MurtiRawat et al. (2020)	ML algorithms such as KNN, Logistic Regression, Ensemble learning with PCA are used for BC diagnosis.	Wisconsin BC diagnosis	The accuracy results are: 98.6% KNN, 97.9% using logistic regression, 99.3% using ensemble learning.
Khuriwal et al. (2018)	Convolutional Neural Network is used for classification	MIAS dataset	98% accuracy
Amrane et al (2018).	Naïve Bayes and KNN are used for BC classification	BC Dataset	KNN accuracy-97.51% NB Classifier- 96.19%

Sahu et al. (2023) proposed the Shuffle-Net-Res-Net scheme, rigorously validated on diverse BC modalities, including mini-DDSM, BUS2, and BUSI. The results reveal that it outperforms current approaches, with impressive accuracy percentages of 99.17% and 98.00% for abnormal and malignancy identification in mini-DDSM datasets and 96.52% & 93.18% for BUSI datasets, respectively. The model attains an impressive 98.13% BUS2 malignancy detection accuracy. Kabir et al. (2021) provided unique ways for BC categorization from B-mode UIs based on WCP pictures. The classical feature-based method achieves over 97% accuracy by modelling ultrasound statistics with the RiIG distribution and utilizing various features with low ANOVA p-values. In contrast, the custom-made CNN achieved 98% accuracy.

3. MATERIALS AND METHODOLOGIES

3.1 Proposed model

Figure 1 shows the proposed model for BC identification through the BC Image dataset. Gather a comprehensive dataset of malignant and benign breast cancer images, forming the basis for model training and evaluation. Perform essential pre-processing tasks such as resizing, augmenting, and cropping on the breast cancer images. These steps will ensure uniformity and enhance data quality for subsequent analysis. Utilize Convolutional Neural Network (CNN) and VGG-16 deep learning models to extract meaningful features from breast cancer images. Fine-tune these pre-trained models using the gathered data to optimize feature extraction.

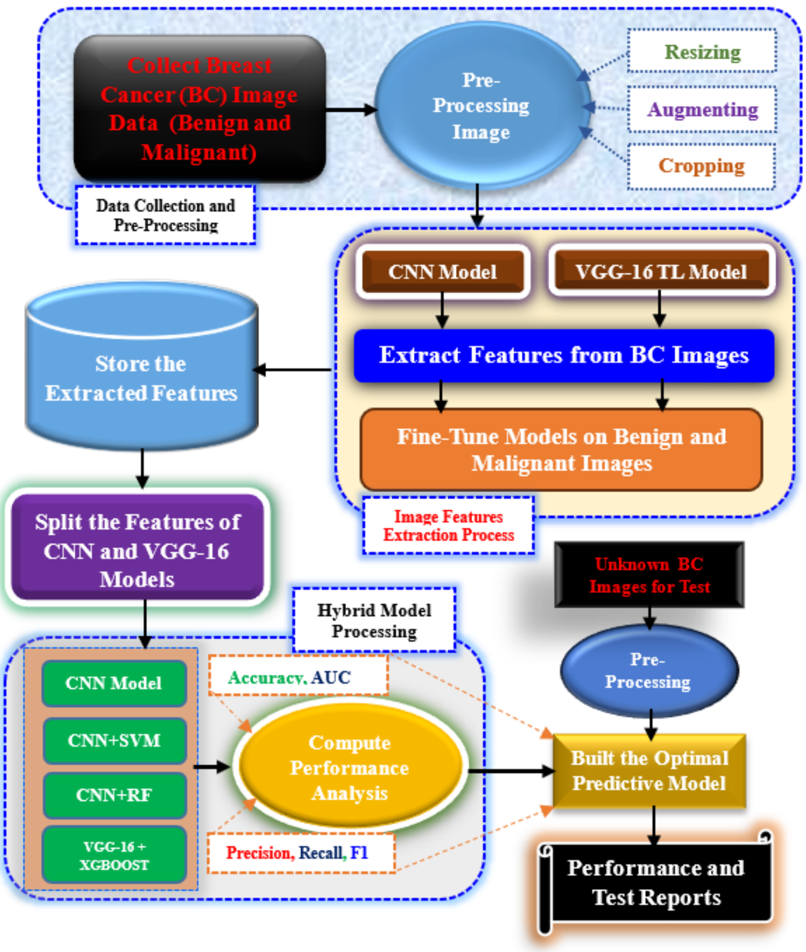

Figure 1. Proposed model for Identification of BC using Image Dataset

Store the extracted features and divide them into distinct training and testing sets, preparing the data for model training and validation. To assess the extracted features, employ various classification algorithms, including CNN, CNN + SVM, and VGG-16 + XGBOOST. Evaluate their performance by calculating key metrics such as Accuracy, Precision, Recall, F1 Score, and AUC values. Identify the most effective predictive model based on the algorithm that delivers the highest classification performance, considering the unique characteristics of the breast cancer dataset. Challenge the model with previously unseen malignant and benign breast cancer images to assess its real-world predictive capability. Generate comprehensive reports that detail the model's performance and test outcomes. These reports will be invaluable tools for assessing the model's practicality and reliability in clinical applications.

3.2 Dataset Description

The dataset (shown in table 3) includes statistics on the magnification level of microscope pictures and counts of malignant and benign images. This information is useful for various applications, notably medical image analysis, where distinguishing between benign and cancerous cells or tissues is critical. The dataset has four magnification levels: 40x, 100x, 200x, and 400x. These levels denote the magnification of the microscope lens used to take photographs. There are 2480 benign photos and 5429 malignant images, totalling 7909 across all magnification settings. It is vital to highlight that the dataset contains a considerable class imbalance between benign and cancerous photos. The dataset includes several magnification levels, which might be useful for jobs that require pictures at various magnification levels. Figure 2 shows the some of the samples of benign (figure 2 (A)) and malignant (figure 2 (B)) breast cancer BreaKHis image dataset.

Table 3. Dataset Description

Magnifica-tion	Malig-nant	Benign	Total Images
400x	1232	588	1820
200x	1390	623	2013
100x	1437	644	2081
40x	1370	625	1995
Total	5429	2480	7909

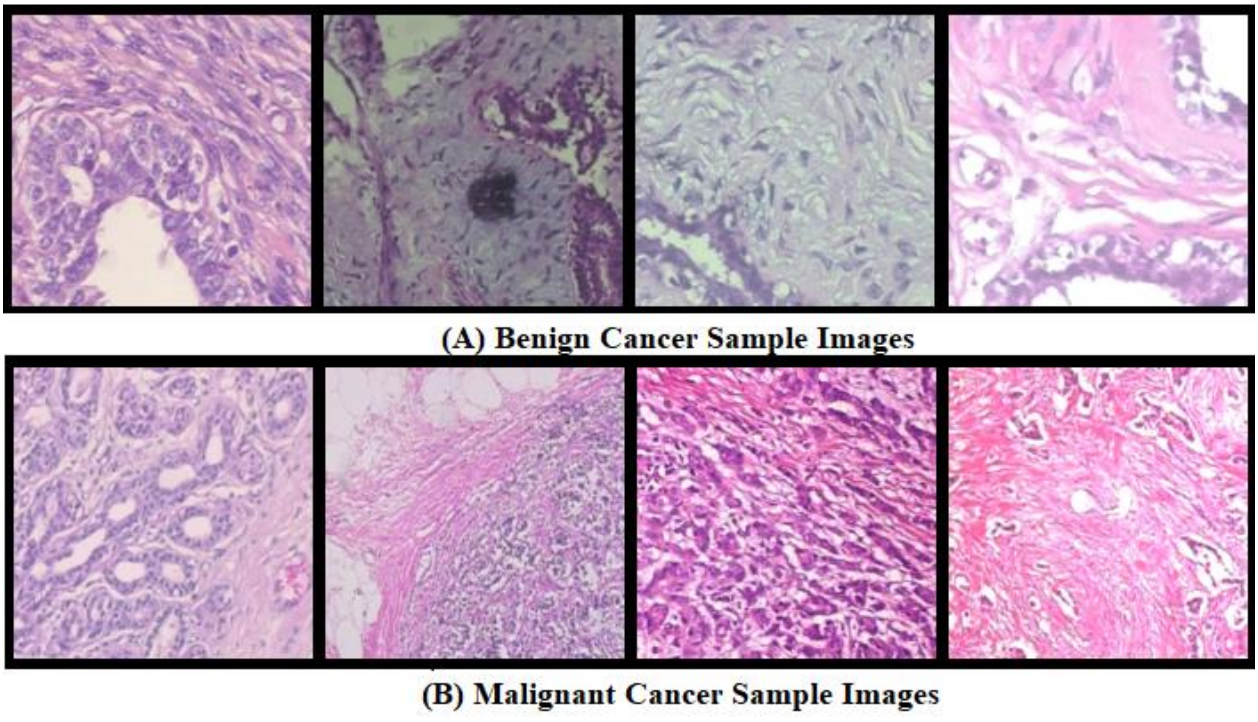

Figure 2. Sample Images of Benign and Malignant Breast Cancer

3.3 Convolutional Neural Network

CNN (shown in Figure 3), sometimes known as convnet, is a subset of Machine Learning, a subset of AI. It is one of several artificial neural networks for diverse applications and data kinds. CNN is a machine learning organization used explicitly for image recognition and operations that entail managing pixel information. Compared to other algorithms, the predicted pre-handling in a CNN is substantially smaller. While channels in crude approaches are hand-designed, CNNs may become familiar with these channels/attributes with adequate practice. CNN's architecture is like the network example of Neurons in the Human Cerebrum and was energized by the addition of the Visual Cortex (Albawi et al., 2017). Individual neurons respond to enhancements only in a small visual field area called the Responsive Field. A variety of such regions cross over to span the entire viewable region. Convolutional, pooling, and fully linked layers make up CNN. A 3 x 3 x 1 convolved feature map will now be constructed from a 5 x 5 x 1 input image by applying a 3 x 3 x 1 filter. Convolution is a method to distinguish significant level features, like edges, from information images. It is okay to limit Convnets to a Convolutional Layer. The very first Convolution layer frequently finds itself in charge of recording Edges, variation, inclination orientation, and so on, examples of low-level elements. With further layers, the architecture adjusts to the Significant Level components, resulting in a network that knows the graphics of the dataset as well as we do. The activity has two alternative outcomes: one in which the dimensionality of the convolved highlight is lowered compared to the data and one in which it is either increased or remains the same. Applying Significant Cushioning considering the prior option or Similar Cushioning considering the final choice completes this. When the 5x5x1 picture is expanded into a 6x6x1 picture, after which the 3x3x1 filter is applied, we see that the convolved framework has 5x5x1-sized components. The name is, hence, the same padding. In the unlikely event that we perform the same activity without cushioning, we are provided with a grid that contains elements of the Bit (3x3x1) itself.

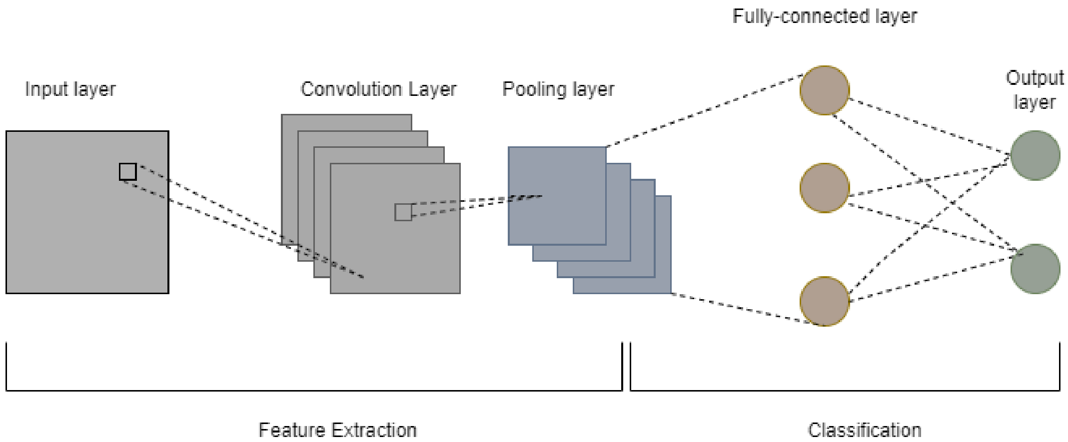

Figure 3. Convolutional Neural Network

It is known as valid padding. The Pooling layer oversees reducing the spatial size of the Convolved Component. Fewer computer resources are expected to handle the information to minimize its complexity. It also aids in extracting current aspects that are rotational and positional invariant, improving the model-creation process. Pooling is available in two varieties: Regular and Max Pooling (MP). MP returns to the most valuable part of the image the Bit covers. Typical pooling produces the average of the proportionally vast number of values from the region of the image that the Bit covers. MP is also a Commotion Suppressant. It eliminates the loud initiations and de-noises and reduces the aspect ratio. Standard Pooling uses dimensionality reduction to suppress noise. Therefore, MP outperforms Average Pooling. Because of the convolutional layer's output, a fully connected (FC) layer is commonly used to generate non-linear mixes of the top-level highlights. There, the FC layer is experimenting with a possible non-linear capability—the pooled characteristics, after flattening, are delivered to the wholly integrated layers.

3.4 Convolutional Neural Network with Support Vector Machine

The goal of SVM computation is to find the best line or choice limit for categorizing the n-layered space to quickly categorize fresh information of relevance later. The presence of a hyperplane limits this best-case situation. SVM selects and concentrates the most bizarre vectors to help create the hyperplane. These absurd cases are known as support vectors, and the following calculation is known as an SVM. SVMs can be divided into two different groups: linear and non-linear (Terlapu et al., 2021). Linear SVM: It is employed to classify directly divisible data, meaning if a database can be broken down into two categories using a single perfect line, it is uniformly detachable data. Non-Linear SVM: It is used for non-directly isolated information, which is used when a database cannot be sorted using a perfect line. The classifier used for this type of information is known as a Non-linear SVM classifier.

3.5 Convolutional Neural Network with Random Forest (RF)

An RF (shown in Figure 4) classifier employs several decision trees on different portions of the input dataset and applies the normal to the dataset's accuracy in forecasting the likely future. Instead of depending just on one decision tree, it forecasts the outcome based on estimates from all trees and most ballots from expectations (Vital et al., 2021). The more trees in the forest, the higher the accuracy and the lower the overfitting. RF operates in two stages: first, it creates the RF by combining N decision trees, and then it makes predictions for each tree created in the first step. These are the two assumptions for a good random forest classifier: It should be noted that the attribute variable in the dataset should contain concrete numerical values for the classifier to do accurate predictions, instead of relying on estimated results. Additionally, to improve the accuracy of the predictions, it is desirable that the predictions from every tree have minimal correlation with one another. RF requires a comparatively shorter training time than other algorithms. It can predict output with a high degree of accuracy, even when handling large datasets in an efficient manner. It can maintain its accuracy even when a notable proportion of data is missing.

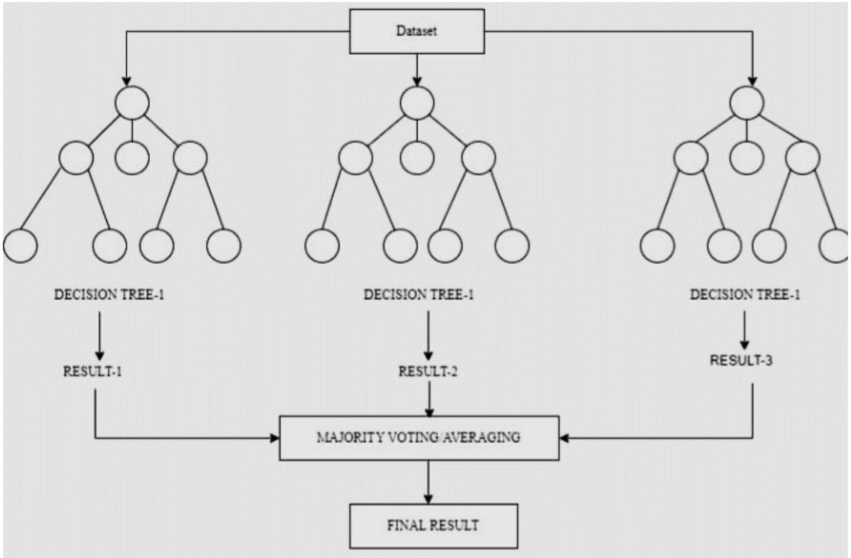

Figure 4. Random Forest Analysis

3.6 XGBoost

The boosting ensemble approach combines several unsuccessful classifiers to create a powerful classifier. A model is first generated using training data, and then further models are created to fix any mistakes in the first model. This approach is repeated until either the maximum number of models is formed, or the full training dataset can be properly predicted. Gradient boosting is a popular boosting approach in which each estimate corrects the error of its predecessor. In contrast to Adaboost, the weight of the training examples is not modified. Instead, the labels from the ancestor's residual mistakes are used to teach each estimator. Gradient boosting is implemented by XGBoost, which generates decision trees successively. The feature weights are critical to XGBoost since they are assigned to each independent variable and supplied into the decision tree, which anticipates the outcome. The factors that the decision tree incorrectly forecasted are given greater weight and put into the decision tree that follows. Then, by integrating these distinct classifiers, a more accurate model is formed. XGBoost can tackle regression, classification, ranking, and particularly specified forecasting problem.

4. RESULTS AND ANALYSIS

4.1 Confusion Matrix Analysis for Experimental Models:

Figure 5 (A) shows the CNN confusion matrix in classifying breast tumours. It correctly identifies a substantial number of benign cases (478) while also effectively distinguishing malignant cases (1135), with only a limited number of misclassifications in both categories (18 benign and 24 malignant).

The CNN + SVM confusion matrix is shown in Figure 5 (A) and indicates a mixed performance in classifying breast tumours. It correctly identifies many malignant cases (939) but challenges distinguishing benign cases (322). The model has many false positives (174 benign cases) and false negatives (220 malignant cases). The CNN + RF confusion matrix is shown in Figure 5 (C) and demonstrates excellent performance in breast tumour classification. It correctly identifies all malignant cases (1159) and does not misclassify any benign patients, resulting in perfect precision for both categories. This indicates a reliable model for distinguishing between benign and malignant cases, with no false positives or negatives. The VGG-16 + XGBoost confusion matrix shows exceptional performance in breast tumour classification. It correctly identifies all malignant issues (1159) and does not misclassify any benign cases, achieving perfect precision for both categories. This reflects a highly reliable model for distinguishing between benign and malignant cases.

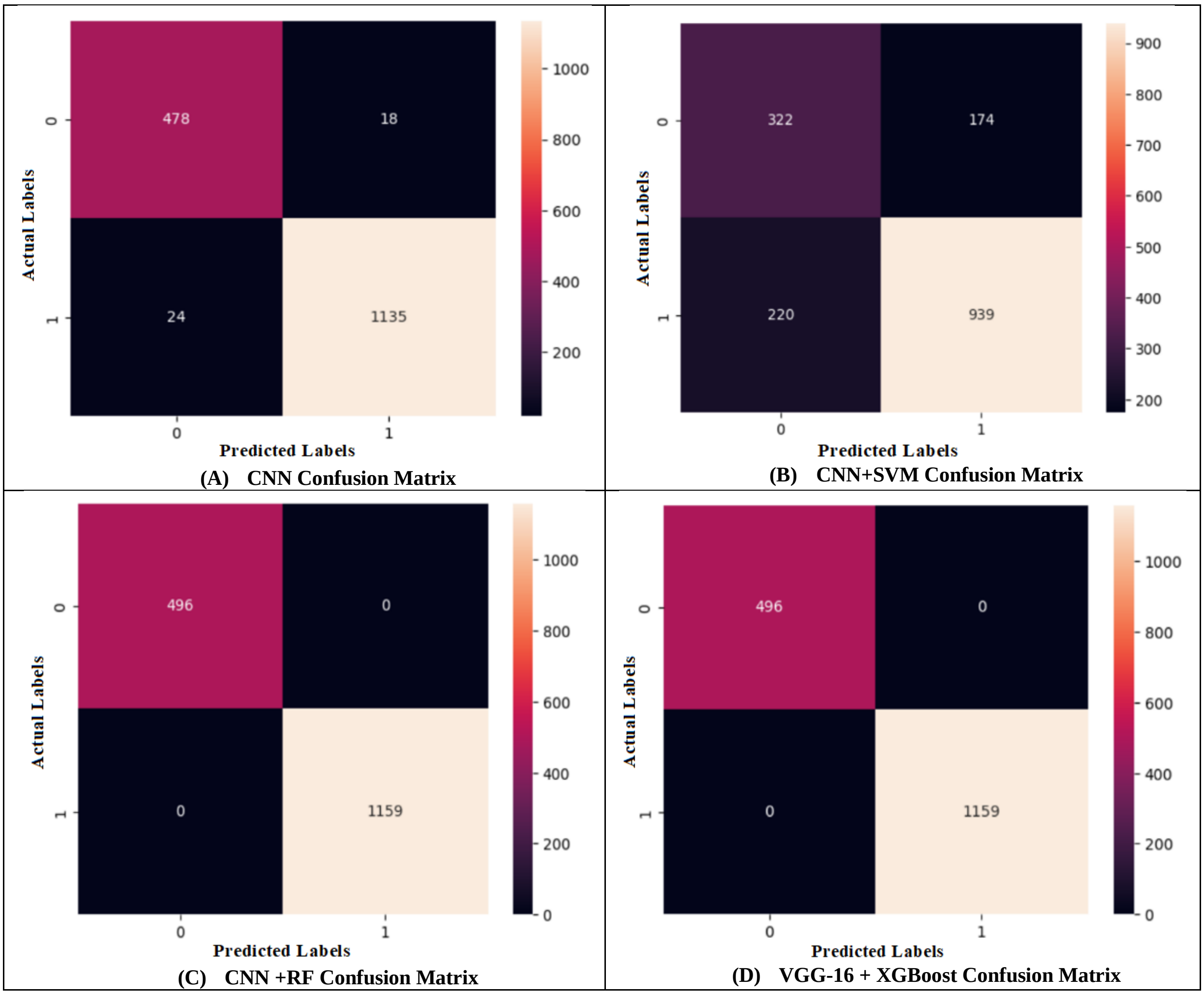

Figure 5. Confusion Matrix for all Experimental Models

4.2 Performance Parameters Analysis for each Experimental Models

Table 4 provides a comprehensive overview of breast cancer classification performance by various algorithms, explicitly focusing on distinguishing between benign (B) and malignant (M) cases. CNN algorithm performs robustly in classifying breast cancer, with an overall accuracy of nearly 97.5%. AUC (Area Under the Curve): 98.3% - The AUC score signifies a high level of discrimination power in separating malignant and benign cases. CNN + SVM Accuracy is 76.19% - The combination of CNN and SVM yields a lower accuracy than the standalone CNN model. CNN + RF Accuracy: 100% - CNN combined with Random Forest achieves a perfect accuracy score, indicating flawless classification. AUC: 100% - The AUC score also reaches the maximum value. Accuracy: 100% - The VGG-16 model combined with XGBOOST achieves perfect accuracy, suggesting flawless classification. AUC: 100%, indicating ideal discrimination. The table reveals that the combination of deep learning models (CNN and VGG-16) with gradient boosting algorithms (XGBOOST) and Random Forest (RF) results in near-perfect or perfect classification performance, particularly in distinguishing between malignant and benign breast cancer cases.

Table 4. Performance parameters evaluations

Algorithm	Accuracy	AUC	Precision	Recall	F1 Score
CNN	0.9746	0.983	B - 0.95 M -0.98	B - 0.96 M - 0.98	B - 0.96 M - 0.98
CNN+SVM	0.7619	0.842	B - 0.59 M - 0.84	B - 0.65 M - 0.81	B - 0.62 M - 0.83
CNN +RF	**1.0**	**1.0**	B - 1.00 M - 1.00	B - 1.00 M - 1.00	B - 1.00 M - 1.00
VGG-16 +XGBOOST	**1.0**	**1.0**	B - 1.00 M - 1.00	B - 1.00 M - 1.00	B - 1.00 M - 1.00
***Note : B for 'Benign,' and M for 'Malignant'**					

4.3 DISCUSSIONS

Figure 6 shows the Performance parameters comparative analysis of CA and ROC values. Accuracy of the CNN and CNN + SVM models achieve accuracies of 97.46% and 76.19%, respectively, indicating a solid performance by CNN in correctly classifying breast tumours. The CNN + RF and VGG-16 + XGBOOST models both achieve perfect accuracies of 100%, demonstrating the highest level of accuracy in breast cancer classification.The AUC values for the CNN and CNN + SVM models are 0.983 and 0.842, respectively, with the CNN model outperforming CNN + SVM in discriminating between benign and malignant cases. Both CNN + RF and VGG-16 + XGBOOST models achieve perfect AUC scores of 1.0, indicating ideal discrimination power and the highest level of Performance in distinguishing between the two classes.

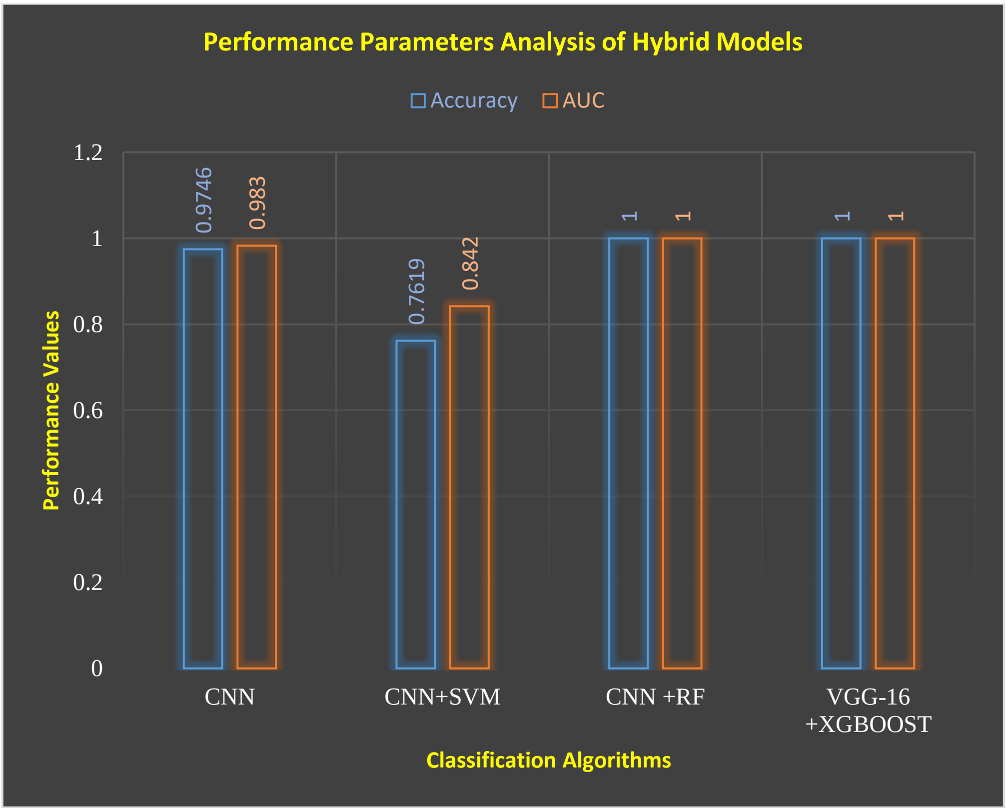

Figure 6: Comparative Analysis performance attributes about experimental models

The table shows the comprehensiveness of the present study with other works related to breast cancer image datasets. In comparison with different experimental results, the present study demonstrates outstanding performance with perfect classification accuracy (1.0) and AUC (1.0) for both the CNN + RF and VGG-16 + XGBOOST models, surpassing previous research efforts that achieved accuracies ranging from 90.02% to 98.27% and F1 scores between 90.49% and 99%.

Table 5. Comparative analysis and evaluations present work with other existing research works

Author (Year)	Description	Dataset Used	Result
Zhang et al. (2020)	A deep learning framework that combines Linear discriminant analysis and auto encoder neural network is used for classification.	Different real-time datasets	98.27% accuracy
Prakash & Visakha (2020)	A neural network consisting of all dense layers are used. It is optimized using early stopping and dropout layers.	Wisconsin BC diagnosis	Benign F1 Score-98 Malignant F1 score-99
Aryan& Saha (2020)	Deep learning-based stacking ensemble framework is used for classification	MetaBric dataset	90.02% accuracy
Gour et al. (2019)	A residual-learning based approach is used for BC classification i.e., ResHist.	BreaKHis dataset	92.52% accuracy. F1 scores are 90.49% and 93.45%
Present Study	**CNN +RF and VGG-16 +XGBOOST Proposal Models.**	**BreaKHis dataset**	**CNN + RF: Classification accuracy 1.0 and AUC 1.0, VGG-16+XGBOOST: Classification accuracy 1.0 and AUC 1.0**

5. CONCLUSION

We took the renowned BreaKHis dataset for assessment. We planned to foster decreased time and cost elements of the patients as well as to limit crafted by specialists. We have utilized basic and justifiable models to finish this work. Our techniques should be used for preparing data, and testing data should be utilized to check, assuming the results are adequately exact. For each ensuing calculation we applied, we worked on the productivity of the model. Along these lines, we created and performed BC identification model. Our research focused on identifying the type of tumour, whether it is benign or malignant. That means we have classified the main types of cancer tumours. Our research doesn't recognize the kind of images as ductal carcinoma, lobular carcinoma, etc. The future work should focus on it. Along with these, they can also concentrate on the stage of breast cancer. By taking input images, the model should estimate the tumour's size, report its respective phase, and prescribe the necessary treatment.

References:

Albawi, S., Mohammed, T. A., & S. Al-Zawi (2017). Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6). doi: 10.1109/ICEngTechnol.2017.8308186

Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast cancer classification using machine learning. In *2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT)* (pp. 1-4). IEEE. doi: 10.1109/EBBT.2018.8391453

Arya, N., & Saha, S. (2020). Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. *IEEE/ACM transactions on computational biology and bioinformatics, 19*(2), 1032-1041. doi: 10.1109/TCBB.2020.3018467

Balasubramaniam, S., Velmurugan, Y., Jaganathan, D., & Dhanasekaran, S. (2023). A Modified LeNet CNN for Breast Cancer Diagnosis in Ultrasound Images. *Diagnostics*, *13*(17), 2746. doi:10.3390/diagnostics13172746

Chiao, J. Y., Chen, K. Y., Liao, K. Y. K., Hsieh, P. H., Zhang, G., & Huang, T. C. (2019). Detection and classification the breast tumors using mask R-CNN on sonograms. *Medicine, 98*(19). doi: 10.1097/MD.0000000000015200

Gour, M., Jain, S., & Sunil Kumar, T. (2019). Residual-learning based CNN for breast cancer histopathological image classification. *Wiley*, 2019. doi:10.1002/ima.22403

Kabir, S. M., Bhuiyan, M. I., Tanveer, M. S., & Shihavuddin, A. S. M. (2021). RIIG modeled WCP image-based CNN Architecture and feature-based approach in breast tumor classification from B-Mode Ultrasound. *Applied Sciences*. 11(24), 12138. doi:10.3390/app112412138

Karthik, R., Menaka, R., Kathiresan, G. S., Anirudh, M., & Nagharjun, M. (2022). Gaussian dropout based stacked ensemble CNN for classification of breast tumor in ultrasound images. *Irbm*, *43*(6), 715-733. doi: 10.1016/j.irbm.2021.10.002

Khosasi, S., Sriwidyani, N. P., Sumadi, I. W. J., Winarti, N. W., Ekawati, N. P., & Muliarta, I. M. (2023). High expression of CD73 as a negative predictor of clinical neoadjuvant chemotherapy response in triple-negative breast cancer. *Indonesia Journal of Biomedical Science*, *17*(1), 94-99. doi: 10.15562/ijbs.v17i1.418

Khuriwal, N., & Mishra, N. (2018, November). Breast cancer detection from histopathological images using deep learning. In *2018 3rd international conference and workshops on recent advances and innovations in engineering (ICRAIE)* (pp. 1-4). IEEE. doi: 10.1109/ICRAIE.2018.8710426

Kim, J., & Villadsen, R. (2018). Expression of luminal progenitor marker CD117 in the human breast gland. *Journal of Histochemistry & Cytochemistry, 66*(12), 879-888. doi:10.1369/0022155418788845

Lukong, K. E. (2017). Understanding breast cancer–The long and winding road. *BBA clinical*, *7*, 64-77. doi:10.1016/j.bbacli.2017.01.001

MurtiRawat, R., Panchal, S., Singh, V. K., & Panchal, Y. (2020, July). Breast Cancer detection using K-nearest neighbors, logistic regression and ensemble learning. In *2020 international conference on electronics and sustainable communication systems (ICESC)* (pp. 534-540). IEEE. doi: 10.1109/ICESC48915.2020.9155783

PanduRanga Vital, T., Murali Krishna, M., Narayana, G. V. L., Suneel, P., & Ramarao, P. (2019). Empirical analysis on cancer dataset with machine learning algorithms. In *Soft Computing in Data Analytics: Proceedings of International Conference on SCDA 2018* (pp. 789-801). Springer Singapore. doi:10.1007/978-981-13-0514-6_75

Prakash, S. S., & Visakha, K. (2020, July). Breast cancer malignancy prediction using deep learning neural networks. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 88-92). IEEE. doi:10.1109/ICIRCA48905.2020.9183378

Sahu, A., Das, P. K., & S. Meher (2023). High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. *Biomedical Signal Processing and Control*, *80*, 104292. doi: 10.1016/j.bspc.2022.104292

Terlapu, P. V., Sadi, R. P. R., Pondreti, R. K., & Tippana, C. R. (2021). Intelligent Identification of Liver Diseases Based on Incremental Hidden Layer Neurons ANN Model. *International Journal of Computing and Digital System*. doi: 10.12785/ijcds/110183

Vital, T. P. (2021). Empirical study on Uddanam chronic kidney diseases (UCKD) with statistical and machine learning analysis including probabilistic neural networks. In Handbook of Computational Intelligence in Biomedical Engineering and Healthcare. 283-314. *Academic Press*. doi: 10.1016/B978-0-12-822260-7.00013-3

Vital, T. P., Prasada Raju, G. S. V., Kaladhar, D. S. V. G. K., Sriram, T. V., Rayavarapu, K. A., Nageswara Rao, P. V., & Appala Raju, S. (2014). Data Collection, Statistical Analysis and Clustering Studies of Cancer Dataset from Viziayanagaram District, AP, India. In ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II: Hosted by CSI Vishakhapatnam Chapter (pp. 423-430). *Springer International Publishing*. doi: 10.1007/978-3-319-03095-1_45

Vital, T. P., Raju, G. P., Rao, I. S., & Kumar, A. P. (2015). Data collection, statistical analysis and machine learning studies of cancer dataset from north costal districts of AP, India. *Procedia Computer Science*, *48*, 706-714. doi :10.1016/j.procs.2015.04.205

Xie, X. Z., Niu, J. W., Liu, X. F., Li, Q. F., Wang, Y., Han, J., & Tang, S. (2022). DG-CNN: Introducing margin information into convolutional neural networks for breast cancer diagnosis in ultrasound images. *Journal of Computer Science and Technology*, *37*(2), 277-294. doi: 10.1007/s11390-020-0192-0

Zhang, X., He, D., Zheng, Y., Huo, H., Li, S., Chai, R., & Liu, T. (2020). Deep learning based analysis of breast cancer using advanced ensemble classifier and linear discriminant analysis. *IEEE access*, *8*, 120208-120217. doi: 10.1109/ACCESS.2020.3005228

Nageswara Rao Gali
Department of Information Technology, Aditya Institute of Technology and Management, Tekkali, India-532 201
gnraoaitam@gmail.com
ORCID 0000-0001-7636-7499

Panduranga Vital Terlapu
Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali, India-532 201
vital2927@gmail.com
ORCID 0000-0002-4393-0692

Yasaswini Mandavakuriti,
Department of Information Technology, Aditya Institute of Technology and Management, Tekkali, India-532 201
asaswinimandavakuriti@gmail.com

Sai Manoj Somu
Department of Information Technology, Aditya Institute of Technology and Management, Tekkali, India-532201,
saimanoj2019@gmail.com

Madhavi Varanasi
Department of Information Technology, Aditya Institute of Technology and Management, Tekkali, India-532201,
madhavivaranasi28@gmail.com

Vijay Telugu
Department of Information Technology, Aditya Institute of Technology and Management, Tekkali, India-532201,
vijaytelugu123@gmail.com

Maheswara Rao V V R
Shri Vishnu Engineering College for Women (A), Bhimavaram,
India
mahesh_vvr@yahoo.com
ORCID 0000-0002-0503-7211