Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

A deep learning approach: To detect the breast cancer.

Ahsan Habib Sami*[1], Tasfia Tabassum [2], Mohammad Saydul Alam[3], Sahana Akter[4]
[1,2,3,4]*Department of Computer Science and Engineering*,
[1,2]*American International University - Bangladesh (AIUB) Dhaka, Bangladesh*,
[3]*Southeast University (SEU) Dhaka, Bangladesh*,
[4]*International Islamic University Chittagong*, Chattogram, Bangladesh
Email: [1]ah626185@gmail.com, [2]tasfiatabassum418@gmail.com, [3]saydul.alam@seu.edu.bd, [4]sahana4143@outlook.com,

***Abstract*—In contemporary healthcare, breast cancer represents one of the most nebulous and formidable fields. In recent years, the issue has progressively escalated, leading to a substantial increase in mortality, with approximately fifty per cent of the impacted female patients perishing from the sickness. A machine learning technique has been used to facilitate the early and precise identification of breast cancer, addressing this essential problem. This research examines the effectiveness of three deep learning models—Convolutional Neural Network (CNN), VGG16, and ResNet-50—in classifying breast cancer photos. The models underwent training and evaluation on a labelled dataset over 40 epochs to measure their performance. The CNN model attained an accuracy of 81%, demonstrating its capability to identify critical characteristics. VGG16 surpassed the other models' accuracy of 96%, owing to its more profound architecture and pre-trained weights. ResNet-50, using residual connections to address vanishing gradients, achieved an accuracy of 85.97%. The results indicate the efficacy of various deep learning methods in enhancing breast cancer diagnosis, with VGG16 identified as the most precise model. Future research aims to improve these models' robustness and explore multimodal approaches to increase diagnostic precision. It also plans to compare various architectures to assess their advantages and disadvantages.**

***Index Terms*—Convolutional Neural Networks (CNN), VGG16, Ultrasound Image Classification, Deep Learning, Machine Learning in Healthcare.**

I. Introduction

Breast cancer is a primary worldwide health concern, especially among women, characterized by increasing incidence rates and considerable effects on morbidity and death [1]. Early identification is essential for enhancing treatment results and survival rates, thereby serving as a fundamental aspect of breast cancer care. Although practical, conventional diagnostic methods, such as clinical assessments, imaging, and biopsies, often include invasive and resource-demanding procedures [1]. Thanks to developments in cancer research and computers, diagnostics have improved. KHCC data demonstrate the importance of comprehensive datasets in patient care and healthcare assessment [2]. By differentiating IDC from normal tissues, novel biomarkers such as the electromechanical coupling factor hold the potential for non-invasive breast cancer screening in conjunction with improvements in healthcare [3]. AI and deep learning have transformed cancer diagnosis. The ELM model's increased accuracy by integrating CNN-based characteristics, tissue structure, and texture has opened up new avenues for computer-assisted diagnosis [4]. The combination of regional inductive moderate hyperthermia (RIMH) and neoadjuvant chemotherapy has decreased the risk of mastectomy by improving tumor response, changing the microenvironment, and increasing breast-conserving procedures. This demonstrates how innovative medicines may be used with traditional therapy to enhance patient results. [5]. The research states comparative analysis of three deep learning architectures—VGG16, CNN, and ResNet50—for the given classification task. By benchmarking their performance using various metrics, the study provides a comprehensive evaluation of their effectiveness. This comparison highlights the strengths and weaknesses of each model, offering valuable insights into their suitability for the task and reinforcing the uniqueness of the approach.

II. LITERATURE REVIEW

Early detection improves breast cancer survival rates and treatment outcomes. Numerous studies have examined different early detection methods, emphasizing the need for innovative approaches to detecting cancer in its early stages. The essential papers listed below discuss breast cancer diagnosis and therapy.

A. Cancer Registry and Data Utilization

Data from cancer registries is essential for evaluating healthcare. KHCC research emphasized the need to improve data gathering for improved therapy and demonstrated good completeness [2].PCA helps identify and model risk factors, while machine learning improves diagnosis and prediction [6].

B. Innovative Biomarkers and Diagnostic Methods

Instead of evaluating stiffness, advanced diagnostics use the electromechanical coupling factor to differentiate IDC from normal tissue and detect breast cancer [3]. In conjunction with deep learning models, infrared thermal imaging has developed as a non-invasive diagnostic instrument, using temperature fluctuations in breast tissue to identify cancers with enhanced precision [7].

C. Applications of Machine Learning and Deep Learning

Machine learning and deep learning improve the detection of breast cancer, with CNN-based feature fusion increasing specificity and accuracy [4]. DWT, MLP, and SVM-RFE reduced needless operations by achieving 97.4% accuracy in cancer detection. ANNs fared better than SVM and Random Forest, with 99% accuracy [8].

D. Progress in Imaging Methodologies

While microwave methods based on Vivaldi antennas increased tumor identification, deep learning improved breast cancer imaging [9]. BC-DROID reduces misdiagnosis by combining deep learning and imaging to provide precise tumor location and diagnosis [10]. Recent studies improved classification accuracy by converting gene expression data into two-dimensional images using a novel ensemble learning technique, such as the Empirical Wavelet Transform [11].

E. Psychological and Behavioral Determinants

Psychological and behavioral variables influence breast cancer. A Taiwanese study emphasizes the need for public health initiatives by linking female smoking, obesity, and cancer risk [12]. Research on anxiety disorders indicated its frequency among breast cancer patients, indicating that prioritizing psychological wellness might enhance coping techniques and results [13].

F. Spatiotemporal and Hybrid Modeling

Advanced modelling tools have shown the potential to enhance breast cancer risk evaluation. Siamese neural networks integrating spatiotemporal information surpassed traditional CNNs, attaining superior risk prediction accuracy [14]. Likewise, deep learning models that include classification and localization exhibited exceptional performance, with an AUC of 98.6% for mammographic image processing [15].

Nonetheless, there are still significant obstacles facing the present breast cancer detection technologies. Nevertheless, Diagnostic accuracy is impacted by false positives and nega tives, resulting in missed diagnoses and needless treatments. Furthermore, most models only use imaging data, ignoring crucial clinical and genetic variables that could enhance clas sification accuracy. Deep learning models' generalizability is still restricted because their performance frequently deterio rates when used on various datasets. Furthermore, inefficient feature extraction can make it more difficult for deep models to recognize intricate tumor features correctly. Our study suggests an integrated strategy using Convolutional Neural Networks (CNN), VGG16, and ResNet50 to overcome these restrictions. Each model contributes to the detection process: VGG16 enhances accuracy by extracting pre-trained features, CNN captures complex information, and ResNet50's resid ual connections allow for deeper learning while addressing vanishing gradient problems. Our approach improves model generalization, accuracy, and resilience, providing a more complete breast cancer diagnosis solution.

III. METHODOLOGY

This research used the BUSI dataset, which consists of breast ultrasound pictures classified as benign, malignant, and routine. The photos underwent preprocessing, which included downsizing to 224×224×3, standardizing pixel values, and using augmentation methods to increase unpredictability. The dataset was divided into training 80% and validation 20% subsets. Three models were executed: a bespoke CNN, VGG16, and ResNet50. The CNN was constructed using convolutional, pooling, and dense layers, including dropout for regularization. The pre-trained VGG16 and ResNet50 models were fine-tuned by unfreezing specific layers and adding categorization layers. All models underwent training for 40 epochs using the Adam optimizer, sparse categorical cross-entropy loss and included callbacks for a learning rate reduction and early stopping. The evaluation revealed that VGG16 had the most incredible validation accuracy at 96%, surpassing CNN at 81% and ResNet50 at 85.97%.

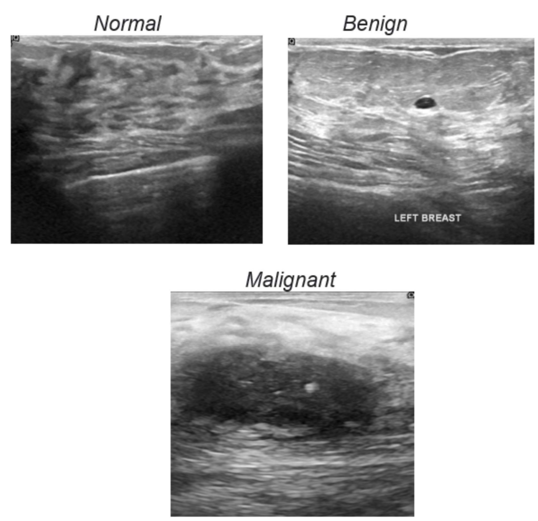

Fig. 1. Some Images of Kaggle dataset. [16]

A. Dataset Description

This research uses the 780 ultrasound photos from Kaggle [16] that have been categorized as normal, malignant, or benign. To solve class imbalance and improve model robustness, images were scaled (224×224), normalized, and enhanced (rotation, zoom, flipping) to help diagnose breast cancer accurately. **Fig. 1** shown some images each type.

B. Data Preprocessing

To enhance balance and generalization, images were enlarged (224×224), normalized ([0,1]), and enhanced (rotation, shifts, zoom, flipping). Data was divided (80% training, 20% validation), and labels were encoded. Dynamic preprocessing was managed by 'ImageDataGenerator,' although stratified splitting and oversampling might improve robustness even further.

IV. RESULT ANALYSIS & DISCUSSION

The assessment of the suggested models—CNN, VGG16, and ResNet50—revealed differing performance levels, providing insights into their advantages and drawbacks for ultrasound picture categorization. The CNN model attained an accuracy of 81%, indicating robust baseline performance with a bespoke architecture; however, it revealed deficiencies in feature extraction capabilities. The VGG16 model, using pre-trained weights and fine-tuning, achieved 96% accuracy, demonstrating the efficacy of transfer learning for this task. Nevertheless, the ResNet50 model attained just 85.97% accuracy, which may be attributable to overfitting or inadequate fine-tuning for the dataset.

The training and validation graphs compile these conclusions. VGG16's performance consistently reduces loss and enhances accuracy while validation measures stabilize, suggesting potential overfitting **Fig 2**. Training and validation losses diminish in the CNN model **Fig 4**, but accuracy steadily increases throughout epochs. ResNet50 exhibits increasing accuracy and decreasing loss **Fig 6**, with validation surpassing training, suggesting overfitting or problems particular to the dataset.

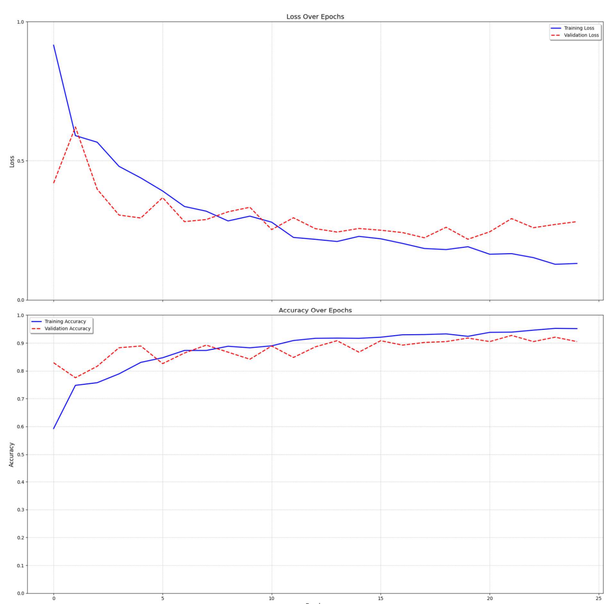

Fig. 2. Accuracy and Loss curve of the VGG16

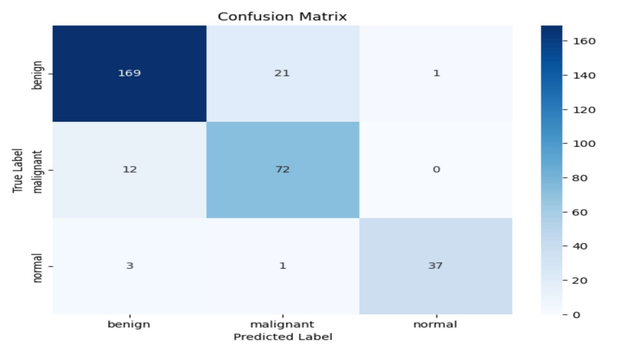

Fig. 3. Confusion Matrix of the VGG16

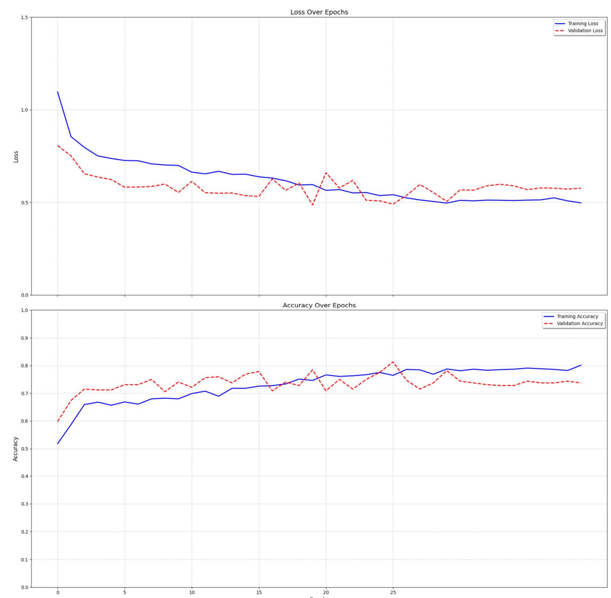

Fig. 4. Accuracy and Loss curve of the CNN

The comparison findings unequivocally indicate that VGG16 is most suited for this job owing to its resilient pre-trained feature maps and versatility. In contrast, ResNet50's deeper design may need more specific modifications or a bigger dataset to use its capabilities thoroughly. Subsequent research may enhance these results by rectifying class imbalance and using sophisticated augmentation or transfer learning methodologies.

TABLE I
PERFORMANCE METRICS OF VARIOUS MODELS FOR BREAST CANCER CLASSIFICATION

Model	Accuracy (%)	Precision	Recall	F1-Score
CNN	81.00	0.7506	0.7373	0.7419
VGG16	96.00	0.89	0.88	0.88
ResNet50	85.97	0.86	0.86	0.85

According to the comparison study **TABLE I**, VGG16 is the best model for this task, while CNN offers dependable results as a baseline, and ResNet50 needs further slight tuning.

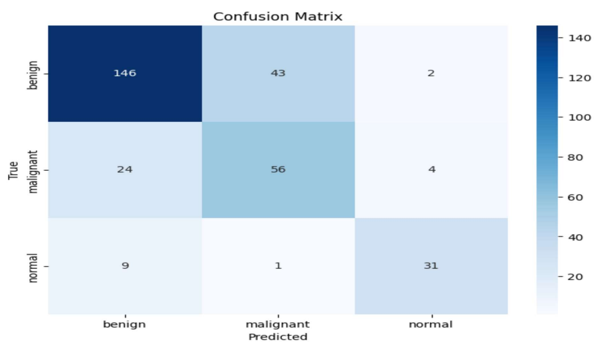

Fig. 5. Confusion Matrix of the CNN

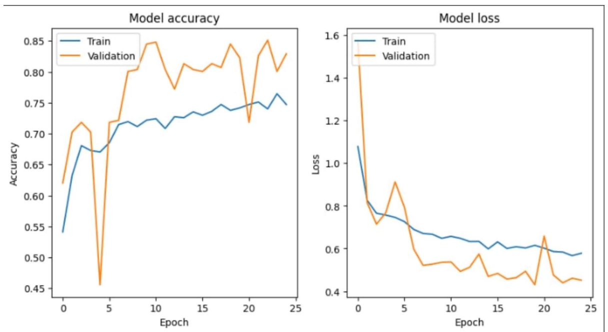

Fig. 6. Accuracy and Loss curve of the ResNet50

V. CONCLUSION AND FUTURE DIRECTIONS

This work assessed three deep learning models—CNN, VGG16, and ResNet50—for categorizing breast cancer from ultrasound pictures. With an accuracy of 96%, the findings demonstrated that VGG16 performed better than the other models, demonstrating the value of transfer learning. With an accuracy of 81%, the CNN model offered a good starting point, but ResNet50's 85.97% performance this indicates that more fine-tuning or hyperparameter changes may be necessary for deeper architectures to respond appropriately to ultrasound imaging using a consistent training technique of 40 epochs per model.

Future research will expand the dataset size, utilize sophisticated data augmentation methods like GANs, and include multimodal data to increase model robustness and diagnostic accuracy.

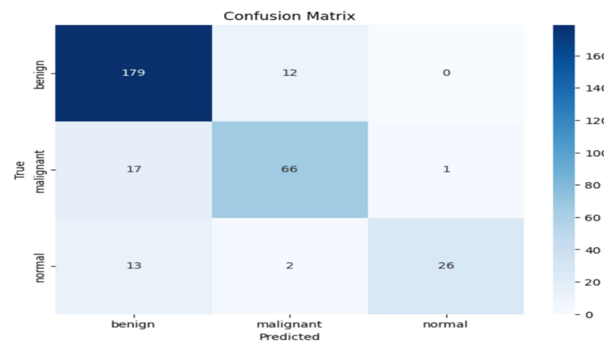

Fig. 7. Confusion Matrix of the ResNet50

REFERENCES

[1] A. K. Das, S. K. Biswas, A. Bhattacharya, and E. Alam, "Introduction to breast cancer and awareness," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1. IEEE, 2021, pp. 227–232.

[2] A. A. Hmaidan, E. Boutou, K. Jamal, and A. Al-Omari, "Availability and usability of the hospital-based cancer registry data for measuring the quality outcome indicators of healthcare provided to breast and colorectal cancer patients at king hussein cancer center," in *2018 1st International Conference on Cancer Care Informatics (CCI)*. IEEE, 2018, pp. 195–204.

[3] K. Park, W. Chen, M. A. Chekmareva, D. J. Foran, and J. P. Desai, "Electromechanical coupling factor of breast tissue as a biomarker for breast cancer," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 96–103, 2017.

[4] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, "Breast cancer detection using extreme learning machine based on feature fusion with cnn deep features," *IEEE Access*, vol. 7, pp. 105 146–105 158, 2019.

[5] V. E. Orel, O. Rykhalskyi, L. Syvak, I. Smolanka, A. Loboda, A. Lyashenko, I. Dosenko, A. Mokhonko, T. Golovko, A. Ganich *et al.*, "Computer-assisted inductive moderate hyperthermia planning for breast cancer patients," in *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO)*. IEEE, 2020, pp. 474–477.

[6] C. Kaur and R. Madaan, "Risk factor analysis in breast cancer using principal component analysis," in *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*. IEEE, 2023, pp. 482–487.

[7] S. J. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," *Sensors*, vol. 18, no. 9, p. 2799, 2018.

[8] A. Rovshenov and S. Peker, "Performance comparison of different machine learning techniques for early prediction of breast cancer using wisconsin breast cancer dataset," in *2022 3rd International Informatics and Software Engineering Conference (IISEC)*. IEEE, 2022, pp. 1–6.

[9] S. Qanoune, H. Ammor, and Z. Er-Reguig, "Revolutionizing early breast cancer detection: Insights from radiofrequency-based imaging and simulation studies," in *2024 International Conference on Computing, Internet of Things and Microwave Systems (ICCIMS)*. IEEE, 2024, pp. 1–5.

[10] R. Platania, S. Shams, S. Yang, J. Zhang, K. Lee, and S.-J. Park, "Automated breast cancer diagnosis using deep learning and region of interest detection (bc-droid)," in *Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics*, 2017, pp. 536–543.

[11] A. Das, M. N. Mohanty, P. K. Mallick, P. Tiwari, K. Muhammad, and H. Zhu, "Breast cancer detection using an ensemble deep learning method," *Biomedical Signal Processing and Control*, vol. 70, p. 103009, 2021.

[12] K.-C. Chu, M.-Y. Xiao, C.-H. Chang, C.-H. Hsiao, Y.-C. Jiang, and P.-Y. Tsai, "Preliminary study of relationship between health behavior and breast cancer," in *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 2019, pp. 410–413.

[13] A. Okaidat, A. Hmedat, and A. Al-Badarneh, "Breast cancer and anxiety: A relationship study," in *2020 11th International Conference on Information and Communication Systems (ICICS)*. IEEE, 2020, pp. 308–312.

[14] A. Melek, S. Fakhry, and T. Basha, "Spatiotemporal mammography-based deep learning model for improved breast cancer risk prediction," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2023, pp. 1–4.

[15] M. M. Rahman, M. Z. B. Jahangir, A. Rahman, M. Akter, M. A. A. Nasim, K. D. Gupta, and R. George, "Breast cancer detection and localizing the mass area using deep learning," *Big Data and Cognitive Computing*, vol. 8, no. 7, p. 80, 2024.

[16] T. A. Hajj, "Breast cancer detection using vgg16 model," 2023, accessed: 2025-01-09. [Online]. Available: https://www.kaggle.com/code/toufikalhajj/breast-cancer-detection-vgg16-model/input?select=Dataset$_{B}USI_{w}ith_{G}T$