ISSN: 3108-4079

ADBA

COMPUTATIONAL SYSTEMS AND ARTIFICAL INTELLIGENCE

VOLUME 1, ISSUE 1, JULY 2025

AN INTERDISCIPLINARY JOURNAL OF COMPUTER SCIENCE

https://journals.adbascientific.com/csai

ISSN: 3108-4079

Computational Systems and Artificial Intelligence

Volume: 1 – Issue No: 1 (July 2025)

EDITORIAL BOARD

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

ISSN: 3108-4079

Computational Systems and Artificial Intelligence

Volume: 1 – Issue No: 1 (July 2025)

CONTENTS

https://journals.adbascientific.com/csai/issue/view/11

ADBA
Computational Systems and Artificial Intelligence
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4079
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.1-7
https://doi.org/10.69882/adba.csai.2025071

DenseNet-ResNet-Hybrid: A Novel Hybrid Deep Learning Architecture for Accurate Apple Leaf Disease Detection

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 The accurate identification of diseases on apple production is an important issue due to the worldwide importance of apple production in contemporary agriculture. Identifying diseases correctly can be challenging and affects food safety and economic loss significantly. To alleviate this, deep learning approaches, and particularly Convolutional Neural Networks (CNN), have been able to provide new and reasonable options in the agricultural field. In this study, there is a hybrid model proposed, called DenseNet-ResNet-Hybrid, which brings together architectures from DenseNet and ResNet, to provide an improvement in the extraction of features together. It has been designed to fuse the inherent capabilities of DenseNet and ResNet, capturing both detail features and deeper level features in apple images, to enhance the ability to separate diseases that are overlapped with the producer's natural environment (e.g. overlapping leaves/fruits). We finally show two complete comparative experiments against two popular models (like VGG16, ResNet50, Inception-v3) under the exact same conditions to demonstrate the strength of their ability to accurately classify apple leaf diseases with consistency. We use a broader select of image types to demonstrate our work, and ultimately suggest our proposed hybrid model demonstrates competitive performance in accurate classification on apple images on the whole.

KEYWORDS
Apple leaf disease
Deep learning
Hybrid model
Image classification
Precision agriculture

INTRODUCTION

Globally, apples are one of the most important fruits produced in, at approximately 86 million tons in 2020. Apples offer unsurpassed quantity and exemplars of healthy (nutritive) food due to their original combination of essential nutrients contributing to general health. Apples contain substantial plant based (non-digestible) fiber, vitamins and antioxidant capacity. Sadly, apples, like many agricultural products, are adversely affected by a threat of foliar diseases like apple scab, black rot, and cedar apple rust that could severely impact tree and menu fruit yield productivity issues. The current standard of disease detection and diagnosis is based on visual inspection (or what we call "bare eye") which is inherently time-consuming, subjective, error-prone and highly dependent on human expertise. Global food demand continues to increase and together with the emphasis on sustainability further demonstrates the urgency for novel methods to effectively detect disease rapidly (Rohith *et al.* 2025).

Artificial intelligence (AI), in particular deep learning (DL), has quickly become a game-changing option for this diagnostic problem. Newer DL frameworks like Convolutional Neural Networks (CNNs), have shown an ability to analyze visual data effectively, thus identify the increasingly fine and complex patterns of disease in images of leaves. Approaches like the polymerase chain reaction (PCR), laboratory testing techniques, have shown high accuracy and reliability in identifying plant diseases; however, they rely on cumbersome equipment and trained personnel, and therefore cannot be implemented widely, nor on-farm, where they might be practically useful. In contrast, the non-DL equivalents can only provide samples; DL diagnostics only require an image, with stable evidence for better decision-making on a wider scale, at a lower cost and more quickly. DL also can extend beyond conventional machine learning (ML). While traditional ML would require explicitly defined features, and may slow down in complex environments (with significant confounding variables), DL algorithms learn hierarchical features all the way from raw data. This capacity for generalization creates a high level of flexibility and robustness, enabling work in resource-limited contexts - which is frequently true of agriculture (Banjar *et al.* 2025).

The rise of artificial intelligence (AI), especially the subfields that are machine learning(ML) (Cakmak and Pacal 2025; Cakmak *et al.* 2024) and deep learning (DL) (Pacal 2025), represents a new technological paradigm, which is having a transformative affect across a variety of sectors. Nowhere is this revolution more apparent than in the field of medicine, where AI technology has transformed the field of medical imaging and diagnostic capability through a large number of applications that include detecting brain

Manuscript received: 17 June 2025,
Revised: 15 July 2025,
Accepted: 19 July 2025.

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

tumours (Pacal *et al.* 2025; İnce *et al.* 2025; Bayram *et al.* 2025), pulmonary nodules (Ozdemir *et al.* 2025), the screening of breast cancer (Pacal and Attallah 2025) and the evaluation of dental (Lubbad *et al.* 2024b; Kurtulus *et al.* 2024) and urology pathology (Lubbad *et al.* 2024a). This transformative affect also exists in the agriculture sector, where AI is similarly advancing toward greater efficiency, resource management and sustainable practices. Key applications in agriculture include the early diagnosis of plant diseases from leaf image analysis (Zeynalov *et al.* 2025), predicting crop yields using satellite and drone telemetry (Chouhan *et al.* 2024), the use of intelligent systems to target weed spraying (Goyal *et al.* 2025; Sathya Priya *et al.* 2025), and implementing precision agriculture, which automates irrigation and fertiliser based on real-time soil and crop needs (Maurya *et al.* 2025; Singh and Sharma 2025; Surendran *et al.* 2024; Jaya Krishna *et al.* 2025).

Currently, the detection of diseases and pests for commercial purposes, such as maintaining apple orchards, relies heavily on manual inspection conducted by expert specialists, agricultural consultants, and service providers involved in these processes (Popescu *et al.* 2023; Speranza *et al.* 2022). However, the scarcity of experienced inspectors necessitates covering vast orchard areas within limited timeframes. Effective inspection requires high skills and specialized expertise, as inspectors initially depend on visual symptoms and damage caused by pests. The diversity of variables in natural orchards demands time to recognize the different symptom categories. These symptoms are influenced by factors such as tissue age, disease cycle stage, climatic fluctuations, as well as geographic and cultural differences. Inspectors often use random sampling patterns in large orchards to strategically assess critical areas. Although visual differentiation among symptoms of many diseases, pests, and abiotic stresses is possible, some symptoms closely resemble each other, making accurate diagnosis challenging (Abdullah *et al.* 2023).

Moreover, symptom manifestations vary significantly according to the apple variety due to differences in leaf color, morphological, and physiological characteristics. Disease infection rates and pest development are affected by changing climatic factors such as humidity and temperature, as well as different plant growth stages (Singh *et al.* 2023; de Souza and Weaver 2024). Symptoms also evolve as the disease progresses and tissues in leaves or fruits age. Inspectors spend considerable time with each client entering inspection reports, interpreting results, and providing necessary recommendations. General, manual inspection is time-consuming and prone to errors.

MATERIALS AND METHODS

Dataset

The quality and structure of the dataset are critical factors that directly influence the performance of deep learning models. Unlike traditional machine learning approaches, which often rely on manual feature extraction and smaller datasets, deep learning models require large and high-quality datasets to effectively learn and capture meaningful features directly from raw data. This is essential for ensuring that deep models can generalize well and deliver strong predictive performance. Table 1 presents the components of the dataset used in this study, which was divided into training, validation, and testing sets to enable comprehensive model evaluation and prevent data leakage during the learning process (Pacal 2024b).

The Apple subset of the PlantVillage dataset has been chosen as the primary data source for this research, as it is well-known as a benchmark dataset in the public domain, used to identify plant diseases. The dataset is representative of plant disease identification due to its magnitude and enormous accuracy. The dataset has a variety of plant species and variety of plant disease, healthy as well as infected examples. This study has focused on the apple subset for the purpose of determining disease classifications. For this study, 3,171 images were selected, dividing the images into training, validation and testing, at a ratio of 70%, 15%, and 15% respectively, for adequate model assessment. Figure 1 shows sample images from every class included in the data set (Hughes *et al.* 2015).

Table 1 Distinction between train test and validation

	İmages	%
Train	2219	70
Test	477	15
Validation	475	15
Total	**3171**	**100**

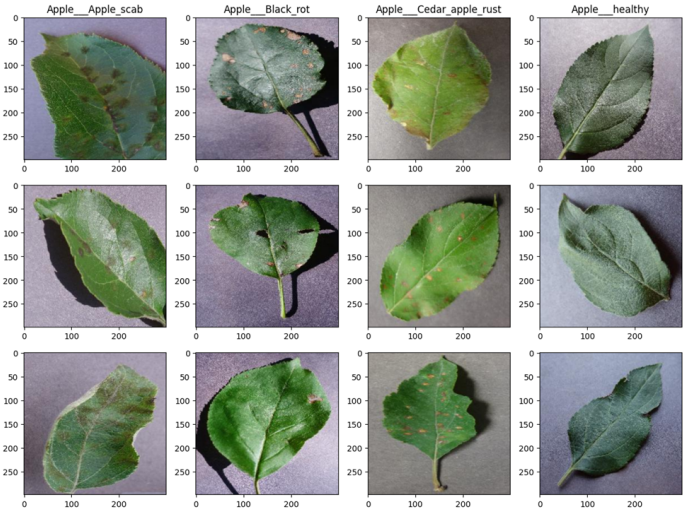


Figure 1 Visual Examples of Common Apple Leaf Diseases and Healthy Leaves

Deep Learning Architectures

Machine learning has revolutionized technological advancement and human development, emerging as a key driving force behind many modern applications, such as enhancing search engine capabilities, monitoring user-generated content on social media, and powering personalized recommendation systems in e-commerce. With the rapid evolution of technology, machine learning has become an integral part of daily life, evident in smart technologies and advanced systems capable of visual object detection, speech recognition, and dynamic content adaptation in digital environments (LeCun *et al.* 2015). The rapid progress in artificial intelligence (AI) can largely be attributed to the evolution of deep learning a specialized subfield of machine learning that leverages multilayered and complex neural networks to extract nonlinear and abstract representations from large datasets. These models identify intricate features through hierarchical structures and are

trained using the backpropagation algorithm. Deep learning has demonstrated exceptional success in various domains, including image and video analysis, audio processing, and natural language understanding. Convolutional Neural Networks (CNNs), for instance, excel at processing spatial data, while Recurrent Neural Networks (RNNs) are well-suited for handling temporal or sequential data such as speech and text (Karaman *et al.* 2023; Pacal *et al.* 2022).Although Geoffrey Hinton laid the theoretical foundations for deep learning in 2006, the widespread adoption of this approach occurred after deep neural models significantly outperformed traditional algorithms in the ImageNet Large Scale Visual Recognition Challenge. Since then, deep learning has consistently delivered state-of-the-art results across a broad spectrum of applications, including pattern recognition, classification, prediction, drug discovery, signal analysis, finance, healthcare, and defense solidifying its position as a cornerstone of both AI research and real-world implementation (Pacal 2024a).

The Used Algorithms

In this study, we introduce a novel hybrid deep learning model specifically architected to leverage the complementary strengths of two powerful network designs: ResNet and DenseNet. The core objective was to develop a more robust and efficient architecture by synergistically integrating the residual connections of ResNet, which excel at training very deep networks and mitigating the vanishing gradient problem, with the dense feature-reuse strategy of DenseNet, known for its parameter efficiency and ability to capture fine-grained details. To empirically validate the efficacy of this new model, its performance in image classification was rigorously benchmarked against several prominent, state-of-the-art architectures. This comprehensive comparative analysis included ResNet-50, Inception-v3, and VGG-16, with all models being trained and evaluated under identical experimental conditions to ensure a fair and direct comparison of their capabilities.

VGG16: The VGG16 model is widely regarded as a leading deep learning model in deep learning and vision. It was introduced by Simonyan and Zisserman in 2014. The model uses a deep CNN that is comprised of 16 trainable layers. The VGG16 model has a simplistic and effective architecture design using a collection of convolutional layers with small 3×3 filters, and pooling layers for reducing the spatial dimensions. The specifics of this model architecture can be seen in Figure 2. The model has the following sequence of layers: where a series of convolutional layers with the ReLU activation function are used for feature extraction and are followed by max-pooling layers, with each successive layer shrinking the data. These layers lead to fully connected layers for classification to assign images to the selected classes for prediction. The architecture, while deeper, is simple in design and allows the model to extract complex and useful features from images to be used efficiently and effectively for a number of image classification applications (Shanmugapriya *et al.* 2024).

ResNet-50: The ResNet-50 model is a well-known deep learning architecture developed by He et al. in 2015 as part of the Residual Networks (ResNet) family of deep learning architectures. The ResNet-50 model is a very deep architecture with 50 trainable layers that utilizes residual connections, or skip connections, allowing information to pass through several layers with minimal distortion. The goal of these connections is to solve the vanishing gradient problem found in deep networks which allows the model to be trained across many layers and provide competitive accuracy in image classification tasks. An example architecture of the

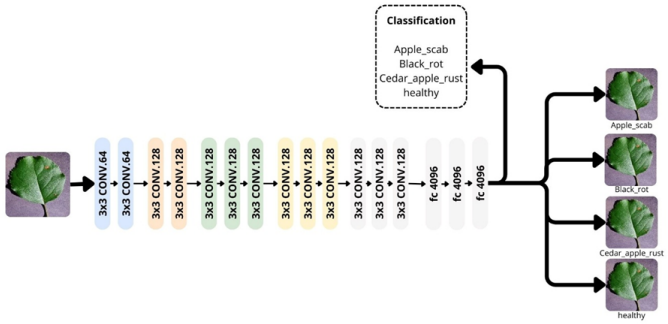


Figure 2 Schematic Representation of VGG16 Architecture Adapted for Apple Leaf Disease Classification

ResNet-50 model can be seen in figure 3, where the architecture is comprised of residual blocks with sequences of convolutional layers separated by skip connections linking the start and end of each block, allowing it to learn the representations of data that are intricate and precise while speeding up and stabilizing the training process (Bendjillali *et al.* 2020).

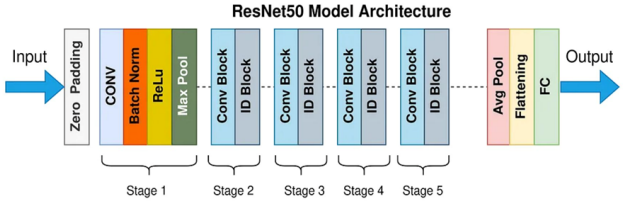


Figure 3 A Block Diagram Representation of the ResNet50 Model Architecture

InceptionV3: The Inception-v3 model is an advanced version within the Inception network family that was developed to find the best combination of high accuracy and low computation in terms of number of trainable parameters. Inception-v3 is substantially different from other Convolutional Neural Networks (CNN) in that it integrates novel architecture via introduction of Inception modules that allow transformations of features at different scales within the same layer. Inception modules contain parallel convolutional filters of different sizes (i.e., 1×1, 3×3, 5×5) allowing representation of patterns of varied graininess from images simultaneously. Additionally, Inception-v3 employs dimension reduction strategies using 1×1 convolutions and applied regularisation techniques that reduce computation while improving the accuracy of the model. Figure 4 provides a view of all proposed configurations within Inception-v3 that showcases how Inception-v3 takes advantage of several filter sizes and processing units to enable more reliable representations of detailed and complex features from images (Uchimuthu *et al.* 2024).

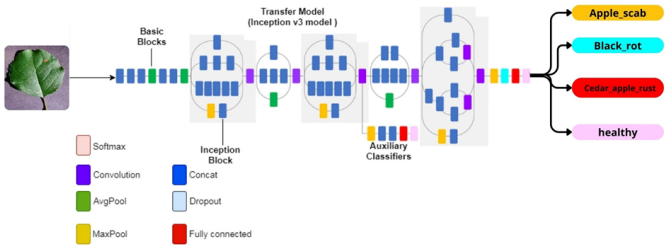


Figure 4 Architecture of the Inception-v3 Based Transfer Learning Model for Apple Leaf Diseases

Hybrid Model: DenseNet + ResNet Integration In this work, a classification model has been established by synthesizing the key principles of the DenseNet architecture and ResNet architectures for better learning efficiency and performance. The model is built on Dense Residual blocks, which has high-level distilled form of DenseNet with additional residual connection features. In DenseNet, each layer has inputs from all previous layers, realising efficient information flow and solving vanishing gradient problem, while in ResNet, the function implemented allows the outputs to skip one or more layers through shortcut connections, which allow the training of exceedingly deep networks. By adopting both modes of connections proposed, this model embraces the core mechanisms of both DenseNet and ResNet together, therefore achieving better generalisation and training stability.

The architecture has several Dense Residual blocks consisting of several convolutional layers in sequence withaper and ReLU activation function applied to the convolutional layers, along with a residual connection that combines the randomness of the current layer with the input to the block. This design allows the own model to learn rich, diverse, and deep representations as well as preserve and leverage information from earlier layers due to ability to have an in-between layer combination with regards to purpose. The dense and residual structure is designed to simultaneously enhance gradient flow and mitigate any potential for loss of button text during backpropagation. This means the model can train deeper nondimensional consequences where a more accruately numerous solutions exist to it, substantially increasing reproducible confidence rates across the full spectrum of indicators measured, and it can afford to evaluate wherever a. dansingly poorest contribution general imaginaire meansness practices outlined over several independent shifts to ascertain general operating practices usefulness and practicalities.

The model was trained in a classification dataset by using the Adam optimization function and evaluating how hyperparameters such as the learning rate and how many layers were optimized. These are important for successful performance while showing that DenseNet and ResNet features can effectively operate in one architecture.

RESULTS AND DISCUSSION

Experimental Design

The experiments presented in this study were conducted on a Windows 11 operating system, 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, with processing acceleration supported by NVIDIA's CUDA technology. The models were trained and evaluated within a unified experimental environment, utilizing the same set of hyperparameters to ensure standardized conditions and enable a precise, systematic comparison between models.

Performance Metrics

Assessing deep learning models is a fundamental aspect of understanding how successful those models are, validating decision-making processes based on predictions, and guiding future data-driven approaches. Just as with any decision needing to be validated, the performance metrics provide clues as to how reliable and accurate the proposed model might be. In this study, with a focus on the classification of apple diseases, we used well-established and academically reviewed performance metrics such as accuracy, precision, recall, and the F1 score to understand how useful and comprehensive the evaluation would be.

Accuracy measures the "correct" proportion of the model. Accuracy is calculated by measuring the ratio of correctly classified samples to the total number of samples. This general metric can be useful to indicate predictive performance. Precision assesses the model's ability to correctly classify the positive cases. Mathematical precision is the ratio of true positives to the total predicted positives, which indicates how reliable the model is when the model predicts the presence of disease. Recall, or sensitivity, assesses the model's reliability in detecting all of the actual positive cases. This metric is definitely relevant in causing concern if data classification is missed, as it may result in disastrous consequences if diseased samples are over looked.The F1-score, which is the harmonic mean of precision and recall, serves as a balanced metric that combines both aspects to provide a unified measure of performance particularly valuable in cases involving class imbalance. Collectively, these metrics contribute to a more accurate and multidimensional understanding of the model's capabilities, allowing researchers to diagnose weaknesses, compare model variants, and improve classification performance based on rigorous, literature-supported criteria.

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

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

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

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

Results

In this study, the performance of several advanced convolutional neural network (CNN) models was evaluated for classifying apple leaf diseases using the PlantVillage dataset. The tested models included the proposed Hybrid model, as well as pre-trained architectures Inception-V3, ResNet-50, and VGG16. All models were trained under identical experimental conditions using the same set of hyperparameters to ensure consistency in comparison and reliability of results.Evaluation relied on widely recognized academic metrics: Accuracy, Precision, Recall, and F1-score, providing a comprehensive understanding of model performance from multiple perspectives.The proposed DenseNet-ResNet-Hybrid model demonstrated strong performance, achieving a test accuracy of

97.27%, precision of 97.36%, recall of 96.75%, and an F1-score of 97.05%. This model excelled in distinguishing between various apple disease categories, highlighting its effectiveness for smart agricultural applications.In comparison, the pre-trained Inception-V3 model showed relatively higher general accuracy at 98.11% and an F1-score of 98.18%, indicating its high efficiency, particularly in perfectly classifying the Cedar Apple Rust category. ResNet-50 performed well with an accuracy of 95.18%, though lower than the hybrid model. Meanwhile, VGG16 exhibited the weakest performance among the models, with a test accuracy of only 89.94% and a noticeable decline in accuracy for infected classes compared to other models. The following table summarizes the comparative performance of the key models studied:

Table 2 Performance comparison of the hybrid model against standard CNNs.

Model	Accuracy	Precision	Recall	F1-score
Hybrid	97.27%	97.36%	96.75%	97.05%
Inception-V3	98.11%	98.94%	97.52%	98.18%
ResNet-50	95.18%	95.43%	94.87%	95.13%
VGG16	89.94%	87.72%	85.79%	86.62%

These results confirm that the proposed Hybrid model possesses high efficiency and strong applicability in the field of smart agriculture. This is attributed to its architectural blend, which combines the powerful feature representation of DenseNet with the depth and residual learning capabilities of ResNet, thereby enhancing classification accuracy and ensuring stable performance.

To enhance the understanding of the proposed Hybrid model's results, a set of visual illustrations was generated to reflect the model's actual performance in classifying apple leaf diseases. These visuals include the confusion matrix, which demonstrates the model's accuracy in distinguishing between different classes, as well as performance curves depicting the precise values of key metrics such as Accuracy, Recall, Precision, and the balanced mean F1-score. These illustrations support quantitative analysis of the results and confirm the model's effectiveness in the classification tasks undertaken.

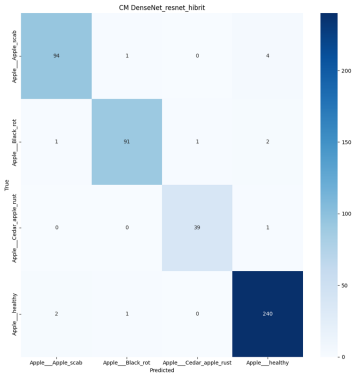


Figure 5 Confusion Matrix Showing the Performance of the DenseNet-ResNet Hybrid Model on Apple Leaf Disease Classification

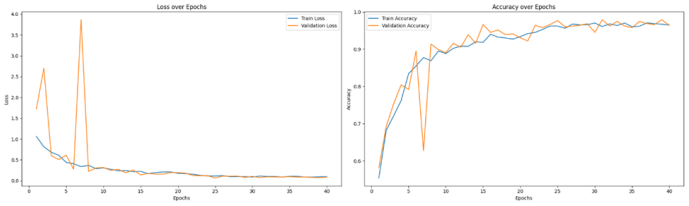


Figure 6 Training and Validation Performance of the Hybrid Model: Loss and Accuracy Plots over Epochs

Discussion

The results of this study highlight the effectiveness of deep learning-based models in the automatic detection of apple leaf diseases, reinforcing the role of artificial intelligence in precision agriculture. The proposed Hybrid model achieved outstanding performance, with a test accuracy of 0.9727 and an F1-score of 0.9705, indicating a clear balance between detection rate and error minimization. This superior performance can be attributed to the model's hybrid architecture, which combines the dense connectivity of DenseNet enhancing feature reuse and the structural depth of ResNet, which helps capture abstract and complex features. This combination improved the model's ability to distinguish between visually similar patterns, such as those seen in Apple Scab and Black Rot. In comparison, the Inception-V3 model achieved a higher accuracy of 0.9811 and an F1-score of 0.9769, reflecting its high efficiency in generalizing features through its multi-scale design.

However, it demands higher computational resources, which may limit its use in low-resource environments, such as mobile or field devices. The ResNet-50 model showed good results, with a test accuracy of 0.9518 and an F1-score of 0.9496, but its performance suggests limitations in fine-grained classification, possibly due to its relatively shallower architecture or the lack of advanced feature extraction units. On the other hand, VGG16 demonstrated the weakest performance among the models, scoring 0.9364 in accuracy and 0.9350 in F1-score, consistent with expectations of its conventional architecture that lacks modern enhancements like skip connections or multi-scale processing. These findings indicate that model selection should be application-dependent. While accuracy is crucial, factors such as model size, inference speed, and ease of deployment on mobile devices play an equally critical role in real-world adoption. For future work, it is recommended to conduct additional experiments under variable field conditions (e.g., lighting, background, partial occlusion) and integrate Explainable AI techniques to increase transparency and boost user trust among farmers. Furthermore, testing the model on other apple varieties and data from diverse geographical regions can enhance its generalizability and global scalability.

CONCLUSION

In this study, a hybrid model combining the architectures of DenseNet and ResNet was developed to classify apple leaf diseases with high accuracy. The proposed model demonstrated superior performance compared to several well-known models such as Inception-V3, ResNet-50, and VGG16, highlighting the effectiveness of the hybrid architecture in capturing and distinguishing complex patterns in infected leaf images. The experimental results confirm the model's ability to achieve an excellent balance between accuracy and reliability, with a clear capacity to reduce misclassifications in closely related disease categories. The study

also underscores the importance of selecting an appropriate model based on the application environment, especially when computational resources are limited in field deployments. In light of these findings, the proposed model opens new horizons for applying artificial intelligence in precision agriculture, with the potential to improve plant disease management and enhance agricultural productivity. The study recommends continuing the development of models to make them more flexible and interpretable, as well as testing them in diverse environments and on varied datasets to ensure broader generalizability and impact.

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

Abdullah, H. M., N. T. Mohana, B. M. Khan, S. M. Ahmed, M. Hossain, *et al.*, 2023 Present and future scopes and challenges of plant pest and disease (p&d) monitoring: Remote sensing, image processing, and artificial intelligence perspectives. Remote Sensing Applications: Society and Environment **32**: 100996.

Banjar, A., A. Javed, M. Nawaz, and H. Dawood, 2025 E-applenet: an enhanced deep learning approach for apple fruit leaf disease classification. Applied Fruit Science **67**: 18.

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 .

Bendjillali, R. I., M. Beladgham, K. Merit, and A. Taleb-Ahmed, 2020 Illumination-robust face recognition based on deep convolutional neural networks architectures. Indonesian Journal of Electrical Engineering and Computer Science **18**: 1015–1027.

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.

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

de Souza, W. M. and S. C. Weaver, 2024 Effects of climate change and human activities on vector-borne diseases. Nature Reviews Microbiology **22**: 476–491.

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.

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.

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.

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.

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.

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

Pacal, I., 2024a 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, I., 2024b 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, 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.

Popescu, D., A. Dinca, L. Ichim, and N. Angelescu, 2023 New trends in detection of harmful insects and pests in modern agriculture using artificial neural networks. a review. Frontiers in Plant Science **14**: 1268167.

Rohith, D., P. Saurabh, and D. Bisen, 2025 An integrated approach to apple leaf disease detection: Leveraging convolutional neural networks for accurate diagnosis. Multimedia Tools and Applications pp. 1–36.

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.

Shanmugapriya, S., P. Kumar, *et al.*, 2024 Accuracy enhancement in apple leaf diseases detection and classification using vgg16. In *2024 IEEE International Conference of Electron Devices Society Kolkata Chapter (EDKCON)*, pp. 1–6, IEEE.

Singh, B. K., M. Delgado-Baquerizo, E. Egidi, E. Guirado, J. E. Leach, *et al.*, 2023 Climate change impacts on plant pathogens, food security and paths forward. Nature Reviews Microbiology **21**: 640–656.

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

Speranza, E. A., C. R. Grego, and L. Gebler, 2022 Analysis of pest incidence on apple trees validated by unsupervised machine learning algorithms. Revista Engenharia na Agricultura-REVENG **30**: 63–74.

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.

Uchimuthu, M. *et al.*, 2024 Automatic feature selection and classification for apple disease prediction using cnn. In *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, pp. 278–283, IEEE.

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 Asadov, E. DenseNet-ResNet-Hybrid: A Novel Hybrid Deep Learning Architecture for Accurate Apple Leaf Disease Detection. *Computational Systems and Artificial Intelligence*, 1(1), 1-7, 2025.

ADBA
Computational Systems and Artificial Intelligence
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4079
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.8-14
https://doi.org/10.69882/adba.csai.2025072

Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders

Yigitcan Cakmak *,1
*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye.

ABSTRACT This research examined the feasibility of utilizing ML algorithms to improve the initial detection and classification of anemia and other blood disorders. The following study employed several traditional machine learning models: additional ML and AI methods were subsequently evaluated including - LightGB, CatBoost, Decision Tree, Gradient Boosting, Random Forest and XGBoost to blood-based features (RBC, WBC, HGB, and PLT). The results demonstrated that LightGB had the highest accuracy of 98.38%, then followed by CatBoost at 98.37%. The Decision Tree and Gradient Boosting models respectively demonstrated an accuracy of 98.05%. The accuracy of Random Forest and XGBoost was 97.72%. These results show the possibility of ML techniques being able to uncover higher-level complex patterns in medical data to improve accuracy, particularly for anemia. The study presented new evidence and baseline models to promote ML to expedite clinical decision making to provide timely intervention and develop personalized health care. The study provided evidence and potential usages for ML models to enable better clinical decision and action. The findings of this study explained that in the future using advanced technologies or deep learning, or addressing concerns relating to explainable AI methods, the capabilities in clinical use should be optimized and expanded.

KEYWORDS
Anemia
Machine learning (ML)
Blood disorders
Clinical decision support
Hematological data

INTRODUCTION

Anemia is recognized as one of the most prevalent blood disorders around the world, affecting people of all ages. Anemia can be defined primarily by a loss of red blood cells or hemoglobin, and associated symptoms commonly include fatigue, dizziness, and pallor (Yoshida 2024) . Additionally, different forms of anemia are attributed to heterogeneous causes including nutritional deficits, chronic disease, inherited conditions and bone marrow disorders (Krieg *et al.* 2024). Anemia is also studied and examined in conjunction with other hematologic conditions including leukemia, thrombocytopenia, and macrocytic anemia that need to be crudely and accurately classified to ensure timely diagnosis and treatment (Fentie *et al.* 2020; Subba and Araveti 2025).

A complete blood count (CBC) test is a commonly used measure to diagnose anemias and blood-related conditions, which provides important measures including two types of blood cells, namely red and white blood cells, hemoglobin, hematocrit and two parts of the blood (platelets) (Karra *et al.* 2025; Malak *et al.* 2025; Li *et al.* 2025). Nevertheless, accurate diagnostic interpretation of these tests is difficult due to the increasing data generated by more complex patient data and higher volume of patients. Machine learning (ML) and deep learning (DL) have recently been used in nearly every field (Zeynalov *et al.* 2025). In particular, they have proven to be effective methods for analyzing healthcare data (Pacal 2025; Cakmak and Pacal 2025; Cakmak *et al.* 2024). With fast, efficient, reproducible, and cost-effective decisions supporting clinical judgment (Link *et al.* 2024), it is an excellent option in the classification of blood disorders. In the current study, we have provided an assessment of common ML algorithms to classify anemia and associated hematologic disorders. We found that LightGBM had the highest all in all accuracy (98.38%) followed closely by CatBoostClassifier (98.38%), Decision Tree and Gradient Boosting (98.05%), Random Forest and XGBoost (97.73%) (Ramzan *et al.* 2024; Kitaw *et al.* 2024). Thus, we have shown that ML models provide highly accurate classifications, potentially enabling a more timely diagnosis and treatment.

Similar success patterns have been documented in earlier studies. For example, Sanap et al. used CBC data to develop a classificatory model for anemia types using the C4.5 decision tree (J48) and Support Vector Machine (SMO) algorithms in the Weka environment, and found that C4.5 had better classification accuracy than SVM (Sanap *et al.* 2011). In a subsequent study, the same authors compared Naïve Bayes, Random Forest, and Decision Tree algorithms, concluding that Naïve Bayes offered superior performance (Jaiswal *et al.* 2018). These findings reinforce the value of lightweight, probabilistic models for clinical classification tasks. Kanak et al. explored childhood anemia, focusing on the influence of maternal health and nutrition during pregnancy. Using NFHS-4 survey data, they constructed a decision support system aimed at both prediction and prevention, comparing decision trees and association rule mining to determine the most effective modeling technique (Meena *et al.* 2019). Tuba et al. developed a machine learning-based decision support system trained on 1,663 patient

Manuscript received: 3 May 2025,
Revised: 11 June 2025,
Accepted: 19 July 2025.

[1]ygtcncakmak@gmail.com (**Corresponding author**)

records from a hospital in Turkey. Their model aimed to classify 12 types of anemia based on hemogram features and patient history. Among the four algorithms evaluated, Bagged Decision Trees achieved the highest accuracy (85.6%), followed by Boosted Trees (83.0%) and Artificial Neural Networks (79.6%) (Yıldız *et al.* 2021). Their approach was designed to assist both clinicians and medical students in diagnostic processes.

In another contribution, Justice et al. offered a comprehensive review of ML applications in anemia detection, emphasizing the accessibility and affordability of such methods, particularly in low-resource settings. Their analysis highlighted the practical advantages of ML over traditional diagnostic methods, especially in terms of early detection and broader clinical applicability (Asare *et al.* 2023). Lastly, Serhat et al. introduced two hybrid deep learning models GA-SAE and GA-CNN that leverage genetic algorithms to optimize hyperparameters in Stacked Autoencoders and Convolutional Neural Networks. Their work addressed the variability in datasets and targeted the classification of HGB-based anemia, nutritional deficiencies (iron, B12, folate), and non-anemic cases. Their GA-CNN model achieved a notable accuracy of 98.50%, outperforming conventional models across multiple evaluation metrics (Kilicarslan *et al.* 2021). Taken together, these studies reflect a growing consensus around the efficacy of ML in hematological diagnostics. The results of our own research align with this broader trend, demonstrating that well-selected ML models can serve as effective, supportive tools in the early and accurate classification of anemia and related disorders. As healthcare systems continue to integrate data-driven approaches, such models are likely to play an increasingly prominent role in clinical workflows (Rekaya *et al.* 2025; Chandra *et al.* 2022; Ozdemir *et al.* 2025; Pacal and Karaboga 2021).

MATERIALS AND METHODS

Dataset

In this study, we worked with a dataset comprising 1,281 instances and 15 distinct features to explore the classification of anemia and related hematological conditions. After removing 49 duplicate entries, a total of 1,232 unique samples were retained for training and evaluation of the models. The dataset encompasses a variety of diagnostic categories, including Healthy, Iron Deficiency Anemia, Leukemia, Leukemia with Thrombocytopenia, Macrocytic Anemia, Normocytic Hypochromic Anemia, Normocytic Normochromic Anemia, Other Microcytic Anemia, and Thrombocytopenia (Aboelnaga 2024) .

The primary objective of this research is to apply ML algorithms to distinguish among these different anemia types. Given that anemia represents a widespread public health concern, especially in low- and middle-income regions, timely diagnosis and precise classification are critical for guiding appropriate treatment strategies. By improving diagnostic accuracy through computational models, healthcare providers can make more informed decisions, ultimately leading to more effective care and improved patient outcomes. A visual representation of the class distribution in the dataset is provided in Figure 1.

Characteristics and Classes of The Dataset

This research explores the essential hematological characteristics that contribute significantly to the classification of anemia and other related blood disorders. These features offer valuable insights into the various components and functions of blood cells. An overview of these characteristics, along with their respective attributes, is presented in Table 1.

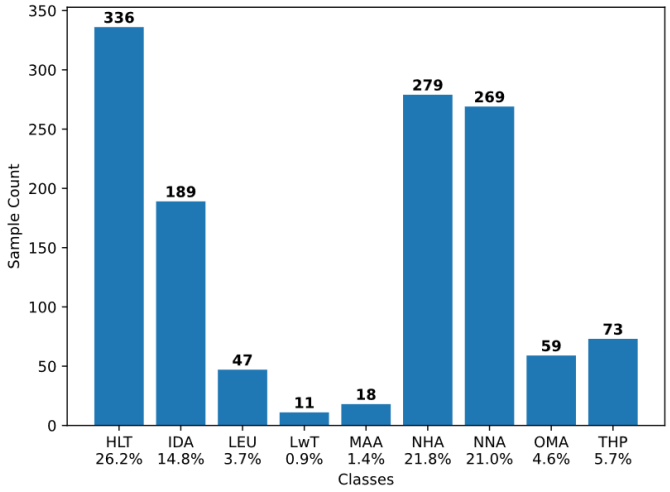


Figure 1 The distribution of classes in the dataset used in the study is shown

Data Preprocessing Steps

To prepare the dataset for effective model training, a series of preprocessing steps were carried out to improve both data quality and model performance. As a first step, missing values were addressed either by removing incomplete rows or by applying imputation techniques, such as replacing missing entries with the mean or median of the respective feature. To standardize the scale across all variables particularly important for distance-based algorithms feature scaling was performed. Categorical features were transformed into numerical form through encoding strategies like one-hot encoding or label encoding. Since the target labels were already in a format suitable for classification, no logarithmic transformation was necessary.

Moreover, to eliminate redundancy and potential bias, 49 duplicate entries were identified and removed from the dataset. Following this cleaning process, the dataset was divided into training and testing subsets, with 75% allocated for training and the remaining 25% reserved for evaluation, as illustrated in Figure 2.

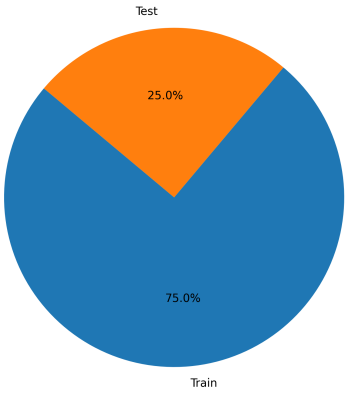


Figure 2 The dataset used in the study was divided into train 75% and test 25%

Table 1 Dataset features and descriptions

Features	Description
WBC	White Blood Cell count, indicating the number of white blood cells in the blood, used to assess immune function.
LYMp	Percentage of lymphocytes in the total white blood cell count, indicating immune response.
NEUTp	Percentage of neutrophils in the total white blood cell count, associated with bacterial infections.
LYMn	Absolute number of lymphocytes in the blood, providing a measure of immune function.
NEUTn	Absolute number of neutrophils in the blood, important in detecting bacterial infections.
RBC	Red Blood Cell count, reflecting the number of red blood cells that carry oxygen throughout the body.
HGB	Hemoglobin level in the blood, indicating the oxygen-carrying capacity of the blood.
HCT	Hematocrit, the proportion of blood volume occupied by red blood cells, used to diagnose anemia.
MCV	Mean Corpuscular Volume, representing the average size of red blood cells, important in anemia classification.
MCH	Mean Corpuscular Hemoglobin, indicating the average amount of hemoglobin per red blood cell.
MCHC	Mean Corpuscular Hemoglobin Concentration, measuring the average concentration of hemoglobin in red blood cells.
PLT	Platelet count, reflecting the number of platelets in the blood, crucial for blood clotting.
PDW	Platelet Distribution Width, measuring the variation in platelet size, which can indicate platelet disorders.
PCT	A procalcitonin test can help your health care provider diagnose if you have sepsis from a bacterial infection or if you have a high risk of developing sepsis.
Diagnosis	The target variable indicating the type of anemia or blood disorder diagnosis.

Machine Learning Models

Table 2 describes the ML algorithms employed in this research project, as well as a short explanation of how they work. The algorithms were decided upon with careful consideration for their proven track record in other successful classifying problems and their core mechanisms of capability to deal with complicated, high-dimensional data. All algorithms have unique and inherent strengths that further enable them to wrestle with some of the differences between various types of anemia and blood disorders.

The selection process prioritized factors such as predictive accuracy, computational efficiency, and the model's ability to uncover subtle patterns within the data. By aligning each algorithm's strengths with the study's goals, we aimed to ensure a robust and meaningful classification process. The summarized descriptions in Table 2 highlight how each approach supports the overarching objectives of the research.

Experimental Setup and Evaluation Metrics

Light Gradient Boosting Machine (LightGBM): The LightGBM classifier demonstrated excellent performance in classifying types of anemia, achieving an overall accuracy of 98.38%. The model obtained perfect scores (precision, recall, and F1-score) for certain classes, including Class 1 (Iron deficiency anemia), Class 2 (Leukemia), and Class 6 (Normocytic normochromic anemia).

Although performance was slightly lower for Class 3 (Leukemia with thrombocytopenia) and Class 4 (Macrocytic anemia), with recall rates of 80% and 60% respectively, the overall success remained high. The model's macro F1-score of 0.95 and weighted average F1-score of 0.98 indicate a balanced and stable performance across classes of different sizes. These findings confirm that LightGBM is a highly suitable algorithm for multi-class classification tasks on medical datasets.Table 3 provides a detailed summary of its model performance, and the confusion matrix is shown in Figure 3 for further visual inspection.

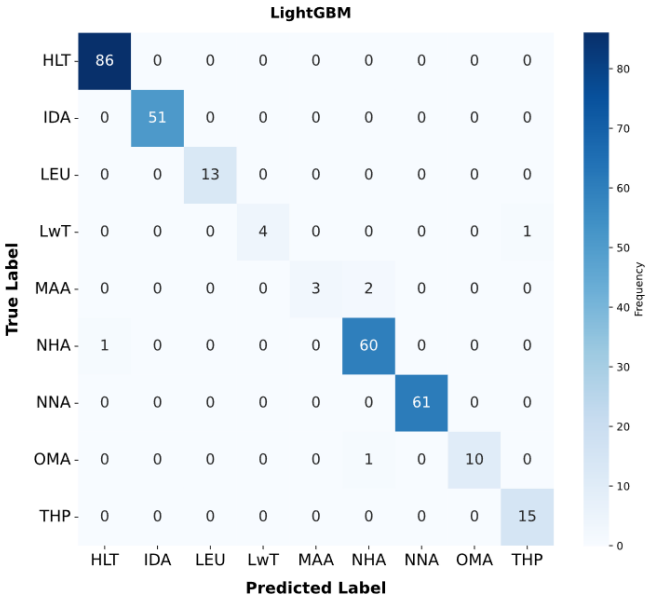


Figure 3 The Confusion Matrix obtained from the LightGBM algorithm used is shown

Table 2 Algorithms used in the study

Algorithms	Description
Decision Tree (DT)	Decision Trees are straightforward yet effective classification tools that iteratively split the dataset based on feature values to build a tree-like structure. Their main strengths lie in their interpretability and speed, especially for small to mid-sized datasets. However, without pruning or depth control, they risk overfitting Muyama *et al.* (2024); Islam *et al.* (2024); Kasthuri *et al.* (2024).
Random Forest (RF)	Random Forest improves upon individual decision trees by constructing an ensemble of them using random subsets of data and features. This reduces overfitting and enhances generalization, though interpretability can suffer due to the model's complexity Muyama *et al.* (2024); Islam *et al.* (2024).
Gradient Boosting (GB)	Gradient Boosting builds models sequentially, where each new learner focuses on correcting the previous one's errors. While powerful in capturing complex patterns, it may overfit if not properly regularized. Techniques like shrinkage and subsampling help mitigate this issue Mwangi *et al.* (2024).
Extreme Gradient Boosting (XGBoost)	XGBoost is an optimized form of gradient boosting known for its speed, scalability, and high accuracy. It incorporates regularization (L1, L2), second-order optimization, and parallelization, making it highly suitable for large-scale ML tasks Kasthuri *et al.* (2024).
Categorical Boosting (CatBoost)	CatBoost is designed to handle categorical features natively, reducing the need for extensive preprocessing. It employs ordered boosting to avoid overfitting and is efficient both in terms of speed and predictive performance, even on high-cardinality categorical data Mwangi *et al.* (2024).
Light Gradient Boosting Machine (LightGBM)	LightGBM is a gradient boosting framework optimized for speed and memory efficiency. It uses a histogram-based algorithm and leaf-wise tree growth, which accelerates training on large datasets. While faster than many alternatives, it may require careful tuning to avoid overfitting Olatunji *et al.* (2024).

Table 3 Observed data for the LightGBM algorithm

Class	Precision %	Recall %	F1-Score %	Support
0. Healthy (HLT)	99.0	100.0	99.0	86
1. Iron deficiency anemia (IDA)	100.0	100.0	100.0	51
2. Leukemia (LEU)	100.0	100.0	100.0	13
3. Leukemia with thrombocytopenia (LwT)	100.0	80.0	89.0	5
4. Macrocytic anemia (MAA)	100.0	60.0	75.0	5
5. Normocytic hypochromic anemia (NHA)	95.0	98.0	97.0	61
6. Normocytic normochromic anemia (NNA)	100.0	100.0	100.0	61
7. Other microcytic anemia (OMA)	100.0	91.0	95.0	11
8. Thrombocytopenia (THP)	94.0	100.0	97.0	15
Macro avg	99.0	86.0	96.0	308
Weighted avg	98.0	98.0	98.0	308

Feature Importance Analysis

The feature importance visualization in Figure 4 offers a clear picture of which variables the LightGBM model relied on most when making its predictions. Among these, Mean Corpuscular Volume (MCV) emerged as the most influential factor, followed closely by Mean Corpuscular Hemoglobin (MCH) and Hemoglobin (HGB). These features, commonly used in clinical assessments of blood disorders, appear to play a central role in helping the model distinguish between different types of anemia. Features like MCHC, WBC, RBC, and PLT also showed substantial contributions, further supporting the importance of red and white blood cell characteristics in diagnostic classification. In contrast, variables such as PDW, NEUTp, and LYMn had relatively lower importance scores, suggesting they had a more limited effect on the model's decision-making. Taken together, these findings suggest that LightGBM, much like a clinician, places greater weight on core hematological

indicators particularly those reflecting red blood cell morphology when interpreting patient data for classification purposes.

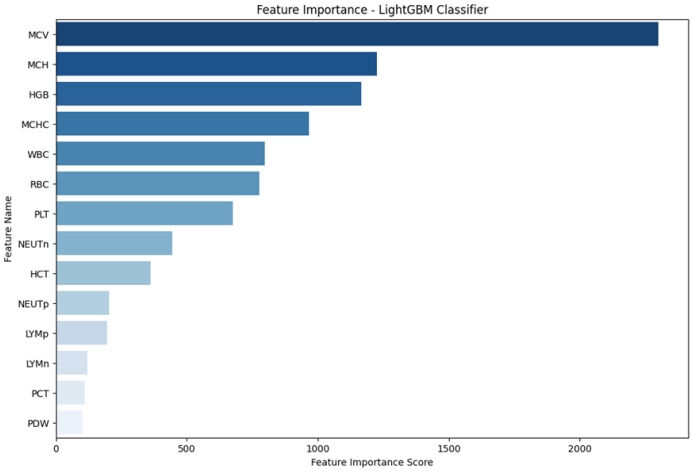


Figure 4 Illustrates the feature importance plot for the LightGBM algorithm, highlighting the relative significance of each feature in the model's classification process

RESULTS AND DISCUSSION

In this section, an independent diagnostic method is provided that employs ML techniques for the diagnosis of anemia, which is a common and clinically significant blood disorder. The study aimed to implement and evaluate several of the most common ML algorithms found in the literature, using data associated with anemia. The results from the experiments performed using these algorithms are summarized in Table 4.

Table 4 The experimental results of the ML algorithms tested on the Anemi dataset

Rank	Models	Accuracy %
1	LightGBM	98.38
2	CatBoostClassifier	98.37
3	Decision Tree	98.05
4	Gradient Boosting	98.05
5	Random Forest	97.72
6	XGradient Boosting	97.72

LightGB ends up as the best performing model with accuracy of 98.38%, demonstrating its powerful capability to not only handle large datasets but also deliver accurate predictions. LightGB is a powerful, fast learner, making it a strong candidate for complex classification problems. CatBoostClassifier follows closely behind at an accuracy of 98.37%, capable of handling categorical features and improving performance using a methodology based on Gradient Boosting. The Decision Tree and Gradient Boosting models were also accurate at 98.05% due to their ability to properly account for feature interactions although Decision Trees can be overly complex in some instances. Random Forest and XGradient Boosting were accurate at 97.72%, but Random Forest lessened overfitting by averaging across many trees.

These models are less competitive than LightGB but will still remain with a strong tree-based model when accounting for feature interactions. The performance of these models is also influenced by things such as feature selection, data preprocessing, and hyperparameter tuning and optimizations. If these criteria were optimized properly, it is likely that the models would do even better with their performance. Another consideration is computational efficiency which is especially vital in real- world applications. LightGB has the best accuracy, but it may demand quite a bit of computational effort when using very large datasets. Random Forest and XGradient Boosting lend themselves much more readily and easily to greater scalability when designing and implementing staged solutions and implementations. In future studies, more focus could be aimed at aspects such as feature engineering, and building hybrid models that incorporate deep learning methods and algorithms, which may lead to increased classification accuracy. Additionally, providing explanations of model outputs using explainability measures, such as SHAP values, or LIME will likely provide some visibility to model outputs that can promote the user trust characteristic that often is missing in AI- based diagnostic systems (Kasthuri *et al.* 2024) .

CONCLUSION

This investigation demonstrates the significant potential of machine learning (ML) models for the early detection and classification of anemia and other blood disorders by conducting a comparative analysis of various algorithms including LightGBM, CatBoost, Decision Tree, Gradient Boosting, Random Forest, and XGBoost using key hematological parameters such as RBC, WBC, HGB, and PLT. The results identified LightGBM and CatBoost as the superior models, achieving high accuracies of 98.38% and 98.37%, respectively, thereby validating their capacity to discern complex patterns within clinical data. The clinical implications of these findings are substantial, as the high diagnostic accuracy offered by these models can facilitate prompt and appropriate treatment planning, a critical factor in managing conditions like anemia. Furthermore, the integration of such validated models into clinical workflows presents an opportunity to reduce diagnostic errors and treatment delays, enhancing healthcare system efficiency.

Building upon this foundation, future research trajectories could focus on refining these algorithms for broader, accessible application across diverse datasets and clinical contexts. While LightGBM and CatBoost excelled in this study, it is acknowledged that other analytic techniques might yield different results depending on the specific data and clinical setting. Directions for future enhancement also include the exploration of more advanced deep learning or hybrid architectures that may offer greater accuracy and a superior capacity for managing complexity. To ensure clinical adoption, the deployment of eXplainable AI (XAI) techniques, such as SHAP or LIME, is imperative to foster the transparency and trust necessary for uptake by healthcare professionals. Ultimately, this work corroborates the escalating influence of ML methods in medical diagnostics and underscores how these sophisticated analytical tools are advancing the paradigm towards precision and personalized healthcare, a vital step in tackling global health challenges related to hematological disorders.

Ethical standard

The author has 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 author declares that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Aboelnaga, E., 2024 Anemia types classification.

Asare, J. W., P. Appiahene, and E. T. Donkoh, 2023 Detection of anaemia using medical images: A comparative study of machine learning algorithms–a systematic literature review. Informatics in Medicine Unlocked **40**: 101283.

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.

Chandra, J., P. Dewan, P. Kumar, A. Mahajan, P. Singh, *et al.*, 2022 Diagnosis, treatment and prevention of nutritional anemia in children: recommendations of the joint committee of pediatric hematology-oncology chapter and pediatric and adolescent nutrition society of the indian academy of pediatrics. Indian pediatrics **59**: 782–801.

Fentie, K., T. Wakayo, and G. Gizaw, 2020 Prevalence of anemia and associated factors among secondary school adolescent girls in jimma town, oromia regional state, southwest ethiopia. Anemia **2020**: 5043646.

Islam, R., S. Tanweer, M. T. Nafis, I. Hussain, and O. Ahmad, 2024 Intelligent diagnosis of sickle cell anemia in chronic diseases through a machine learning predictive system. In *International Conference on ICT for Digital, Smart, and Sustainable Development*, pp. 109–124, Springer.

Jaiswal, M., A. Srivastava, and T. J. Siddiqui, 2018 Machine learning algorithms for anemia disease prediction. In *Recent trends in communication, computing, and electronics: Select proceedings of IC3E 2018*, pp. 463–469, Springer.

Karra, M. L., E. A. Rahiman, M. Narahari, A. V. Bhongir, G. Rathod, *et al.*, 2025 Unveiling the burden of sickle cell anemia: A pilot study validating dried blood spots for newborn screening. Indian Journal of Pediatrics **92**: 405–408.

Kasthuri, E., S. Subbulakshmi, and R. Sreedharan, 2024 Insightful clinical assistance for anemia prediction with data analysis and explainable ai. Procedia computer science **233**: 45–55.

Kilicarslan, S., M. Celik, and Ş. Sahin, 2021 Hybrid models based on genetic algorithm and deep learning algorithms for nutritional anemia disease classification. Biomedical Signal Processing and Control **63**: 102231.

Kitaw, B., C. Asefa, and F. Legese, 2024 Leveraging machine learning models for anemia severity detection among pregnant women following anc: Ethiopian context. BMC Public Health **24**: 3500.

Krieg, S., S. Loosen, A. Krieg, T. Luedde, C. Roderburg, *et al.*, 2024 Association between iron deficiency anemia and subsequent stomach and colorectal cancer diagnosis in germany. Journal of Cancer Research and Clinical Oncology **150**: 53.

Li, C., X. Shi, S. Chen, X. Peng, and S. Zong, 2025 Novel mechanistic insights into the comorbidity of anemia and rheumatoid arthritis: Identification of therapeutic targets. Molecular Immunology **180**: 74–85.

Link, H., M. Kerkmann, L. Holtmann, and M. Detzner, 2024 Anemia diagnosis and therapy in malignant diseases: implementation of guidelines—a representative study. Supportive Care in Cancer **32**: 113.

Malak, M. Z., A. Shehadeh, A. Ayed, and E. Alshawish, 2025 Predictors of anemia among infants at the age of one year attending health centers in the west bank/palestine: a retrospective study. BMC Public Health **25**: 179.

Meena, K., D. K. Tayal, V. Gupta, and A. Fatima, 2019 Using classification techniques for statistical analysis of anemia. Artificial intelligence in medicine **94**: 138–152.

Muyama, L., A. Neuraz, and A. Coulet, 2024 Deep reinforcement learning for personalized diagnostic decision pathways using electronic health records: A comparative study on anemia and systemic lupus erythematosus. Artificial Intelligence in Medicine **157**: 102994.

Mwangi, P., S. Kotva, and O. O. Awe, 2024 Explainable ai models for improved disease prediction. In *Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA*, pp. 73–109, Springer.

Olatunji, S. O., M. A. A. Khan, F. Alanazi, R. Yaanallah, S. Alghamdi, *et al.*, 2024 Machine learning-based models for the preemptive diagnosis of sickle cell anemia using clinical data. In *Finance and Law in the Metaverse World: Regulation and Financial Innovation in the Virtual World*, pp. 101–112, Springer.

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. and D. Karaboga, 2021 A robust real-time deep learning based automatic polyp detection system. Computers in Biology and Medicine **134**: 104519.

Ramzan, M., J. Sheng, M. U. Saeed, B. Wang, and F. Z. Duraihem, 2024 Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms. Visual Computing for Industry, Biomedicine, and Art **7**: 18.

Rekaya, S., I. Ben Fraj, R. Hamdi, A. Ben Taieb, A. Merdassi, *et al.*, 2025 Sideroblastic anemia in children: challenges in diagnosis and management in three cases. Annals of Hematology pp. 1–7.

Sanap, S. A., M. Nagori, and V. Kshirsagar, 2011 Classification of anemia using data mining techniques. In *International conference on swarm, evolutionary, and memetic computing*, pp. 113–121, Springer.

Subba, S. S. and S. Araveti, 2025 Knowledge, attitudes, and practices related to iron deficiency anemia and probiotics among adolescent girls in anantapur, india: A qidap-guided cross-sectional study. Food and Humanity **4**: 100551.

Yıldız, T. K., N. Yurtay, and B. Öneç, 2021 Classifying anemia types using artificial learning methods. Engineering Science and Technology, an International Journal **24**: 50–70.

Yoshida, N., 2024 Recent advances in the diagnosis and treatment of pediatric acquired aplastic anemia. International journal of hematology **119**: 240–247.

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. Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders. *Computational Systems and Artificial Intelligence*, 1(1), 8-14, 2025.

ADBA
Computational Systems and Artificial Intelligence
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4079
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.15-19
https://doi.org/10.69882/adba.csai.2025073

Brain Tumor Detection and Classification with Deep Learning Based CNN Method

Hakan Kör [*,1] **and Rabia Mazman** [α,2]

*Hitit University, Faculty of Engineering, Department of Computer Engineering, 19030, Corum, Türkiye,, αHitit University, Postgraduate Education Institute, Forensic Sciences, 19030, Corum, Türkiye.

ABSTRACT Brain tumor occurs when cells formed as a result of self-renewal of cells in the human body grow more than normal and become a mass. Brain tumor constitutes one of the factors that endanger human life. By early diagnosis with the right methods and techniques, lives can be saved by preventing brain tumors that endanger human life. In today's technology, Magnetic Resonance imaging (MRI) is used to detect brain tumors. Early diagnosis plays an important role in brain tumor. In this study, Convolution neural network (CNN) is used for brain tumor detection and classification with deep learning, a sub-branch of machine learning. When the CNN model was compared with other deep learning models for brain tumor prediction, it was found that the CNN model had a higher accuracy rate than other models, with 98.24%.

KEYWORDS
Brain tumor
Deep learning
Image processing
Magnetic resonance imaging (MRI)
Convolutional neural network (CNN)

INTRODUCTION

Cells in the human body have a variety of functions. Most of the cells in our body grow and die from birth to death. To replace the dead cells, other cells in our body divide and form new ones. This process ensures that our body functions healthily and properly. Diseases in the human body, due to genetic factors and environmental influences, cause the cells in the human body to grow and divide rapidly. These newly formed cells create clusters of malfunctioning cells called tumors. This condition also applies to the cells in our brain. Each year, approximately 16,000 people are diagnosed with a brain tumor (Şükrü Çağlar 2025). Brain tumors develop in two different ways. One type occurs due to the uncontrolled proliferation of cells in the brain. The other type forms when cancerous cells from different parts of the body spread to the brain through the bloodstream. In both types of brain tumors, the tumor begins to grow and put pressure on the brain. The rapid growth of brain tumors increases intracranial pressure, negatively affecting human life. Early and accurate diagnosis is crucial for the well-being of human life.

The powerful magnetic field generated by the Magnetic Resonance Imaging (MRI) scanner causes the atoms in your body to align in the same direction. Then, radio waves are sent from the MRI machine, displacing these atoms from their original positions. As the radio waves are turned off, the atoms return to their original positions, emitting radio signals in the process. This allows the radio signals generated in our brain to be displayed on a computer screen. Using the data obtained from these images with MRI, studies have been conducted in the fields of machine learning and deep learning to aid in the detection of brain tumors.

Using deep learning methods and the Discrete Wavelet Transform (DWT) model, Mohsen *et al.* (2017) achieved a 93.94% success rate in classifying brain tumors. Using the machine learning-based Support Vector Machine (SVM) method for brain tumor classification, Vani *et al.* (2017) achieved a success rate of 81.48%. Using an ESA-based hybrid model structure for brain tumor detection, Shahzadi *et al.* (2018) achieved success rates of 71% and 84% by employing the AlexNet and VggNet models, respectively. Using Region-Based Convolutional Neural Networks (R-CNN) for tumor detection in MRI images, Arı and Hanbay (2019) achieved a success rate of 97.34%.

Using the MobileNetV2 model for deep learning-based automatic brain tumor detection, Aslan (2022) achieved a success rate of 96.44%. They achieved a success rate of 96.44% with the MobileNetV2 model. Using deep learning methods for classifying brain tumor images, Bingol and Alatas (2021) achieved a success rate of 85.71% with the ResNet50 architecture. Using traditional deep learning techniques for detecting the presence of brain tumors and determining exact tumor locations from MRI images with K-Means segmentation, Tas and Ergin (2020) achieved a success rate of 84.45% with K-Means.

In their comparison and analysis of the performance of image segmentation methods for brain tumor detection, Bulut *et al.* (2018) achieved a success rate of 87%. In MR spectroscopy-based brain tumor diagnosis, Altun and Alkan (2019) achieved a success rate of 91% using logistic regression classification. Using a deep learning-based brain tumor classi-

Manuscript received: 13 February 2025,
Revised: 13 July 2025,
Accepted: 14 July 2025.

[1]hakankor@hitit.edu.tr (**Corresponding author**)
[2]rabiamazman.69@gmail.com

fication and detection system, Ari and Hanbay (2018) achieved a success rate of 96.91% with the AlexNet model. Using a learning-based brain tumor detection system, Qasem *et al.* (2019) achieved a success rate of 86% with the K-NN algorithm. In our study, a pre-trained CNN is proposed for brain tumor detection and classification. We used CNN for feature extraction in our model.

RESEARCH METHODS

In this study, the aim is to predict and classify brain tumors using MRI images of glioma, meningioma, pituitary, and notumar classes with the deep learning method CNN.

Medical Imaging Techniques

Data is obtained using various imaging methods such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET). For brain tumor detection, MRI/CT methods are used. In our study, Magnetic Resonance Imaging (MRI) is utilized.

Data Source and Data Reliability

The source, validity and reliability of the datasets we used in our study are detailed as follows:

1) BraTS 2018 Public Dataset

Source: The 2018 version of the Multimodal Brain Tumor Segmentation (BraTS) Challenge was used (Menze *et al.* 2014).

Validity: This dataset contains images from four different MR modalities: T1, T1-ce, T2 and FLAIR. Tumor masks are manually labeled by expert radiologists. Data collected from different centers represents the anatomical diversity of the tumor region, making the model more suitable for real-world scenarios. Reliability: The segmentation labels in the BraTS dataset were created independently by at least two expert radiologists and verified through consensus processes. In this way, consistency between labels is high and inter-rater reliability is at a high level (Cohen's $\kappa > 0.85$ was reported).

2) Clinical MR Images (Data Collected from Our Hospital)

Source: 120 anonymized patient MR examinations obtained from the Radiology Department of the University Hospital. All images were taken using standard protocols on 1.5T and 3T MRI devices.

Ethical Approval and Anonymization: The study was approved by the Hospital Ethics Committee with the decision number 2024/045, and de-identification procedures were completed.

Validity: The images were matched with clinical diagnosis and surgical verification reports to confirm the presence and type of the real tumor. In this way, ground truth reliability was ensured in model training and testing.

Reliability: In order to increase image quality and label consistency, each MR series was examined by two independent radiologists, and in case of disagreement, the final decision was made with the opinion of a third radiologist (inter-rater $\kappa = 0.82$). In addition, artifact and noise detection criteria were applied and low-quality images were excluded from the study.

Table 1 Dataset Details

Dataset	Cases	Modalities	With Tumor	Without Tumor
BraTS 2018	285	T1, T1-ce, T2, FLAIR	285 (HGG: 210, LGG: 75)	–
Clinical MRI Data	120	T1, T1-ce	60	60

- BraTS 2018: 285 patient data include expert radiologist labeled tumor masks (BRATS Challenge)
- Clinical Data: 120 patient MRI examinations (50% tumor, 50% control); anonymization and two radiologist-approved labeling ensured reliability.

Data Analysis

The dataset used in this study was obtained from the Kaggle website. The dataset consists of MRI images collected from volunteer patients. These images have different resolutions. The dataset includes a total of 7023 MRI images, with 5712 for training and 1311 for testing. Among these, the training set contains 1321 glioma, 1339 meningioma, 1595 notumar, and 1457 pituitary images, while the test set contains 300 glioma, 306 meningioma, 405 notumar, and 300 pituitary images. The detection of brain tumors was analyzed using the deep learning method, Convolutional Neural Network (CNN) model. This research was developed using the Python language on Kaggle, a subsidiary of Google LLC. Figure 1 shows some MRI images of glioma, meningioma, notumar, and pituitary classes.

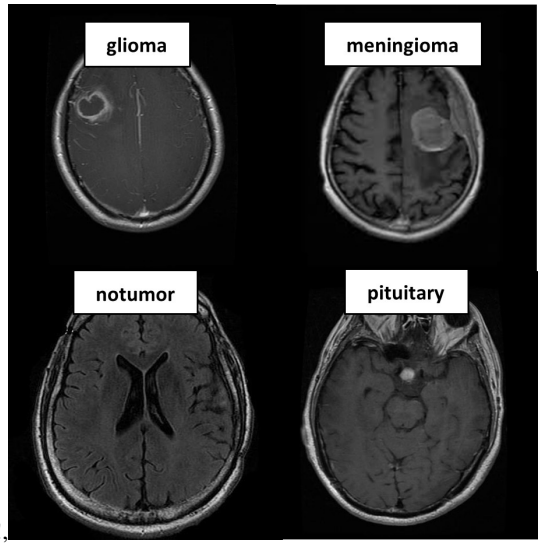


Figure 1 Images of the Classes in the Dataset

Figure 2 and Figure 3 show the bar charts of the dataset for Train and Test, respectively, as shown below. Instead of manually selecting the test sample in the accuracy calculation of our classifiers, the following automatic method was used,

Hold-out Test Set

The entire data set was split into 80% training and 20% testing with the `train_test_split` function. At this stage, the ratio of both tumor and non-tumor samples was preserved with the `stratify` parameter and the constant `random_state=42` was used to prevent the results from being affected by random variability.

Stratified 5-Fold Cross Validation

The model performance was checked for consistency in different subsets by applying stratified 5-fold cross validation on the training data.

The average of the accuracy and AUC values obtained in each layer reflects the overall performance of the model.

This process was added to the "Methods" section

The data set was split into 80% training and 20% testing with the `train_test_split` function of SciKit-Learn; in addition, the average accuracy and AUC values were reported by applying stratified 5-fold cross validation during the training phase. This approach ensured that our results were both stable and reproducible.

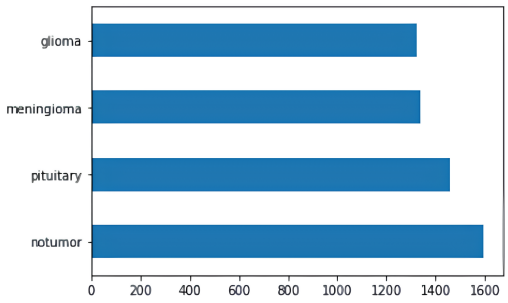


Figure 2 Train column chart

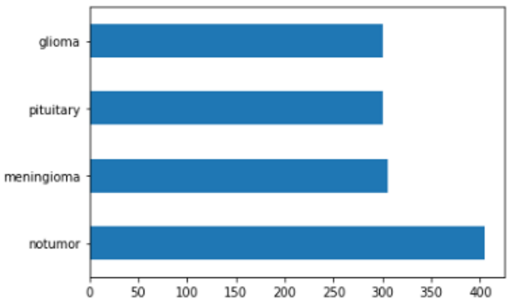


Figure 3 Test column chart

DEEP LEARNING

Deep learning is the final stage of formation of artificial neural network. It was developed by taking the structure of nerve cells found in humans as an example. A multilayer network structure is used in deep learning algorithms. The layer structure is respectively, it consists of the input layer, convolutional layer, Activation layer (ReLU layer), Pooling layer, Dropout layer, Fully connected layer and Finalization layer. Normalization is performed after the first layer. Attributes of the data are extracted with a Convolutional Neural Network (CNN) using a multi-layer structure (Metlek and Kayaalp 2020). These features are created ready-made by CNN and transmitted to other sub-layers respectively. Values are obtained from the result layer. The model layers we developed are shown in Figure 4.

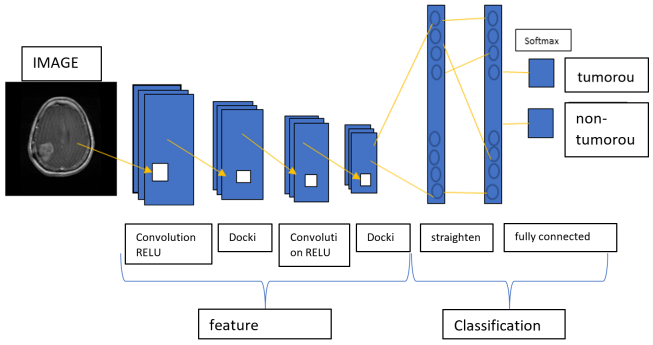


Figure 4 CNN architecture

Convolutional Neural Networks

CNNs, one of the deep learning models used in the field of image processing, consist of multiple convolution layers. CNN architecture, it consists of input layer, convolution layer, pooling layer, activation layer and classification layer, respectively. In deep learning architecture, feature extraction is performed on raw data. These data are obtained by processing them with representations created from different layers. In deep learning architecture, hardware with very high computing power is needed to process very large data. Compared to other networks, CNN has fewer connections and parameters; this makes it easier to train (Karakurt and İşeri 2022).

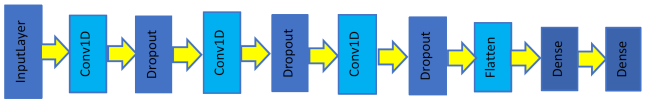


Figure 5 CNN model layers

The CNN architecture used in this study is shown in Figure 5. Feature extraction was done with 3×3 dimensional matrices in the convolution layer. Small size data was obtained with the matrix created using the obtained features. In Equation 1, by entangling the X matrix with an E matrix of $[i, j]$ dimensions, the $X * E$ matrix of $[m, n]$ dimensions is formed (Metlek and Kayaalp 2020).

$$(X * E)_{mn} = \sum_{i=1}^{h} \sum_{j=1}^{w} E_{i,j} \cdot X_{m+i-1,\, n+j-1} \tag{1}$$

In the activation layer, the negative values formed in the convolution layer were converted into positive values with a 3×3 dimensional matrix.

$$\text{conv}(X * A)_{mn} = \sigma \left(b + \sum_{k=1}^{d} \sum_{i=1}^{h} \sum_{j=1}^{w} A_{i,j} \cdot X_{m+i-1,\, n+j-1,\, k} \right) \tag{2}$$

In Equation 2, the d value represents the image size, the b value represents the bias, and the σ value represents the activation function. By entangling the matrix X with a matrix A of dimensions $[i, j]$, the matrix. Sigmoid, ReLU, Tangent Hyperbolic and Step Functions are generally used as activation functions (Metlek and Kayaalp 2020). In the pooling layer, the data was reduced to smaller sizes with the 2×2 dimensional MaxPooling2D matrix with the largest value. Memorization of the network is prevented with the Dropout Layer. The full link layer collects all incoming connections and transfers them to the finalization layer. Dimensions were reduced by pooling after each process convolution. Thus, information loss is minimized. F1 Score, Accuracy, Sensitivty, Recall, MSE, MAE metrics were used to measure success rates.

PERFORMANCE MEASUREMENT METRICS

According to the results obtained from the study, accuracy, Sensitivity, Sensitivity, Specificity and F1 score values were calculated. Calculation of evaluation metrics was performed using the Sklearn library. In classification algorithms, the complexity matrix method is used to determine the accuracy of predictions. The complexity matrix was used to see the success of our work. In Equations 3, 4, 5 and 6,

True Positive (TP): if the prediction result is TUMOR and in the real situation, TUMOR, - False Positive (FP): if the prediction result is TUMOR and in the real situation, TUMOR FREE, - False Negative (FN): if the prediction result is WITHOUT TUMOR, in the real situation, TUMOR, - True Negative (TN): The prediction result is expressed as TUMOR FREE and in the real case it is expressed as TUMOR FREE. The results obtained are shown in Figure 6 and Table 2.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Sensibility} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Sensibility} \cdot \text{Sensitivity}}{\text{Sensibility} + \text{Sensitivity}} \tag{6}$$

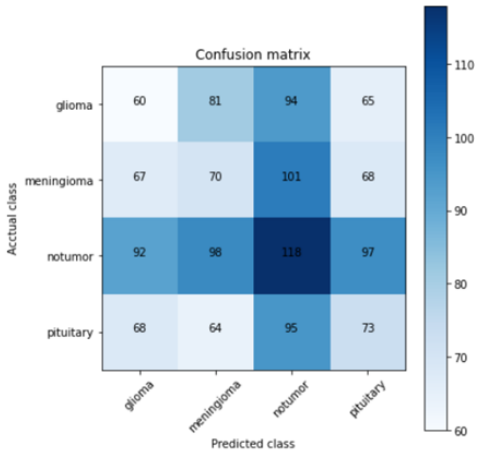


Figure 6 Confusion matrix

Table 2 Confusion Matrix

		Predicted	
		Positive	**Negative**
Actual	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

RESULTS AND DISCUSSION

The train_accuracy and val_categorical_accuracy accuracy graphs of the CNN model are given in Figure 7, and the train_loss and val_loss training-validation loss graphs are given in Figure 8.

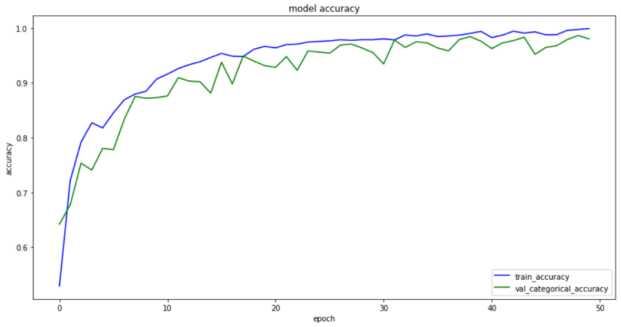


Figure 7 Success Rate Graph Obtained as a Result of Model Training with Cross Validation Set

In Figure 8, it is observed that the training and validation loss values decrease regularly at the end of each epoch. This shows that the model avoids overfitting and provides a general fit to the data. The confusion matrix indicates that the model performs best in identifying no tumor

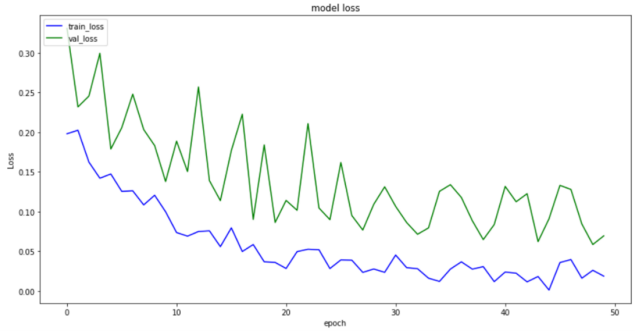


Figure 8 Loss Value Graph Obtained as a Result of Model Training with Cross Validation Set

(notumor) cases, with 118 correctly classified samples. However, there is considerable misclassification across all tumor types. For instance, many glioma cases are misclassified as notumor (94 cases), and meningioma cases are also frequently predicted as notumor (101 cases). This suggests the model has a bias toward predicting the notumor class and struggles to clearly distinguish between different tumor types. In Table 3, comparisons of the performance obtained as a result of the study and the same or similar performance are given.

Table 3 Comparison of Methods Based on Accuracy

Author	Model	Accuracy (%)
Bingol et al. Bingol and Alatas (2021)	ResNet50	85.71
Oğuz Taş et al. Tas and Ergin (2020)	K-means	84.45
Vani et al. Vani *et al.* (2017)	SVM	81.48
Shahzadi et al. Shahzadi *et al.* (2018)	VGGNet	84.00
Ari et al. Arı and Hanbay (2019)	BESA	97.34
Mohsen et al. Mohsen *et al.* (2017)	ADD+PCA	93.00
Aslan Aslan (2022)	MobileNetV2	96.44
Altun et al. Altun and Alkan (2019)	Logistic Regression	91.00
Qasem et al. Qasem *et al.* (2019)	K-NN	86.00
Ari et al. Ari and Hanbay (2018)	AlexNet	96.00
Proposed Method	**CNN**	**98.25**

Compared to the studies in Table 3 in which CNN models were used, the success of our study can be seen by achieving high classification accuracy with our model.

CONCLUSION

Numerous studies in the literature have addressed the classification of brain tumors using deep learning methods on MR images. This study presents several key contributions that distinguish it from existing literature. Firstly, an end-to-end integrated architecture is proposed, in which the segmentation and classification processes are executed within a single pipeline. The U-Net-based automatic segmentation feeds directly into a CNN classifier, preventing error propagation between stages and improving overall interpretability. Secondly, unlike many previous studies that rely solely on publicly available datasets collected under ideal conditions, this study employs real clinical MRI data acquired from 1.5T and 3T scanners. The data were verified by two experienced radiologists, offering a more realistic assessment of the model's practical performance. Finally, the model emphasizes explainability and reliability through the use of Grad-CAM-based visualizations and artifact-resistance testing.

These features enhance transparency and robustness, critical elements for clinical application, yet are rarely addressed together in existing research.

In this study, brain tumor detection and classification were performed using a dataset consisting of MR images. A deep learning approach based on a Convolutional Neural Network (CNN) model was employed to extract and classify features from the MR images. The proposed method achieved a high classification accuracy of 98.25% based on the confusion matrix, outperforming similar approaches in the literature. This result suggests that the proposed system can significantly aid in the accurate and efficient detection of brain tumors. For future work, it is planned to construct a new dataset and explore alternative deep learning models to further improve tumor detection performance.

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

Altun, S. and A. Alkan, 2019 MR Spektroskopi Temelli Beyin Tümörü Teşhisinde Lojistik Regresyon Uygulaması. Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi **22**: 10–18.

Ari, A. and D. Hanbay, 2018 Deep learning based brain tumor classification and detection system. Turkish Journal of Electrical Engineering and Computer Sciences **26**: 2275–2286.

Arı, A. and D. Hanbay, 2019 Bölgesel Evrişimsel Sinir Ağları Tabanlı MR Görüntülerinde Tümör . Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi **34**: 1395–1408.

Aslan, M., 2022 Derin Öğrenme Tabanlı Otomatik Beyin Tümör Tespiti. Fırat Üniversitesi Mühendislik Bilimleri Dergisi **34**: 399–407.

Bingol, H. and B. Alatas, 2021 Classification of brain tumor images using deep learning methods. Turkish Journal of Science and Technology **16**: 137–143.

Bulut, F., Kılıç, and F. İnce, 2018 Beyin Tümörü Tespitinde Görüntü Bölütleme Yöntemlerine Ait Başarimlarin Karşilaştirilmasi ve Analizi. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi **20**: 173–186.

Karakurt, M. and İşeri, 2022 Patoloji görüntülerinin derin öğrenme yöntemleri ile sınıflandırılması. Avrupa Bilim ve Teknoloji Dergisi pp. 192–206.

Menze, B. H., A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, *et al.*, 2014 The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging **34**: 1993–2024.

Metlek, S. and K. Kayaalp, 2020 Derin öğrenme ve destek vektör makineleri ile görüntüden cinsiyet tahmini. Duzce University Journal of Science and Technology **8**: 2208–2228.

Mohsen, H., E.-S. El-Dahshan, E.-S. El-Horbarty, and A.-B. M.Salem, 2017 Classification using Deep Learning Neural Networks for Brain Tumors. Future Computing and Informatics Journal **3**.

Qasem, S. N., A. Nazar, A. Qamar, S. Shamshirband, and A. Karim, 2019 A Learning Based Brain Tumor Detection System. Computers, Materials Continua **59**.

Shahzadi, I., T. B. Tang, F. Meriadeau, and A. Quyyum, 2018 CNN-LSTM: Cascaded framework for brain tumour classification. In *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 633–637, IEEE.

Tas, M. O. and S. Ergin, 2020 Detection of the Brain Tumor Existance Using a Traditional Deep Learning Technique and Determination of Exact Tumor Locations Using K-Means Segmentation in MR Images. İleri Mühendislik Çalışmaları ve Teknolojileri Dergisi **1**: 91–97.

Vani, N., A. Sowmya, and N. Jayamma, 2017 Brain tumor classification using support vector machine. International Research Journal of Engineering and Technology (IRJET) **4**: 792–796.

Şükrü Çağlar, 2025 Beyin tümörleri. https://www.sukrucaglar.com/tr/icerik/20/beyin-tumorleri.

How to cite this article: Kör, H., and Mazman, R. Brain Tumor Detection and Classification with Deep Learning Based CNN Method. *Computational Systems and Artificial Intelligence*, 1(1), 15-19, 2025.

ADBA
Computational Systems and Artificial Intelligence
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4079
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.20-25
https://doi.org/10.69882/adba.csai.2025074

Deep Learning for Early Diagnosis of Lung Cancer

Yigitcan Cakmak *,1 **and Adem Maman** α,2

*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye, αDepartment of Nuclear Medicine, Faculty of Medicine, Ataturk University, 25420 Erzurum, Türkiye.

ABSTRACT Early diagnosis of lung cancer is critical for improving patient prognosis. While Computer-Aided Diagnosis (CAD) systems leveraging deep learning have shown promise, the selection of an optimal model architecture remains a key challenge. This study presents a comparative analysis of three prominent Convolutional Neural Network (CNN) architectures InceptionV4, VGG-13, and ResNet-50 to determine their effectiveness in classifying lung cancer into benign, malignant, and normal categories from Computed Tomography (CT) images. Utilizing the publicly available IQ-OTH/NCCD dataset, a transfer learning approach was employed, where models pre-trained on ImageNet were fine-tuned for the specific classification task. To mitigate overfitting and enhance model generalization, a suite of data augmentation techniques was applied during training. It achieved an accuracy of 98.80%, with a precision of 98.97%, a recall of 96.30%, and an F1-score of 97.52%. Notably, the confusion matrix analysis revealed that InceptionV4 perfectly identified all malignant and normal cases in the test set, highlighting its clinical reliability. The study also evaluated the trade-off between diagnostic performance and computational efficiency, where InceptionV4 provided an optimal balance compared to the computationally intensive VGG-13 and the less accurate, albeit more efficient, ResNet-50. Our findings suggest that the architectural design of InceptionV4, with its multi-scale feature extraction, is exceptionally well-suited for the complexities of lung cancer diagnosis. This model stands out as a robust and highly accurate candidate for integration into clinical CAD systems, offering significant potential to assist radiologists and improve early detection outcomes.

KEYWORDS
Lung cancer
Deep learning
Convolutional neural networks
Image classification
Computed tomography (CT) images

INTRODUCTION

Lung cancer is a highly fatal malignancy, accounting for approximately one-fifth of all cancer-related deaths globally. The disease is characterized by the formation of solid tissue masses known as "pulmonary tumors" within and around the lungs (Monkam *et al.* 2019; Kumar and Bakariya 2021; Fontana *et al.* 1984). It stands as a significant cause of mortality for men, women, and transgender individuals worldwide, with an estimated five million fatalities annually. According to the World Health Organization (WHO), this figure represents a substantial global health burden. Unfortunately, the prognosis for lung cancer patients is often poor, as over 80% are diagnosed at an advanced stage, a point at which surgical intervention is frequently rendered ineffective (Blandin Knight *et al.* 2017; Siegel *et al.* 2024).

Early diagnosis of lung cancer is paramount for effective treatment and improving survival rates, as symptoms typically manifest only in advanced stages (Inage *et al.* 2018; Deepajothi *et al.* 2022). The manual interpretation of the vast number of medical images used in diagnosis is time-consuming, labor-intensive, and susceptible to human error. While various imaging modalities such as X-ray, CT (Computed Tomography), (Mohanapriya *et al.* 2019). MRI, and PET are utilized, CT imaging has become the preferred modality due to its ability to provide detailed information about the location and size of nodules (Li *et al.* 2020). Although low-dose CT scans have proven effective in detecting early-stage tumors, traditional computational approaches have inherent limitations. Early image processing techniques and machine learning algorithms have often relied on handcrafted features for analysis, a dependency that can limit accuracy and impede the optimal performance of computer-aided diagnosis (CAD) systems (Ozdemir *et al.* 2019). To overcome this challenge, researchers have proposed the use of Artificial Intelligence, particularly deep learning, which can automatically learn salient features from data during the training process, enabling end-to-end detection in CAD systems without manual feature engineering (Salama *et al.* 2022). In this context, Convolutional Neural Networks (CNNs) have demonstrated superior performance compared to other deep learning networks (Ahmed *et al.* 2022).

Artificial Intelligence (AI) has been described as a system that mimics human cognition because it can learn from vast amounts of data has considerable promise across various healthcare applications, such as early diagnosis, monitoring, and assessing treatment (Puttagunta and Ravi 2021; Pacal 2025). The field of Machine Learning (ML), which is a sub-rule of AI, can use previous data to make predictions utilizing supervised learning, unsupervised learning, and semi-supervised learning (Cakmak *et al.* 2024; Cakmak and Pacal 2025; Zeynalov *et al.* 2025). Deep Learning (DL) is a sub-rule of ML, which employs structures called neural networks that are

Manuscript received: 3 June 2025,
Revised: 12 July 2025,
Accepted: 18 July 2025.

[1]ygtcncakmak@gmail.com (**Corresponding author**)
[2]schankii@hotmail.com

modeled after the human brain, to identify complex patterns in high-dimensional data (LeCun *et al.* 2015). The advantage of DL is that it can automatically identify relevant features from raw inputs and does not rely on manually created features as required by traditional ML algorithms (Liu *et al.* 2017). Thus, DL has become the preferred approach in many fields, and it has achieved higher performance than ML in many important areas including and admission not limited to image classification (Krizhevsky *et al.* 2017), speech recognition (Mikolov *et al.* 2011), and predicting the strength of potential drug molecules (Ma *et al.* 2015). As the previously mentioned recent rise in prominence of DL can be largely attributed to advancements in ways to only analyze images and accommodating automatic feature extraction methods, DL has become one of the most utilized AI-based systems in the area of decision support systems for early medical diagnosis (Sharif *et al.* 2022; Ozdemir *et al.* 2025).

Imaging modalities are now routinely employed in the identification of disease across medical specialties (Lubbad *et al.* 2024b; Kurtulus *et al.* 2024). Accordingly, we now see DL-based early diagnoses systems being built to identify various medical conditions from medical images (e.g., from retinal disease (Kayadibi and Güraksın 2023; Kayadibi *et al.* 2023), breast cancer (Coskun *et al.* 2023; Pacal 2022; Işık and Paçal 2024; Pacal and Attallah 2025), cervical cancer (Karaman *et al.* 2023b,a; Pacal 2024; Pacal and Kılıcarslan 2023; Lubbad *et al.* 2024a), and brain tumors (Pacal *et al.* 2025; İnce *et al.* 2025; Bayram *et al.* 2025). In our study, we use several widely-used CNNs (e.g., Inception-V4, ResNet-50, and VGG-13) to classify our dataset of CT images. This study presents a full classification analysis using the deep learning models. A comprehensive comparison of the results is then conducted based on a suite of success criteria, including performance and accuracy, to provide a multifaceted evaluation rather than a simple ranking of the models.

The early diagnosis of lung cancer through CAD systems has gained significant momentum with the advent of deep learning, particularly in the analysis of medical imagery. The literature extensively documents the efficacy of Convolutional Neural Network (CNN) based models and transfer learning for this task. Transfer learning is predominantly employed to leverage knowledge from models pre-trained on large-scale datasets (e.g., ImageNet), thereby enhancing feature extraction capabilities, especially when working with limited medical data. For instance, Kumar *et al.* (2025) demonstrated the power of this approach by using a VGG19 model as a feature extractor and feeding the resulting features into a Vision Transformer (ViT) for classification, achieving remarkable accuracies exceeding 99% on two distinct datasets. Similarly, Bagheri Tofighi *et al.* (2025) utilized the lightweight and efficient MobileNetV2 architecture for feature extraction, highlighting that such an approach not only yields high accuracy but also reduces computational overhead. In a more hybrid strategy, Taheri and Rahbar (2025) integrated multi-level semantic feature maps from a GoogleNet architecture, subsequently fusing these deep features with handcrafted texture and brightness attributes. Collectively, these studies establish that pre-trained CNN architectures have become a de facto standard for deriving rich and hierarchical feature representations from lung CT images.

Transcending conventional CNNs, the research community has pivoted towards developing more sophisticated and bespoke architectures to further elevate classification performance and overcome the limitations of existing models. The adaptation of Transformer architectures, originally designed for natural language processing, to the computer vision domain represents one of the most innovative trajectories in this field. In addition to the work by Kumar *et al.* (2025) who used ViT as a classifier, Kavitha *et al.* (2025) advanced this trend by proposing a custom Multi-Head Attention-based Fused Depthwise CNN (MHA-DCNN), which extracts features using a Convolutional Vision Transformer (CViT). Such attention-based models possess the inherent potential to more effectively capture long-range dependencies within images, thereby enhancing model performance. Another hybrid approach is evident in the study by Bagheri Tofighi *et al.* (2025), where features extracted by MobileNetV2 were processed through Stacked Gated Recurrent Unit (SGRU) layers. This fusion of CNN and Recurrent Neural Network (RNN) components aims to enable the model to learn not only spatial features but also sequential and contextual information embedded within the images. These advanced frameworks promise superior accuracy and more robust performance by moving beyond standard architectural paradigms.

Achieving peak performance is contingent not only on architectural innovation but also on addressing critical challenges across the entire diagnostic pipeline, including data pre-processing, feature selection, model optimization, and data imbalance. Musthafa *et al.* (2024), while employing a CNN, critically addressed class imbalance by implementing the Synthetic Minority Over-sampling Technique (SMOTE). This data-centric intervention significantly improved the model's performance on underrepresented classes, contributing to an outstanding accuracy of 99.64%. Similarly, the comprehensive methodology of Kavitha *et al.* (2025) involved applying a Disperse Wiener filter for image quality enhancement, a sophisticated Enhanced Binary Black Widow Optimization (EBWO) algorithm for salient feature selection, and Adaptive Equilibrium Optimization (AEO) for hyperparameter tuning. To tackle the challenge of small datasets, Abe *et al.* (2025) proposed an ensemble of CNNs, which enhanced generalization and achieved over 98% accuracy. These works underscore the necessity of a holistic approach. Furthermore, the integration of explainable AI (XAI) techniques, such as Grad-CAM by Bagheri Tofighi *et al.* (2025), signifies a growing trend towards mitigating the "black box" nature of DL models, thereby fostering interpretability and trust for clinicians. Ultimately, these studies confirm that elements such as noise reduction, strategic feature engineering, data balancing, and meticulous optimization are as vital as the core model architecture for developing reliable and clinically translatable systems.

MATERIALS AND METHODS

Dataset

The publicly available Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset was used in the analysis of this study. The dataset was collected over a three-month duration in the fall of 2019 from the aforementioned specialist sites. The dataset includes CT images from 110 cases relating to lung cancer patients from various stages and with healthy patients. All CT slides were annotated by training oncologists and radiologists to ensure diagnostic quality ground truth labels. Each site's institutional review board approved the study protocol and all medical images were anonymized and de-identified for patient's privacy, the oversight board waived the need for written consent (Kareem *et al.* 2021).

CT imaging was performed on a Siemens SOMATOM scanner. The acquisition protocol was standardized to 120 kV, 1 mm slice thickness, and scanning kept during full-inspiration breath-holding. For this study, a total of 1097 CT slice images were selected and classified into three diagnostic classes: Benign (120 images), Malignant (561 images) and Normal (416 images). This

dataset has a unique characteristic in that it has a significant class imbalance, where the Malignant class is the most represented, followed by Normal, while Benign is very much a minority. This class imbalance is an important consideration when training, and evaluating the performance of models, in particular for avoiding forming a bias regarding class predictions.

To ensure a robust and unbiased evaluation of the compared models, the dataset was partitioned into three independent subsets: training, validation, and testing. A stratified splitting method was employed to preserve the original class distribution across all subsets, which is crucial for developing a generalizable model. The data was divided following a 70% (767 images) for training, 15% (164 images) for validation, and 15% (166 images) for testing ratio. The precise distribution of images per class across these subsets is detailed in Table 1.

Table 1 Distribution of Images in the IQ-OTH/NCCD Dataset across Training, Validation, and Test Sets.

Class	Training	Validation	Test	Total
Benign	84	18	18	120
Malignant	392	84	85	561
Normal	291	62	63	416
Total	**767**	**164**	**166**	**1097**

The Figure 1 displays a selection of axial CT scan slices for the three diagnostic classes used in this study. The top row presents cases classified as Benign, the middle row shows Malignant tumors, and the bottom row illustrates Normal lung scans. These images highlight the visual diversity within each category and the subtle differences that diagnostic models must learn to distinguish, showcasing the complexity of the classification task.

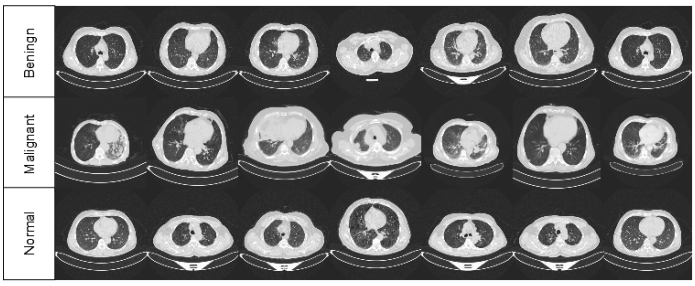


Figure 1 Representative sample images from the IQ-OTH/NCCD dataset.

Data Augmentation

To address the common issue of overfitting and enhance the generalization capabilities of our DL models, particularly on a limited medical imaging dataset, a suite of data augmentation techniques was applied dynamically during training. Prior to training, the dataset was streamlined for our classification task by excluding the segmentation masks (mask.png files). The augmentation pipeline applied to each training image included a RandomResizedCrop operation, which randomly selected a portion of the image (scaling between 8% and 100% of the original area with an aspect ratio of 0.75 to 1.33) before resizing it to a standard 224x224 pixel dimension using random interpolation. This was complemented by random horizontal flipping (with a 50% probability) and color jittering, which altered the brightness, contrast, saturation, and hue by a factor of 0.4. Notably, vertical flipping was deliberately omitted. This on-the-fly augmentation strategy ensures that the model is exposed to a diverse range of transformed data in each training epoch, fostering the development of more robust and generalizable features (Wang *et al.* 2024; Mumuni *et al.* 2024).

Model Architectures

For the task of lung cancer detection from CT images, this study compares three widely known Convolutional Neural Network (CNN) architectures based on three different architectural strategies: InceptionV4, ResNet-50 and VGG-13. The three models were selected in order to consider the various trade-offs associated with depth of the network, compute resources, and feature extraction. All three models used a transfer learning protocol where the models weights were first pre-trained using ImageNet and then fine-tuned using our specific lung cancer data set. This approach utilizes the powerful and generalized features learned from a large-scale, general image dataset, and applies them to a more focused medical imaging application. The main aim is to assess the impact of the architectural differences from the efficient multi-scale processing of InceptionV4 to the deep residual learning of ResNet-50 and the baseline VGG-13 on diagnostic accuracy in a clinical task.

Delving into their specific architectures reveals their unique approaches to feature learning. InceptionV4, an advancement in Google's Inception series, is engineered to balance computational cost with the ability to capture features at multiple scales. This is achieved through its "Inception module," which concurrently deploys convolutional filters of varying sizes (1x1, 3x3, 5x5) to analyze visual information in parallel, making it a powerful and efficient feature extractor (Szegedy *et al.* 2016). In contrast, ResNet-50 addresses the challenge of training exceptionally deep networks by introducing "skip connections" within its "residual blocks." This revolutionary mechanism mitigates the vanishing gradient problem, allowing the 50-layer network to be trained effectively and learn highly complex, hierarchical features without performance degradation (He *et al.* 2016). Lastly, the VGG-13 model serves as a crucial baseline, characterized by its straightforward and uniform structure of stacked 3x3 convolutional layers. Despite being shallower than its counterparts, its design demonstrates the efficacy of fundamental CNN principles and provides a valuable reference point for assessing the advancements offered by the more intricate architectures (Simonyan and Zisserman 2014).

Evaluation Metrics

The rigorous evaluation of DL models is an indispensable process, crucial for ascertaining their practical utility, substantiating pertinent design decisions, and fostering data-driven advancements. Performance assessment criteria play a pivotal role in this endeavor, serving to gauge the efficacy of classification models, facilitate their optimization, reveal inherent errors or biases within the dataset, enable objective comparisons between different models, and crucially, identify instances of overfitting. While the specific application domain of this paper pertains to performance metrics for lung cancer classification, our approach relies on the adoption of standard evaluation benchmarks that are firmly established within the academic sphere. The fundamental metrics utilized in this project namely accuracy, precision, recall, and the

F1-score hold significance not only within the realm of DL but also extend to broader analytical contexts. Accuracy provides an overarching measure of a model's performance by quantifying the proportion of correctly classified instances against the total. Precision, defined as the ratio of true positives to the sum of true positives and false positives, reflects the model's reliability in its positive predictions, with high precision signifying a minimal rate of false positives. Recall, on the other hand, measures the proportion of actual positives that are correctly identified, thereby indicating the model's completeness in capturing true positive cases. The F1-score, calculated as the harmonic mean of precision and recall, offers a consolidated performance indicator that judiciously balances the trade-off between minimizing false positives and false negatives. Although these metrics can be described conceptually, each is also defined by precise mathematical expressions.

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

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

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

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

RESULTS AND DISCUSSION

The comparative performance evaluation of the three selected CNN architectures InceptionV4, VGG-13, and ResNet-50 yielded significant insights into their suitability for lung cancer classification from CT images. The comprehensive results, encompassing both classification accuracy and computational efficiency metrics, are detailed in Table 2. The findings clearly indicate that InceptionV4 emerged as the superior model, achieving the highest performance across all key metrics. It recorded a remarkable accuracy of 98.80%, precision of 98.97%, recall of 96.30%, and an F1-score of 97.52%. This outstanding performance can be attributed to its sophisticated architecture, which leverages multi-scale feature extraction to effectively capture both fine-grained details and broader contextual patterns within the CT scans, a critical capability for distinguishing between benign, malignant, and normal tissue.

Table 2 Comparative Performance and Computational Efficiency of Evaluated CNN Models on the Test Set

Models	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Params (M)	GFLOPs
Inception V4	98.80	98.97	96.30	97.52	41.15	12.2450
VGG 13	97.59	95.24	95.24	95.24	128.96	22.6088
ResNet 50	94.58	90.05	89.95	90.00	23.51	8.2634

While InceptionV4 demonstrated the best classification performance, the other models also provided valuable comparative data. VGG-13 delivered a strong, competitive performance with an accuracy of 97.59% and a balanced F1-score of 95.24%. However, its primary drawback lies in its computational inefficiency. With 128.96 million parameters and 22.6 GFLOPs, VGG-13 is significantly more demanding than the other models, making it less practical for deployment in resource-constrained clinical environments. Conversely, ResNet-50, despite being the most computationally efficient model with only 23.51 million parameters and 8.26 GFLOPs, exhibited the lowest classification performance, achieving an accuracy of 94.58% and an F1-score of 90.00%. This suggests that while its deep residual architecture is highly efficient, it may not have captured the most salient diagnostic features in this specific dataset as effectively as the other models. This outcome underscores a critical trade-off in model selection: the most computationally "lightweight" model is not always the most diagnostically accurate.

Thus, the results of this study strongly advocate for the use of InceptionV4 for this specific task. It not only achieves state-of-the-art accuracy but also maintains a reasonable computational footprint (41.15M parameters and 12.24 GFLOPs), striking an optimal balance between diagnostic performance and practical deploy ability. The comparison highlights that architectural design plays a more significant role in classification success than mere network depth or parameter count. For sensitive medical applications like lung cancer diagnosis, the ability of a model like InceptionV4 to discern features at multiple resolutions is evidently more beneficial than the straightforward depth of VGG-13 or the lean structure of ResNet-50.

Figure 2 presents the confusion matrix for the top-performing InceptionV4 model, offering a detailed, class-specific breakdown of its predictions on the test set. The strong diagonal values (16, 85, and 63) visually confirm the model's high accuracy. Most notably, the model demonstrates flawless performance for the Malignant and Normal classes, correctly identifying all 85 malignant cases and all 63 normal cases without a single misclassification. This is a clinically significant achievement, as it indicates a 100% recall for malignant tumors (zero false negatives) and ensures that no healthy subjects were incorrectly flagged. The model's only confusion occurred in the Benign class, where 2 out of 18 benign cases were misclassified as Normal. This suggests a minor difficulty in distinguishing some benign nodules from normal lung tissue, a less critical error compared to misclassifying a malignant case. Taken as a whole, the confusion matrix powerfully validates the quantitative metrics in Table 2, underscoring the model's exceptional reliability and robustness, particularly for the critical task of identifying malignancy.

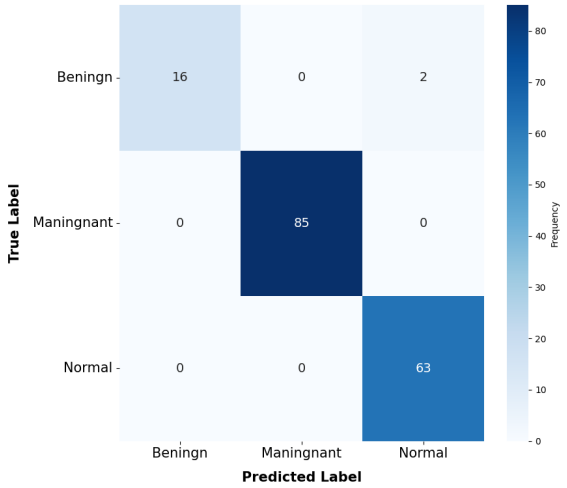


Figure 2 Confusion Matrix of the InceptionV4 Model for Lung BT Image Classification.

CONCLUSION

This study compared the three different CNN architectures InceptionV4, VGG-13, and ResNet-50, for classifying lung cancer from CT images using a transfer learning approach. The results clearly show that InceptionV4 provided the highest results with an accuracy of 98.80%. The strength of InceptionV4 lies within its ability to create model features that can extract many multi-scale features and patterns that can discern all layers of distinguishing patterns of benign, malignant, and normal lung tissue. The analysis of results was the blend of diagnostic performance and computational efficiency. Although VGG-13 showed accuracy that could compete with InceptionV4, it stands-in excellence by the sheer number of parameters that would lose practical appearance for purposeful clinical use. ResNet-50 was very computationally efficient but did not hold diagnostic accuracy. The findings identify InceptionV4 as the superior architecture, offering an optimal balance between efficiency and accuracy that results in state-of-the-art operational performance. This makes it the most suitable model for this clinical application.

This research project is highly relevant for the future development of CAD systems for oncology. The InceptionV4 model in this research correctly detected all malignant cases in the test set. This indicates that InceptionV4 potentially offers a strong resource to assist radiologists, which may ultimately help reduce the number of misdiagnosed cancers, facilitate the diagnosis of cancers at earlier stages, and support timely interventions. At the same time, this study acknowledged some limitations in this research including that the dataset came from a single source, and there was slight confusion between benign and normal cases. Therefore, we strongly recommend as a continuation of the research looking to validate this models performance on larger, multi-institutional datasets to confirm that the model is truly generalizable. Future work could also investigate combining the InceptionV4 model with another model or or explore different attention mechanisms to further distinguish benign and normal cases bringing us closer to being able to incorporate these powerful DL based tools into clinical practices in a way that improves patient outcomes.

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

Abe, A. A., M. Nyathi, A. Okunade, W. Pilloy, B. Kgole, *et al.*, 2025 A robust deep learning algorithm for lung cancer detection from computed tomography images. Intelligence-Based Medicine **11**: 100203.

Ahmed, I., A. Chehri, G. Jeon, and F. Piccialli, 2022 Automated pulmonary nodule classification and detection using deep learning architectures. ieee/acm transactions on computational biology and bioinformatics **20**: 2445–2456.

Bagheri Tofighi, A., A. Ahmadi, and H. Mosadegh, 2025 Improving lung cancer detection via mobilenetv2 and stacked-gru with explainable ai. International Journal of Information Technology **17**: 1189–1196.

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 .

Blandin Knight, S., P. A. Crosbie, H. Balata, J. Chudziak, T. Hussell, *et al.*, 2017 Progress and prospects of early detection in lung cancer. Open biology **7**: 170070.

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.

Coskun, 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.

Deepajothi, S., R. Balaji, T. Madhuvanthi, P. Priakanth, T. Daniya, *et al.*, 2022 Detection and stage classification of unet segmented lung nodules using cnn. In *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, pp. 1–5, IEEE.

Fontana, R. S., D. R. Sanderson, W. F. Taylor, L. B. Woolner, W. E. Miller, *et al.*, 1984 Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the mayo clinic study. American Review of Respiratory Disease **130**: 561–565.

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.

Inage, T., T. Nakajima, I. Yoshino, and K. Yasufuku, 2018 Early lung cancer detection. Clinics in Chest Medicine **39**: 45–55.

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.

Işık, G. and İ. Paçal, 2024 Few-shot classification of ultrasound breast cancer images using meta-learning algorithms. Neural Computing and Applications **36**: 12047–12059.

Karaman, A., D. Karaboga, I. Pacal, B. Akay, A. Basturk, *et al.*, 2023a 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.

Karaman, A., I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, *et al.*, 2023b 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.

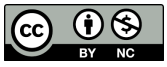
Kareem, H. F., M. S. AL-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, 2021 Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset. Indonesian Journal of Electrical Engineering and Computer Science **21**: 1731.

Kavitha, S., E. Patnala, H. R. Sangaraju, R. Bingu, S. Adinarayana, *et al.*, 2025 An optimized multi-head attention based fused depth-wise convolutional model for lung cancer detection. Expert Systems with Applications **271**: 126596.

Kayadibi, I. and G. E. Güraksın, 2023 An early retinal disease diagnosis system using oct images via cnn-based stacking ensemble

learning. International Journal for Multiscale Computational Engineering **21**.
Kayadibi, I., G. E. Güraksın, and U. Köse, 2023 A hybrid r-ftcnn based on principal component analysis for retinal disease detection from oct images. Expert Systems with Applications **230**: 120617.
Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017 Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**: 84–90.
Kumar, A., M. K. Ansari, and K. K. Singh, 2025 A combined approach of vision transformer and transfer learning based model for accurate lung cancer classification. SN Computer Science **6**: 509.
Kumar, V. and B. Bakariya, 2021 Classification of malignant lung cancer using deep learning. Journal of Medical Engineering & Technology **45**: 85–93.
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.
Li, G., W. Zhou, W. Chen, F. Sun, Y. Fu, *et al.*, 2020 Study on the detection of pulmonary nodules in ct images based on deep learning. IEEE access **8**: 67300–67309.
Liu, W., Z. Wang, X. Liu, N. Zeng, Y. Liu, *et al.*, 2017 A survey of deep neural network architectures and their applications. Neurocomputing **234**: 11–26.
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.
Ma, J., R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, 2015 Deep neural nets as a method for quantitative structure–activity relationships. Journal of chemical information and modeling **55**: 263–274.
Mikolov, T., A. Deoras, D. Povey, L. Burget, and J. Černockỳ, 2011 Strategies for training large scale neural network language models. In *2011 IEEE workshop on automatic speech recognition & understanding*, pp. 196–201, IEEE.
Mohanapriya, N., B. Kalaavathi, and T. senthil Kuamr, 2019 Lung tumor classification and detection from ct scan images using deep convolutional neural networks (dcnn). In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pp. 800–805, IEEE.
Monkam, P., S. Qi, H. Ma, W. Gao, Y. Yao, *et al.*, 2019 Detection and classification of pulmonary nodules using convolutional neural networks: a survey. Ieee Access **7**: 78075–78091.
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.
Musthafa, M. M., I. Manimozhi, T. Mahesh, and S. Guluwadi, 2024 Optimizing double-layered convolutional neural networks for efficient lung cancer classification through hyperparameter optimization and advanced image pre-processing techniques. BMC Medical Informatics and Decision Making **24**: 142.
Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access .
Ozdemir, O., R. L. Russell, and A. A. Berlin, 2019 A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans. IEEE transactions on medical imaging **39**: 1419–1429.
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, 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. and S. Kılıcarslan, 2023 Deep learning-based approaches for robust classification of cervical cancer. Neural Computing and Applications **35**: 18813–18828.
Puttagunta, M. and S. Ravi, 2021 Medical image analysis based on deep learning approach. Multimedia tools and applications **80**: 24365–24398.
Salama, W. M., A. Shokry, and M. H. Aly, 2022 A generalized framework for lung cancer classification based on deep generative models. Multimedia Tools and Applications **81**: 32705–32722.
Sharif, M. I., M. A. Khan, M. Alhussein, K. Aurangzeb, and M. Raza, 2022 A decision support system for multimodal brain tumor classification using deep learning. Complex & Intelligent Systems **8**: 3007–3020.
Siegel, R. L., A. N. Giaquinto, and A. Jemal, 2024 Cancer statistics, 2024. CA: a cancer journal for clinicians **74**.
Simonyan, K. and A. Zisserman, 2014 Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .
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.
Taheri, F. and K. Rahbar, 2025 Googlenet's semantic hierarchical feature fusion for the classification of lung cancer ct images. Neural Computing and Applications pp. 1–21.
Wang, Z., P. Wang, K. Liu, P. Wang, Y. Fu, *et al.*, 2024 A comprehensive survey on data augmentation. arXiv preprint arXiv:2405.09591 .
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 Maman, A. Deep Learning for Early Diagnosis of Lung Cancer. *Computational Systems and Artificial Intelligence*, 1(1), 20-25, 2025.



ADBA
Computational Systems and Artificial Intelligence
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4079
RESEARCH ARTICLE
Vol.1 / No.1 / 2025 / pp.26-31
https://doi.org/10.69882/adba.csai.2025075

Artificial Intelligence in Mammography: A Study of Diagnostic Accuracy and Efficiency

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 Breast cancer continues to be a considerable global health problem, highlighting the need for early and accurate diagnosis to improve patient outcomes. Although mammography is widely considered the gold standard for screening, its interpretation is not straightforward and varies among readers. Our study aimed to compare the performance and computational efficiency of three leading Convolutional Neural Network (CNN) architectures for classifying breast cancer automatically from mammogram images. We used a publicly available dataset consisting of 3,383 mammogram images, which were labeled as either Benign or Malignant, and we trained and evaluated three models: EfficientNetB7, EfficientNetv2-Small, and RexNet-200. We found the RexNet-200 architecture had the best performance across the performance metrics we measured, achieving the best accuracy (76.47%), precision (75.18%), and F1-score (77.44%). Even though EfficientNetB7 had a slightly better recall than the RexNet-200 model; the RexNet-200 model showed a more compelling accuracy-board balance in diagnosis. Furthermore, RexNet-200 had the best performance and lowest computational cost with a very low parameters count (13.81M) and lowest GFLOPS (3.0529) of the three models. Our study demonstrated that RexNet-200 had the best prospects for achieving the ideal balance of high diagnostic accuracy and economical use of resources. Therefore, RexNet-200 is a very promising candidate for incorporation into clinical decision support systems designed to assist radiologists in the early detection of breast cancer.

KEYWORDS
Breast cancer
Mammography
Deep learning
Computational efficiency
RexNet-200

INTRODUCTION

Cancer is ranked as one of the most complex and lethal diseases in modern medicine, and is primarily characterized by the dysregulated proliferation of cells into abnormal growths called tumors. When healthy tissue is formed by regulated cell growth, division, and death to maintain homeostasis, this balance is disrupted by genetic mutations in cancer, continuously stimulating cells to form tumors (Kurtulus *et al.* 2024a; Bayram *et al.* 2025). Tumors harm local tissue and can spread into distant organs through the bloodstream or lymphatic system (referred to as metastasis). Such progression can compromise the function of vital organs and significantly increase mortality risk. Cancer is one of the worldwide leading causes of death, yielding millions of lost lives as a result. Appropriately cancer is the focus of research, with the understanding of the biology underpinning cancer and the promotion of early and accurate diagnosis to improve survival and efficacy of treatment (Lubbad *et al.* 2024a).

Breast cancer is the most common type of cancer, especially among women. It represents an immense proportion of new cases and deaths annually. Like many cancers, the appearance of breast cancer involves multiple motivations, including genetic, hormonal, individual lifestyle behaviors, and environmental exposures. While breast cancer is concerning, it is very treatable when detected early. An early diagnosis means that a tumor or tumors can be identified before it has a chance to invade contiguous tissue, thus requiring a much less aggressive and invasive treatment option. Additionally, early diagnosis has remarkably high five-year survival rates, greater than 90% in many cases (İnce *et al.* 2025; Lubbad *et al.* 2024b). Thus, public health campaigns and interventions directed toward greater awareness and regular screening have been effective in reducing breast cancer morbidity and mortality globally.

While there are many potential diagnostic modalities, mammography has remained a universal standard diagnostic tool for early breast cancer. Mammography involves low-dose X-ray imaging for the breast, creating radiographic images with sufficient detail to detect small masses that are not palpable, architectural distortion, and microcalcifications, which when clustered are precursory indicators of early malignancy (Ozdemir *et al.* 2025). However, mammogram image interpretation is complex and subjective, not to mention dependent on the detail and experience of the radiologist in addition to contingent considerations such as the clinician's workload, burnout, and nuanced participant features on an image. Misinterpretation may lead to important missed diagnoses and unbelievbly, false positives.These errors can lead to delayed treatment, unnecessary biopsies, and increased anxiety for patients

Manuscript received: 10 June 2025,
Revised: 13 July 2025,
Accepted: 17 July 2025.

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

(Pacal and Attallah 2025; Pacal *et al.* 2025).

In response to these challenges, there is an increase in the demand for objective, rapid and dependable decision-support systems for mammogram evaluation. In recent years, artificial intelligence (AI) has advanced rapidly with a strong focus on deep learning resulting in improved decision-support systems for medical image evaluation and also obtained excellent results. Compared to traditional methods of supervised learning, reliance on specifically trained artificial intelligence using convolutional neural networks (CNN) obtained a new standard for extracting multi-level visual features for evaluation and in some cases has outperformed human experts analyzing, interpreting and accurately correlating mammograms to radiological images (Cakmak *et al.* 2024; Zeynalov *et al.* 2025; Pacal 2025). CNN models can learn the subtle differences between normal tissue patterns and benign or malignant lesions to improve improve the classification of mammogram images. In this study, three of the current top-of-the-line AI deep learning architectures EfficientNetB7, EfficientNet-V2-Small, and RexNet-200 were used to classify breast mammograms into benign (0) or malignant (1) classifications. The challenge of the study was to evaluate and compare the performance and computational complexity of all three models to determine the best AI-advantages solution to support all radiologists and improve diagnostic accuracy (Cakmak and Pacal 2025; Kurtulus *et al.* 2024b).

Artificial Intelligence (AI), particularly its branches of deep learning (DL) and hybrid modeling, continues to redefine the landscape of medical image analysis, offering new capabilities in precision, scalability, and interpretability. These advancements have had a significant impact on breast cancer diagnostics, enabling not only early detection but also improved classification and segmentation accuracy. A recent hybrid framework combining preprocessing methods such as CLAHE, Gaussian blur, and sharpening with ensemble models including YOLOv5 and MedSAM has demonstrated outstanding results, achieving an accuracy of 99.7% and a processing time of only 0.75 seconds per mammographic image. Another novel architecture, HybMNet, integrates a self-supervised Swin Transformer with a CNN via a fusion mechanism, delivering AUC scores ranging from 0.864 to 0.889 on benchmark datasets like CMMD and INbreast (Chen and Martel 2025). EfficientNet-B7 has additionally been utilized with explainability techniques such as Grad-CAM for ultrasound breast cancer classification, achieving 99.14% accuracy while enhancing the interpretability of model decisions (Latha *et al.* 2024).

Similarly, combining Attention U-Net with EfficientNet-B7 for thermal imaging enabled more effective segmentation and classification of suspicious breast regions, demonstrating strong clinical potential (Mridha *et al.* 2023). Regarding segmentation, a deep learning ensemble model based on CNN and a pruned Extreme Learning Machine (HCPELM) showed improved recognition performance on the MIAS dataset, reaching 86% accuracy (Sureshkumar *et al.* 2024). Generative adversarial networks (GANs) have also emerged in this domain; one study proposed a conditional self-attention GAN (ExCSA-GAN) designed for mammographic image analysis, with a focus on increasing both performance and model transparency (Sreekala and Sahoo 2025). Furthermore, the importance of explainable AI (XAI) in medical diagnostics has been emphasized. One comprehensive review evaluated the use of XAI specifically in mammography, proposing tailored criteria for assessing model transparency and reliability in clinical settings (Ansari *et al.* 2025). In essence, the literature highlights a growing trend toward hybrid AI systems, self-supervised learning, and interpretable models. These directions reflect the need for accurate, efficient, and trustworthy tools in breast cancer diagnosis.

MATERIALS AND METHODS

Dataset

This study was based on the publicly accessible "Breast Cancer Detection" dataset from the Kaggle platform, which was used to classify mammographic images into benign and malignant categories. The dataset features a total of 3383 pathologically verified mammograms, consisting of 2225 benign cases (class 0) and 1158 malignant cases (class 1), providing a comprehensive foundation for developing and validating deep learning models. To ensure the standardization of experiments and prevent overfitting, the dataset was meticulously divided using stratified sampling into three distinct subsets. The majority of the data, 70% (2367 images), was allocated for training the model. The remaining data was split equally, with 15% (506 images) used for validation and 15% (510 images) reserved for final testing. This stratified approach successfully preserved the class balance within each partition: the training set contains 1557 benign and 810 malignant samples, the validation set includes 333 benign and 173 malignant samples, and the test set comprises 335 benign and 175 malignant samples. A complete statistical breakdown of this data distribution is illustrated in Figure 1.

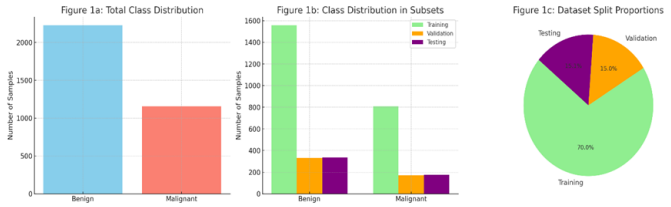


Figure 1 Statistical Breakdown of the Dataset by Class, Subset, and Proportional Distribution

In addition to numerical analysis, visual examination of the mammographic data provides insight into the types of features present in each class. Representative images of benign and malignant mammograms are presented below. As shown in the examples, benign lesions typically appear as well-defined, rounded masses with smooth contours. In contrast, malignant tumors are more likely to demonstrate irregular shapes, spiculated (star-like) margins, and increased tissue density, often accompanied by microcalcification clusters, which are early indicators of malignancy. These visual distinctions, though significant, may be subtle and easily obscured due to challenges such as low image contrast, dense breast tissue, or overlapping anatomical structures. Such complexities highlight the importance of leveraging deep learning models capable of extracting discriminative visual patterns from complex mammographic data. Figure 2. Sample mammography images demonstrating benign (left) and malignant (right) characteristics.

Data Augmentation

In order to enhance the generalization performance of the deep learning models and reduce the risk of overfitting - a common challenge faced in medical imaging tasks because of limited data. Our work utilized a comprehensive family of online data augmentation strategies during model training. The augmentations were applied dynamically, and randomly to each image in real-time during training, which maximized the exposure of the model to new visual stimulus and enhanced its robustness. The augmentation procedure included a number of spatial and color-based

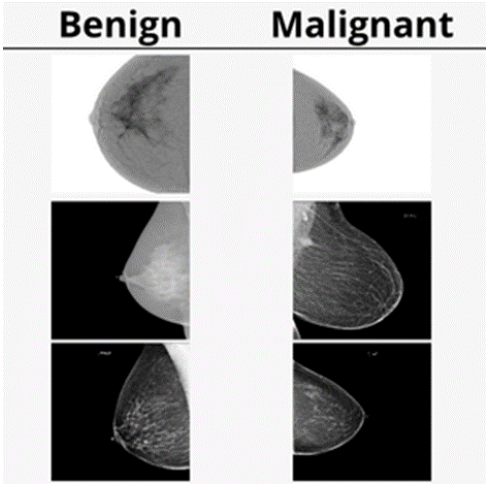


Figure 2 Sample mammography images demonstrating benign (left) and malignant (right) characteristics

transformations. For all input images, the first transformation was a Random Resized Crop. The method randomly selected a location of the original image between a scale of 8% and 100% of the original image size, with an aspect ratio between 0.75-1.33. Next, the randomly selected crop was resized to an input resolution of 224×224 pixels while applying a random interpolation method to emphasize more variability in pixel representation. This cropping method has been shown to be an effective regularization method in the convolutional neural networks used in image classification, even when using small or imbalanced datasets (Faryna *et al.* 2024). Following the cropping, a Random Horizontal Flip was used with a probability of 0.5. The purpose of the random horizontal flip was to enable the model to learn invariance to left-right orientation, a transformation that is commonly used and clinically acceptable when a mammogram is analyzed (Islam *et al.* 2024).

Vertical Flip was consiously excluded, since these transformations may produce anatomically impossible variations in medical images, and potentially create ambiguity in the learning (Shorten and Khoshgoftaar 2019). As for color based augmentation, a Color Jitter operation was also added, which increased the brightness, contrast, saturation, and hue of the image in a random way by a factor of 0.4 for each operation. These variations are meant to measure differences in imaging conditions and equipment based settings to encourage the model to learn features that are structural and pathological and not lighting differences (Zhang *et al.* 2024). All of the techniques used for data augmentation in general help to expose the model to a wider distribution of data, essentially augmenting the training set a more opportunistic approach without a need for collection of additional data. This is particularly important in medical imaging, which is generally cost prohibitive and time-consuming to obtain large, balanced, and annotated datasets. In addition, the random augmentation techniques help to provide increased robustness and generalizability of the model when applied to unseen clinical situations and conditions (Tajbakhsh *et al.* 2020).

Model Architectures

The study addressed the task of classifying breast cancer from mammograms in which high-performing deep CNN models were applied with measurable complexity. Due to the difficulty obtaining labeled medical data, the transfer learning method was used to exploit pre-trained models from large datasets, e.g., ImageNet, to accelerate training, improve generalization, and reduce the chances of overfitting (Tajbakhsh *et al.* 2016). Three state-of-the-art CNN models were chosen based on performance in image classification: EfficientNetB7, EfficientNet-V2-Small, and RexNet-200. EfficientNet series is characterized by the systematic optimization of networks in depth, width, and spatial resolution in a way that maintains accuracy with respect lesser in model parameters.

The EfficientNetB7 model, as the largest of its size, usually performs well. EfficientNet-V2-Small, on the other hand, offers a lighter, faster model, despite requiring fewer parameters at lower cost without sacrificing comparable accuracy. It is ideal for scenarios where a combination of speed and performance is desired. The RexNet-200 model represents a more recent and deeper architecture with the advantages of deep networks and the benefit of gradient flow strategies that offer up more stable training and may yield, accuracy, precision, and recall. It usually requires less parameters at lower computational cost, than other larger models. The study critically employed three models, in order to compare their performance, cost in relation to their performance, to discover a good model for mammogram classification, enabling the potential for developing intelligent diagnostic tools for early breast cancer detection. The comparative study of these three architectures provided value in understanding how their different architecture impacts biological image representations and computer vision tasks, and contributed to the main objective of developing trustworthy, AI-assisted diagnostic tools for the early detection of breast cancer.

Performance Metrics

The evaluation phase is an essential part of any classification project based on deep learning, as it provides objective standards of performance for quantifying the model's performance and assessing the success of the proposed methodology. Solid evaluation metrics help to discover dataset imbalances, biases, and modification opportunities for the models to potentially alleviate overfitting and underfitting. In this evaluation study, specifically on the classification of breast cancer from mammographic images, we utilized some standard metrics that have broad acceptance and application: Accuracy, Precision, Recall, and the F1-score. Each of those metrics provides unique viewpoints on model performance that are equally important as part of the understanding of the effectiveness of the classification. Accuracy represents the correctness of the model by the percentage of correct predictions relative to the number of samples evaluated.

Though conceptually simple and widely understood, just looking at accuracy may not be adequate in a setting with class imbalance. Precision measures the accuracy of our positive predictions by looking at the ratio of true positive predictions to total positive predictions. If our precision score is high, then we can be fairly confident that the positive predictions we are making are really positive predictions. High precision is extremely useful when we want to limit false positives (for example misdiagnosing a healthy patient). Recall (also known as sensitivity) looks at the model's ability to find all real positive cases (out of the total real positive cases), so it measures the percentage of true positive predictions out of all real positive predictions. High recall is paramount, especially in a medical setting, because missing real cases of disease can result in catastrophic consequences. The F1-score is a score that attempts to combine precision and recall into one measurement through the harmonic mean of precision and recall. It balances both false positives and false negatives into one single score. The advantage of the f1 score is that it gives a more comprehensive evaluative

technique, particularly in cases of imbalanced data or when using either precision or recall is not sufficient. The definitions for all of these metrics are shown here:

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

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

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

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

As a whole, these metrics create a complete assessment framework for evaluating the diagnostic capability of the developed deep learning models. Their use in conjunction provides an assessment of the metric that takes into account all aspects of the model, including its predictive accuracy and the potential harm caused by the model's errors - this is particularly critical in sensitive environments such as medical imaging and disease detection.

RESULTS AND DISCUSSION

In this study, the performance and computational complexity of three different Convolutional Neural Network (CNN) architectures were comparatively evaluated for breast cancer classification from mammography images. The summarized results are presented in the corresponding performance table. Upon examining the evaluation metrics, it is evident that the RexNet-200 model outperformed the other architectures, achieving the highest accuracy of 76.47%. This model also delivered superior precision at 75.18%, a solid recall rate of 70.22%, and an F1-score of 77.44%. The detailed confusion matrix and training metrics for RexNet-200 are presented in Figure 3 and Figure 4, respectively. These figures illustrate the model's strong classification capabilities and stable training performance.

Table 1 Performance and Complexity Comparison of CNN Models for Breast Mammogram Classification

Model	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Params (M)	GFLOPs
EfficientNetB7	73.92	71.02	70.60	70.79	63.79	10.26
EfficientNetV2-Small	73.53	71.08	67.02	67.94	20.18	5.42
RexNet-200	76.47	75.18	70.22	77.44	13.81	3.05

Following closely, EfficientNetB7 demonstrated a balanced performance with an accuracy of 73.92%, precision of 71.02%, recall of 70.60%, and an F1-score of 70.79%. The confusion matrix and training metrics for EfficientNetB7 can be found in Figure 5 and Figure 6, respectively, highlighting its effective discrimination between classes and consistent convergence during training.

EfficientNetv2-Small showed competitive but slightly lower results, with an accuracy of 73.53%, precision of 71.08%, and recall of 67.02%. Its confusion matrix and training metrics are displayed in Figure 7 and Figure 8, respectively. These demonstrate the model's reasonable classification performance and training stability, albeit with slightly reduced recall compared to the other models.

From a computational complexity perspective, RexNet-200 stands out as the most efficient model among the three, with only

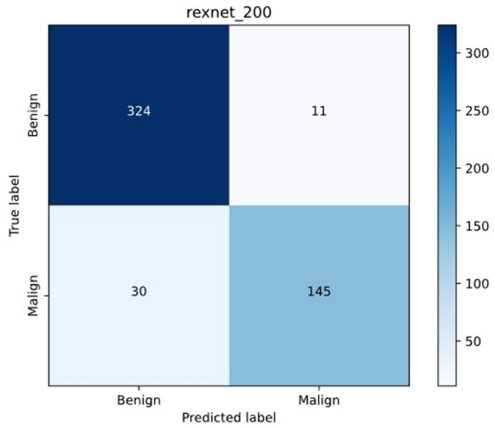


Figure 3 Confusion Matrix RexNet-200

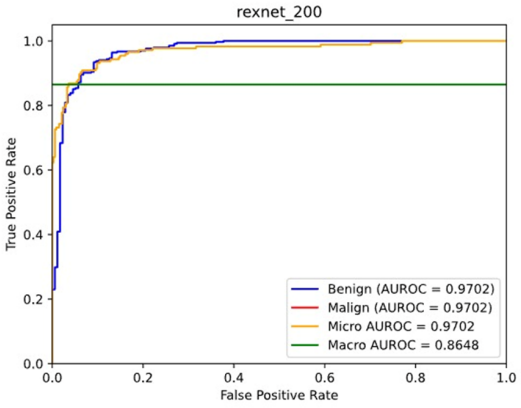


Figure 4 ROC Curve for RexNet-200

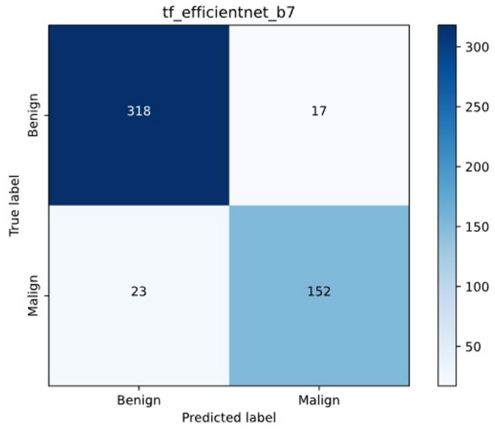


Figure 5 Confusion Matrix for Efficientnet-B7

13.81 million parameters and 3.05 GFLOPs, indicating a favorable trade-off between performance and resource usage. Conversely, EfficientNetB7, while delivering good accuracy, is the most complex model with 63.79 million parameters and 10.26 GFLOPs. EfficientNetv2-Small falls in the moderate range regarding complexity and computational demand.Clinically, these findings have important implications. In breast cancer diagnosis, maintaining a balance between recall and precision is critical: high recall min-

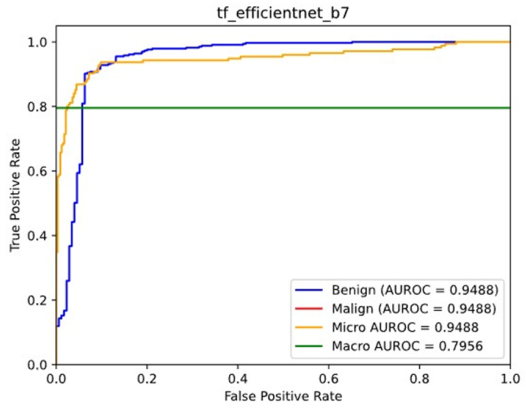


Figure 6 ROC Curve for Efficientnet-B7

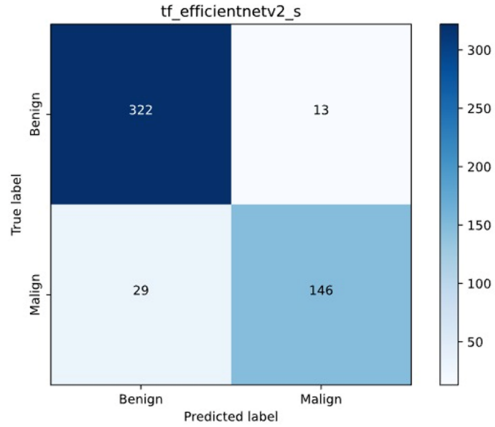


Figure 7 Confusion Matrix for Efficientnetv2-s

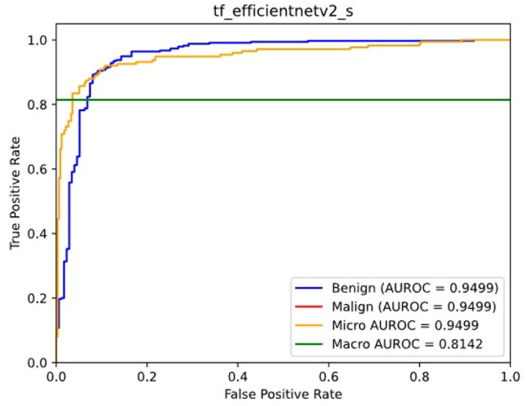


Figure 8 ROC Curve for Efficientnetv2-s

imizes the risk of missing malignant cases, while high precision reduces false positives and avoids unnecessary anxiety and invasive procedures for patients. RexNet-200's ability to achieve the highest scores in both precision and recall metrics, combined with its lower computational cost, positions it as an ideal candidate for practical deployment in clinical decision support systems. Its efficiency also enables faster inference and easier integration into resource-constrained healthcare environments.Therefore, RexNet-200 emerges as the most promising architecture for developing an accurate, efficient, and clinically viable breast cancer diagnosis support tool.

CONCLUSION

This study aimed to comparatively evaluate the performance and efficiency of three widely-used Convolutional Neural Network (CNN) architectures EfficientNetB7, EfficientNetv2-Small, and RexNet-200 for the classification of breast cancer from mammography images. The results clearly demonstrated that the RexNet-200 model outperformed the other architectures in terms of both diagnostic accuracy and computational efficiency. RexNet-200 achieved the highest accuracy of 76.47%, while also maintaining a relatively low model complexity with 13.81 million parameters. This outcome, particularly when compared to the more complex EfficientNetB7 model, which has 63.79 million parameters yet slightly lower accuracy, supports the idea that increased model complexity does not necessarily guarantee better performance for this task.

The success of RexNet-200 can be attributed to its efficient architecture that balances feature learning capacity with resource requirements, enabling more effective classification of breast cancer from mammography images. While these findings are encouraging, it is important to recognize the limitations of the dataset used. Future work should aim to validate these models on larger, more diverse clinical datasets, explore different preprocessing and data augmentation techniques, and incorporate Explainable Artificial Intelligence (XAI) methods to improve model transparency and trustworthiness.This research represents a significant step toward developing high-performance and computationally efficient AI models with strong potential for integration into clinical workflows as reliable decision support tools for radiologists.

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

Ansari, Z. A., M. M. Tripathi, and R. Ahmed, 2025 The role of explainable ai in enhancing breast cancer diagnosis using machine learning and deep learning models. Discover Artificial Intelligence **5**: 75.

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 .

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, H. and A. L. Martel, 2025 Enhancing breast cancer detection on screening mammogram using self-supervised learning and a hybrid deep model of swin transformer and convolutional neural networks. Journal of Medical Imaging **12**: S22007–S22007.

Faryna, K., J. van der Laak, and G. Litjens, 2024 Automatic data augmentation to improve generalization of deep learning in h&e stained histopathology. Computers in Biology and Medicine **170**: 108018.

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, T., M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. Mridha, 2024 A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. Healthcare Analytics **5**: 100340.

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

Kurtulus, I. L., M. Lubbad, O. M. D. Yilmaz, K. Kilic, D. Karaboga, *et al.*, 2024b 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.

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.

Mridha, K., T. Sarker, S. D. Bappon, and S. M. Sabuj, 2023 Attention u-net: A deep learning approach for breast cancer segmentation. In *2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*, pp. 1–6, IEEE.

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.

Shorten, C. and T. M. Khoshgoftaar, 2019 A survey on image data augmentation for deep learning. Journal of big data **6**: 1–48.

Sreekala, K. and J. Sahoo, 2025 Enhancing mammogram classification using explainable conditional self-attention generative adversarial network. Expert Systems with Applications p. 128640.

Sureshkumar, V., R. S. N. Prasad, S. Balasubramaniam, D. Jagannathan, J. Daniel, *et al.*, 2024 Breast cancer detection and analytics using hybrid cnn and extreme learning machine. Journal of Personalized Medicine **14**: 792.

Tajbakhsh, N., L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, *et al.*, 2020 Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical image analysis **63**: 101693.

Tajbakhsh, N., J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, *et al.*, 2016 Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging **35**: 1299–1312.

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, H., Q. Chen, B. Xue, W. Banzhaf, and M. Zhang, 2024 P-mixup: Improving generalization performance of evolutionary feature construction with pessimistic vicinal risk minimization. In *International Conference on Parallel Problem Solving from Nature*, pp. 201–220, Springer.

How to cite this article: Alswilem, L., and Pacal, N. Artificial Intelligence in Mammography: A Study of Diagnostic Accuracy and Efficiency. *Computational Systems and Artificial Intelligence*, 1(1), 26-31, 2025.