ISSN: 3108-4060

ADBA

ARTIFICIAL INTELLIGENCE IN APPLIED SCIENCES

VOLUME 1, ISSUE 1,
JULY 2025

AN INTERDISCIPLINARY JOURNAL OF COMPUTER SCIENCE

https://journals.adbascientific.com/aiapp

ISSN: 3108-4060

Artificial Intelligence in Applied Sciences

Volume: 1 – Issue No: 1 (July 2025)

EDITORIAL BOARD

https://journals.adbascientific.com/aiapp/about/editorialTeam

ISSN: 3108-4060

Artificial Intelligence in Applied Sciences

Volume: 1 – Issue No: 1 (July 2025)

CONTENTS

https://journals.adbascientific.com/aiapp/issue/view/10

ADBA
Artificial Intelligence in Applied Sciences
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4060
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.1-6
https://doi.org/10.69882/adba.ai.2025071

Computational Efficiency and Accuracy of Deep Learning Models for Automated Breast Cancer Detection in Ultrasound Imaging

Luaay Alswilem *,1 **and Nurettin Pacal** α,2

*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye, αDepartment of Biology, Faculty of Arts and Sciences, Igdir University, 76000, Igdir, Türkiye.

ABSTRACT This study explores the trade-off between diagnostic performance and computational efficiency in deep learning models for the classification of breast cancer in ultrasound images. To this end, we evaluate three contemporary CNN architectures EfficientNetB7, EfficientNetV2-Small, and RexNet-200 in a multiple comparative study with standardized performance and complexity metrics. Our evaluations provide evidence that all three models achieved an identical high accuracy of 95.00%, but there were sizeable differences in the computational resources required to achieve that accuracy. RexNet-200 demonstrated tremendous computational efficiency, achieving identical performance with the least amount of resources (13.81M parameters; 3.05 GFLOPs) required compared to EfficientNetB7 which is much more computationally intensive. An examination of the confusion matrix for the models enhances the models clinical validity, as there are no malignant lesions misclassified as normal. Ultimately, our study clearly demonstrates that diagnostic accuracy is not a good metric for practical clinical deployment. RexNet-200, by representing high performance, with minimal resource utilization, is the most pragmatic and clinically applicable model, creating the opportunity to develop scalable and accessible CAD systems in resource-limited settings.

KEYWORDS
Breast cancer
Breast ultrasound
Deep learning
Computational efficiency
RexNet-200

INTRODUCTION

Breast cancer represents one of the most prevalent malignancies among women globally, characterized by the uncontrolled proliferation of epithelial cells within the breast tissue (Kim *et al.* 2025; Xiong *et al.* 2025). The etiology of the disease is rooted in a complex interplay of genetic predisposition with hormonal, environmental, and lifestyle factors (Obeagu and Obeagu 2024). As early detection significantly enhances treatment success and survival rates, the development of effective screening and diagnostic methodologies is of paramount importance. In this context, non-invasive medical imaging modalities assume a fundamental role in identifying pathological changes within the tissue (Kiani *et al.* 2025; Alshawwa *et al.* 2024; Begum *et al.* 2024). Although mammography is the cornerstone of standard screening, its diagnostic efficacy can be diminished, particularly in women with dense breast tissue, underscoring the need for supplementary imaging techniques (Katsika *et al.* 2024; Trentham-Dietz *et al.* 2024; Abeelh and AbuAbeileh 2024).

Owing to advantages such as its non-ionizing nature, widespread accessibility, and cost-effectiveness, ultrasonography is regarded as a valuable instrument for evaluating breast lesions (Iacob *et al.* 2024). It provides distinct benefits in clarifying suspicious mammographic findings, differentiating between cystic and solid masses, and guiding biopsy procedures, enhancing the characterization of lesions in women with dense parenchyma (Gordon *et al.* 2025). Nevertheless, its utility is constrained by certain limitations, including operator dependency, inter-observer variability in interpretation, and an inadequate capacity to detect microcalcifications. These challenges necessitate the development of more objective and standardized methodologies for the interpretation of ultrasound images (Vogel-Minea *et al.* 2025; Rana *et al.* 2024).

Recently, artificial intelligence (AI), and specifically deep learning (DL) techniques, have prompted a paradigm shift in the analysis of medical images (Karaman *et al.* 2023; Pacal *et al.* 2025; Pacal and Attallah 2025a; Zeynalov *et al.* 2025). Architectures such as Convolutional Neural Networks (CNNs) are delivering groundbreaking results in fields like radiology and pathology, attributed to their capacity to autonomously extract hierarchical features from large-scale datasets (Pacal 2024; Ozdemir *et al.* 2025; Lubbad *et al.* 2024b). In the context of breast cancer, DL models have demonstrated high success rates in the detection, classification, and segmentation of lesions across various imaging modalities, including mammography, ultrasound, and MRI (Pacal and Kılıcarslan 2023; COŞKUN *et al.* 2023; İnce *et al.* 2025; Lubbad *et al.* 2024a).

Manuscript received: 25 May 2025,
Revised: 5 June 2025,
Accepted: 12 June 2025.

[1]luaayalswilem3@gmail.com (**Corresponding author**)
[2]nurettin.pacal@igdir.edu.tr

The present study aims to enhance the effectiveness and accuracy of early breast cancer diagnosis through DL-based analysis of ultrasound images (Pacal and Attallah 2025b; Cakmak *et al.* 2024; Kurtulus *et al.* 2024; Bayram *et al.* 2025; Pacal 2025). Accordingly, a comparative performance evaluation is conducted by training three modern CNN architectures EfficientNetB7, EfficientNetv2-Small, and RexNet-200 on a publicly available breast ultrasound dataset. The ultimate objective of this research is to identify the architecture that offers the highest efficiency and performance for this specific diagnostic task (Pacal 2022; Cakmak and Pacal 2025).

The field of medicine is undergoing a profound transformation through the integration of artificial intelligence (AI) and its sub-disciplines, machine learning (ML) and deep learning (DL) (Obuchowicz *et al.* 2024; Koçak *et al.* 2025). These technologies present groundbreaking opportunities across a broad spectrum, from the early diagnosis of diseases to the personalization of treatment protocols and from drug discovery to the decryption of complex biological data (Li *et al.* 2024; Islam *et al.* 2024). The domain of medical imaging, in particular, holds significant potential due to the capacity of AI algorithms to detect subtle patterns beyond human perception and to rapidly analyze vast volumes of data (Chambi *et al.* 2025; Meng *et al.* 2024). Disciplines such as radiology, pathology, and oncology are swiftly adopting AI-powered systems for their potential to enhance diagnostic accuracy and improve patient outcomes. In the management of prevalent health issues like breast cancer, the fusion of AI with accessible imaging modalities such as ultrasound (US) opens new horizons for advancing early detection capabilities (Rajkumar *et al.* 2024).

Current research in the literature is focused on both developing integrated clinical decision support systems for breast cancer diagnosis and enhancing model performance through hybrid approaches. For instance, Gagliardi et al. proposed a holistic system that provides radiologists with both a segmentation mask and a classification result, reporting clinically valuable outcomes on the BUSI dataset with over 90% accuracy, 92% precision, and 90% recall (Gagliardi *et al.* 2024). In a different approach, Abhisheka et al. achieved an accuracy of 89.02% and an AUC of 0.8717 with a hybrid model (HBCPS) that combines deep learning (ResNet50) and handcrafted (HOG) features, using an SVM as the classifier (Abhisheka *et al.* 2025). Along similar lines, Latha et al. leveraged an EfficientNet-B7 architecture augmented with advanced data augmentation and interpretability (XAI) techniques like Grad-CAM, attaining a superior classification accuracy of 99.14%, particularly in recognizing minority classes (Latha *et al.* 2024).

Other lines of investigation are directed towards exploring segmentation performance, computational efficiency, and the potential offered by next-generation architectures like Mamba. In this context, Umer et al. focused on the segmentation task with a U-shaped autoencoder featuring a multi-attention mechanism, achieving high Dice scores of 90.45% and 89.13% on the UDIAT and BUSI datasets, respectively (Umer *et al.* 2024). With the objective of reducing computational cost, Cai et al. developed SC-Unext, a lightweight architecture, demonstrating the importance of model efficiency with 97.09% accuracy and a 75.29% Dice score (Cai *et al.* 2024). Finally, Sarvi et al. revealed that Mamba-based architectures can deliver significant performance gains over traditional CNNs and Transformers up to a 1.98% increase in AUC and 5.0% in accuracy by better capturing long-range dependencies in limited data scenarios (Nasiri-Sarvi *et al.* 2024). These collective efforts indicate that the field is in a state of continuous evolution towards more accurate, efficient, and innovative models.

MATERIALS AND METHODS

Dataset

For this study, we utilized the publicly available dataset "Breast Ultrasound Images Dataset" provided by sabahesaraki on Kaggle, for classifying breast ultrasound (US) images (Kaggle 2025). This dataset contains pathologically proven breast lesions with three basic classes of benign, malignant, and normal breast tissue. The heterogenous dataset serves as a valuable resource to evaluate deep learning models' ability to differentiate tissues with different morphologies and lesion types.

To ensure standardization and reproducibility in the model development and evaluation phases, the collection of 780 images (437 benign, 210 malignant, 133 normal) was methodically partitioned into training, validation, and testing subsets. This division allocated 70% of the collection (545 images) for model training, 15% (115 images) for the validation process, and the remaining 15% (120 images) for the test phase to impartially assess final model performance. This strategic partitioning aims to facilitate model training on sufficient data while reliably measuring generalization capabilities and mitigating the risk of overfitting. Furthermore, potential model bias towards any specific class was addressed by ensuring that the class distribution within each subset mirrored the proportions of the original dataset. Accordingly, the training, validation, and test sets were structured to contain (305B, 147M, 93N), (65B, 31M, 19N), and (67B, 32M, 21N) samples, respectively. A schematic of this dataset partitioning is also visualized in Figure 1.

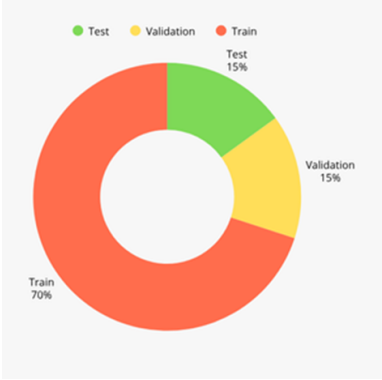


Figure 1 The Breast Ultrasound Images dataset was divided into three subsets: 70% for training, 15% for validation, and 15% for testing.

To elucidate the composition of the dataset and the visual distinctions among its classes, representative ultrasound images for each category (benign, malignant, and normal) are presented in Figure 2. Upon examination, benign lesions typically exhibit well-defined contours and a homogeneous internal structure. In contrast, malignant lesions often display morphological characteristics such as irregular borders, spiculated margins, and a heterogeneous internal echo pattern. Normal breast tissue, for its part, reflects characteristic fibroglandular and adipose tissue patterns. These examples not only highlight the morphological differences between the classes but also expose the inherent challenges associated with ultrasound imaging, such as speckle noise and low contrast. This visual presentation provides a valuable context for understanding the key distinguishing features that the models must learn to identify, and for appreciating the diversity encapsulated within the dataset.

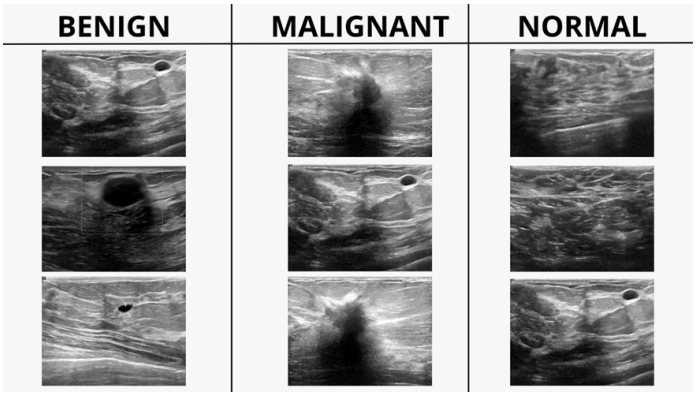


Figure 2 Sample ultrasound images from the dataset belonging to the benign, malignant, and normal classes.

Data Augmentation

To enhance the generalization performance of the deep learning models and mitigate the risk of overfitting a prevalent challenge in limited datasets typical of medical imaging this study incorporated a suite of on-the-fly data augmentation techniques into the training pipeline. As the focus of the task was on classification, the segmentation masks (mask.png) included in the original dataset were excluded from the analysis. The primary augmentation strategies applied to each image during the training loop included the following: images were first subjected to a "Random Resized Crop," where they were randomly cropped to a scale of 8% to 100% of the original area with an aspect ratio between 0.75 and 1.33, and subsequently resized to 224x224 pixels using a random interpolation method. Additionally, a random horizontal flip was applied with a 50% probability. To introduce color diversity, "Color Jitter" was employed, randomly altering the brightness, contrast, and saturation of the images by a factor of 0.4. Vertical flipping was not utilized in this work. This on-the-fly augmentation methodology ensures that the model encounters diverse variations of the data in each training epoch. This approach is designed to prevent the model from becoming overly dependent on the specific artifacts of the training set, thereby fostering a more robust and reliable performance on unseen data (Wang *et al.* 2024; Mumuni *et al.* 2024).

The Used Algorithms

In this research, we use deep Convolutional Neural Network (CNN) architectures, which have been shown to be effective in the computer vision research literature, to automatically classify breast ultrasound (US) images. In fields like medical imaging where there tends to be a small amount of labelled data, the advantage of transfer learning - as opposed to training a model from scratch - is considerable. By using transfer learning in particular fields, it is possible to leverage the feature extraction ability of models pre-trained on large data sets like ImageNet, and then challenge those features in a smaller and more specific target dataset. The use of a transfer learning approach is also intended to facilitate quicker convergence, better generalization, and a reduced chance of overfitting. For our project, all the CNN architectures we used were loaded with pre-trained weights from a model trained on ImageNet, and then we fine-tuned the models on our target breast ultrasound dataset.

The first set of architectures considered includes the EfficientNet family. EfficientNet families of architectures changed how researchers think about model scaling. Tan and Le proposed these architectures that scale model dimensions (Depth, Width, and Resolution) in a systematic way using what they call 'compound scaling' rather than just scaling in a random way. This principle allows for higher efficiency and accuracy using fewer parameters. EfficientNetB7, being a large and performant member of the family and the largest and most performant of the versions (that was scaled in this compound fashion) stands as a baseline for image classification tasks. The second architecture considered, EfficientNetV2-S, is a next-generation architecture that builds on the first and offers both faster training and a more efficient parameter.zip. It uses both MBConv and Fused-MBConv blocks and improved its training strategy to optimally achieve a good balance of speed and accuracy, especially for the S (Small) version (Tan and Le 2019, 2021).

RexNet-200, another modern architecture that we evaluated, was created for addressing the 'representational bottleneck' problem raised by standard designs. Rank eXpansion Networks (RexNets), as introduced by Han et al., are based on the idea that in standard convolutional blocks, channel narrowing-and-widening operations can lose information. RexNets work around this issue by providing blocks for networks to preserve and build the 'rank' of inter-layer channel representations, or the amount of unique information. This construction can facilitate a fuller and more varied flow of features between the layers, and thus increase the model's representational capabilities. RexNet-200, which we used in the study, is a higher-performing type of this architecture (with a 2.0 scaling factor) (Han *et al.* 2021).

Three distinct and contemporary CNN architectures EfficientNetB7, EfficientNetV2-S, and RexNet-200 were chosen to compare their respective performance in classifying breast ultrasound image classes as benign, malignant, and normal. Each model has its own design philosophies and contributions, including compound scaling, training optimization, and overcoming representational bottlenecks, which provides a broad view of the variability in CNN approaches to this complex medical classification task. The models are evaluated using comprehensive metrics to derive meaningful conclusions regarding the most suitable architecture for this task.

Performance Metrics

Measuring the performance of deep learning models is a critical step for assessing the practical value of these models, justifying methodological choices, and allowing data-driven choices. Relying on performance measures can have different purposes, such as evaluating the effectiveness of a model, guiding the optimization process, guarding against data errors or biases, allowing for an objective comparison between models, and identifying phenomena such as overfitting. The current paper adopts conventional evaluation criteria that are established and accepted in the academic literature that is specific to the issue of breast cancer classification.

The primary metrics employed within this project accuracy, precision, recall, and F1-score are indicators of central importance not only in deep learning evaluations but also in other disciplines. Accuracy, which offers an initial impression of general performance, is the ratio of correct predictions to the total number of instances. Precision, which measures the exactness of positive predictions, reflects the reliability of the model's positive labeling; high precision implies a low false positive rate. Recall, which measures the model's ability to identify all actual positive cases, indicates its success in detecting events that should not be missed. The F1-score, which combines these two metrics into a single measure, is the harmonic mean of precision and recall, serving as a balanced performance criterion that reflects the trade-off between false positives and false negatives. Conceptually, these definitions may also

be formulated through the mathematical expressions presented below.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

RESULTS AND DISCUSSION

The comparative analysis, as detailed in Table 1, reveals that the three evaluated architectures each achieved an identical accuracy of 95.00% in classifying breast ultrasound images. Behind this uniform accuracy score, however, lie significant divergences in other performance and complexity metrics. While EfficientNetV2-Small led in precision at 94.89%, EfficientNetB7 yielded the best results for recall and F1-score, at 95.38% and 94.41%, respectively. A striking paradox emerges when these performance data are considered alongside the computational costs of the models: EfficientNetB7, despite possessing some of the highest metrics, is the most resource-intensive model with 63.79 million parameters and 10.26 GFLOPs. In contrast, RexNet-200 attains the same high accuracy with only 13.81 million parameters and 3.05 GFLOPs, proving to be a remarkably efficient alternative that requires approximately 4.6 times fewer parameters and 3.4 times less computational power.

Table 1 A Comparison of Performance and Complexity in CNN Models for Classifying Breast Ultrasound Images

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Params (M)	GFLOPs
EfficientNetB7	95.00	93.58	95.38	94.41	63.79	10.26
EfficientNetV2-Small	95.00	94.89	93.75	94.28	20.18	5.42
RexNet-200	95.00	93.38	93.75	93.54	13.81	3.05

These findings demonstrate that for the evaluation of modern deep learning architectures, computational efficiency is a critical factor alongside diagnostic accuracy. RexNet-200, by achieving high accuracy with minimal resources, emerges as the most practical and convenient solution for Computer-Aided Diagnosis (CAD) systems intended for deployment in resource-constrained clinical environments or on local devices. With its high precision and balanced efficiency, EfficientNetV2-Small presents a strong option for scenarios where minimizing false positives is vital. On the other hand, despite its highest recall rate, the heavy computational burden of EfficientNetB7 significantly limits its scalability and practicality for real-world applications. This work clearly establishes the potential of efficient and lightweight architectures like RexNet-200 to enable the development of sustainable and accessible systems for the early diagnosis of breast cancer, without compromising on accuracy compared to their larger, more complex counterparts.

A detailed breakdown of the classification performance for the RexNet-200 model is provided in the confusion matrix presented in Figure 3. The concentration of values along the matrix's diagonal axis is an indicator of the model's success; it correctly classified a total of 116 samples (64 Benign, 31 Malignant, and 21 Normal). An analysis of the errors reveals that 2 Benign cases were misclassified as Malignant, and 1 Malignant case was misclassified as Benign. A particularly noteworthy finding that reinforces the model's clinical reliability is that no errors were made in the 'Normal' class, and crucially, no 'Malignant' case was overlooked as 'Normal' the most critical error scenario. This error profile corroborates the robust performance underlying the model's high accuracy rate.

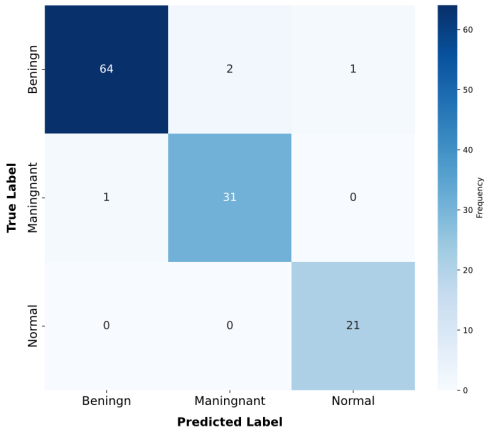


Figure 3 RexNet-200 model confusion matrix for breast ultrasound classification.

CONCLUSION

This study, by comparing three distinct deep learning models for the classification of breast ultrasound images, has demonstrated that computational efficiency is a decisive differentiating factor, even amidst an identical accuracy rate of 95.00%. The results beginning to clearly outline RexNet-200, a model that can know the diagnostic performance equivalent to a more complex architecture such as EfficientNetB7 according to leading metrics, but that comes with far less resource usage. Specifically, there is clearly a large resource-use advantage to the design of this architecture; the model operated with roughly 4.6 times fewer parameters than EfficientNetB7 and 3.4 times lower computational demands on the host system. More importantly, RexNet-200 was confirmed to be clinically robust based on the confusion matrix analysis, especially in terms of not misclassifying diagnoses of malignant as normal.

Thus, at minimum, this study provides evidence suggesting that simply pursuing the highest accuracy metric is not effective for the development of any future modern Computer-Aided Diagnosis (CAD) systems. However, one fundamental change is to move to architectures that maximize the trade-off between diagnostics accuracy and efficiency. Because of the success of RexNet-200 and the rest of our study, it is evident that it is possible to develop a system for the early diagnosis of breast cancer that is high-performance, scalable, sustainable, and can be used in settings where hardware resources are limited. There are several reasons why an efficient CAD model is best for the real world.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Abeelh, E. A. and Z. AbuAbeileh, 2024 Comparative effectiveness of mammography, ultrasound, and mri in the detection of breast carcinoma in dense breast tissue: a systematic review. Cureus **16**.

Abhisheka, B., S. K. Biswas, B. Purkayastha, and S. Das, 2025 Integrating deep and handcrafted features for enhanced decision-making assistance in breast cancer diagnosis on ultrasound images. Multimedia Tools and Applications pp. 1–23.

Alshawwa, I. A., H. Q. El-Mashharawi, F. M. Salman, M. N. A. Al-Qumboz, B. S. Abunasser, *et al.*, 2024 Advancements in early detection of breast cancer: Innovations and future directions .

Bayram, B., I. Kunduracioglu, S. Ince, and I. Pacal, 2025 A systematic review of deep learning in mri-based cerebral vascular occlusion-based brain diseases. Neuroscience .

Begum, M. M. M., R. Gupta, B. Sunny, and Z. L. Lutfor, 2024 Advancements in early detection and targeted therapies for breast cancer; a comprehensive analysis. Asia Pacific Journal of Cancer Research **1**: 4–13.

Cai, F., J. Wen, F. He, Y. Xia, W. Xu, *et al.*, 2024 Sc-unext: A lightweight image segmentation model with cellular mechanism for breast ultrasound tumor diagnosis. Journal of Imaging Informatics in Medicine **37**: 1505–1515.

Cakmak, Y. and I. Pacal, 2025 Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the wisconsin dataset. Journal of Operations Intelligence **3**: 175–196.

Cakmak, Y., S. Safak, M. A. Bayram, and I. Pacal, 2024 Comprehensive evaluation of machine learning and ann models for breast cancer detection. Computer and Decision Making: An International Journal **1**: 84–102.

Chambi, E. A., D. G. Alzamora, and A. A. Salas, 2025 Ultrasonic image processing for the classification of benign and malignant breast tumors: Comparative study of convolutional neural network architectures. Engineering Proceedings **83**: 15.

COŞKUN, D., D. KARABOĞA, A. BAŞTÜRK, B. Akay, Ö. U. NALBANTOĞLU, *et al.*, 2023 A comparative study of yolo models and a transformer-based yolov5 model for mass detection in mammograms. Turkish Journal of Electrical Engineering and Computer Sciences **31**: 1294–1313.

Gagliardi, M., T. Ruga, E. Vocaturo, and E. Zumpano, 2024 Predictive analysis for early detection of breast cancer through artificial intelligence algorithms. In *International Conference on Innovations in Computational Intelligence and Computer Vision*, pp. 53–70, Springer.

Gordon, P. B., L. J. Warren, and J. M. Seely, 2025 Cancers detected on supplemental breast ultrasound in women with dense breasts: update from a canadian centre. Canadian Association of Radiologists Journal p. 08465371251318578.

Han, D., S. Yun, B. Heo, and Y. Yoo, 2021 Rethinking channel dimensions for efficient model design. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 732–741.

Iacob, R., E. R. Iacob, E. R. Stoicescu, D. M. Ghenciu, D. M. Cocolea, *et al.*, 2024 Evaluating the role of breast ultrasound in early detection of breast cancer in low-and middle-income countries: a comprehensive narrative review. Bioengineering **11**: 262.

İnce, S., I. Kunduracioglu, B. Bayram, and I. Pacal, 2025 U-net-based models for precise brain stroke segmentation. Chaos Theory and Applications **7**: 50–60.

Islam, M. R., M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, *et al.*, 2024 Enhancing breast cancer segmentation and classification: An ensemble deep convolutional neural network and u-net approach on ultrasound images. Machine Learning with Applications **16**: 100555.

Kaggle, 2025 Breast ultrasound images dataset(busi).

Karaman, A., D. Karaboga, I. Pacal, B. Akay, A. Basturk, *et al.*, 2023 Hyper-parameter optimization of deep learning architectures using artificial bee colony (abc) algorithm for high performance real-time automatic colorectal cancer (crc) polyp detection. Applied Intelligence **53**: 15603–15620.

Katsika, L., E. Boureka, I. Kalogiannidis, I. Tsakiridis, I. Tirodimos, *et al.*, 2024 Screening for breast cancer: a comparative review of guidelines. Life **14**: 777.

Kiani, P., H. Vatankhahan, A. Zare-Hoseinabadi, F. Ferdosi, S. Ehtiati, *et al.*, 2025 Electrochemical biosensors for early detection of breast cancer. Clinica Chimica Acta **564**: 119923.

Kim, J., A. Harper, V. McCormack, H. Sung, N. Houssami, *et al.*, 2025 Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nature Medicine pp. 1–9.

Koçak, B., A. Ponsiglione, A. Stanzione, C. Bluethgen, J. Santinha, *et al.*, 2025 Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. Diagnostic and interventional radiology **31**: 75.

Kurtulus, I. L., M. Lubbad, O. M. D. Yilmaz, K. Kilic, D. Karaboga, *et al.*, 2024 A robust deep learning model for the classification of dental implant brands. Journal of Stomatology, Oral and Maxillofacial Surgery **125**: 101818.

Latha, M., P. S. Kumar, R. R. Chandrika, T. Mahesh, V. V. Kumar, *et al.*, 2024 Revolutionizing breast ultrasound diagnostics with efficientnet-b7 and explainable ai. BMC Medical Imaging **24**: 230.

Li, X., L. Zhang, J. Yang, and F. Teng, 2024 Role of artificial intelligence in medical image analysis: A review of current trends and future directions. Journal of Medical and Biological Engineering **44**: 231–243.

Lubbad, M., D. Karaboga, A. Basturk, B. Akay, U. Nalbantoglu, *et al.*, 2024a Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. Neural Computing and Applications **36**: 6355–6379.

Lubbad, M. A., I. L. Kurtulus, D. Karaboga, K. Kilic, A. Basturk, *et al.*, 2024b A comparative analysis of deep learning-based approaches for classifying dental implants decision support system. Journal of Imaging Informatics in Medicine **37**: 2559–2580.

Meng, X., J. Ma, F. Liu, Z. Chen, and T. Zhang, 2024 An interpretable breast ultrasound image classification algorithm based on convolutional neural network and transformer. Mathematics **12**: 2354.

Mumuni, A., F. Mumuni, and N. K. Gerrar, 2024 A survey of synthetic data augmentation methods in machine vision. Machine Intelligence Research **21**: 831–869.

Nasiri-Sarvi, A., M. S. Hosseini, and H. Rivaz, 2024 Vision mamba for classification of breast ultrasound images. In *Deep Breast Workshop on AI and Imaging for Diagnostic and Treatment Challenges in Breast Care*, pp. 148–158, Springer.

Obeagu, E. I. and G. U. Obeagu, 2024 Breast cancer: A review of risk factors and diagnosis. Medicine **103**: e36905.

Obuchowicz, R., M. Strzelecki, and A. Piórkowski, 2024 Clinical applications of artificial intelligence in medical imaging and

image processing—a review.
Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access .
Pacal, İ., 2022 Deep learning approaches for classification of breast cancer in ultrasound (us) images. Journal of the Institute of Science and Technology **12**: 1917–1927.
Pacal, I., 2024 Maxcervixt: A novel lightweight vision transformer-based approach for precise cervical cancer detection. Knowledge-Based Systems **289**: 111482.
Pacal, İ., 2025 Diagnostic analysis of various cancer types with artificial intelligence .
Pacal, I., O. Akhan, R. T. Deveci, and M. Deveci, 2025 Nextbrain: Combining local and global feature learning for brain tumor classification. Brain Research p. 149762.
Pacal, I. and O. Attallah, 2025a Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.
Pacal, I. and O. Attallah, 2025b Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.
Pacal, I. and S. Kılıcarslan, 2023 Deep learning-based approaches for robust classification of cervical cancer. Neural Computing and Applications **35**: 18813–18828.
Rajkumar, R., S. Gopalakrishnan, K. Praveena, M. Venkatesan, K. Ramamoorthy, *et al.*, 2024 Darknet-53 convolutional neural network-based image processing for breast cancer detection. Mesopotamian Journal of Artificial Intelligence in Healthcare **2024**: 59–68.
Rana, A. S., J. Rafique, and H. Riffat, 2024 Advances in breast ultrasound imaging: Enhancing diagnostic precision and clinical utility. In *Latest Research on Breast Cancer-Molecular Insights, Diagnostic Advances and Therapeutic Innovations*, IntechOpen.
Tan, M. and Q. Le, 2019 Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114, PMLR.
Tan, M. and Q. Le, 2021 Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pp. 10096–10106, PMLR.
Trentham-Dietz, A., C. H. Chapman, J. Jayasekera, K. P. Lowry, B. M. Heckman-Stoddard, *et al.*, 2024 Collaborative modeling to compare different breast cancer screening strategies: a decision analysis for the us preventive services task force. JAMA **331**: 1947–1960.
Umer, M. J., M. Sharif, and M. Raza, 2024 A multi-attention triple decoder deep convolution network for breast cancer segmentation using ultrasound images. Cognitive Computation **16**: 581–594.
Vogel-Minea, C. M., W. Bader, J.-U. Blohmer, V. Duda, C. Eichler, *et al.*, 2025 Best practice guidelines–degum recommendations on breast ultrasound. Ultraschall in der Medizin-European Journal of Ultrasound **46**: 245–258.
Wang, Z., P. Wang, K. Liu, P. Wang, Y. Fu, *et al.*, 2024 A comprehensive survey on data augmentation. arXiv preprint arXiv:2405.09591 .
Xiong, X., L.-W. Zheng, Y. Ding, Y.-F. Chen, Y.-W. Cai, *et al.*, 2025 Breast cancer: pathogenesis and treatments. Signal transduction and targeted therapy **10**: 49.
Zeynalov, J., Y. Çakmak, and İ. Paçal, 2025 Automated apple leaf disease classification using deep convolutional neural networks: A comparative study on the plant village dataset. Journal of Computer Science and Digital Technologies **1**: 5–17.

How to cite this article: Alswilem, L., and Pacal, N. Computational Efficiency and Accuracy of Deep Learning Models for Automated Breast Cancer Detection in Ultrasound Imaging. *Artificial Intelligence in Applied Sciences*, 1(1), 1-6, 2025.

ADBA
Artificial Intelligence in Applied Sciences
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4060
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.7-12
https://doi.org/10.69882/adba.ai.2025072

Income Level Estimation with Light-GBM: Understanding Model Decisions with Explainable AI Techniques Shap and Lime

Cem Özkurt [*,α,1], **Fatmir Garri** [α,2], **Bilal Emre Yahyaoğlu** [α,3], **Onur Ağca** [α,4], **Necip Furkan Bildiren** [α,5] **and Sergen Kaynak** [α,6]

*AI and Data Science Research and Application Center, Sakarya University of Applied Science, Türkiye, αDepartment of Computer Engineering, Sakarya University of Applied Sciences, Türkiye.

ABSTRACT This study examines the use of machine learning and artificial intelligence algorithms to predict individuals' annual incomes. In analyses conducted using the Python programming language, the best performance was achieved in models utilizing the "Synthetic Minority Over-sampling Technique (SMOTE)" for imbalanced data sets, with an accuracy of 87.45%, precision of 85.74%, recall of 89.31%, and an F1 score of 87.30, using the "Light Gradient Boosting Machines" algorithm. Additionally, the impact of parameters and variables on income prediction was examined using interpretable artificial intelligence algorithms. The results of the study emphasize the importance of employing effective methods and explaining machine learning model predictions, as well as addressing imbalanced data sets.

KEYWORDS
Revenue prediction
Machine learning
Explainable artificial intelligence
SMOTE

INTRODUCTION

The estimation of individuals' incomes is crucial for financial planning and resource management. This study aims to explore practical applications of machine learning and artificial intelligence algorithms to predict individuals' annual incomes. The research seeks to develop income prediction models and achieve more accurate estimations. Several studies focus on estimating individual income levels using various approaches. One study employed a machine learning approach to predict individual income and highlighted the issue of individuals misreporting their earnings (Matkowski 2021). Another study explored individual-level income prediction using Facebook profiles, examining the density distributions of annual income and comparing them with U.S. Census data (Matz *et al.* 2019). Additionally, the UCI Adult Dataset, a common resource for predicting annual income levels in the U.S., has been used to classify whether a person's income exceeds a certain threshold. This dataset has been applied to predict whether an individual's annual income surpasses $50,000 based on demographic data (Becker and Kohavi 2023).

An analysis was conducted on subjective expectations about future income changes using household panel data. This study found that income changes strongly depended on past changes. It also observed that expected income changes were significantly influenced by factors such as employment status, family structure, permanent income, and past expectations. The study concluded that expectations were not rational, particularly noting that households with decreasing past incomes underestimated future income growth (Das and Van Soest 1999). Another research examined weekly earnings expectations reported in subjective probabilities by participants in a national household survey during the spring of 1994. This study assessed the potential of obtaining expectations in future surveys, suggesting that such data could be more informative than typical economic expectations reports.

It also analyzed revisions in expectations and the relationship between expectations and actual earnings, providing positive findings on the validity of the data (Dominitz 1998). One study investigated the use of principal component analysis and support vector machines to create and evaluate income prediction data based on the U.S. Census Bureau's Current Population Survey. This research demonstrated the effectiveness of detailed statistical studies for rel-

Manuscript received: 25 November 2024,
Revised: 6 June 2025,
Accepted: 12 June 2025.

[1]cemozkurt@subu.edu.tr **(Corresponding author)**
[2]23502405002@subu.edu.tr
[3]24502405018@subu.edu.tr
[4]24502405013@subu.edu.tr
[5]24502405013@subu.edu.tr
[6]24502405008@subu.edu.tr

evant feature selection and their impact on improving classification accuracy. It also emphasized that shaping computational methods around specific real datasets is a critical factor in enhancing the power of algorithms (Lazar 2004).

MATERIALS AND METHODS

Dataset

In this study, the dataset was examined across 14 parameters (variables or features) (Becker and Kohavi 2023). Initially containing 48,842 observations, the dataset was reduced to 45,223 observations by removing those with missing information. The dataset was organized to serve as a suitable input for all machine learning algorithms used in the study. As part of this organization, a proper format conversion was applied to parameters that did not have numerical values.

Table 1 Parameters and Data Types of the Dataset Used

Parameter	Data Type
Age	Integer
Work class	Categorical
Education Level	Categorical
Marital Status	Categorical
Occupation	Categorical
Relationship	Categorical
Race/Ethnicity	Categorical
Gender	Categorical
Capital Gain	Integer
Capital Loss	Integer
Weekly Work Hours	Integer
Native Country	Categorical
Income	Binary

Machine Learning

Machine learning algorithms are computational models that enable computers to make data-driven predictions. Supervised machine learning involves training algorithms using a labeled dataset. This type of learning allows the model to learn a mapping from input data to output labels, enabling it to make predictions or classifications on unseen data (Michalski *et al.* 2013). In this study, six different machine learning algorithms were employed, and the results were obtained.

Decision Trees: Decision trees are a machine learning algorithm used to solve classification and regression problems. This algorithm analyzes the features in a dataset to reach a conclusion through a series of decisions. The most significant advantage of decision trees is their ease of interpretation. The general formula for a decision tree is as shown in (1):

$$f(x) = \sum_{m=1}^{M} c_m \cdot I(x \in R_m) \tag{1}$$

In this equation, $f(x)$ represents the predicted output for the input feature vector x. M enotes the total number of nodes in the tree. R_m represents the region at node m. The indicator function $I(x \in R_m)$ takes the value 1 if x belongs to the region R_m and 0 otherwise. c_m denotes the predicted value at node, m.

Random Forest: A random forest is an ensemble algorithm that combines multiple decision trees. The main idea behind the mathematical formulation of the random forest model is to aggregate the predictions of each decision tree, either by averaging them (for regression tasks) or by voting (for classification tasks) (Erdem *et al.* 2018).

$$f(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \tag{2}$$

In this formula, $f(x)$ represents the predicted target variable. N denotes the total number of trees, and $f_i(x)$ represents the prediction of the i-th tree for the input dataset x. This formula takes the predictions of each tree and then averages these predictions or performs voting to arrive at the final prediction. This approach helps make the model more stable and generally improves performance, as the error tendency of one tree can offset the errors of other trees. Additionally, the algorithm introduces randomness in tree construction, ensuring that each tree is different from the others.

Gradient Boosting : Gradient boosting is a machine learning method often employing tree-based algorithms. Its primary goal is to create a strong predictive model by combining weak learners (usually decision trees) (Atasoy and Demiröz 2021; Friedman 2001). Let the dataset consist of points (x_i, y_i) , $i = 1, 2, 3, ..., N$. Here x_i represents the input features, and, y_i represents the target variable. If the model's initial prediction is set to zero:

$$F_0(x) = 0 \tag{3}$$

In each iteration, a new prediction model is added to minimize the error function.

$$F_m(x) = F_{m-1}(x) + \rho \cdot h_m(x) \tag{4}$$

In this equation, m represents the number of iterations, ρ represents the learning rate and $h_m(x)$ represents the newly added weak model.

Extreme Gradient Boosting (XGBoost): Extreme Gradient Boosting (XGBoost) is a machine learning algorithm and fundamentally a tree-based model. This algorithm is an ensemble model that combines a series of weakly learned Decision Trees. These trees are structured to complement each other and correct errors (Chen *et al.* 2019; Mitchell and Frank 2017). The Extreme Gradient Boosting algorithm is frequently used to solve regression and classification problems, and its mathematical function is generally as shown in (5):

$$F(x) = L(\theta) + \Omega(\theta) \tag{5}$$

In this equation, $L(\theta)$ represents the loss function, measuring how far the model's predictions deviate from the actual values. For classification problems, cross-entropy functions can be used as the loss function.

$$L(\theta) = \sum_{i=1}^{n} \left(-y_i \log(\hat{y}_i) + (1 - y_i) \log(\hat{y}_i)\right) \tag{6}$$

In this equation n represents the total number of data points, $\hat{y}_i$, represents the actual value of the i-th data point, and $\hat{y}_i$, represents the model's prediction for the i-th data point.

$$\Omega(\theta) = \gamma T + \frac{\lambda}{2} \sum_{j=1}^{T} w_j^2 \tag{7}$$

In this equation T represents the number of trees and w_j represents the node weights of the j-th tree. γ adds a regularization term to each tree and controls the addition of trees. λ controls the complexity of the tree by penalizing the square of the node weights. The last term of Equation (5), θ, represents the parameters of the model. These parameters include the decision rules at the nodes of each tree, the weights, and other features.

Adaptive Boosting (AdaBoost): Adaptive Boosting is an ensemble learning algorithm that combines multiple weak learners to create a strong learner. This algorithm weights each learner based on its misclassification rate after training it, using a weighted error function (Bulut 2016). This error represents the difference between the actual label y_i and the predicted label $(h_t(x_i))$.

$$\epsilon = \sum_{i=1}^{N} w_{(t,i)} \cdot \prod (h_t(x_i) \neq y_i) \tag{8}$$

In this equation, N represents the number of data points, $w_{(}(t,i))$ is the weighting factor of the t-th learner, and $\Pi(\cdot)$ is the representative indicator function.

Weights are assigned to the learners using the formula in (9).

$$a_t = \frac{1}{n} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right) \tag{9}$$

In this formula, ϵ_t, represents the weighted error rate. The assigned weights depend on the performance of the learner. To update the weights, the formula in (10) is used.

$$w_{t+1,i} = \frac{w_{t,i} \cdot e^{-a_t \cdot y_t \cdot h_t(x_i)}}{Z_t} \tag{10}$$

In this equation, Z_t is the normalization factor that ensures the sum of the weights equals 1. With the contributions of all learners, a strong learner is created using the formula below. In this way, the Adaptive Boosting algorithm combines a series of weak learners to form a strong learner.

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} a_t \cdot h_t(x)\right) \tag{11}$$

Light Gradient Boosting Machines (LightGBM): Light Gradient Boosting Machines (LightGBM) is an implementation of the Gradient Boosting framework, a machine learning framework. Therefore, the mathematical formula of the LightGBM algorithm generally resembles the formula of the Gradient Boosting algorithm. LightGBM stands out with features such as histogram-based learning and scaled gradient descent. Its basic mathematical formula is as shown in (12).

$$F_m = F_{m-1}(x) + \eta \cdot h_m(x) \tag{12}$$

In this equation, $F_m(x)$, represents the total prediction after adding the m-th . $F_{(m-1)}(x)$, is the prediction after m-1 trees have been added. η represents the learning rate and $h_m(x)$, is the contribution of the m-th tree. Light Gradient Boosting Machines accelerate the learning process and allow for reduced memory usage, particularly due to their use of histogram-based learning.

eXplainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) refers to a set of processes and methods aimed at providing clear and understandable explanations for the decisions offered by Machine Learning models. The architecture of XAI consists of three main components (Adadi and Berrada 2018; Vilone and Longo 2020): 1) Machine Learning Algorithm, 2) Explanation Algorithm 3) Interface The explanation algorithm is used to provide information about the most relevant and influential factors in the process. The interface component presents the information generated by the explanation algorithm.In this study, the two most popular algorithms of Explainable Artificial Intelligence were examined.

Local Interpretable Model-Agnostic Explanations (LIME): LIME is a popular Explainable Artificial Intelligence approach that uses the local behavior of a model to provide interpretable and explainable information about the most relevant and influential factors in predictions. The LIME algorithm generally follows the steps of categorizing numerical variables, generating new observations similar to the dataset's distribution, and developing an explainable model based on this dataset to determine the effects of variables on observations (Ribeiro *et al.* 2016; ElShawi *et al.* 2021). The general mathematical representation of the LIME model is as follows:

$$e(x) = \underset{g \in \varsigma}{\text{argmin}}(f, g, \pi_x) + \Omega(g) \tag{13}$$

In this equation x is the example being explained. The explanation of x (f, g, π_x) is the result of maximizing the fidelity term while considering the complexity $\Omega(g)$.Here, f represents a black-box model being explained, and g represents the interpretable model that explains (Molnar 2018).

Shapley Additive Explanations (SHAP): SHAP is an Explainable Artificial Intelligence approach that uses Shapley values, derived from game theory, to provide interpretable information about the most important and influential factors in predictions. Shapley values originate from cooperative game theory and represent a concept that fairly measures a player's contribution. SHAP provides a framework for understanding how a model makes predictions using these values (Lundberg and Lee 2017; Antwarg *et al.* 2021).The general mathematical formula of the SHAP algorithm is as follows:

$$\phi_i(f) = \frac{1}{N!} \sum_{\pi} \left[f(x_{\pi(i)}) - f(x_{\pi})\right] \tag{14}$$

In this equation, $f(x)$ represents the output of the model (where x is the input features.). Here π represents all N! permutations, and $x_{(\pi(i))}$ is thei-th permutation of x's feature according to π . The SHAP value adapts Shapley values to understand the contribution of each parameter to the model's prediction.

Model Performance Metrics

Model performance metrics are measurements used to evaluate the performance of a machine learning model (Cıhan and Coşkun 2021). These metrics can be used to understand how well a model performs, compare different models, or tune the model's hyperparameters.

Confusion Matrix: A confusion matrix is used to interpret the results of a classification model and to cross-examine the errors in the relationship between actual and predicted values.

Table 2 Confusion Matrix showing prediction outcomes versus actual values

		Actual Values	
		Positive (1)	**Negative (0)**
Predicted Values	**Positive (1)**	TP [1,1]	FP [1,0]
	Negative (0)	FN [0,1]	TN [0,0]

- True Positive (TP): Correctly predicting the positive condition.
- True Negative (TN): Correctly predicting the negative condition.
- False Positive (FP): Incorrectly predicting the positive condition.
- False Negative (FN): Incorrectly predicting the negative condition.

Accuracy, Precision, Recall, F1-Score: These scores are derived from the confusion matrix and help provide a clearer understanding of the model's success.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{16}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{17}$$

$$F1\text{-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

DISCUSSION AND RESULTS

In this study, the operations were performed using Python programming language version 3.11.4. In the dataset used, the predicted parameter was classified based on whether the annual income of a person/observation is less than or greater than $ 50,000. Among the 34,014 observations, the annual income is less than $ 50,000, while for 11,208 observations, it is more than $ 50,000. The percentage distribution of the target variable in the dataset is presented in Figure 1.

Figure 1 Class Distribution of the Target Variable in the Dataset.

After the other parameters in the dataset were adjusted to be suitable inputs for machine learning algorithms, the data was trained using six different machine learning algorithms. Since the target variable in the dataset is binary, classification algorithms were preferred over regression algorithms. The performance metrics of the models used are presented in Figure 2.

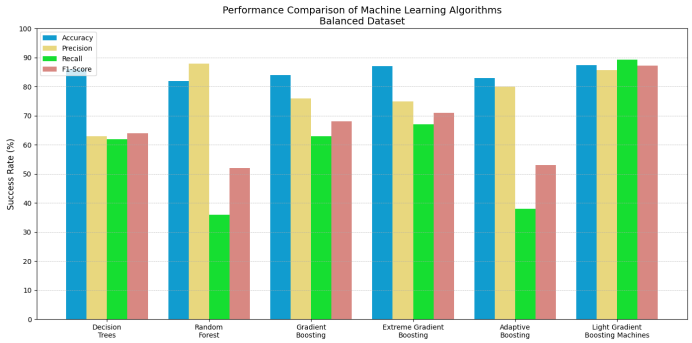


Figure 2 Performance Comparison of Machine Learning Algorithms Balanced Dataset.

Since the target variable's class distribution in the dataset is imbalanced, it is classified as an imbalanced dataset. Performance metrics based on the correct classification rate are unsuitable for this scenario (Chawla *et al.* 2002), as confirmed by the results in Figure 2. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied, generating synthetic examples for the minority class by creating artificial instances along the line segments joining each minority class instance with its nearest neighbors (Blagus and Lusa 2013). After applying SMOTE, the target variable "Income" was balanced, consisting of 23,756 examples earning more than $ 50,000 and 23,756 examples earning less.

After addressing the imbalance in the dataset, the machine learning algorithms were retrained and tested. The results are presented in Figure 3.

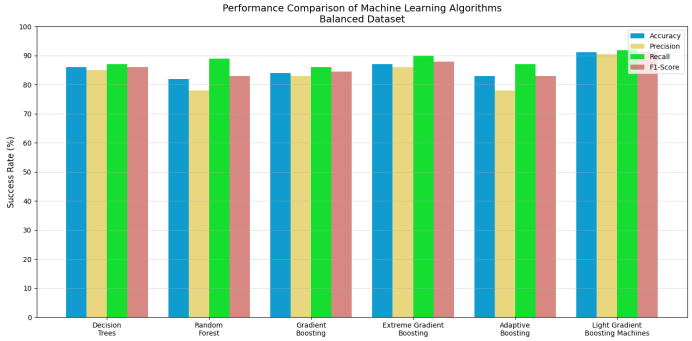


Figure 3 Comparison of Machine Learning Algorithm Performances.

The performance comparisons of the machine learning algorithms used in the study for the imbalanced dataset and the dataset adjusted with SMOTE are shown in Figures 2 and 3. The algorithm that demonstrated the best performance, as seen in these figures, was the Light Gradient Boosting Machines (LightGBM) algorithm. The results obtained from this algorithm were explained using Explainable Artificial Intelligence algorithms.

As shown in Figure 4, the "Capital Gain" parameter has the most significant impact on determining which class a sample belongs to. Following this, "Marital Status" is observed to be another key parameter influencing an individual's annual income. In contrast, the parameters "Workclass" and "Sex" are seen to have the least effect on an individual's annual income.

In Figure 5, we can examine in greater detail the impact of the parameters/variables in the dataset on the classes of the target variable. In the graph in Figure 5a, we observe the SHAP values of the parameters/variables when the annual income of our example

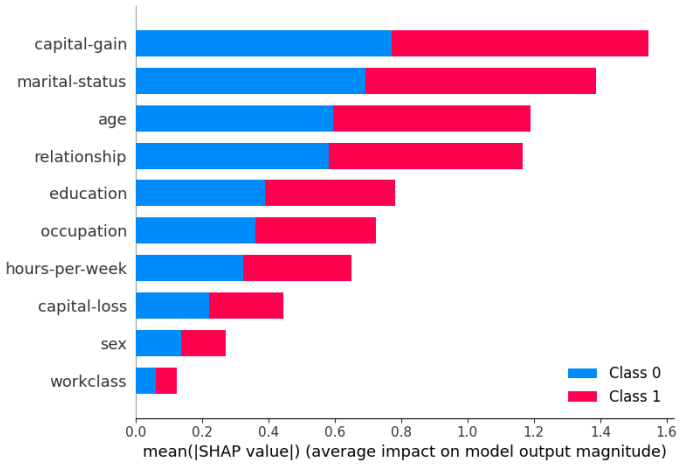
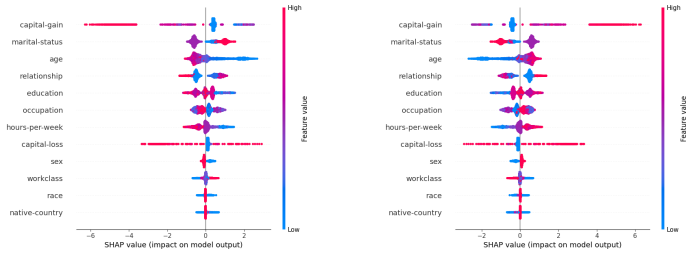


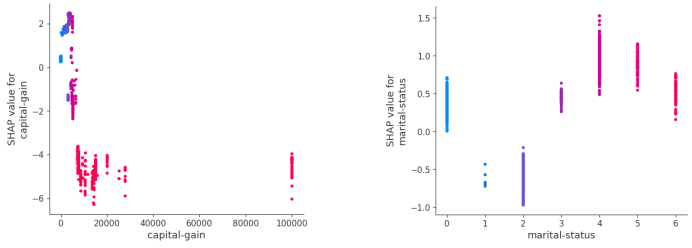
Figure 4 Class Distribution of the Target Variable in the Dataset.



(a) **(b)**

Figure 5 SHAP value distribution for different models. (a) Model A, (b) Model B.

(individual) is less than $ 50,000, while in Figure 5b, we observe the SHAP values when the annual income exceeds $ 50,000. For instance, the SHAP value of the "Capital Gain" parameter ranges between -2 and 6 (or between -6 and 2 for the "0" label), and when the target variable's label is 1, the SHAP value ranges from 0 to 6.



(a) **(b)**

Figure 6 Parameter Dependence Plot of the SHAP Algorithm.

Another interface provided by the SHAP algorithm, the Parameter Dependence Plot (Figure 6), allows us to observe the classification of the observation/individual based on the value of the relevant parameter. In Figure 6a, the dependence of the "Capital Gain" parameter on the class to which the observation/individual belongs is shown, while in Figure 6b, the dependence of the "Marital Status" parameter on the class of the observation/individual is illustrated. From this, it can be inferred that if the "Capital Gain" parameter value exceeds 100,000, the annual income of the observation/individual will be more than $ 50,000. With the SHAP algorithm, an Explainable Artificial Intelligence algorithm, we can clearly observe the impact of parameters/variables and their values on the classes of the target variable. On the other hand, with another Explainable Artificial Intelligence algorithm, LIME, we examine locally which parameters influence the classification of the observation/individual.

(a) **(b)**

Figure 7 Variable Importance Plot of the LIME Algorithm.

The graphs in Figure 7 show the probability of an individual belonging to a specific class, the parameters influencing this classification, and the actual values of those parameters for the observation. For example, the observation/individual in Figure 7a has a 97% probability of having an annual income greater than $ 50,000. In contrast, the observation/individual in Figure 7b has a 93 % probability of having an annual income less than $ 50,000.

CONCLUSION

Studies investigating the practical use of machine learning and artificial intelligence algorithms to predict individuals' annual income generally aim to develop income prediction models and achieve more accurate predictions. In line with this goal, this study conducted a comprehensive analysis of several machine learning algorithms using the Python programming language. The imbalanced distribution of the dataset was corrected using the SMOTE method, followed by a comparison of the performance of the machine learning algorithms. At this stage, the best performance was achieved with the Light Gradient Boosting Machines algorithm. Additionally, the impact of parameters/variables on the classes was analyzed using explainable artificial intelligence algorithms. These analyses helped us better understand the study's results and explain the decisions made by the model. These findings underscore the importance of employing effective methods to handle imbalanced datasets and interpret the predictions of machine learning models in data science applications.

Acknowledgments

This work was supported by AI and Data Science Research and Application Center, Sakarya University of Applied Science University Scientific Research Projects Unit.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

Not applicable.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Adadi, A. and M. Berrada, 2018 Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access **6**: 52138–52160.

Antwarg, L., R. M. Miller, B. Shapira, and L. Rokach, 2021 Explaining Anomalies Detected by Autoencoders Using Shapley Additive Explanations. Expert Systems with Applications **186**: 115736.
Atasoy, N. A. and A. Demiröz, 2021 Makine Öğrenmesi Algoritmaları Kullanılarak Prostat Kanseri Tümör Oluşumunun İncelenmesi. Avrupa Bilim ve Teknoloji Dergisi pp. 87–92.
Becker, B. and R. Kohavi, 2023 Adult, UCI Machine Learning Repository (1996). https://doi.org/10.24432/C5XW20.
Blagus, R. and L. Lusa, 2013 Improved Shrunken Centroid Classifiers for High-Dimensional Class-Imbalanced Data. BMC Bioinformatics **14**: 1–13.
Bulut, F., 2016 Determining Heart Attack Risk Ratio Through AdaBoost. Celal Bayar University Journal of Science **12**: 459–472.
Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 2002 SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research **16**: 321–357.
Chen, X., Z. X. Wang, and X. M. Pan, 2019 HIV-1 Tropism Prediction by the XGBoost and HMM Methods. Scientific Reports **9**: 9997.
Cıhan, P. and H. Coşkun, 2021 Performance Comparison of Machine Learning Models for Diabetes Prediction. In *2021 29th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE.
Das, M. and A. Van Soest, 1999 A Panel Data Model for Subjective Information on Household Income Growth. Journal of Economic Behavior & Organization **40**: 409–426.
Dominitz, J., 1998 Earnings Expectations, Revisions, and Realizations. Review of Economics and Statistics **80**: 374–388.
ElShawi, R., Y. Sherif, M. Al-Mallah, and S. Sakr, 2021 Interpretability in Healthcare: A Comparative Study of Local Machine Learning Interpretability Techniques. Computational Intelligence **37**: 1633–1650.
Erdem, F., M. A. Derinpinar, R. Nasirzadehdizaji, O. Y. Selen, D. Z. Şeker, *et al.*, 2018 Rastgele Orman Yöntemi Kullanılarak Kıyı Çizgisi Çıkarımı: İstanbul Örneği. Geomatik **3**: 100–107.
Friedman, J. H., 2001 Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics pp. 1189–1232.
Lazar, A., 2004 Income Prediction via Support Vector Machine. In *Proceedings of the International Conference on Machine Learning and Applications (ICMLA)*, pp. 143–149.
Lundberg, S. and S.-I. Lee, 2017 A Unified Approach to Interpreting Model Predictions. CoRR **abs/1705.07874**: 1–15.
Matkowski, M., 2021 Prediction of Individual Income: A Machine Learning Approach. Preprint or working paper.
Matz, S. C., J. I. Menges, D. J. Stillwell, and H. A. Schwartz, 2019 Predicting Individual-Level Income from Facebook Profiles. PLOS ONE **14**: e0214369.
Michalski, R. S., J. G. Carbonell, and T. M. Mitchell, 2013 *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media.
Mitchell, R. and E. Frank, 2017 Accelerating the XGBoost Algorithm Using GPU Computing. PeerJ Computer Science **3**: 1–20.
Molnar, C., 2018 *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub, https://christophm.github.io/interpretable-ml-book/.
Ribeiro, M. T., S. Singh, and C. Guestrin, 2016 "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, ACM.
Vilone, G. and L. Longo, 2020 Explainable Artificial Intelligence: A Systematic Review. arXiv preprint **arXiv:2006.00093**: 1–39.

How to cite this article: Özkurt, Ö., Garri, F., Yahyaoğlu, B. E., Ağca, O., Bildiren, N. F., and Kaynak, S. Income Level Estimation with Light-GBM:Understanding Model Decisions with Explainable AI Techniques Shap and Lime. *Artificial Intelligence in Applied Sciences*, 1(1), 7-12, 2025.

ADBA
Artificial Intelligence in Applied Sciences
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4060
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.13-19
https://doi.org/10.69882/adba.ai.2025073

Deep Learning for Automated Breast Cancer Detection in Ultrasound: A Comparative Study of Four CNN Architectures

Yigitcan Cakmak *,1 **and Nurettin Pacal** α,2

*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye, αDepartment of Biology, Faculty of Arts and Sciences, Igdir University, 76000, Igdir, Türkiye.

ABSTRACT Breast cancer is one of the most common malignancies among women globally, and it constitutes a significant public health problem in terms of morbidity and mortality. Since early-stage diagnosis significantly increases treatment success and survival rates, effective screening and diagnostic methods are of great importance. Various imaging modalities, such as mammography, ultrasonography (US), and magnetic resonance imaging, play a critical role in the detection of breast cancer. Ultrasound, in particular, is a valuable imaging method due to its non-ionizing nature, its accessibility, and its role as a complementary tool in dense breast tissue. In recent years, deep learning (DL) algorithms, particularly Convolutional Neural Networks (CNNs), have exhibited promising results in medical image analysis, especially in cancer detection. The aim of this research is to investigate and compare the four most common CNN architectures, ResNet50, DenseNet169, InceptionV3 and InceptionV4, for breast ultrasound images to classify breast cancer automatically. We have utilized publicly available breast ultrasound image datasets for the models and reported results in metrics of accuracy, precision, sensitivity, and F1-score. The InceptionV3 architecture had the best performance across the models examined with metrics of accuracy: 96.67%, precision: 96.55%, sensitivity: 96.38%, and F1-score: 96.41%. It was also noticed that the DenseNet169 model performed similarly to the InceptionV3 model but had substantially fewer parameters. The results of this study suggest that the InceptionV3 DL architecture may have significant potential for accuracy in the classification of cancer from breast ultrasound images and can contribute to the development of computer aided diagnosis systems for the early detection of breast cancer.

KEYWORDS
Breast cancer
Deep learning
Breast ultrasound
Image classification
Computer-aided diagnosis (CAD)

INTRODUCTION

Breast cancer is one of the leading cancers that affect women's health around the world and is the abnormal and unregulated growth of mammary epithelial cells (Kim *et al.* 2025; Xiong *et al.* 2025). The origin of breast cancer is a multifactorial process mediated by genetic susceptibility, hormones, lifestyle, and environmental factors (Obeagu and Obeagu 2024). Given that the prognosis for treatment response and survival rate improve drastically if the cancer is found at an early stage, better screening and diagnostic strategies are crucial. Therefore, it is important to study medical imaging techniques that are the least invasive way to characterize abnormal changes in the breast (Kiani *et al.* 2025; Alshawwa *et al.* 2024; Begum *et al.* 2024). Breast cancer screening programs have relied on mammography as the definitive tool of choice (Katsika *et al.* 2024; Trentham-Dietz *et al.* 2024). However, mammography may lack diagnostic sensitivity particularly with women with dense breast tissue and in women who are younger. This raises the need for complementary or alternative forms of imaging (Abeelh and AbuAbeileh 2024).

Ultrasonography is a valuable component in evaluating breast lesions, primarily due to its non-ionizing nature, availability, affordability, and real-time images (Iacob *et al.* 2024). Specifically, it has advantages for the evaluation of breast lesions, especially in determining whether suspicious findings on mammography are cystic or solid masses, and facilitates biopsy procedures. For women with dense breast parenchyma, ultrasonography is essentially an adjunct that improves mammographic diagnostic performance and provides clarity in graphically characterizing lesions (Gordon *et al.* 2025). However, ultrasonography has disadvantages, including operator-dependency that introduces inter-observer variability in the detection and interpretation of lesions. Additionally, ultrasonography is limited in its ability to detect microcalcifications. Research has been conducted to evaluate new ways to provide more objective and standardized analysis of ultrasound images (Vogel-Minea *et al.* 2025; Rana *et al.* 2024).

Artificial intelligence (AI), and deep learning (DL) algorithms in particular, have generated paradigm shifting advances in medical image analysis in recent years (Pacal *et al.* 2025; Pacal and Attallah 2025a). DL architectures, such as convolutional neural networks (CNNs), have shown significant potential in many med-

Manuscript received: 29 May 2025,
Revised: 15 June 2025,
Accepted: 19 June 2025.

[1]ygtcncakmak@gmail.com (**Corresponding author**)
[2]nurettin.pacal@igdir.edu.tr

ical specialties, including radiology and pathology, due to their ability to automatically learn complex patterns and hierarchical features from massive image datasets (Ozdemir *et al.* 2025; Lubbad *et al.* 2024b; Pacal 2025). With respect to breast cancer, DL models have demonstrated excellent performance to achieve high levels of accuracy in the detection, classification, and segmentation of suspicious lesions from mammograms, ultrasound images, and magnetic resonance scans (İnce *et al.* 2025; Lubbad *et al.* 2024a). In this paper, we plan to train several CNN algorithms (e.g. ResNet50, DenseNet169, InceptionV3, InceptionV4) with breast ultrasound image dataset publicly available online and then analyze and compare the results to determine their potential to assist clinicians in breast cancer diagnosis (Pacal and Attallah 2025b; Cakmak *et al.* 2024; Bayram *et al.* 2025). The hope is that by performing a DL based analysis of ultrasound images we can assist with the early diagnosis process and also improve diagnostic accuracy (Cakmak and Pacal 2025; Zeynalov *et al.* 2025; Kurtulus *et al.* 2024).

The field of medicine is undergoing a transformative evolution through the integration of AI, particularly its sub-disciplines of DL and machine learning (ML) (Obuchowicz *et al.* 2024; Koçak *et al.* 2025). These technologies offer revolutionary advancements across a broad spectrum, ranging from the early diagnosis of diseases to the development of personalized treatment protocols, from drug discovery to the analysis of complex biological data (Li *et al.* 2024; Islam *et al.* 2024). Medical imaging, in particular, holds immense potential due to the capacity of AI algorithms to detect subtle patterns and anomalies imperceptible to the human eye and to rapidly process and interpret large volumes of data (Chambi *et al.* 2025; Meng *et al.* 2024). Disciplines such as radiology, pathology, and oncology are rapidly adopting these innovations with the promise of enhancing diagnostic accuracy, optimizing workflows, and ultimately improving patient outcomes. In the management of prevalent and serious health issues like breast cancer, the combination of AI with accessible imaging modalities such as ultrasound is opening promising avenues for early-stage detection and effective treatment strategies (Rajkumar *et al.* 2024).

In studies on breast cancer classification and segmentation, various AI approaches have gained prominence. Abhisheka et al., highlighting the importance of breast cancer in the healthcare sector, noted the insufficiency of traditional ML or DL models alone and, accordingly, proposed the Hybrid Breast Cancer Prediction System (HBCPS) model. This system combines deep CNN features (obtained via ResNet50) with handcrafted features (Histogram of Oriented Gradients - HOG) and uses a Support Vector Machine (SVM) for classification. The system also incorporates a Block-Matching and 3D filtering (BM3D) filter to reduce noise in Breast Ultrasound (BUS) images, achieving satisfactory results on the BUSI dataset, such as 89.02% accuracy and an AUC of 0.8717 (Abhisheka *et al.* 2025). Similarly, Latha *et al.* (2024) combined a scalable CNN architecture, EfficientNet-B7, with advanced data augmentation techniques to address low accuracy in minority classes, particularly malignant tumors. They also integrated eXplainable AI (XAI) techniques like Grad-CAM to enhance the interpretability of the model's predictions. With this approach, they achieved a high classification accuracy of 99.14%, significantly outperforming existing CNN-based approaches. These studies underscore the potential of both hybrid modeling and the integration of advanced CNN architectures with XAI techniques in breast cancer classification.

Other notable contributions in the literature have focused on improving segmentation accuracy and computational efficiency. Umer *et al.* (2024) proposed a U-shaped autoencoder-based CNN model featuring a multi-attention mechanism and a triple decoder, focusing on capturing multi-scale spatial features and highlighting the tumor region, particularly in BC segmentation from U/S images. Their proposed model achieved Dice scores of 90.45% and 89.13% on the UDIAT and BUSI datasets, respectively. On the other hand, Cai *et al.* (2024), as a solution to the challenges of high computational complexity and large model parameters in existing segmentation methods, developed the SC-Unext model. This model, based on the Unext network and inspired by the mechanisms of cellular apoptosis and division, not only improved segmentation performance but also reduced model parameters and computational resource consumption, achieving a 75.29% Dice score and 97.09% accuracy on the BUSI dataset; it also demonstrated fast inference times on CPUs. These studies demonstrate the importance of developing not only complex architectures but also efficient and lightweight models, especially for segmentation tasks.

Finally, the comparison of next-generation architectures and the development of holistic systems for clinical application also hold a significant place on the research agenda. Cai *et al.* (2024) compared Mamba-based models (VMamba and Vim) with traditional CNNs and Vision Transformers (ViTs), demonstrating that some Mamba-based architectures offer statistically significant performance improvements, particularly due to their ability to capture long-range dependencies in limited data. For instance, on dataset B, Mamba-based models were reported to provide an improvement of 1.98% in mean AUC and 5.0% in mean Accuracy. Nasiri-Sarvi *et al.* (2024) adopted an approach aimed at presenting the radiologist with both the tumor mask and its classification. They examined different DL models and identified the best-performing one, which achieved over 90% accuracy, 92% precision, 90% sensitivity, and a 90% F1-score on the BUSI dataset. This study emphasizes that DL architectures are effective in the classification and segmentation of ultrasound breast images and could be used in clinical trials in the near future. Such comparative studies and proposals for integrated systems further solidify the role of AI in breast cancer diagnosis and pave the way for its clinical adaptation (Gagliardi *et al.* 2024).

MATERIALS AND METHODS

Dataset

In this research, a publicly available dataset, the "Breast Ultrasound Images Dataset", was used to classify and analyze breast ultrasound images. This dataset was made available through the Kaggle platform by Sabah Saraki (Kaggle 2025), and contains ultrasound images which demonstrate different appearances of breast cancer. The dataset contains samples of ultrasound images grouped into three main classes based on pathologically confirmed diagnoses of benign tumors, malignant tumors, and normal breast tissue images. This variety gives a solid ground for evaluating the capability of the DL models to distinguish other tissue structures and lesion types.

In order to ensure a standardized and reproducible model development and evaluation process, the dataset, comprising a total of 780 samples (437 benign, 210 malignant, and 133 normal), was meticulously partitioned into training, validation, and testing subsets. This partitioning allocated 70% of the data (545 samples) to the training set, 15% (115 samples) to the validation set, and the remaining 15% (120 samples) to the test set. These proportions were selected to ensure the model is trained on sufficient data, while simultaneously allowing for a reliable assessment of its generalization capability and mitigating the risk of overfitting. Furthermore, a stratified sampling approach was employed to ensure that the

class distribution within each subset precisely mirrors that of the original dataset, a crucial step to prevent the model from developing a bias towards any particular class. Consequently, the training set was composed of 305 benign, 147 malignant, and 93 normal samples; the validation set contained 65 benign, 31 malignant, and 19 normal samples; and the test set consisted of 67 benign, 32 malignant, and 21 normal samples. This dataset partitioning is also illustrated in Figure 1.

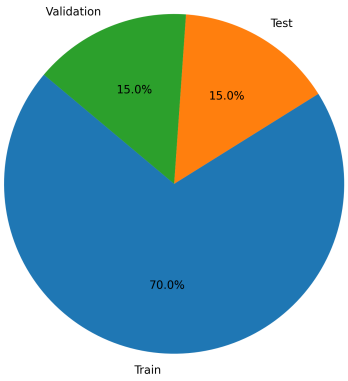


Figure 1 Distribution of the Breast Ultrasound Images Dataset into Training, Validation, and Test Sets (70%-15%-15%)

To better visualize the structure of the dataset and the types of images it contains, representative ultrasound images from each class (benign, malignant, and normal) are presented in Figure 2. As can be seen in Figure 2, benign lesions generally present with regular borders and a homogeneous internal echo, whereas malignant lesions may exhibit more irregular margins, spiculated extensions, and heterogeneous internal structures. Normal breast tissue images, in turn, show typical fibroglandular and adipose tissue patterns. These examples reflect not only the visual differences between the classes but also the inherent challenges of ultrasound imaging, such as speckle noise and low contrast. These visual representations help in understanding the fundamental morphological features that our models must learn and differentiate, and they provide an insight into the diversity of the dataset.

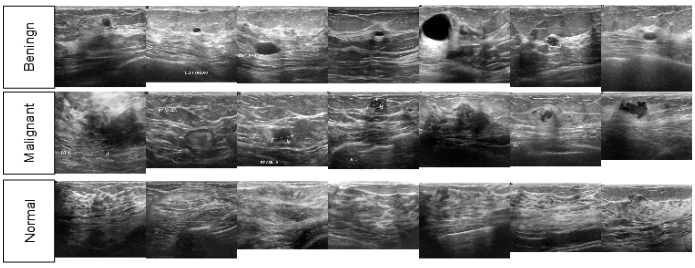


Figure 2 Sample Ultrasound Images Illustrating the Three Classes in the Breast Cancer Dataset: Benign, Malignant, and Normal.

Data Augmentation

To enhance the generalization capability of the DL models and to mitigate the problem of overfitting, a frequent challenge in limited datasets such as medical images, various on-the-fly data augmentation techniques were integrated into the training process in this study. Initially, as the focus was on the classification task, the mask.png files, which were included in the original dataset for segmentation purposes, were excluded from the analysis. During the training phase, the primary augmentation methods randomly applied to each image were as follows: images were first subjected to a "Random Resized Crop," where they were cropped to a random size with a scale ranging from 8% to 100% of the original area (scale: [0.08, 1.0]) and an aspect ratio between 0.75 and 1.33 (ratio: [0.75, 1.3333333333333333]), and subsequently resized to 224x224 pixels (img-size: 224) using a random interpolation method (train interpolation: random). Additionally, random horizontal flipping was applied to each image with a 50% probability (hflip: 0.5). For color-based augmentations, random alterations were made to the color properties of the images, including brightness, contrast, saturation, and hue, with a factor of 0.4 (color-jitter: 0.4). Vertical flipping was not utilized in this study (vflip: 0.0). These on-the-fly augmentation strategies were intended to ensure that the model encounters differentiated data samples during each training epoch, thereby preventing it from becoming overly dependent on the specific features of the training data and aiming for a more robust and reliable performance on unseen data (Wang *et al.* 2024; Mumuni *et al.* 2024).

Model Architectures

In this study, for the automatic classification of breast cancer from breast ultrasound images, well-established and widely recognized deep CNN architectures from the field of computer vision were utilized. In domains such as medical imaging, where the amount of labeled data is often limited, adopting a transfer learning approach rather than training a model from scratch presents significant advantages. Transfer learning enables the transfer of the rich feature extraction capabilities of models pre-trained on large-scale, general-purpose datasets (e.g., ImageNet) to a more specific and smaller target dataset. This approach aims to achieve faster model convergence, improved generalization performance, and a reduced risk of overfitting, particularly when working with limited data. Within the scope of this study, all selected CNN architectures were initialized with weights pre-trained on the ImageNet dataset and were subsequently subjected to a fine-tuning process on our target dataset comprising breast ultrasound images.

First, the ResNet50 architecture, based on the principle of residual learning, was employed. Developed by He et al., ResNet architectures addressed the vanishing gradient problem encountered in the training of very deep networks through the use of "residual blocks" containing "shortcut connections," which allow the input to be passed directly to subsequent layers. ResNet50, a 50-layer deep implementation of this structure, is frequently preferred as a robust baseline model for image classification tasks (He *et al.* 2016). Another architecture of choice was DenseNet169. Proposed by Huang et al., Densely Connected Networks (DenseNets) introduce a "dense connectivity" structure where each layer receives the feature maps from all preceding layers as input and passes on its own feature maps to all subsequent layers. This architecture strengthens feature propagation, encourages feature reuse, reduces the number of parameters, and improves gradient flow, making it particularly prominent for its parameter efficiency; DenseNet169 is a 169-layer version of this architecture (Huang *et al.* 2017).

The study also evaluated two models from the Inception architecture family, developed by Google, which are capable of capturing features at multiple scales simultaneously. InceptionV3, through its "Inception modules," applies convolutional filters of different sizes (e.g., 1x1, 3x3, 5x5) and pooling operations in parallel at the same layer level and concatenates their outputs. This structure allows the model to analyze complex visual patterns at

various scales, while performance is optimized through techniques such as factorizing larger convolutions into smaller ones and using auxiliary classifiers. InceptionV4, as an advancement over InceptionV3, aims to deliver improvements in both performance and computational efficiency by presenting the Inception modules in a more uniform and simplified structure. This model is characterized by deeper and more optimized Inception blocks (Szegedy *et al.* 2016).

These four distinct CNN architectures (ResNet50, DenseNet169, InceptionV3, and InceptionV4) were selected to compare their effectiveness in the task of differentiating between benign, malignant, and normal tissue classes in breast ultrasound images. The distinct architectural approaches and feature extraction strategies of each model are expected to approach this challenging medical image classification problem from different perspectives, thereby providing valuable insights into which architecture or architectural features are more suitable for this specific task. The performance of the models is carefully analyzed using various evaluation metrics, and the results contribute to the literature on the development of deep learning-based automated systems for the early diagnosis of breast cancer.

Evaluation Metrics

Assessing how well DL models work is a vital process to assess their usefulness, provide rationale for relevant decisions, and support data-driven decisions. Performance evaluation criteria can fulfill many important roles such as assessing the effectiveness of a classification models, helping them to be optimized, revealing errors or biases in the data, comparing models, and detecting overfitting. This paper focuses specifically on performance metrics for breast cancer classification, at the same time, we have decided to utilize standard evaluation criteria that are clearly entrenched in the academic literature.

The basic metrics that are used in this project (accuracy, precision, recall, and F1-score) are important in not only DL but other areas. Accuracy can be defined as the number of correctly classified instances over the total number of instances, giving insight into the performance as a whole. Precision (true positives / (true positives + false positives)) tells how reliable the model is in classifying positive instances; if the model has a high precision, it means there are few if any false positives. Recall tells us about the number of actual positives correctly identified the measure of completeness. The F1-score is defined as the harmonic mean of precision and recall, thus making it a single measure of performance that balances the trade-off between false positives and false negatives. While these definitions may seem complicated, they can also be defined mathematically:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}} \tag{1}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

RESULTS AND DISCUSSION

In this section, we present and analyze the performance results of the different CNN architectures we evaluated for the purposes of classifying breast ultrasound images. We compared the performance of the ResNet50, DenseNet169, InceptionV3 and InceptionV4 models used in this work using some fundamental classification performance metrics: Accuracy, Precision, Recall, and F1 score - as well as Quantification metrics including the number of parameters (Params M) and GFLOPs (Giga Floating Point Operations per Second), which estimate the complexity of models and computational resources required for both model training and inference. We consider that exploring such metrics is vital to gaining insight into models' diagnostic performance and important use cases.

The results obtained are summarized in Table 1. Table 1 illustrates the performance metrics reached by each model on the test dataset, along with information on model complexity. These data show the strengths and weaknesses of the various architectures and demonstrate the trade-off between performance and computational expense.

Table 1 Comparative Performance and Complexity of CNN Models for Breast Ultrasound Image Classification

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Params (M)	GFLOPs
Inception V3	96.67	96.55	96.38	96.41	21.79	5.67
DenseNet 169	94.17	92.71	95.43	93.91	12.49	6.72
Inception V4	94.17	92.01	95.97	93.60	41.15	12.25
ResNet 50	90.83	90.29	89.08	89.10	23.51	8.26

The data from Table 1 clearly illustrates that the InceptionV3 model showed the best performance. With an accuracy of 96.67%, precision of 96.55%, sensitivity of 96.38% and F1-score of 96.41%, InceptionV3 was the most capable of successfully classifying breast ultrasound images. It is expected that InceptionV3 performs so well because of the architecture's ability to capture features at different scales and learn complex patterns. Also worth noting, is that InceptionV3 (21.79 million parameters, 5.6719 GFLOPs) delivered the best results from a model complexity standpoint not because it was the most complex model. Having the lowest GFLOPs value means that it was performing at a high level while using a relatively low amount of computational cost.

The DenseNet169 model also achieved highly competitive results. With 94.17% accuracy, 92.71% precision, 95.43% sensitivity, and a 93.91% F1-score, it exhibited the second-best performance after InceptionV3. The most striking feature of DenseNet169 is its model complexity; with 12.49 million parameters, it has the lowest parameter count among the evaluated models, and with 6.7169 GFLOPs, it has the second-lowest GFLOPs value after InceptionV3 (there may be an error in the table, as the GFLOPs for InceptionV3 is lower). This indicates that, as a result of its dense connectivity structure that enhances feature propagation and increases parameter efficiency, DenseNet169 offers a favorable performance-to-efficiency balance. DenseNet169 could be an attractive alternative, especially for scenarios where computational resources are constrained.

The InceptionV4 model, despite having an accuracy rate of 94.17% similar to DenseNet169, along with 92.01% precision, 95.97% sensitivity, and a 93.60% F1-score, is the model with the

highest complexity and computational cost among those evaluated, at 41.15 million parameters and 12.2450 GFLOPs. The fact that it did not surpass InceptionV3, despite having a deeper and more complex structure, suggests that for this specific task and dataset, increased complexity does not invariably translate to better performance. ResNet50, in contrast, exhibited a more modest performance compared to the other three models, with 90.83% accuracy and an 89.10% F1-score. Although it is a strong baseline model, it lagged behind the other more modern and complex architectures used in this study. It possesses a moderate level of complexity with 23.51 million parameters and 8.2634 GFLOPs.

The findings of this study indicate that the InceptionV3 architecture offers a compelling combination of high diagnostic accuracy and balanced computational efficiency. In contrast, DenseNet169 presents itself as a potent alternative for resource-constrained environments, owing to its lower parameter size and computational cost. These findings represent a major contribution to the choice of DL architectures in the context of developing automated solutions for the early diagnosis of breast cancer, and possibilities for future involvement of real-world clinical applications. In all cases, the choice of architecture must be assessed relative to the intended application's requirements (e.g., maximum accuracy versus fast inference time). To further assess the classification performance of the best performing InceptionV3, its confusion matrix is shown in Figure 3.

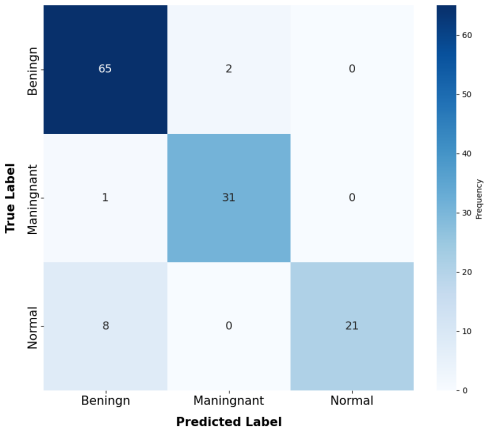


Figure 3 Confusion Matrix of the InceptionV3 Model for Breast Ultrasound Image Classification.

CONCLUSION

In this study, the performance of four different deep Convolutional Neural Network (CNN) architectures (ResNet50, DenseNet169, InceptionV3, and InceptionV4) was comprehensively compared and evaluated for the classification of breast cancer from breast ultrasound (US) images. The findings clearly demonstrated that the InceptionV3 model exhibited the highest classification performance compared to the other evaluated architectures, with superior metric values of 96.67% accuracy, 96.55% precision, 96.38% sensitivity, and a 96.41% F1-score. This high performance can be attributed to the Inception architecture's ability to effectively capture multi-scale features and learn complex visual patterns, while it is also noteworthy that the model offers a relatively efficient computational cost with 21.79 million parameters and 5.6719 GFLOPs.

The DenseNet169 architecture also stood out as a promising alternative for resource-constrained environments, drawing attention with its 94.17% accuracy rate and particularly its low parameter counts of 12.49 million. While InceptionV4 could not surpass InceptionV3 despite its high complexity, ResNet50 yielded more modest results. This study demonstrates that InceptionV3 is a strong candidate for the classification of breast US images in terms of both high diagnostic accuracy and acceptable computational efficiency. The obtained results offer valuable insights for the selection of appropriate DL architectures for the development of automated systems for the early diagnosis of breast cancer and underscore the potential for the integration of these technologies into future clinical applications. Validating these models on larger and more diverse datasets, investigating the impact of different data augmentation strategies and fine-tuning techniques, and integrating eXplainable AI (XAI) methods to enhance model interpretability represent critical next steps for advancing research in this field. Ultimately, such deep learning-based approaches have great potential to support the decision-making processes of radiologists, thereby improving the accuracy and efficiency of breast cancer diagnosis.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Abeelh, E. A. and Z. AbuAbeileh, 2024 Comparative effectiveness of mammography, ultrasound, and mri in the detection of breast carcinoma in dense breast tissue: a systematic review. Cureus **16**.

Abhisheka, B., S. K. Biswas, B. Purkayastha, and S. Das, 2025 Integrating deep and handcrafted features for enhanced decision-making assistance in breast cancer diagnosis on ultrasound images. Multimedia Tools and Applications pp. 1–23.

Alshawwa, I. A., H. Q. El-Mashharawi, F. M. Salman, M. N. A. Al-Qumboz, B. S. Abunasser, *et al.*, 2024 Advancements in early detection of breast cancer: Innovations and future directions .

Bayram, B., I. Kunduracioglu, S. Ince, and I. Pacal, 2025 A systematic review of deep learning in mri-based cerebral vascular occlusion-based brain diseases. Neuroscience .

Begum, M. M. M., R. Gupta, B. Sunny, and Z. L. Lutfor, 2024 Advancements in early detection and targeted therapies for breast cancer; a comprehensive analysis. Asia Pacific Journal of Cancer Research **1**: 4–13.

Cai, F., J. Wen, F. He, Y. Xia, W. Xu, *et al.*, 2024 Sc-unext: A lightweight image segmentation model with cellular mechanism for breast ultrasound tumor diagnosis. Journal of Imaging Informatics in Medicine **37**: 1505–1515.

Cakmak, Y. and I. Pacal, 2025 Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the wisconsin dataset. Journal of Operations Intelligence **3**: 175–196.

Cakmak, Y., S. Safak, M. A. Bayram, and I. Pacal, 2024 Comprehensive evaluation of machine learning and ann models for breast cancer detection. Computer and Decision Making: An International Journal **1**: 84–102.

Chambi, E. A., D. G. Alzamora, and A. A. Salas, 2025 Ultrasonic image processing for the classification of benign and malignant breast tumors: Comparative study of convolutional neural network architectures. Engineering Proceedings **83**: 15.

Gagliardi, M., T. Ruga, E. Vocaturo, and E. Zumpano, 2024 Predictive analysis for early detection of breast cancer through artificial intelligence algorithms. In *International Conference on Innovations in Computational Intelligence and Computer Vision*, pp. 53–70, Springer.

Gordon, P. B., L. J. Warren, and J. M. Seely, 2025 Cancers detected on supplemental breast ultrasound in women with dense breasts: update from a canadian centre. Canadian Association of Radiologists Journal p. 08465371251318578.

He, K., X. Zhang, S. Ren, and J. Sun, 2016 Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 2017 Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Iacob, R., E. R. Iacob, E. R. Stoicescu, D. M. Ghenciu, D. M. Cocolea, *et al.*, 2024 Evaluating the role of breast ultrasound in early detection of breast cancer in low-and middle-income countries: a comprehensive narrative review. Bioengineering **11**: 262.

İnce, S., I. Kunduracioglu, B. Bayram, and I. Pacal, 2025 U-net-based models for precise brain stroke segmentation. Chaos Theory and Applications **7**: 50–60.

Islam, M. R., M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, *et al.*, 2024 Enhancing breast cancer segmentation and classification: An ensemble deep convolutional neural network and u-net approach on ultrasound images. Machine Learning with Applications **16**: 100555.

Kaggle, 2025 Breast ultrasound images dataset(busi).

Katsika, L., E. Boureka, I. Kalogiannidis, I. Tsakiridis, I. Tirodimos, *et al.*, 2024 Screening for breast cancer: a comparative review of guidelines. Life **14**: 777.

Kiani, P., H. Vatankhahan, A. Zare-Hoseinabadi, F. Ferdosi, S. Ehtiati, *et al.*, 2025 Electrochemical biosensors for early detection of breast cancer. Clinica Chimica Acta **564**: 119923.

Kim, J., A. Harper, V. McCormack, H. Sung, N. Houssami, *et al.*, 2025 Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nature Medicine pp. 1–9.

Koçak, B., A. Ponsiglione, A. Stanzione, C. Bluethgen, J. Santinha, *et al.*, 2025 Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. Diagnostic and interventional radiology **31**: 75.

Kurtulus, I. L., M. Lubbad, O. M. D. Yilmaz, K. Kilic, D. Karaboga, *et al.*, 2024 A robust deep learning model for the classification of dental implant brands. Journal of Stomatology, Oral and Maxillofacial Surgery **125**: 101818.

Latha, M., P. S. Kumar, R. R. Chandrika, T. Mahesh, V. V. Kumar, *et al.*, 2024 Revolutionizing breast ultrasound diagnostics with efficientnet-b7 and explainable ai. BMC Medical Imaging **24**: 230.

Li, X., L. Zhang, J. Yang, and F. Teng, 2024 Role of artificial intelligence in medical image analysis: A review of current trends and future directions. Journal of Medical and Biological Engineering **44**: 231–243.

Lubbad, M., D. Karaboga, A. Basturk, B. Akay, U. Nalbantoglu, *et al.*, 2024a Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. Neural Computing and Applications **36**: 6355–6379.

Lubbad, M. A., I. L. Kurtulus, D. Karaboga, K. Kilic, A. Basturk, *et al.*, 2024b A comparative analysis of deep learning-based approaches for classifying dental implants decision support system. Journal of Imaging Informatics in Medicine **37**: 2559–2580.

Meng, X., J. Ma, F. Liu, Z. Chen, and T. Zhang, 2024 An interpretable breast ultrasound image classification algorithm based on convolutional neural network and transformer. Mathematics **12**: 2354.

Mumuni, A., F. Mumuni, and N. K. Gerrar, 2024 A survey of synthetic data augmentation methods in machine vision. Machine Intelligence Research **21**: 831–869.

Nasiri-Sarvi, A., M. S. Hosseini, and H. Rivaz, 2024 Vision mamba for classification of breast ultrasound images. In *Deep Breast Workshop on AI and Imaging for Diagnostic and Treatment Challenges in Breast Care*, pp. 148–158, Springer.

Obeagu, E. I. and G. U. Obeagu, 2024 Breast cancer: A review of risk factors and diagnosis. Medicine **103**: e36905.

Obuchowicz, R., M. Strzelecki, and A. Piórkowski, 2024 Clinical applications of artificial intelligence in medical imaging and image processing—a review.

Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access .

Pacal, İ., 2025 Diagnostic analysis of various cancer types with artificial intelligence .

Pacal, I., O. Akhan, R. T. Deveci, and M. Deveci, 2025 Nextbrain: Combining local and global feature learning for brain tumor classification. Brain Research p. 149762.

Pacal, I. and O. Attallah, 2025a Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.

Pacal, I. and O. Attallah, 2025b Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.

Rajkumar, R., S. Gopalakrishnan, K. Praveena, M. Venkatesan, K. Ramamoorthy, *et al.*, 2024 Darknet-53 convolutional neural network-based image processing for breast cancer detection. Mesopotamian Journal of Artificial Intelligence in Healthcare **2024**: 59–68.

Rana, A. S., J. Rafique, and H. Riffat, 2024 Advances in breast ultrasound imaging: Enhancing diagnostic precision and clinical utility. In *Latest Research on Breast Cancer-Molecular Insights, Diagnostic Advances and Therapeutic Innovations*, IntechOpen.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2016 Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Trentham-Dietz, A., C. H. Chapman, J. Jayasekera, K. P. Lowry, B. M. Heckman-Stoddard, *et al.*, 2024 Collaborative modeling to compare different breast cancer screening strategies: a decision analysis for the us preventive services task force. JAMA **331**: 1947–1960.

Umer, M. J., M. Sharif, and M. Raza, 2024 A multi-attention triple decoder deep convolution network for breast cancer segmentation using ultrasound images. Cognitive Computation **16**: 581–594.

Vogel-Minea, C. M., W. Bader, J.-U. Blohmer, V. Duda, C. Eichler, *et al.*, 2025 Best practice guidelines–degum recommendations on breast ultrasound. Ultraschall in der Medizin-European Journal of Ultrasound **46**: 245–258.

Wang, Z., P. Wang, K. Liu, P. Wang, Y. Fu, *et al.*, 2024 A comprehensive survey on data augmentation. arXiv preprint arXiv:2405.09591 .

Xiong, X., L.-W. Zheng, Y. Ding, Y.-F. Chen, Y.-W. Cai, *et al.*, 2025 Breast cancer: pathogenesis and treatments. Signal transduction and targeted therapy **10**: 49.

Zeynalov, J., Y. Çakmak, and İ. Paçal, 2025 Automated apple leaf disease classification using deep convolutional neural networks: A comparative study on the plant village dataset. Journal of Computer Science and Digital Technologies **1**: 5–17.

How to cite this article: Cakmak, Y., and Pacal, N. Deep Learning for Automated Breast Cancer Detection in Ultrasound: A Comparative Study of Four CNN Architectures. *Artificial Intelligence in Applied Sciences*, 1(1), 13-19, 2025.

ADBA
Artificial Intelligence in Applied Sciences
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4060
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.20-27
https://doi.org/10.69882/adba.ai.2025074

Deep Learning in Maize Disease Classification

Luaay Alswilem *,1 **and Elsevar Asadov** α,2

*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye, αDepartment of Basic Medical, Faculty of Architecture and Engineering, Nakhchivan State University, AZ 7012, Nakhchivan, Azerbaijan.

ABSTRACT As a strategic global crop, maize productivity is directly threatened by leaf diseases such as Southern Leaf Blight and Gray Leaf Spot, making early and accurate detection crucial for food security. Artificial intelligence, particularly deep learning, provides a powerful solution for the automated classification of plant diseases from images. This study developed an intelligent system to address this challenge, utilizing the publicly available PlantVillage dataset to evaluate five leading Convolutional Neural Network (CNN) architectures: DenseNet121, InceptionV3, MobileNetV2, ResNet-50, and VGG16. The models were optimized with established techniques, including transfer learning, data augmentation, and hyper-parameter tuning, while a Soft Voting Ensemble strategy was used to enhance combined performance. Evaluation across multiple metrics showed that InceptionV3 achieved the highest test accuracy at 94.47%. However, MobileNetV2 demonstrated the strongest performance across all metrics with a 95% cumulative accuracy and proved highly efficient, making it ideal for deployment on mobile devices. These findings confirm the significant potential of deep learning for building cost-effective and efficient diagnostic systems in agriculture, ultimately contributing to the reduction of crop losses and the promotion of sustainable farming practices.

KEYWORDS
Maize leaf disease
Deep learning
Image classification
Transfer learning
Sustainable agriculture

INTRODUCTION

Maize is a major crop important to the global economy, being the most produced and consumed cereal in the word (Willer *et al.* 2024a). It has a multitude of essential functions: food production for human consumption, livestock production for animal feed, food industry, and industrial products like ethanol, oils, and starch and thus has considerable relevance in the context of food security and energy supply (Ranum *et al.* 2014; Willer *et al.* 2024b). Latest 2023 FAO data indicate global production of maize at over 1.2 billion tons a year, with production on about 200 million hectares (Committee *et al.* 2023; Fang and Katchova 2023). The US is the leading maize producing country, producing over 350 million tons annually, making 30% of global production. Major maize production states are Iowa and Nebraska, Illinois, Minnesota, and Indiana. China is the second leading supplier of maize at roughly 270 million tons, followed by Brazil at about 125 million tons, while Argentina, India ,and Ukraine are also major contributors (Philpott 2020; Demanyuk *et al.* 2023; Pignati 2018).

Despite this substantial production volume, maize remains vulnerable to several severe plant diseases affecting leaves, stalks, and roots, which lead to significant yield and quality deterioration. The most notable diseases include: Gray Leaf Spot, Northern Leaf Blight, Common Rust, Powdery Mildew, and Stalk and Root Rot. These diseases cause enormous economic losses exceeding 10 billion US dollars annually on a global scale (Bickel and Koehler 2021; Dinh and Joyce 2007). Studies indicate that losses can reach up to 60% in severely infected regions, especially under humid and warm climatic conditions that facilitate the spread of fungal and bacterial infections (Teixeira *et al.* 2021a,b). Early detection of maize leaf diseases is essential to limit the spread of infections and preserve crop yield. However, traditional manual inspection requires agricultural expertise, is costly, and often lacks precision. For these reasons, recent scientific research has increasingly relied on artificial intelligence (AI) and deep learning techniques for plant disease diagnosis from images (Mahlein 2016; Kamilaris and Prenafeta-Boldú 2018; Pacal 2025).Among these techniques, convolutional neural networks (CNNs) have proven to be particularly effective. CNNs are widely used in image classification and have demonstrated high accuracy in recognizing complex visual patterns (Sladojevic *et al.* 2016; Ferentinos 2018).

The advent of artificial intelligence (AI), and more specifically its sophisticated subfields of machine learning (ML) (Cakmak and Pacal 2025; Cakmak *et al.* 2024) and deep learning (DL) (Pacal 2025), has ignited a foundational transformation, redefining the operational landscape across a multitude of global sectors. This technological revolution is profoundly demonstrated in healthcare, where AI has revolutionized diagnostic medicine by enhancing the interpretation of medical imagery. Its applications are extensive, powering breakthroughs in oncology through the early identification of brain tumors (Pacal *et al.* 2025; İnce *et al.* 2025; Bayram *et al.* 2025), pulmonary nodules (Ozdemir *et al.* 2025), and breast cancer (Pacal and Attallah 2025), while also advancing specialized fields like dental diagnostics (Lubbad *et al.* 2024b; Kurtulus *et al.* 2024) and urological pathology (Lubbad *et al.* 2024a). In a paral-

Manuscript received: 25 June 2025,
Revised: 15 July 2025,
Accepted: 21 July 2025.

[1]luaayalswilem3@gmail.com (**Corresponding author**)
[2]asadoves@mail.ru

lel trend, this same technological momentum is spearheading a movement towards a more sustainable, efficient, and data-driven agricultural industry. AI is becoming pivotal to modern farming by enabling critical functions such as the early diagnosis of plant ailments via leaf image analysis (Zeynalov *et al.* 2025) and the accurate forecasting of crop yields using data from satellites and drones (Chouhan *et al.* 2024). Furthermore, it powers the deployment of intelligent robotic systems for highly targeted weed elimination (Goyal *et al.* 2025; Sathya Priya *et al.* 2025) and underpins precision agriculture, where automated systems adjust irrigation and fertilization in real-time according to immediate soil and crop conditions, thereby optimizing resource management and promoting sustainability (Maurya *et al.* 2025; Singh and Sharma 2025; Surendran *et al.* 2024; Jaya Krishna *et al.* 2025). In this project, the "Corn Leaf Disease" dataset was utilized to train and test five of the most prominent CNN models for maize leaf disease classification, namely: DenseNet121, InceptionV3, MobileNetV2, ResNet50, and VGG16. The performance of these models was evaluated using four primary metrics: accuracy, recall, precision, and F1-score (Dong *et al.* 2023; Brahimi *et al.* 2024).

Despite the considerable potential of deep learning for plant disease classification, research specifically targeting the identification of corn leaf diseases using convolutional neural network (CNN) methodologies remains limited (Rui *et al.* 2022). Among the early notable contributions, Priyadharshini et al. applied a modified LeNet architecture to the PlantVillage dataset, successfully classifying four corn disease categories with a high accuracy rate of 97.89% (Zhang *et al.* 2018). Complementing this, Zhang et al. explored the classification of eight distinct corn diseases by fine-tuning hyperparameters within the GoogleNet and Cifar10 frameworks, laying foundational work in this domain (Ahila Priyadharshini *et al.* 2019).

Further advancing this field, Wang et al. employed customized hyperparameter optimization on the ResNet-50 model, attaining an impressive classification accuracy of 98.52% across five corn disease classes (Waheed *et al.* 2020). In a related study, Waheed et al. focused on DenseNet models optimized through hyperparameter tuning to distinguish four corn diseases. While their DenseNet-based approach achieved slightly lower accuracy (98.06%) compared to the EfficientNet-B0 model, it notably reduced the model size and parameter count, highlighting an effective trade-off between performance and computational efficiency (Chen *et al.* 2020).

Addressing transfer learning strategies, Chen et al. proposed a hybrid model combining pre-trained ImageNet weights within VGGNet and Inception modules. Their model, termed INC-VGGN, achieved a minimum validation accuracy of 91.83% when tested on corn images from the PlantVillage dataset, illustrating the promise of integrated architectures (Meng *et al.* 2020). Recognizing the challenges inherent in real-world deployment, Zeng et al. introduced LDSNet, a highly lightweight CNN designed specifically for corn disease diagnosis under complex backgrounds and dilation issues. This model attained a test accuracy of 95.4%, demonstrating practical applicability in field conditions (Pacal *et al.* 2024).

MATERIALS AND METHODS

Dataset

In this research, five of the leading Convolutional Neural Network (CNN) architectures were used. The PlantVillage is one of the most significant and well-known open-source datasets in the context of plant disease diagnosis based on digital images. The PlantVillage dataset was created as part of a research effort designed to support farmers and researchers, with a large collection of high-quality images, all annotated with the assistance of scholars in botany and plant pathology. The PlantVillage dataset contains images of leaves spanning several different crops, such as tomato, potato, maize, grapevine and many more, as well as a wide range of plant diseases (Hughes *et al.* 2015). For this study, only images of maize leaves were used from the PlantVillage database because the focus was on diagnosing and ultimately classifying diseases associated with this important crop. The images were further divided into four main categories leading to healthy maize leaves, and leaves afflicted by Gray Leaf Spot, Common Rust and Northern Leaf Blight. The dataset was chosen based on the quality of the images and number of distinct disease cases that facilitate effective training on deep learning model and advance classification accuracy. Table 1 details the composition of the dataset used in this study, which was divided into training, validation, and testing sets to ensure comprehensive model evaluation and to prevent data leakage during the learning process. The figure 1 illustrates a number of selected samples for each category from the utilized dataset.

Table 1 Distribution of Images for Training, Testing, and Validation

Subset	Number of Images	Percentage (%)
Train	2,696	70
Test	579	15
Validation	577	15
Total	**3,852**	**100**

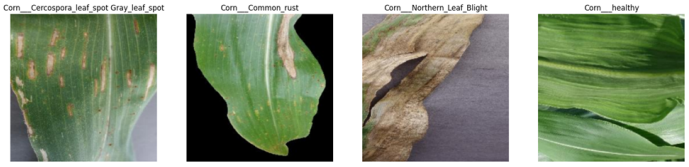


Figure 1 Visual Examples of Different Corn Leaf Conditions

Deep Learning Architectures

Machine learning has revolutionized technological advancement and human development, becoming a major driving force behind many modern applications such as improving search engine capabilities, monitoring user-generated content on social media, and enabling personalized recommendation systems in e-commerce. With rapid technological progress, machine learning has become an integral part of our daily lives, manifesting in intelligent technologies and advanced systems featuring capabilities like visual object detection, speech recognition, and dynamic content adaptation in digital environments (Lecun *et al.* 2015).The rapid progress in artificial intelligence is largely attributed to the development of deep learning, a specialized branch of machine learning that relies on multilayered, complex neural networks to extract nonlinear and intricate representations from large datasets. These models enable the identification of fine-grained features through hierarchical structures and are trained using the backpropagation algorithm. Deep learning has demonstrated exceptional success across many domains such as image and video analysis, audio processing, and natural language understanding. Convolutional Neural Networks (CNNs) excel in handling spatial data, while Recurrent Neural

Networks (RNNs) are more suitable for temporal or sequential data like speech and text (Karaman *et al.* 2023; Pacal *et al.* 2022).

Although Geoffrey Hinton laid the theoretical foundations of deep learning in 2006, widespread adoption of this technology surged after deep models significantly outperformed traditional algorithms in the ImageNet Large Scale Visual Recognition Challenge. Since then, deep learning has consistently achieved state-of-the-art results across a wide array of applications including pattern recognition, classification, prediction, drug discovery, signal analysis, finance, healthcare, and defense, making it the leading paradigm in AI research and practical applications alike (Pacal 2024).

The Used Algorithms

In this study, five of the most prominent Convolutional Neural Network (CNN) architectures were utilized, each renowned for its high efficiency and foundational role in image classification tasks. These models are particularly well-suited for this project due to their proven success in domains requiring nuanced visual analysis, such as medical imaging and precision agriculture. Their selection was based not only on their widespread popularity in recent academic research but also on their validated ability to extract deep, hierarchical features from complex images with remarkable accuracy and efficiency. By leveraging these powerful architectures, which have set benchmarks on large-scale datasets, this work aims to build upon their established feature extraction capabilities to achieve robust classification of maize leaf diseases.

The VGG16 model, developed at the University of Oxford by Simonyan and Zisserman (Pacal and Attallah 2025), is considered a classical and highly influential architecture in the field of computer vision. Its defining characteristic is a simple yet profound design homogeneity: it is constructed by stacking multiple convolutional layers that exclusively use small (3×3) kernels. This strategy demonstrated that a significant increase in network depth, rather than the use of larger, more complex filters, was a key to improving performance. These convolutional blocks are systematically followed by max-pooling layers, which reduce the spatial dimensions of the feature maps, thereby decreasing computational load and creating invariance to the position of features. Despite its structural elegance, VGG16 is a very large model containing approximately 138 million parameters, the majority of which are in its final fully connected layers. This large capacity allows it to learn rich representations but also makes it computationally intensive and prone to overfitting, establishing it as a critical benchmark for both performance and resource management in deep learning (Simonyan and Zisserman 2014). The structural layout of the VGG16 model is illustrated in Figure 2.

ResNet (Residual Network), developed by the Microsoft Research team led by He et al., is a revolutionary architecture that won the ILSVRC 2015 competition and fundamentally changed the landscape of deep learning (He *et al.* 2016a,b). Its primary motivation was to solve the "degradation" problem, a counter-intuitive phenomenon where adding more layers to a deep network would cause its accuracy to saturate and then rapidly decline. ResNet masterfully addresses this challenge with the ingenious concept of "residual connections," also known as "skip connections." This structure allows the input of a layer block to be added directly to its output, effectively creating a shortcut. By doing this, the network is reframed to learn the residual mapping rather than the entire underlying transformation. If a certain block is not useful, the network can easily learn to make the residual zero, essentially "skipping" the block by turning it into an identity mapping, thus

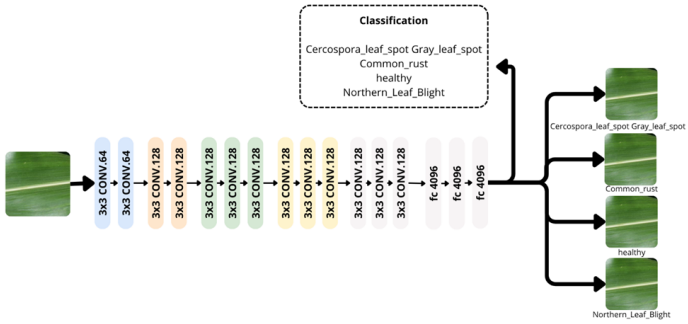


Figure 2 VGG16 Architecture Used for Corn Leaf Disease Classification

preventing performance loss.

The ResNet-50 model, used in this research, is a 50-layer version of the ResNet architecture. It uses an even more efficient "bottleneck" structure in its residual block, applying 1x1, 3x3, and another 1x1 convolutional filters in the residual block to compress then provide dimension back. There is sufficient depth for good feature extraction while reducing the parameter count to approximately 25 million, which is considerably lower than earlier models like VGG16. Because the ResNet architecture is very successful at solving the degradation problem, ResNet-50 enables training of much deeper networks. It also serves as a baseline model that provides state-of-the-art accuracy and significant computational cost savings during training and inference for applicable computer vision tasks. The ResNet-50 architecture is shown as Figure 3.

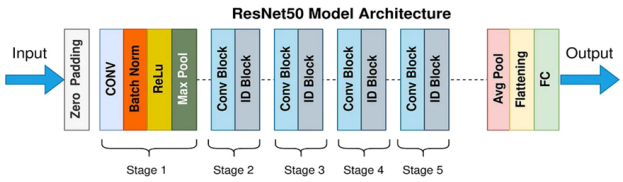


Figure 3 Block Diagram of the ResNet-50 Processing Pipeline

DenseNet121, introduced by Huang et al. (Huang *et al.* 2017a,b; Kaur *et al.* 2024), represents a significant evolution in network architecture designed to maximize information flow between layers. The model is built upon the core concept of "dense connectivity," a powerful alternative to the residual connections found in ResNet. Instead of summing features, DenseNet concatenates them. In this paradigm, each layer receives the feature maps from all preceding layers as its input, creating a direct and deep channel for information transfer. This architecture ensures that all features, from the earliest low-level ones to more complex high-level ones, are accessible throughout the network.

This dense connectivity yields several critical advantages. Firstly, it strongly encourages feature reuse, which makes the model highly parameter-efficient; since each layer has access to a

"collective knowledge" of all prior features, it only needs to learn a small number of new feature maps. Secondly, this enhanced information flow significantly alleviates the vanishing gradient problem, as gradients can propagate more directly to earlier layers during training. Consequently, DenseNets are not only easier to train but also achieve state-of-the-art performance with considerably fewer parameters compared to models of similar depth. The fundamental building block of this efficient architecture, the dense block, is illustrated in Figure 4.

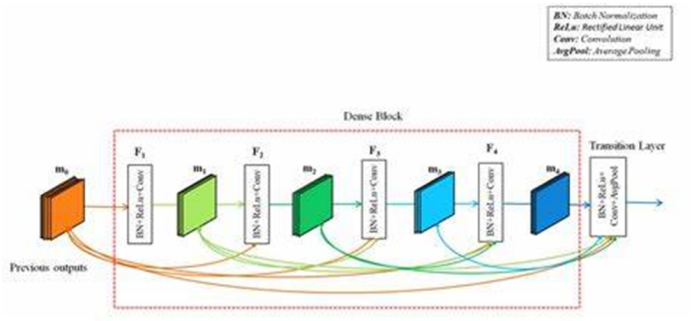


Figure 4 Schematic Diagram of a Dense Block and its Transition Layer

The InceptionV3 model, developed by Szegedy et al. at Google, is a highly influential architecture and a significant iteration within the GoogLeNet family (Szegedy *et al.* 2016a,b). It was designed to tackle the dual challenges of improving classification accuracy while drastically reducing the computational burden of very deep networks. The model's ingenuity lies in its core component, the "Inception Module," which employs a "split, transform, merge" strategy. Instead of choosing a single convolution kernel size for a layer, an Inception module performs multiple convolution operations with different kernel sizes (e.g., 1x1, 3x3, 5x5) and a max-pooling operation in parallel within the same block. This allows the network to capture visual features at multiple scales simultaneously, from fine-grained details to more abstract, larger patterns.

A key to its computational efficiency is the extensive use of 1x1 convolutions as bottleneck layers to reduce the feature map dimensions before the more expensive 3x3 and 5x5 convolutions are applied. InceptionV3 further refines this concept by factorizing larger convolutions into smaller, stacked ones (e.g., replacing a 5x5 filter with two consecutive 3x3 filters), which reduces parameters and increases non-linearity. The outputs from these parallel paths are then concatenated into a single, rich feature map. This sophisticated design enables InceptionV3 to build a deep and wide network with high efficiency. The intricate parallel structure of the Inception module, which is fundamental to the model's success, is detailed in Figure 5.

MobileNetV2, developed by Google, was purposefully engineered to deliver high efficiency for platforms with limited computational and power resources, such as smartphones, drones, and IoT devices (Sandler *et al.* 2018a,b). Its architecture is designed to minimize model size and computational cost with minimal impact on accuracy, relying on two innovative core concepts. The first is the foundational technique of "Depthwise Separable Convolutions," which replaces standard convolutions by splitting the process into two stages: a depthwise convolution that applies a single filter to each input channel for spatial filtering, followed by a pointwise convolution (a 1x1 filter) to combine the channel outputs. This factorization dramatically reduces the number of parameters and the computational load. The second and primary innovation in MobileNetV2 is the "Inverted Residual" block. Contrary to traditional residual blocks that have a wide-to-narrow-to-wide structure, the inverted residual block starts with a narrow, low-dimensional input, expands it to a high-dimensional space, applies the efficient depthwise convolution, and then projects it back to a narrow representation. The skip connection links the narrow bottleneck layers, which improves gradient flow and allows for the construction of deeper, more effective networks. This combination of techniques makes MobileNetV2 an ideal solution for real-world applications like smart agriculture, where it can perform tasks such as on-the-spot disease detection on a mobile device or image processing on an autonomous drone, perfectly balancing performance with the constraints of on-device deployment.

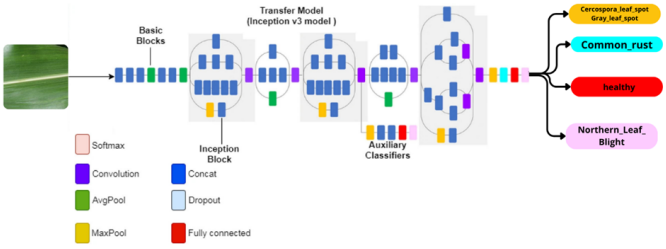


Figure 5 Structural Flowchart of the InceptionV3 Transfer Learning Model

RESULTS AND DISCUSSION

Experimental Design

The experiments conducted in this study were executed on a system running Windows 11, equipped with an Intel Core i7 processor, 32 GB of DDR5 RAM, and an NVIDIA GeForce RTX 4060 Laptop GPU. All models were developed using the PyTorch framework, leveraging NVIDIA's CUDA technology for accelerated computation. Training and evaluation of the models were performed within a unified experimental environment, utilizing identical sets of hyperparameters to ensure consistency and enable a rigorous, systematic comparison among the models.

Performance Metrics

The development of robust intelligent systems for modern agricultural applications hinges on the rigorous assessment of machine learning models. To this end, a comprehensive performance benchmark was conducted to determine the most effective architecture for classifying diseases affecting corn leaves. This study leveraged the well-known PlantVillage dataset to train and validate five of the most influential and powerful convolutional neural network (CNN) architectures: VGG16, ResNet-50, DenseNet121, InceptionV3, and MobileNetV2. The objective was to systematically evaluate these established models in a specific, high-impact agricultural context, thereby providing clear insights into their practical effectiveness.

To ensure a direct and unbiased comparison, a controlled experimental environment was established where all five models were trained under identical conditions, using the same dataset partitions and hyperparameters. The subsequent evaluation was based on a suite of standard quantitative metrics to holistically

measure performance. This included Accuracy, which provides a top-level view of the overall percentage of correct classifications. To gain deeper insight, Precision was used to measure the reliability of positive predictions, while Recall assessed the model's ability to identify all true positive cases of a given disease. Finally, the F1-Score was employed to provide a balanced assessment by calculating the harmonic mean of precision and recall, a particularly crucial metric when dealing with potentially imbalanced class distributions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Results

The results demonstrated that the DenseNet121 model delivered excellent performance compared to the other models, achieving a test accuracy of 96.02%, precision of 95.67%, recall of 95.90%, and an F1-score of 95.78%. This performance reflects DenseNet121's ability to leverage dense connectivity between layers, resulting in effective and accurate discrimination among the disease classes. In second place was the InceptionV3 model, which attained a test accuracy of 94.47%, precision of 93.50%, recall of 93.18%, and an F1-score of 93.33%, highlighting its high efficiency in analyzing multi-scale features within the images. The ResNet-50 model achieved a test accuracy of 90.85%, while the VGG16 model showed relatively lower performance with an accuracy of 89.98%.MobileNetV2, despite having fewer parameters, showed competitive results with an accuracy of 92.10%, making it suitable for applications requiring speed and resource efficiency, such as deployment on mobile devices. The following table (Table 2) summarizes the comparative performance of the models used in this study:

Table 2 Performance Comparison of Deep Learning Models

Model	Accuracy	Precision	Recall	F1-Score
DenseNet121	96.02%	95.67%	95.90%	95.78%
InceptionV3	94.47%	93.50%	93.18%	93.33%
MobileNetV2	92.10%	91.23%	91.45%	91.34%
ResNet-50	90.85%	89.88%	89.30%	89.59%
VGG16	89.98%	88.70%	88.95%	88.82%

Figure 6 displays the confusion matrix for the DenseNet121 model, which demonstrated the best performance among all tested models with an accuracy of 96.02%. This visualization provides a detailed breakdown of the model's classification results, showing the distribution of correct and incorrect predictions for each class. Analyzing the confusion matrix in Figure 6 allows for a deeper understanding of the specific strengths and weaknesses of the DenseNet121 model's predictive capabilities.

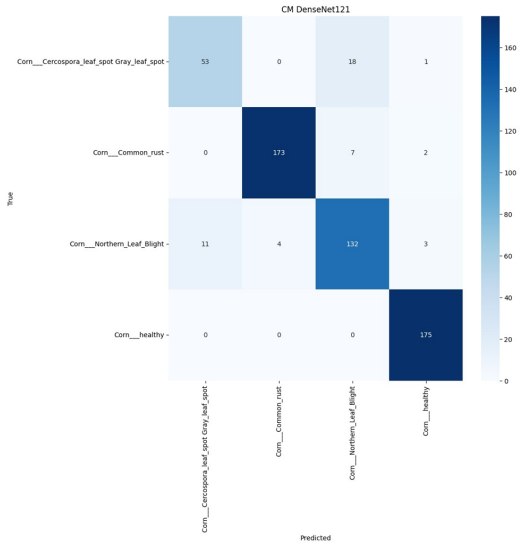


Figure 6 Confusion Matrix of the DenseNet121 Model

To further analyze the performance of the top-performing DenseNet121 model, a detailed classification report is presented in Table 3. The table shows the precision, recall, and F1-score for each individual class: Cercospora/Gray Leaf Spot, Common Rust, Northern Leaf Blight, and Healthy. The model demonstrates exceptional performance in identifying 'Healthy' leaves, achieving a perfect recall of 1.00 and an F1-score of 0.98, which indicates that no healthy leaves were misclassified. Similarly, the 'Common Rust' class is identified with high confidence, posting an F1-score of 0.96. The most challenging category for the model appears to be 'Cercospora/Gray Leaf Spot,' with an F1-score of 0.78. Overall, the weighted average F1-score of 0.92 confirms the model's robust and effective classification capability across the different corn leaf diseases.

Table 3 Detailed Classification Report for the DenseNet121 Model

Class	Precision	Recall	F1-Score	Support
Gray Leaf Spot	0.83	0.74	0.78	72
Common Rust	0.98	0.95	0.96	182
Northern Leaf Blight	0.84	0.88	0.86	150
Healthy	0.97	1.00	0.98	175
Weighted Avg	0.92	0.92	0.92	579
Accuracy			0.92	579

Discussion

The results of this study highlight the effectiveness of deep learning models in accurately classifying corn leaf diseases, confirming the significant role of artificial intelligence in supporting smart agriculture and automating plant disease diagnosis. The DenseNet121 model outperformed the other models, achieving an accuracy of 96.02% and an F1-score of 95.78%, reflecting an excellent balance

between detection rate and error reduction. This superior performance is attributed to the DenseNet architecture, which relies on dense connections that enhance feature reuse and facilitate learning of precise representations of disease patterns. The following table summarizes the performance of the five models used in this study:

The second-best performing model was InceptionV3, which achieved an accuracy of 94.47% and an F1-Score of 93.33%. This strong performance can be attributed to its multi-scale architectural design, enabling it to extract visual features at various levels. However, its relatively high computational resource consumption may limit its suitability for deployment in resource-constrained environments. MobileNetV2 demonstrated good performance with an accuracy of 92.10% and an F1-Score of 91.34%. Due to its lightweight architecture and efficient inference capabilities, it is considered a suitable choice for mobile and embedded applications, although this comes at the cost of somewhat reduced accuracy compared to larger models.

ResNet-50 achieved a moderate performance, with an accuracy of 90.85% and an F1-Score of 89.59%, indicating a fair ability to discriminate between classes. Meanwhile, VGG16 ranked lowest in performance, with an accuracy of 89.98% and an F1-Score of 88.82%, which aligns with its simpler architecture lacking advanced techniques such as residual connections or multi-scale feature extraction. These results suggest that selecting an appropriate model should not rely solely on accuracy metrics but also consider factors such as model size, inference speed, and deployment efficiency in real-world settings, such as agricultural fields or mobile applications. For future work, it is recommended to expand the study by incorporating data from real field environments and diverse imaging conditions (e.g., varying lighting and backgrounds). Additionally, integrating Explainable AI techniques would enhance model transparency and build user trust in model decisions. Evaluating the model across multiple corn varieties and geographic regions is also advised to improve generalizability in broader agricultural contexts.

CONCLUSION

This study concluded that deep learning–based models, particularly the proposed hybrid model, serve as effective and accurate tools for classifying corn leaf diseases from images. The hybrid model demonstrated superior performance compared to conventional models, underscoring the importance of designing network architectures that combine depth with dense internal connections to extract fine-grained features. The results also highlighted that balancing accuracy with computational efficiency is a critical factor when selecting an optimal model for smart agriculture applications, especially in resource-constrained environments. The study affirms that integrating artificial intelligence techniques into the agricultural sector represents a pivotal step towards the digital transformation of plant disease management, contributing to improved crop quality and enhanced early response to disease challenges. Accordingly, it is recommended to continue developing these models and expanding their testing to encompass real-world scenarios and varying imaging conditions, with an emphasis on adopting explainable AI tools to increase trustworthiness and facilitate adoption by agricultural practitioners.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Ahila Priyadharshini, R., S. Arivazhagan, M. Arun, and A. Mirnalini, 2019 Maize leaf disease classification using deep convolutional neural networks. Neural Computing and Applications **31**: 8887–8895.

Bayram, B., I. Kunduracioglu, S. Ince, and I. Pacal, 2025 A systematic review of deep learning in mri-based cerebral vascular occlusion-based brain diseases. Neuroscience .

Bickel, J. T. and A. M. Koehler, 2021 Review of pythium species causing damping-off in corn. Plant Health Progress **22**: 219–225.

Brahimi, F., A. Aid, M. Amad, A. Mehennaoui, and A. Baadache, 2024 Enhanced k-nearest neighbors for smart cardiovascular disease prediction in iot system. Revue d'Intelligence Artificielle **38**.

Cakmak, Y. and I. Pacal, 2025 Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the wisconsin dataset. Journal of Operations Intelligence **3**: 175–196.

Cakmak, Y., S. Safak, M. A. Bayram, and I. Pacal, 2024 Comprehensive evaluation of machine learning and ann models for breast cancer detection. Computer and Decision Making: An International Journal **1**: 84–102.

Chen, J., J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, 2020 Using deep transfer learning for image-based plant disease identification. Computers and electronics in agriculture **173**: 105393.

Chouhan, S. S., U. P. Singh, and S. Jain, 2024 *Applications of computer vision and drone technology in agriculture 4.0*. Springer.

Committee, I. A. P. *et al.*, 2023 Usda agricultural projections to 2032 .

Demanyuk, O., G. Matusevich, S. Mazur, D. Shatsman, S. Bukhtyk, *et al.*, 2023 Wheat, corn, and sunflower are the primary crops of ukrainian exports. Agriculture and plant sciences: theory and practice pp. 41–50.

Dinh, S.-Q. and D. C. Joyce, 2007 Prospects for cut-flower postharvest disease management with host defence elicitors. Stewart Postharvest Review **3**: 1–11.

Dong, X., Q. Wang, Q. Huang, Q. Ge, K. Zhao, *et al.*, 2023 Pdddpretrain: A series of commonly used pre-trained models support image-based plant disease diagnosis. Plant Phenomics **5**: 0054.

Fang, X. and A. L. Katchova, 2023 Evaluating the oecd–fao and usda agricultural baseline projections. Q Open **3**: qoad029.

Ferentinos, K. P., 2018 Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture **145**: 311–318.

Goyal, R., S. Kumari, A. Nath, and G. Kaur, 2025 Iort and ai-driven solution for optimal herbicides spray on weeds in a dynamic agriculture environment. In *International Conference on Advanced Information Networking and Applications*, pp. 297–308, Springer.

He, K., X. Zhang, S. Ren, and J. Sun, 2016a Deep residual learning for image recognition.

He, K., X. Zhang, S. Ren, and J. Sun, 2016b Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Huang, G., Z. Liu, L. van der Maaten, and K. Q. Weinberger, 2017a Densely connected convolutional networks.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 2017b Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Hughes, D., M. Salathé, *et al.*, 2015 An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 .

İnce, S., I. Kunduracioglu, B. Bayram, and I. Pacal, 2025 U-net-based models for precise brain stroke segmentation. Chaos Theory and Applications **7**: 50–60.

Jaya Krishna, V., A. S. Roy, M. Mahato, and S. Das, 2025 Artificial intelligence for precision agriculture and water management. In *Integrated Land and Water Resource Management for Sustainable Agriculture Volume 2*, pp. 1–20, Springer.

Kamilaris, A. and F. X. Prenafeta-Boldú, 2018 Deep learning in agriculture: A survey. Computers and electronics in agriculture **147**: 70–90.

Karaman, A., I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, *et al.*, 2023 Robust real-time polyp detection system design based on yolo algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (abc). Expert systems with applications **221**: 119741.

Kaur, A., V. Kukreja, M. Kumar, A. Choudhary, and R. Sharma, 2024 A fine-tuned densenet model for an efficient maize leaf disease classification. 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2024 .

Kurtulus, I. L., M. Lubbad, O. M. D. Yilmaz, K. Kilic, D. Karaboga, *et al.*, 2024 A robust deep learning model for the classification of dental implant brands. Journal of Stomatology, Oral and Maxillofacial Surgery **125**: 101818.

Lecun, Y., Y. Bengio, and G. Hinton, 2015 Deep learning. Nature **521**: 436–444.

Lubbad, M., D. Karaboga, A. Basturk, B. Akay, U. Nalbantoglu, *et al.*, 2024a Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. Neural Computing and Applications **36**: 6355–6379.

Lubbad, M. A., I. L. Kurtulus, D. Karaboga, K. Kilic, A. Basturk, *et al.*, 2024b A comparative analysis of deep learning-based approaches for classifying dental implants decision support system. Journal of Imaging Informatics in Medicine **37**: 2559–2580.

Mahlein, A.-K., 2016 Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. Plant disease **100**: 241–251.

Maurya, P. K., L. K. Verma, G. Thakur, and Mayank, 2025 Artificial intelligence for precision agriculture and water management. In *Integrated Land and Water Resource Management for Sustainable Agriculture Volume 2*, pp. 185–198, Springer.

Meng, R., Z. Lv, J. Yan, G. Chen, F. Zhao, *et al.*, 2020 Development of spectral disease indices for southern corn rust detection and severity classification. Remote sensing **12**: 3233.

Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access .

Pacal, I., 2024 Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. Expert Systems with Applications **238**: 122099.

Pacal, İ., 2025 Diagnostic analysis of various cancer types with artificial intelligence .

Pacal, I., O. Akhan, R. T. Deveci, and M. Deveci, 2025 Nextbrain: Combining local and global feature learning for brain tumor classification. Brain Research p. 149762.

Pacal, I. and O. Attallah, 2025 Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.

Pacal, I., A. Karaman, D. Karaboga, B. Akay, A. Basturk, *et al.*, 2022 An efficient real-time colonic polyp detection with yolo algorithms trained by using negative samples and large datasets. Computers in biology and medicine **141**: 105031.

Pacal, I., I. Kunduracioglu, M. H. Alma, M. Deveci, S. Kadry, *et al.*, 2024 A systematic review of deep learning techniques for plant diseases. Artificial Intelligence Review **57**: 304.

Philpott, T., 2020 *Perilous bounty: The looming collapse of American farming and how we can prevent it*. Bloomsbury Publishing USA.

Pignati, W., 2018 Use of agrochemicals in brazil: The workers'and environmental health perspective. Revista Brasileira de Medicina do Trabalho **16**: 37–38.

Ranum, P., J. P. Peña-Rosas, and M. N. Garcia-Casal, 2014 Global maize production, utilization, and consumption. Annals of the new York academy of sciences **1312**: 105–112.

Rui, W., W. Cheng, P. Hong-yu, *et al.*, 2022 Intelligent diagnosis of northern corn leaf blight with deep learning model. Journal of integrative agriculture **21**: 1094–1105.

Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, 2018a Mobilenetv2: Inverted residuals and linear bottlenecks.

Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, 2018b Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520.

Sathya Priya, R., N. Jagathjothi, M. Yuvaraj, N. Suganthi, R. Sharmila, *et al.*, 2025 Remote sensing application in plant protection and its usage in smart agriculture to hasten decision making of the farmers. Journal of Plant Diseases and Protection **132**: 84.

Simonyan, K. and A. Zisserman, 2014 Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Singh, G. and S. Sharma, 2025 A comprehensive review on the internet of things in precision agriculture. Multimedia Tools and Applications **84**: 18123–18198.

Sladojevic, S., M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, 2016 Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience **2016**: 3289801.

Surendran, U., K. C. V. Nagakumar, and M. P. Samuel, 2024 Remote sensing in precision agriculture. In *Digital agriculture: A solution for sustainable food and nutritional security*, pp. 201–223, Springer.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2016a Rethinking the inception architecture for computer vision.

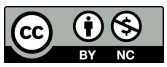
Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2016b Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Teixeira, E., K. C. Kersebaum, A.-G. Ausseil, R. Cichota, J. Guo, *et al.*, 2021a Understanding spatial and temporal variability of n leaching reduction by winter cover crops under climate change. Science of The Total Environment **771**: 144770.

Teixeira, E., K. C. Kersebaum, A.-G. Ausseil, R. Cichota, J. Guo, *et al.*, 2021b Understanding spatial and temporal variability of n leaching reduction by winter cover crops under climate change.

Science of The Total Environment **771**: 144770.
Waheed, A., M. Goyal, D. Gupta, A. Khanna, A. E. Hassanien, *et al.*, 2020 An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. Computers and Electronics in Agriculture **175**: 105456.
Willer, H., J. Trávníček, and B. Schlatter, 2024a The world of organic agriculture. statistics and emerging trends 2024 .
Willer, H., J. Trávníček, and B. Schlatter, 2024b The world of organic agriculture. statistics and emerging trends 2024 .
Zeynalov, J., Y. Çakmak, and İ. Paçal, 2025 Automated apple leaf disease classification using deep convolutional neural networks: A comparative study on the plant village dataset. Journal of Computer Science and Digital Technologies **1**: 5–17.
Zhang, X., Y. Qiao, F. Meng, C. Fan, and M. Zhang, 2018 Identification of maize leaf diseases using improved deep convolutional neural networks. Ieee Access **6**: 30370–30377.

How to cite this article: Alswilem, L., and Asadov, E. Deep Learning in Maize Disease Classification. *Artificial Intelligence in Applied Sciences*, 1(1), 20-27, 2025.



ADBA
Artificial Intelligence in Applied Sciences
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4060
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.28-34
https://doi.org/10.69882/adba.ai.2025075

A Comparative Analysis of Convolutional Neural Network Architectures for Breast Cancer Classification from Mammograms

Yigitcan Cakmak *,1 **and Javanshir Zeynalov** α,2

*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye, αDepartment of Electronics and Information Technologies, Faculty of Architecture and Engineering, Nakhchivan State University, AZ 7012, Nakhchivan, Azerbaijan.

ABSTRACT Breast cancer represents a significant global health challenge, ranking as one of the most prevalent malignancies among women. Early and accurate diagnosis through medical imaging is paramount for improving patient outcomes, with mammography serving as the gold standard for screening. However, the interpretation of mammograms can be challenging and subject to inter-observer variability. This study aims to comparatively evaluate the performance and computational efficiency of four prominent Convolutional Neural Network (CNN) architectures for the automated classification of breast cancer from mammogram images. Utilizing a publicly available dataset comprising 3,383 mammogram images classified as either Benign or Malignant, we trained and evaluated four distinct models: InceptionV3, DenseNet169, InceptionV4, and ResNet50. The results demonstrate that the DenseNet169 architecture achieved superior performance across all evaluated metrics, attaining the highest accuracy (73.33%), precision (70.45%), recall (67.83%), and F1-score (68.60%). Notably, DenseNet169 also exhibited the highest computational efficiency, featuring the lowest parameter count (12.49M) among the tested models. These findings suggest that DenseNet169 offers an optimal balance between diagnostic accuracy and model efficiency, positioning it as a highly promising candidate for integration into clinical decision support systems to aid radiologists in the early detection of breast cancer.

KEYWORDS
Breast cancer
Deep learning
Mammography
Image classification
Computer-aided diagnosis (CAD)

INTRODUCTION

Cancer represents one of the most complex and devastating diseases confronting modern medicine (García Megías *et al.* 2025). It is fundamentally a pathological condition characterized by the uncontrolled division and proliferation of the body's cells (Siqueira *et al.* 2024; Rezaei *et al.* 2025). Normally, healthy cells grow, divide, and die according to the body's needs. However, cancerous cells arise from genetic mutations that disrupt this regulated cycle, proliferating incessantly to form masses known as "tumors" (Sirvi *et al.* 2025; Yousefnia 2024). These tumors not only damage the tissue in which they are located but can also spread to other parts of the body via the blood or lymphatic system, a process known as "metastasis," thereby impairing the function of vital organs (Zuo *et al.* 2024; Li *et al.* 2025). According to World Health Organization (WHO) data, cancer is a leading cause of death globally, responsible for millions of fatalities each year (Mohanti *et al.* 2025; Lin and Park 2024). In the fight against this global health problem, understanding the biology of the disease is as crucial as achieving an early and accurate diagnosis in order to increase survival rates and enhance treatment success (Mundel *et al.* 2023; Aggarwal and Bagri 2025).

Within this broad spectrum of cancer, breast cancer is distinguished as the most prevalent type, particularly among women (Kim *et al.* 2025; Xiong *et al.* 2025). Millions of women worldwide are diagnosed with breast cancer annually, and it is the leading cause of cancer-related mortality in this demographic. Breast cancer, which can develop due to a multitude of risk factors including genetic predisposition, hormonal factors, lifestyle, and environmental influences, is a disease that is highly responsive to treatment when detected at an early stage (Obeagu and Obeagu 2024). Early diagnosis ensures the tumor is identified when it is still small and has not spread to surrounding tissues (Kiani *et al.* 2025; Alshawwa *et al.* 2024; Begum *et al.* 2024). This enables the use of less invasive treatment methods and elevates five-year survival rates to over 90% (Katsika *et al.* 2024; Trentham-Dietz *et al.* 2024). Therefore, raising public awareness and expanding regular screening programs are regarded as the most effective strategies for reducing the mortality of the disease.

Among the various imaging techniques used for the early detection of breast cancer, mammography is the most common and effective method, widely recognized as the "gold standard" (Chandra *et al.* 2025). Mammography is a radiological technique that provides detailed imaging of the breast using low-dose X-rays (Dhamija *et al.* 2025). This method allows for the detection of

Manuscript received: 21 May 2025,
Revised: 19 June 2025,
Accepted: 12 July 2025.

[1]ygtcncakmak@gmail.com (**Corresponding author**)
[2]cavansirzeynalov@ndu.edu.az

masses too small to be palpated, architectural distortions, and especially microcalcification clusters (small calcium deposits), which can be an early sign of cancer (Al-Balas *et al.* 2024). However, the interpretation of mammography is a complex and subjective process that relies heavily on the radiologist's experience and diligence (Nicosia *et al.* 2024). Factors such as high workload, fatigue, or overlooking subtle details in the image can lead to false-negative (missing a cancer) or false-positive (suspecting cancer where there is none) results (Mousa *et al.* 2024; Bahrami *et al.* 2025). This situation can lead to unnecessary biopsies, anxiety for patients, or delays in treatment. These challenges have necessitated the development of a more objective, rapid, and reliable decision support system for the analysis of mammographic images.

In recent years, advancements in artificial intelligence (AI), particularly in the field of deep learning, have instigated a revolution in medicine, especially in medical image analysis (Pacal and Attallah 2025; Pacal 2025). Convolutional Neural Networks (CNNs), owing to their superior capability to learn hierarchical features from visual data, have demonstrated performance comparable to, and in some cases, superior to that of human experts in analyzing radiological images (Pacal *et al.* 2025; İnce *et al.* 2025; Bayram *et al.* 2025). CNN-based models can automatically learn the subtle and complex discriminative features of normal tissue patterns versus benign and malignant lesions from mammograms. Within the scope of this study, the potential of this technology has been leveraged for the detection of breast cancer from mammography images (Lubbad *et al.* 2024b; Kurtulus *et al.* 2024). To this end, four different deep learning (DL) architectures with proven success in the literature ResNet50, DenseNet169, InceptionV3, and InceptionV4 were utilized to classify mammogram images as benign (0) or malignant (1) (Cakmak *et al.* 2024; Ozdemir *et al.* 2025). The objective is to compare the performance of these models to identify the most effective AI approach that can serve as a robust second-opinion and decision support tool for radiologists (Cakmak and Pacal 2025; Zeynalov *et al.* 2025; Lubbad *et al.* 2024a).

The field of medicine is undergoing a transformative evolution with the integration of AI, particularly its subfields of DL and machine learning (ML) (Obuchowicz *et al.* 2024; Koçak *et al.* 2025). These technologies offer significant advancements across a wide range of applications, from the early diagnosis of diseases and the development of personalized treatment protocols to drug discovery and the analysis of complex biological data (Li *et al.* 2024; Islam *et al.* 2024). Medical imaging, in particular, holds immense potential due to the ability of AI algorithms to process vast amounts of data and detect patterns imperceptible to the human eye (Chambi *et al.* 2025; Meng *et al.* 2024). In this context, Sarvi et al. compared Mamba-based models (VMamba and Vim) with CNN and Vision Transformer (ViT) architectures, demonstrating that under limited data conditions, Mamba architectures more effectively capture long-range dependencies, achieving a 1.98% increase in mean AUC and a 5.0% increase in accuracy (Nasiri-Sarvi *et al.* 2024). Gagliardi et al., on the other hand, developed a system that concurrently addresses both classification and segmentation tasks, investigating models that simultaneously provide radiologists with a tumor mask and diagnostic information. They identified models that achieved high-performance metrics on the BUSI dataset, including an accuracy exceeding 90%, 92% precision, 90% recall, and a 90% F1 score (Gagliardi *et al.* 2024).

In studies conducted in the realm of breast cancer diagnosis, it is found that hybrid methods and sophisticated CNN architectures have shown promising outcomes. Abhisheka et al. pointed out that in isolation DL and ML techniques are commonly inadequate, and they proposed the Hybrid Breast Cancer Prediction System (HBCPS). This model fuses deep features obtained with ResNet50 with handcrafted features, such as Histogram of Oriented Gradients (HOG), and performs classification using a Support Vector Machine (SVM). The HBCPS model proved to be effective on the BUSI data set, achieving 89.02% accuracy and a 0.8717 AUC score (Abhisheka *et al.* 2025). Similarly, Latha et al. used the EfficientNet-B7 architecture and innovative data augmentation techniques to address some accuracy concerns discovered with minority classes and applied appropriate XAI techniques (like Grad-CAM) to promote explainability of the model. Using their methods, the authors achieved classification accuracy of 99.14%, far surpassing the results of previous approaches (Latha *et al.* 2024). Thus, the findings of these studies suggest hybrid modelling and deep CNN architectures with the support of explainable AI, can be successful in classifying breast cancer.

On the other hand, efforts in the literature to enhance segmentation accuracy and improve computational efficiency are also prominent. Umer et al. proposed a U-shaped autoencoder-based CNN model equipped with a triple decoder featuring multi-attention mechanisms. They showed that this model, with its ability to capture multi-scale spatial features, achieved Dice scores of 90.45% and 89.13% on the UDIAT and BUSI datasets, respectively (Umer *et al.* 2024). Cai et al. developed the SC-Unext model, based on Unext and inspired by cellular apoptosis and division processes, with the aim of reducing computational complexity and model parameter load. This model achieved a Dice score of 75.29% and an accuracy of 97.09% on the BUSI dataset, and was also noted for its fast inference time in CPU environments (Cai *et al.* 2024). Such efficient and lightweight models are considered significant steps toward increasing usability in clinical applications. Thus, this diversity in the literature indicates that AI-based approaches point to a promising future for breast cancer diagnosis, both in terms of accuracy and operational efficiency.

MATERIALS AND METHODS

Dataset

In this study, the publicly available "Breast Cancer Detection" dataset, published by Hayder17 on the Kaggle platform, was utilized for the purpose of detecting breast cancer from mammography images (Kaggle 2025). The dataset consists of pathologically confirmed mammography images divided into two primary classes: benign lesions (labeled as class 0) and malignant lesions (labeled as class 1). Containing a total of 3383 images, with 2225 being benign and 1158 malignant, this rich dataset provides a robust foundation for evaluating the ability of the developed DL models to learn the subtle structural and textural differences between these two critical classes.

To ensure that the model development and evaluation processes are standardized and reproducible, the dataset, comprising 3383 images, was carefully partitioned into training, validation, and testing subsets. This split was performed by allocating 70% of the total dataset (2367 samples) for training, 15% (506 samples) for validation, and the remaining 15% (510 samples) for testing to independently evaluate the final performance of the model. These proportions are intended to ensure the model is trained with sufficient data while also allowing its generalization capability to be reliably measured without overfitting. Furthermore, care was taken to ensure that the class distribution in each subset reflected the proportions of the original dataset. Accordingly, the training set was composed of 1557 benign (0) and 810 malignant (1) samples; the validation set contained 333 benign (0) and 173 malignant

(1) samples; and the test set included 335 benign (0) and 175 malignant (1) samples. The partitioning of the dataset and the class distributions are also detailed in Figure 1.

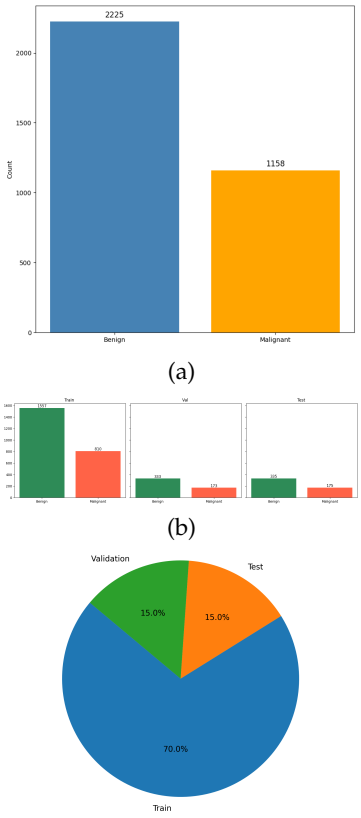


Figure 1 Statistical distribution and partitioning of the "Breast Cancer Detection" dataset. (a) Number of samples for the Benign and Malignant classes in the complete dataset. (b) Class distribution within the training (70%), validation (15%), and testing (15%) subsets. (c) The proportional split of the dataset into training, validation, and test sets.

To better visualize the structure of the dataset and the types of images it contains, representative mammography images for each class (benign and malignant) are presented in Figure 2. As illustrated in Figure 2, benign lesions tend to exhibit smooth and well-defined margins, whereas malignant lesions are more likely to display features such as irregular, indistinct, or spiculated (star-like) borders, higher density, and suspicious microcalcification clusters. In addition to these apparent morphological differences between the classes, these examples also highlight the challenges inherent in mammography, such as low contrast, the potential for dense breast tissue to obscure underlying lesions, and the ambiguities created by overlapping tissue layers. These visual representations aid in understanding the fundamental morphological features that our models must learn and differentiate, and they offer insight into the diversity of the dataset.

Data Augmentation

This study employed a dynamic data augmentation pipeline during training to enhance model generalization and mitigate the risk of overfitting, a common challenge associated with limited medical image datasets. The core strategies of this pipeline, applied

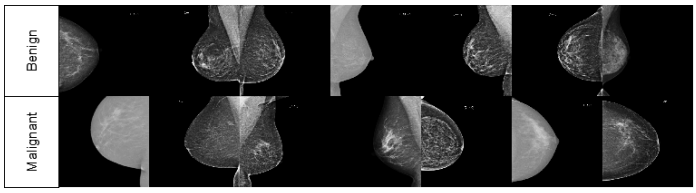


Figure 2 Representative mammogram images illustrating the Benign and Malignant classes.

randomly to each image on-the-fly, are as follows: First, through a "Random Resized Crop" operation, each image was cropped to a random scale of 8% to 100% of its original area with a variable aspect ratio (0.75 to 1.33), and subsequently resized to 224x224 pixels using a random interpolation method. This was complemented by a random horizontal flip, applied with a 50% probability. To introduce chromatic variance, "Color Jitter" was utilized to randomly alter the brightness, contrast, saturation, and hue of the images by a factor of 0.4. Notably, vertical flipping was deliberately excluded from the augmentation strategy.

The purpose of this on-the-fly methodology was to present the model with a diverse and continuously varying stream of data. This approach is designed to discourage the model from memorizing specific artifacts of the training set, thereby fostering a more robust and reliable performance on previously unseen data Wang *et al.* (2024); Mumuni *et al.* (2024).

Model Architectures

In order to tackle the problem of automatically classifying breast cancer using mammographic images, this research utilized several important deep CNN architectures. Since there is a common limitation of having little labelled data for the medical domain, we opted to utilize a transfer learning approach. In particular, we are able to utilize the excellent feature representations provided by the pre-trained models from the very large ImageNet database and transfer that knowledge learned onto the small mammogram classification task. The primary motive for this was to speed up the convergence of the model, improve generalization and reduce overfitting. Therefore, in this transfer learning approach, we loaded the initial weights of each architecture to the pre-trained ImageNet version, and we then fine-tuned each model on the breast mammography image dataset.

The first two models we wish to assess are called ResNet50 and DenseNet169, which are both architectures based on different philosophies. ResNet50 is a 50-layer network developed by He et al, and is based on the concept of residual learning. Residual learning allows ResNet50 to prevent the vanishing gradient problem with a very deep neural network. In ResNet50, the vanishing gradient problem is overcome by the addition of shortcut connections into its layers, or what they call "residual blocks" and is considered a solid baseline for classification tasks (He *et al.* 2015). DenseNet169, a 169-layer architecture proposed by Huang et al is based on dense connectivity. In this design, each layer receives inputs from all previous layers, making it easier to propagate features and reuse features. It is a unique architecture because of the improvement in parameter efficiency and assumption of improved gradient flow during training (Huang *et al.* 2017).

The study included two architectures from the Inception family of Google for their capacity to capture visual information at multiple scales. InceptionV3 uses "Inception modules" which process the input from parallel paths of different sizes of convolutional filters (e.g. 1x1, 3x3, 5x5) and pooling layers. This means that the

output of these paths can be concatenated, allowing the network to learn complex invariances at different resolutions, in addition to optimizations that included the use of factorized convolutions just to name a few. InceptionV4 builds off from InceptionV3 and establishes a more consistent and simple modular structure. It can be thought of as a refinement of the Inception concept, and within InceptionV4, there are a deeper number of blocks that have been more optimized with the interest to achieve greater performance and greater computational efficiency (Szegedy *et al.* 2016).

The strategic selection of these four distinct CNNs ResNet50, DenseNet169, InceptionV3, and InceptionV4 was intended to provide a comprehensive comparison for the task of differentiating between benign, malignant, and normal tissue in mammograms. The rationale is that the unique architectural designs and feature-learning strategies of each model are expected to yield valuable insights into which approach is most effective for this complex medical imaging problem. Through a rigorous analysis of their performance across various evaluation metrics, this study aims to contribute to the growing body of literature on developing automated, deep learning-based systems for the early diagnosis of breast cancer.

Evaluation Metrics

The evaluation of DL models is a fundamental step, indispensable for quantifying their efficacy, justifying methodological choices, and enabling informed, data-centric decision-making. Performance criteria serve multiple critical functions, including gauging the effectiveness of classification models, guiding their optimization, identifying potential errors or biases within the dataset, facilitating comparative analysis between different architectures, and diagnosing issues like overfitting. In the context of this study on breast cancer classification, we have adopted a set of standard evaluation metrics that are well-established and widely accepted within the scientific literature.

The key metrics used in this project i.e. Accuracy, Precision, Recall, and the F1-score have significant importance in scientific fields apart from deep learning. Accuracy is a general measure of performance based on the number of correctly classified instances compared to the total instances given. Precision is simply defined as true positives divided by the number of true positives and false positives and measures the reliability of the model's positive predictions. The higher the precision score, the less the false positives. Recall, or sensitivity, measures whether the model identified all actual positive cases, acting as a measure of completeness. The F1 score is simply the harmonic mean of precision and recall, providing one single metric that weights the trade-off between false positives and false negatives. These definitions are also supported by their mathematical definitions:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}} \tag{1}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

RESULTS AND DISCUSSION

In this study, the performance and computational complexity of four different CNN architectures were comparatively evaluated for the classification of breast cancer from mammography images. The obtained results are summarized in Table 1. Upon examination of the evaluation metrics, it is clearly evident that the DenseNet169 architecture exhibited superior performance compared to all other models. DenseNet169 achieved the highest score with an accuracy of 73.33%. This model also attained the most successful results with 70.45% precision, 67.83% recall, and a 68.60% F1-score. Subsequently, although the InceptionV3 and ResNet50 models presented identical accuracy rates of 72.16%, they exhibited different profiles in their precision and recall metrics. ResNet50 offered higher precision (69.39% versus 68.96%), while InceptionV3 showed higher recall (66.52% versus 65.02%). This suggests that the two models have different error profiles. Among the tested models, the InceptionV4 architecture exhibited the lowest performance with an accuracy of 70.20%. These findings, based solely on classification performance, establish DenseNet169 as the most suitable architecture for this task.

Table 1 Performance and Complexity of CNN Models for Breast Mammography Image Classification

Model	Acc.	Prec.	Rec.	F1	Params (M)	GFLOPs
DenseNet 169	73.33	70.45	67.83	68.60	12.49	6.72
Inception V3	72.16	68.96	66.52	67.22	21.79	5.67
ResNet 50	72.16	69.39	65.02	65.79	23.51	8.26
Inception V4	70.20	66.60	65.58	65.96	41.15	12.25

Beyond performance metrics, the complexity and computational efficiency of the models play a critical role in evaluating their potential for clinical application. In this context, the most striking finding is that DenseNet169, which demonstrated the highest performance, also possesses the lowest number of parameters among the tested models, with 12.49 million. This indicates that the principles of feature reuse and dense connectivity, which form the foundation of the DenseNet architecture, enable the learning of richer and more discriminative features with fewer parameters. In stark contrast, InceptionV4, which exhibited the lowest performance, is the most complex and computationally expensive model with 41.15 million parameters and 12.24 GFLOPs. This result strongly suggests that in deep learning, a larger and more complex model does not always translate to better performance; in fact, for this specific dataset, it may lead to overfitting or optimization challenges, thereby degrading performance. ResNet50 (23.51M parameters) and InceptionV3 (21.79M parameters) are positioned at a moderate level in terms of complexity, offering a balance between performance and efficiency.

Discussing the results from a clinical perspective reveals the practical value of the models. In medical diagnosis, particularly for life-threatening conditions like cancer, the balance between recall and precision metrics is of vital importance. High recall reduces the likelihood of the model missing malignant cases (false negatives), while high precision prevents the application of unnecessary anxiety and invasive procedures, such as biopsies, to a patient by incorrectly diagnosing a benign case as malignant (false positives). The fact that DenseNet169 achieved the highest scores in both recall and precision metrics indicates that it establishes this critical balance most effectively. Furthermore, its low parameter

count and reasonable GFLOPs value (6.71) facilitate its deployment on systems requiring fewer hardware resources and offer faster inference times, making its integration into the radiologist's workflow practical. Consequently, DenseNet169 emerges as the most promising candidate for development as a decision support system for breast cancer diagnosis, not only for its superior diagnostic accuracy but also for its efficiency and balanced error profile. To analyze the classification capabilities of the InceptionV3 model, which exhibited the highest performance, in more detail, its confusion matrix is presented in Figure 3.

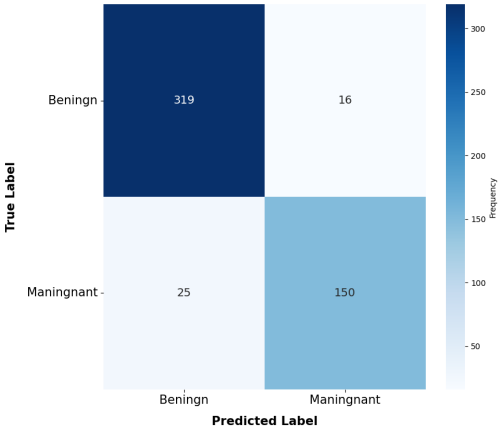


Figure 3 Confusion Matrix of the InceptionV3 Model for Breast Mammography Image Classification.

CONCLUSION

This study aimed to comparatively evaluate the performance and efficiency of four widely-used CNN architectures InceptionV3, DenseNet169, InceptionV4, and ResNet50 for the classification of breast cancer from mammography images. The results obtained unequivocally established that the DenseNet169 model was markedly superior to all other architectures in terms of both diagnostic accuracy and computational efficiency. DenseNet169 exhibited the highest performance with an accuracy of 73.33%, while also proving to be the most lightweight model with 12.49 million parameters. This finding, particularly when contrasted with the lowest performance exhibited by the most complex model, InceptionV4, reinforces the hypothesis that increased model complexity does not invariably lead to better outcomes for this specific task.

Consequently, it is concluded that the principles of dense connectivity and feature reuse inherent in the DenseNet architecture provide a significant advantage in breast cancer classification by enabling the learning of more effective features with fewer resources. While the findings of this study are promising, the limitations of the utilized dataset must be acknowledged. Future work should validate the performance of the DenseNet169 model on larger and more diverse clinical datasets, investigate the impact of different preprocessing and data augmentation techniques, and integrate Explainable Artificial Intelligence (XAI) methods to enhance model interpretability. This research constitutes an important step toward the development of both high-accuracy and efficient AI models and highlights the potential for such systems to be integrated into clinical practice as a reliable decision support tool for radiologists in the future.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Abhisheka, B., S. K. Biswas, B. Purkayastha, and S. Das, 2025 Integrating deep and handcrafted features for enhanced decision-making assistance in breast cancer diagnosis on ultrasound images. Multimedia Tools and Applications pp. 1–23.

Aggarwal, A. and N. Bagri, 2025 Incidence and epidemiology of breast cancer. In *Imaging in Management of Breast Diseases: Volume 1, Overview of Modalities*, pp. 1–9, Springer.

Al-Balas, M., H. Al-Balas, Z. AlAmer, G. Al-Taweel, A. Ghabboun, *et al.*, 2024 Clinical outcomes of screening and diagnostic mammography in a limited resource healthcare system. BMC Women's Health **24**: 191.

Alshawwa, I. A., H. Q. El-Mashharawi, F. M. Salman, M. N. A. Al-Qumboz, B. S. Abunasser, *et al.*, 2024 Advancements in early detection of breast cancer: Innovations and future directions .

Bahrami, M., S. H. Hasani, F. Karami, and M. Karami, 2025 Diagnostic values of ultrasound and mammography on prediction of lymphovascular invasion in breast cancer. Indian Journal of Gynecologic Oncology **23**: 35.

Bayram, B., I. Kunduracioglu, S. Ince, and I. Pacal, 2025 A systematic review of deep learning in mri-based cerebral vascular occlusion-based brain diseases. Neuroscience .

Begum, M. M. M., R. Gupta, B. Sunny, and Z. L. Lutfor, 2024 Advancements in early detection and targeted therapies for breast cancer; a comprehensive analysis. Asia Pacific Journal of Cancer Research **1**: 4–13.

Cai, F., J. Wen, F. He, Y. Xia, W. Xu, *et al.*, 2024 Sc-unext: A lightweight image segmentation model with cellular mechanism for breast ultrasound tumor diagnosis. Journal of Imaging Informatics in Medicine **37**: 1505–1515.

Cakmak, Y. and I. Pacal, 2025 Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the wisconsin dataset. Journal of Operations Intelligence **3**: 175–196.

Cakmak, Y., S. Safak, M. A. Bayram, and I. Pacal, 2024 Comprehensive evaluation of machine learning and ann models for breast cancer detection. Computer and Decision Making: An International Journal **1**: 84–102.

Chambi, E. A., D. G. Alzamora, and A. A. Salas, 2025 Ultrasonic image processing for the classification of benign and malignant breast tumors: Comparative study of convolutional neural network architectures. Engineering Proceedings **83**: 15.

Chandra, M., B. Varghese, and A. Sah, 2025 Mammography and advances. In *Imaging in Management of Breast Diseases: Volume 1, Overview of Modalities*, pp. 35–78, Springer.

Dhamija, E., S. Chandola, and S. Hari, 2025 Contrast-enhanced mammography. In *Imaging in Management of Breast Diseases: Volume 1, Overview of Modalities*, pp. 93–110, Springer.

Gagliardi, M., T. Ruga, E. Vocaturo, and E. Zumpano, 2024 Predictive analysis for early detection of breast cancer through artificial intelligence algorithms. In *International Conference on Innovations in Computational Intelligence and Computer Vision*, pp. 53–70, Springer.

García Megías, I., L. S. Almeida, A. K. Calapaquí Terán, K. M. Pabst, K. Herrmann, *et al.*, 2025 Fapi radiopharmaceuticals in nuclear oncology and theranostics of solid tumours: are we nearer to surrounding the hallmarks of cancer? Annals of nuclear medicine pp. 1–17.

He, K., X. Zhang, S. Ren, and J. Sun, 2015 Deep residual learning for image recognition, nd http://image-net. org/challenges. LSVRC/2015/(accessed May 24, 2021) .

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 2017 Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

İnce, S., I. Kunduracioglu, B. Bayram, and I. Pacal, 2025 U-net-based models for precise brain stroke segmentation. Chaos Theory and Applications **7**: 50–60.

Islam, M. R., M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, *et al.*, 2024 Enhancing breast cancer segmentation and classification: An ensemble deep convolutional neural network and u-net approach on ultrasound images. Machine Learning with Applications **16**: 100555.

Kaggle, 2025 Breast cancer.

Katsika, L., E. Boureka, I. Kalogiannidis, I. Tsakiridis, I. Tirodimos, *et al.*, 2024 Screening for breast cancer: a comparative review of guidelines. Life **14**: 777.

Kiani, P., H. Vatankhahan, A. Zare-Hoseinabadi, F. Ferdosi, S. Ehtiati, *et al.*, 2025 Electrochemical biosensors for early detection of breast cancer. Clinica Chimica Acta **564**: 119923.

Kim, J., A. Harper, V. McCormack, H. Sung, N. Houssami, *et al.*, 2025 Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nature Medicine pp. 1–9.

Koçak, B., A. Ponsiglione, A. Stanzione, C. Bluethgen, J. Santinha, *et al.*, 2025 Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. Diagnostic and interventional radiology **31**: 75.

Kurtulus, I. L., M. Lubbad, O. M. D. Yilmaz, K. Kilic, D. Karaboga, *et al.*, 2024 A robust deep learning model for the classification of dental implant brands. Journal of Stomatology, Oral and Maxillofacial Surgery **125**: 101818.

Latha, M., P. S. Kumar, R. R. Chandrika, T. Mahesh, V. V. Kumar, *et al.*, 2024 Revolutionizing breast ultrasound diagnostics with efficientnet-b7 and explainable ai. BMC Medical Imaging **24**: 230.

Li, X., L. Zhang, J. Yang, and F. Teng, 2024 Role of artificial intelligence in medical image analysis: A review of current trends and future directions. Journal of Medical and Biological Engineering **44**: 231–243.

Li, Y., P. Liu, B. Zhang, J. Chen, and Y. Yan, 2025 Global trends and research hotspots in nanodrug delivery systems for breast cancer therapy: a bibliometric analysis (2013–2023). Discover Oncology **16**: 269.

Lin, H.-Y. and J. Y. Park, 2024 Epidemiology of cancer. In *Anesthesia for oncological surgery*, pp. 11–16, Springer.

Lubbad, M., D. Karaboga, A. Basturk, B. Akay, U. Nalbantoglu, *et al.*, 2024a Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. Neural Computing and Applications **36**: 6355–6379.

Lubbad, M. A., I. L. Kurtulus, D. Karaboga, K. Kilic, A. Basturk, *et al.*, 2024b A comparative analysis of deep learning-based approaches for classifying dental implants decision support system. Journal of Imaging Informatics in Medicine **37**: 2559–2580.

Meng, X., J. Ma, F. Liu, Z. Chen, and T. Zhang, 2024 An interpretable breast ultrasound image classification algorithm based on convolutional neural network and transformer. Mathematics **12**: 2354.

Mohanti, B. K., P. Mathur, U. Jayarajah, B. M. Biswal, and S. Prinja, 2025 Introduction to cancer world. In *Radiation Oncology–Principles, Precepts and Practice: Volume I–Technical Aspects*, pp. 1–30, Springer.

Mousa, W. A. E.-F., D. A. A. E.-R. Tolba, and M. M. Moawad, 2024 Diagnostic value of digital breast tomosynthesis in suspicious lesions detected in screening mammogram. Egyptian Journal of Radiology and Nuclear Medicine **55**: 232.

Mumuni, A., F. Mumuni, and N. K. Gerrar, 2024 A survey of synthetic data augmentation methods in machine vision. Machine Intelligence Research **21**: 831–869.

Mundel, R., S. Dhadwal, S. Bharti, and M. Chatterjee, 2023 A comprehensive overview of various cancer types and their progression. Handbook of oncobiology: from basic to clinical sciences pp. 1–17.

Nasiri-Sarvi, A., M. S. Hosseini, and H. Rivaz, 2024 Vision mamba for classification of breast ultrasound images. In *Deep Breast Workshop on AI and Imaging for Diagnostic and Treatment Challenges in Breast Care*, pp. 148–158, Springer.

Nicosia, L., A. Rotili, F. Pesapane, A. C. Bozzini, O. Battaglia, *et al.*, 2024 Contrast-enhanced mammography (cem) compared to breast magnetic resonance (mri) in the evaluation of breast lobular neoplasia. Breast Cancer Research and Treatment **203**: 135–143.

Obeagu, E. I. and G. U. Obeagu, 2024 Breast cancer: A review of risk factors and diagnosis. Medicine **103**: e36905.

Obuchowicz, R., M. Strzelecki, and A. Piórkowski, 2024 Clinical applications of artificial intelligence in medical imaging and image processing—a review.

Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access .

Pacal, İ., 2025 Diagnostic analysis of various cancer types with artificial intelligence .

Pacal, I., O. Akhan, R. T. Deveci, and M. Deveci, 2025 Nextbrain: Combining local and global feature learning for brain tumor classification. Brain Research p. 149762.

Pacal, I. and O. Attallah, 2025 Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.

Rezaei, F., A. Mazidimoradi, Z. Pasokh, F. Mobasheri, A. Taheri, *et al.*, 2025 Global trend of breast cancer among women aged 55 and older from 2010 to 2019: An analysis by socio-demographic index and geographic regions. Indian Journal of Gynecologic Oncology **23**: 1–18.

Siqueira, P. B., M. M. de Sousa Rodrigues, Í. S. S. de Amorim, T. G. da Silva, M. da Silva Oliveira, *et al.*, 2024 The ape1/ref-1 and the hallmarks of cancer. Molecular Biology Reports **51**: 47.

Sirvi, P. K., V. Jadhav, G. Paul, R. Jain, and A. K. Yadav, 2025 Therapeutic hallmarks of cancer and immunology. In *Nanotechnology Based Strategies for Cancer Immunotherapy: Concepts, Design, and Clinical Applications*, pp. 21–53, Springer.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2016 Rethinking the inception architecture for computer vision. In

Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Trentham-Dietz, A., C. H. Chapman, J. Jayasekera, K. P. Lowry, B. M. Heckman-Stoddard, *et al.*, 2024 Collaborative modeling to compare different breast cancer screening strategies: a decision analysis for the us preventive services task force. JAMA **331**: 1947–1960.

Umer, M. J., M. Sharif, and M. Raza, 2024 A multi-attention triple decoder deep convolution network for breast cancer segmentation using ultrasound images. Cognitive Computation **16**: 581–594.

Wang, Z., P. Wang, K. Liu, P. Wang, Y. Fu, *et al.*, 2024 A comprehensive survey on data augmentation. arXiv preprint arXiv:2405.09591 .

Xiong, X., L.-W. Zheng, Y. Ding, Y.-F. Chen, Y.-W. Cai, *et al.*, 2025 Breast cancer: pathogenesis and treatments. Signal transduction and targeted therapy **10**: 49.

Yousefnia, S., 2024 Breast cancer, subtypes, risk factors, and treatment. In *The Palgrave Encyclopedia of Disability*, pp. 1–14, Springer.

Zeynalov, J., Y. Çakmak, and İ. Paçal, 2025 Automated apple leaf disease classification using deep convolutional neural networks: A comparative study on the plant village dataset. Journal of Computer Science and Digital Technologies **1**: 5–17.

Zuo, W.-F., Q. Pang, X. Zhu, Q.-Q. Yang, Q. Zhao, *et al.*, 2024 Heat shock proteins as hallmarks of cancer: insights from molecular mechanisms to therapeutic strategies. Journal of Hematology & Oncology **17**: 81.

How to cite this article: Cakmak, Y., and Zeynalov, J. A Comparative Analysis of Convolutional Neural Network Architectures for Breast Cancer Classification from Mammograms. *Artificial Intelligence in Applied Sciences*, 1(1), 28-34, 2025.