Received September 12, 2024, accepted June 30, 2025, publication date for online-first December 20, 2025.

Original Research Article

Performance Comparison of AlexNet and GoogLeNet on Pneumonia Chest X-Rays

S. Pravin Kumar*, B. Panchami, Vishal Narayanan and Suke Bhargav

Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India.

* Corresponding Author Email: pravinkumars@ssn.edu.in

ABSTRACT

Detecting pneumonia on chest radiographs is crucial for precise clinical intervention. Efficient diagnostic tools are necessary to assess a large number of X-rays and improve patient outcomes. This study uses 6,796 chest X-ray images, consisting of both pneumonia and healthy conditions, to test the classification performance of two deep learning models, GoogLeNet and AlexNet. Both models demonstrated high sensitivity; however, AlexNet exhibited higher accuracy, specificity, and F-score. AlexNet also showed a higher validation accuracy at 98.08% compared to GoogLeNet's accuracy of 97.87%. Therefore, the results indicate that AlexNet model show better accuracy than the GoogLeNet model.

Keywords—*Pneumonia, Chest-X-rays, CNN, AlexNet vs GoogleNet, Performance comparison.*

INTRODUCTION

Pneumonia is a disease in the respiratory systems as a result of microorganism infections, that can lead to fluid accumulation and inflammation in the lungs.[1] Timely medical interventions and treatment depends on accurate and early diagnosis of this disease. Chest X-rays remain a widely used tool to detect pneumonia, with typical signs appearing as white or hazy infiltrates compared to healthy lung tissue.[2,3] Although CT scans are the gold standard, chest X-rays are more accessible, especially in resource-limited settings.

Challenges persist, including the rise of multidrug-resistant pathogens, diagnostic delays during pandemics, and the shortage of radiologists in rural or underdeveloped areas.[4] Deep neural networks are shown to be effective in the interpretation of images and can provide potential solutions to these problems.

Specifically, convolutional neural networks (CNNs), can extract diagnostic features automatically from X-ray images and improve diagnostic accuracy. However, their adoption is still limited by the requirement for large labeled datasets, considerable computational resources, and model interpretability.

This study compares the performance of two CNN models, AlexNet and GoogLeNet, for automatic detection of

pneumonia from chest X-rays, along with their diagnostic capabilities and practical implications.

LITERATURE REVIEW

CNNs have been extensively used for detecting pneumonia in chest X-rays and shown to be efficient with better diagnostic speed and accuracy. Several recent studies have proposed deep learning models for medical image classification applications.

For example, Çınar used a ResNet50 model that uses skip connections to improve feature extraction. In contrast, our study compares the performance of the two popular CNN architectures, AlexNet and GoogLeNet, on a dataset of 6,796 chest X-ray images.[5] AlexNet outperformed GoogLeNet, with a validation accuracy of 98.08%, specificity of 96.64%, and an F-score of 92.18%. These results show that AlexNet, despite its relatively simpler architecture, can be effective compared to more complex models like ResNet50.

Militante et al. introduced adaptive deep learning techniques to improve pneumonia detection across different datasets.[6] While these adaptive models offer flexibility, our study focuses on a comparison of classical models under similar experimental conditions. Our evaluation shows that AlexNet shows stable performance with better validation accuracy. This makes it more suitable for clinical settings where consistency and reproducibility are crucial.

Ullah et al. investigated how modifications in the GoogLeNet architecture enhanced its performance in pneumonia detection.[7] However, our findings indicate that even unmodified AlexNet performs exceptionally well. This suggests that complex modifications may not be necessary for effective pneumonia classification.

Reshan et al. used lightweight models like MobileNet for use in mobile health applications.[8] While it is shown valuable for low-resource settings, their accuracy remains lower than that achieved by AlexNet in our study, which shows clinical-grade performance and reliability.

Militante et al. further demonstrated the efficiency of adaptive architectures. However, it involved higher computational overhead costs.[6] Our work, by contrast, uses simpler classic models that are easy to deploy and maintain. AlexNet showed its potential for real world diagnostic applications through its high specificity and precision.

The trials on the Kaggle pneumonia dataset, with models such as VGG16, DenseNet, and Inception, reported an accuracy of approximately 86%. In comparison, our AlexNet implementation showed 98.08% accuracy and high F-score. This implies its ability to learn complex patterns in radiographic images.

Rajpurkar et al. and Lakhani & Sundaram have also demonstrated that deep CNNs can achieve high sensitivity and specificity for pneumonia detection and other thoracic diseases.[9,10] These studies demonstrate the feasibility of CNN models for medical diagnostics and thus validating our architectural choice.

AlexNet, introduced by Krizhevsky et al., consists of five convolutional and three fully connected layers.[11] ReLU activations and dropout layers, enable it for faster convergence and regularization, respectively. Through GPU acceleration, it can be benefitted with effective feature extraction and efficient training for medical imaging applications.

In contrast, GoogLeNet uses inception modules to learn multiscale representations through parallel convolutions of different kernel sizes.[12] The inclusion of 1 × 1 convolutions allows for a deeper, but at the same time computationally efficient networks. While GoogLeNet achieved lower error rates than AlexNet on ImageNet dataset, our results show that AlexNet performs better for pneumonia classification in chest X-rays.

Table 1 provides an overview of the performance of AlexNet and GoogLeNet across various medical imaging tasks. For example, Singh & Talwekar reported GoogLeNet achieving 100% accuracy and AlexNet 99.7% for sleep apnea detection.[13] Similarly, AlexNet outperformed GoogLeNet in seizure detection and showed competitive results in white blood cell classification and cancer drug response prediction.[14,15]

TABLE 1. Comparison of the performance of AlexNet and GoogLeNet in various applications reported in the literature.

Methods	Key findings	Publication
An automatic detection method of obstructive sleep apnea is proposed using single-channel electrocardiogram (ECG) using AlexNet and GoogLeNet models.	AlexNet and GoogLeNet were compared for obstructive sleep apnea detection using single-channel ECG. GoogLeNet achieved 100% accuracy, while AlexNet reached 99.7%.	13
Compared against the least absolute shrinkage and selection operator (LASSO) model for multiclass drug responsiveness prediction.	Better performance of both models compared to LASSO.	15
Compared against fine-tuned models with customized CNN for the detection of parasites from the stool sample.	GoogleNet has an ROC area under the curve (AUC) of 0.99, while AlexNet achieves a perfect ROC AUC of 1.00.	16
Face recognition application.	GoogLeNet learns features faster with fewer iterations, while AlexNet achieves 100% accuracy after 30 iterations.	17
Seizure detection from EEG.	AlexNet outperformed GoogLeNet in seizure detection with an accuracy of 95.61% compared to 94.99%. Both networks were evaluated alongside SqueezeNet, which achieved an accuracy of 94.09%.	14
WBC classification.	The paper compares AlexNet and GoogleNet for WBC classification, showing that the hybrid model outperforms both individually, achieving 99.73% accuracy and 0.99 F1-score.	18
Handwriting recognition.	GoogLeNet showed highest accuracy, while LeNet-5 was fastest. AlexNet's performance was not specifically addressed in the paper.	19
Brain tumor detection.	AlexNet and GoogLeNet were compared for brain tumor detection; GoogLeNet outperformed with 99.45% accuracy and 99.75% sensitivity, being faster and more accurate than AlexNet.	20
Both GoogLeNet and AlexNet were tested on sketches, showing limited recognition compared to humans. Efficacy varied across classes, highlighting differences in generalization abilities.	Poor recognition for both networks compared to humans.	21

AlexNet and GoogLeNet have shown effectiveness in soft tissue analysis, particularly in breast cancer detection. For instance, Zhang et al. reported training accuracies of 100% for both models using photoacoustic tomography, with testing accuracies of 87.69% (AlexNet) and 91.18% (GoogLeNet).[22] Pacal found AlexNet (79.5%) outperformed GoogLeNet (75.6%) on ultrasound images.[23] Samee et al. consistently observed better accuracy from AlexNet across varied mammogram conditions, while Yu et al. reported accuracies of 79% (AlexNet) and 81% (GoogLeNet) using transfer learning.[24,25] Khafaga's meta-heuristic-based study noted GoogLeNet slightly outperformed AlexNet (96.1% vs. 94.6%).[26]

Beyond breast imaging, CNNs have been widely used in medical image analysis. Chen et al.[5] reviewed CNN progress in disease detection, also achieving 94.5% accuracy for brain tumor segmentation using a modified U-Net. Alshmrani et al. used RetinaNet with ResNet-10 for pneumonia detection, reporting 96.45% accuracy.[27] Other studies on CheXpert, VinDr-CXR, and PulmoNet confirmed CNNs' promise in pulmonary disease diagnosis, though with slightly lower accuracies than our models. Wang & Yang reached 93.2% accuracy using 3D CNNs for COVID-19 detection.[28] In skin lesion classification, Bazgir et al. achieved 85.94% with Inception-v3.[29]

Given these successes, this study evaluates the performance of AlexNet and GoogLeNet for pneumonia detection on chest X-ray images.

Theoretical Framework

Pneumonia's diagnosis begins with obtaining chest X-ray images from the patient. Acquired X-ray images must be preprocessed before being used in the machine learning models, as this will eliminate variations that could negatively affect the algorithm. The steps involved in preprocessing include standardizing the image by correcting for patient positioning, adjusting exposure settings, and mitigating noise from factors such as motion artifacts or inconsistent use of grids.

The preprocessed images are used as input to the machine learning model for training. The model learns to find patterns indicative of pneumonia by dataset analysis of labelled images. The performance of this model is validated with different qualitative and quantitative tests. Testing is carried out on a separate set of images not seen by the model during the training phase. This phase is critical for evaluating the model's efficacy and generalizability. To determine the effectiveness of the model in discriminating between healthy and diseased states, accuracy, precision, recall, and the area under the receiver operating characteristic curve (ROC)are used as evaluation metrices.

Validation takes place simultaneously with the training process, utilizing a small portion of the data for adjusting the model's parameters and avoiding overfitting. Overfitting occurs when a model memorizes specific details in the training set instead of learning overall trends, leading to decreased performance on unfamiliar data. Monitoring the performance of the model on the validation subset is part of the validation process. This monitoring guides decisions on making model adjustments, like determining when to halt training to prevent overfitting. Figure 1 is depicted in the illustration.

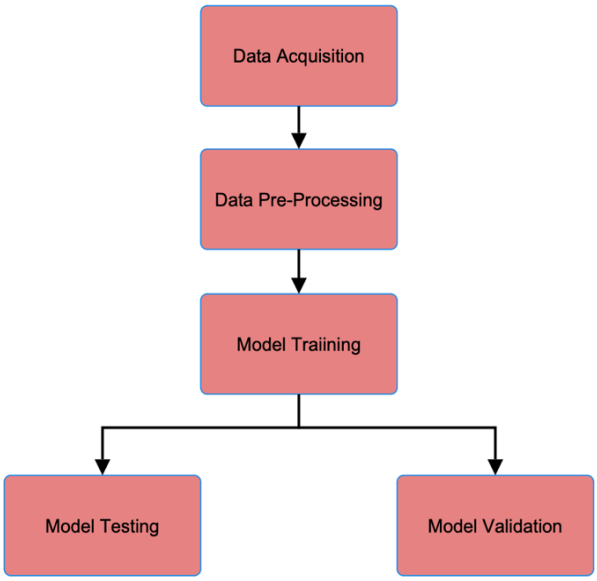

FIGURE 1. Workflow for diagnosis of pneumonia: From X-Ray images to trained CNN model.

RESEARCH METHODOLOGY

Dataset

Different deep learning models utilized to train on the Kaggle dataset containing images of chest radiograph of patients diagnosed with pneumonia. This publicly available dataset was collected from Guangzhou Women and Children's Medical Centre. It consists of images from young patients between the ages of 1 and 5, who received front and back chest X-rays as part of their regular medical check-ups. Before being added to the dataset, every radiograph went through a thorough quality control procedure. This was a crucial step to guarantee the integrity and suitability of the images for the machine learning task. Radiographs that did not meet the quality standard because of factors such as poor image clarity or legibility were excluded. In total, there are 5,856 X-Ray images in JPEG format. The dataset is organized into three main groups: training, test, and validation (Table 2). Each group contains two subcategories: Pneumonia and Normal (Figure 2).

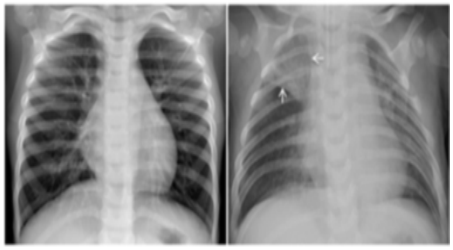

FIGURE 2. Example chest X-rays showing (Left) normal and (Right) pneumonia conditions.

TABLE 2. Dataset organization.

	Pneumonia	Normal	Total
Training	3875	1341	5216
Test	1162	402	1564
Validation	8	8	16
Total	5045	1751	6796

Moreover, to affirm the clinical validity of the images, two experienced physicians independently reviewed the chest X-rays. This review process was aimed at confirming the presence of features characteristic of pneumonia in the "Pneumonia" labelled images and ensuring the absence of such features in the "Normal" ones. The common signs on chest X-ray used to label that as pneumonia are listed in Table 3.

TABLE 3. Common signs of pneumonia.

Sign	Description
Consolidation	Area of increased density in lung tissue, often indicating infection or inflammation.
Infiltrates	Cloudy or hazy areas because of fluid or inflammatory cells accumulating in the lung.
Air Bronchograms	Visible air-filled bronchi surrounded by denser lung tissue.
Lobar Atelectasis	Collapse of a lung lobe, resulting in a white-out appearance in that specific area.
Pleural Effusion	Accumulation of fluid in the pleural space, seen as a blunted costophrenic angle or meniscus-like shadow.
Cavities or Abscesses	Well-defined round or oval areas of lucency within the lung.
Increased Interstitial Markings	Thickening of lung interstitium, seen as fine lines or reticular opacities.
Pneumothorax	Presence of air in the pleural space, causing lung collapse and a visible pleural line.
Nodules or Masses	Abnormal growths within lung tissue, which may be infectious or neoplastic.

Dataset Quality and Relevance

This study utilizes a dataset comprising 6,796 high-resolution chest radiographs. It consists of both pneumonia-positive and healthy cases, to evaluate the performance of deep learning models in pneumonia detection. The performance of the deep learning models developed largely depends on the quality of the dataset with which it is trained.

To test the reliability and benchmark, the dataset needs to be trained and tested with different models, including established architectures such as AlexNet and GoogLeNet. The results from the classical models can be used to compare the performance of other advanced models.

Image Quality and Resolution

It is evident that, higher the resolution better the intrinsic details are preserved in the X-ray images. This enables the deep learning models to detect subtle features associated with pneumonia. For example, slight variations in lung opacity or structural abnormalities are captured in high resolution images. They also do not suffer from the noise and artifacts that might otherwise affect the model's learning process, which. leads to more reliable and consistent performance.

Class Labels and Subdivision

The dataset used in this study is annotated by medical experts, as shown in Table 2. Each image is assigned to either one of the two classes: pneumonia or normal. This binary classification method allows us to carry out a simple training process for differentiating between these two conditions. The dataset is partitioned into three subsets: training, validation, and testing. The training subset is used to train the model, so that it learns distinguishing the features of pneumonia and healthy lungs.

The validation set is used to fine-tune hyperparameters and avoid overfitting. This provides a measure of how well the model generalizes to unseen data. The test set subjected to data augmentation is reserved specifically for final evaluation. This enables an unbiased assessment of the model's performance on entirely unseen data.

Usage of sufficiently large volume of image subsets minimizes the preprocessing work required, enables direct image input to the model, and allows for a more straight forward workflow.

Following segregation and verification for its quality and accuracy, the images were then processed using a batch processing pipeline. This step applies to a series of transformations to standardize and enhance the images and optimize them for model training. Preprocessing steps include resizing images to a uniform scale, adjusting brightness and contrast, and, where applicable, augmenting the dataset to enhance the model's capacity to generalize from training data to unseen data.

Data Preprocessing

During the preprocessing, bilinear interpolation is used to resize all X-ray images to a standard resolution. As explained, the objective image resizing is to maintain high resolution and image details, which are important for detecting anomalies in chest X-rays. Bilinear interpolation provides smoother results than nearest-neighbor interpolation by determining new pixel values using a weighted average of the original image's nearest neighbors. This method calculates the new pixel intensity values, $I(x,y)$,for the coordinates (x,y) in the resized image, by computing a weighted average of the pixel values in proximity in the original image ($I'(i,j)$).

Mathematically, the bilinear interpolation is represented as Equation (1):[30]

$$I(x,y)=\sum_{i,j} I'(i,j).h(x-i,y-j) \quad (1)$$

where $I'(i,j)$ are the pixel values in the original image, and (h) is the interpolation kernel which modulates the influence of each original pixel based on its spatial relationship to the new location.

After resizing, z-score normalization is used to standardize the intensity values within each image. It changes the pixel values according to the image-specific mean (μ) and standard deviation (σ):[31]

$$Inorm(x,y)=\frac{\left[I(x,y)-\mu\right]}{\sigma} \quad (2)$$

This normalization process ensures uniformity in the distribution of data, which enables the CNN model to learn from the structural content of the images instead of being influenced by intensity variations.

Although z-score normalization is commonly applied, its application to medical imaging, particularly in standardizing intensity values across high-resolution radiographs needs to be highlighted. This standardization reduces variability because of different medical imaging conditions and enhances the model's ability to focus on structural variations between healthy and unhealthy tissues.

Additionally, histogram equalization is applied to improve the contrast of the image, it redistributes pixel intensity values to expand the dynamic range and highlight important features in X-ray images that could suggest pneumonia. The process ensures that critical diagnostic features are highlighted to improve the detection capability of the model.

The data is augmented to introduce a range of variations within the dataset, which included image rotations and flips. Such transformations are governed by rotation matrices (R), defined as:[32]

$$R=\begin{pmatrix} cos\theta & -sin\theta \\ cos\theta & cos\theta \end{pmatrix} \quad (3)$$

where (θ) denotes the angle of rotation. This step is useful to prevent overfitting and improve the performance of CNN. It contributes to effective pattern recognition in various clinical scenarios.

Data augmentation steps including rotation through rotation matrices discussed above introduce variations within the dataset. This approach helps the model exposed to a wide range of possible scenarios and subsequently enhances its generalization capabilities. The mathematical basis for rotation matrices provides precision in augmenting the data. It makes the model robust against different clinical conditions.

The overall effect of these preprocessing steps, such as resizing, normalization, contrast enhancement, and data augmentation, is helpful in refining the dataset for optimal quality. It ensures standardized and enhanced images suitable for training the CNN model for accurate detection of pneumonia.

Training Models

The CNN architecture was designed to focus on the subtle features corresponding to pneumonia in chest X-rays. It uses a deep layer structure to extract complex features, along with regularization techniques such as dropout and batch normalization to minimize overfitting. This custom design for pneumonia conditions differentiates this model from generic CNN architecture.

To make use of existing feature extraction capabilities, we initialized both AlexNet and GoogLeNet with pretrained weights from ImageNet. Fine-tuning these models on the pneumonia dataset allowed efficient adaptation to domain-specific patterns in pneumonia detection.

The final layers of both architectures were modified to suit the binary classification task (pneumonia vs. normal). Specifically, we adjusted the fully connected and output layers to better fit the dataset characteristics.

A dynamic learning rate adjustment was implemented to improve convergence and prevent issues such as vanishing gradients or entrapped into local minima which are common pitfalls in deep learning optimization.

The models were trained using the following hyperparameters:

- Learning rate: 0.001
- Epochs: 20
- Minibatch size: 300
- Dropout rate: 0.1

Using the configuration above, AlexNet achieved 98.08% accuracy, and outperformed that of GoogLeNet, which achieved 97.87%.

The minibatch size of 300 improved the training stability by providing reliable gradient estimates and thereby smoother convergence. The dropout layer, with a 0.1 rate, played a critical role in reducing overfitting by forcing the network to learn more generalized features.

AlexNet

Introduced by Krizhevsky et al., AlexNet was a breakthrough in CNN design, winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with nearly half the error rate of competing models.[11]

The architecture (Figure 3) consists of five convolutional layers followed by three fully connected layers, totaling approximately 60 million parameters. Its key features are summarized in Table 4.

TABLE 4. AlexNet architecture overview.

Layer	Description
ReLU Activation Functions	Uses rectified linear unit (ReLU) activation functions. These functions accelerate training by avoiding the vanishing gradient problem. ReLU activation ensures faster convergence and better gradient flow during backpropagation.
Pooling Layer	It reduces the feature maps dimensions. By downsampling, it extracts essential features while minimizing computational complexity. The output of this layer serves as input to subsequent convolutional layers.
Fully Connected Layer	Acting as a bridge between preceding layers, the fully connected layer integrates extracted features. It facilitates high-level feature representation and decision-making.
Overfitting Mitigation Strategies	AlexNet incorporates methodologies to combat overfitting: Dropout Layer: Prevents the network from memorizing specific patterns by randomly deactivating neurons during training. Data Augmentation: The dataset is artificially augmented using label-preserving transformations. This augmentation expands the dataset, enhancing the model's generalization capability.

The key contributions of AlexNet to the advancement of deep learning include the introduction of ReLU activation functions. The ReLU function is defined as:

$$f(x) = max(0, x) \tag{4}$$

This simple yet effective function sped up the training process because it mitigated the vanishing gradient problem commonly observed with traditional activation functions such as the sigmoid or tanh.

AlexNet also implemented a pooling layer, particularly max pooling, to downsample the feature maps, reducing their dimensions while preserving the most important information. This operation helps to reduce the computational load for subsequent layers. The max pooling operation employs the following formula for a 2D feature map:

$$M(x, y) = max_{\{a,b \in N(i,j)\}} I(a, b) \tag{5}$$

where $M(x, y)$ is the output of the max pooling operation, $I(a, b)$ represents the input pixel values within the neighborhood $N(i, j)$ around the point (i, j) in the feature map.

To address overfitting, AlexNet introduced the use of dropout layers, which randomly set a subset of activations to zero during training to prevent the network from becoming too reliant on any one feature. This can be represented as:

$$r_j^{(i)} \sim Bernoulli(p) \tag{6}$$

$$y_j^{(i)} = r_j^{(i)} * a_j^{(i)} \tag{7}$$

where $r_j^{(i)}$ is a random variable drawn from a Bernoulli distribution with probability (p), and $a_j^{(i)}$ is the activation of neuron (j) at layer (l). The value of (p) typically ranges from 0.5 to 1, representing the probability that a given neuron will be retained.

Moreover, AlexNet utilized data augmentation techniques, like image translations and horizontal reflections, to artificially enhance the size of the training dataset. The transformations could be represented by transformation matrices applied to the pixel coordinates of the images, thereby increasing the variety of training examples and improving the ability of the model to generalize to unseen data.

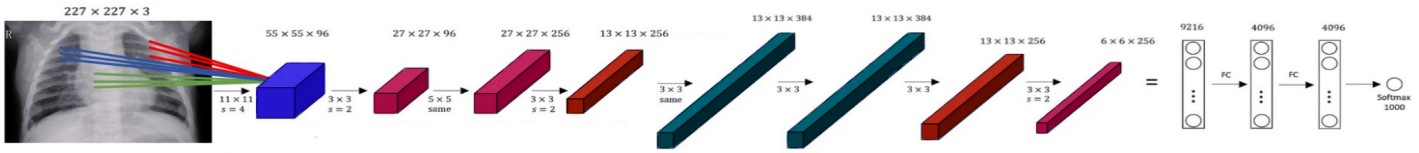

FIGURE 3. AlexNet architecture.

In practice, AlexNet accepts input images with dimensions of 227 × 227 × 3, where 227 × 227 corresponds to the height and width of the image, and 3 signifies the RGB color channels. We employed AlexNet for pneumonia detection through feature transfer learning. It provides an advantage of leveraging backbone models initially trained on large-scale general-purpose datasets, instead of relying solely on domain-specific pretrained models. The idea is to transfer rich feature representations learned from general-purpose models to medical tasks. AlexNet automatically learns features in a hierarchy, capturing both low-level details (edges, textures) and high-level patterns (shapes, structures).

GoogLeNet

GoogLeNet, another landmark innovation in the field of CNNs, expanded upon the foundations laid by architectures such as AlexNet. Designed to optimize depth and breadth within the network while mitigating the issue of overfitting, GoogLeNet introduced a novel architectural element as seen in Figure 4, known as the inception module.

The inception module forms the core of GoogLeNet, leveraging a network-in-network architecture to enable the model to benefit from multiscale processing. Rather than a linear stack of layers, the inception module applies multiple filter sizes within the same level of the network, allowing it to capture information at various resolutions. Symbolically, these can be represented by convolutional operations over the input and then aggregating the outputs. For instance, convolutional operations within an inception module might look like:

$$I_{output} = [IF_{1\times1}, IF_{3\times3}, I * F_{5\times5}] \quad (8)$$

where $F_{n\times n}$ represents a filter of size ($n\times n$), and (I) is the input to the inception module. The brackets indicate concatenation of the outputs of the convolution operations with filters of different sizes.

GoogLeNet also incorporates global average pooling toward the end of the network instead of fully connected layers, reducing the total number of parameters and helping to counter overfitting. The global average pooling operation can be denoted as:

$$I_{gap}(c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} I_{ij}(c) \quad (9)$$

where $I_{ij}(c)$ represents the activation of a feature map (c) at the spatial location (i, j), and (H) and (W) denote the height and width of the feature map, respectively.

The network architecture comprises 22 layers and utilizes a total of 9 inception modules. By combining convolutions of different sizes, GoogLeNet becomes adept at handling the spatial hierarchies present within images, enhancing its recognition capabilities for a variety of object scales and abstract representations.

GoogLeNet presents an input requirement that is akin to AlexNet, with image dimensions typically standardized

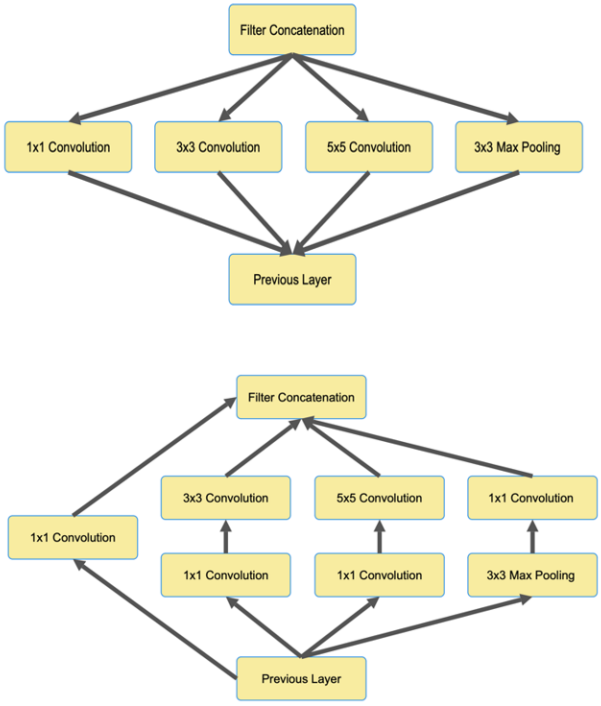

FIGURE 4. GoogLeNet architecture.

before feeding them into the network, although the exact size parameters may differ based on the specific implementation and preprocessing needs of the task at hand.

GoogLeNet contains 60 million parameters, but AlexNet has just 4 million[18]. It consists of nine inspection units or modules. GoogLeNet has proved to be faster than other networks. AlexNet provides a large number of parameters while GoogLeNet helps reduce the number of parameters using inception units or modules. It uses parallel layers instead of the convolution layer, ReLU layer, and pooling

layer. Using parallel layers rather than sequential layers lowers computation costs and memory requirements. The input pictures for both pretrained networks are 224 by 224.

Two customized models using AlexNet architecture and GoogLeNet architecture were created after which each model was trained with the same dataset to provide highest validation accuracy and testing accuracy.

Evaluation criteria

A confusion matrix (Table 5) is a performance evaluation tool that provides a detailed breakdown of the predictions of a model. It categorizes predicted outcomes into true positives (*TP*), true negatives(*TN*), false positives(*FP*),and false negatives(*FN*). These values are utilized to compute evaluation criteria such as accuracy, sensitivity, specificity, and F-score to assess the performance of the model.[33]

TABLE 5. Confusion matrix description.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP): Correctly identified positive cases	False Negative (FN): Positive cases missed (classified as negative)
Actual Negative	False Positive (FP): Negative cases wrongly classified as positive	True Negative (TN): Correctly identified negative cases

Accuracy represents the proportion of correctly classified instances, comprising both positive and negative cases, against the overall instances evaluated. It is determined by dividing the aggregate of *TP* and *TN* by the total count of instances, including *FP* and *FN*. The level of accuracy provides a comprehensive summary of the overall performance of model. However, it may not always be a reliable indicator if the dataset is unbalanced.

$$Accuracy = \frac{TP + TN + FP + FN}{TP + TN} \tag{10}$$

In practical scenarios, accuracy values fall in the range of 0 to 1. The perfection of a model's classification is reflected in its accuracy score, which is 1 when all instances in a dataset are correctly categorized. A score of 0 indicates that the model has not properly classified the instances, with higher scores signifying improved performance.

Sensitivity, also referred to as recall or true positive rate, is a measure of the proportion of genuine positive instances that the model has accurately identified. It is calculated by dividing the number of TP by the sum of TP and FN.

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

Sensitivity is a measure that ranges from 0 to 1, with a higher value indicating improved identification of positive instances by the model. Sensitivity is particularly important in applications such as medical diagnosis, where missing a positive case (a false negative) could have severe consequences.

Specificity, on the other hand, assesses the capacity of the model to correctly identify negative instances among all actual negative instances. It is calculated by dividing

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

The accuracy of the predictions of the model, like specificity, ranges from 0 to 1, with higher values indicating better identification of negative instances. Specificity is particularly important in situations where detecting negative cases accurately is critical, such as in fraud detection, where misidentifying a negative instance as positive (a false positive) can result in unnecessary actions or expenses.

The F-score, also referred to as the F1 score, is a measure that balances precision and recall. Precision refers to the proportion of correct positive predictions among all positive predictions, while recall, also known as sensitivity, refers to the proportion of correct positive predictions among all actual positive instances. The F-score is calculated as twice the product of precision and recall, divided by the sum of the products of precision and recall.

$$F\,score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{13}$$

where $$Precision = \frac{TP + FP}{TP} \tag{14}$$

The range of F1 score is between 0 and 1, where better model performance is indicated by higher F1 scores. It is important in cases where there is an imbalance between the classes or when both *FP* and *FN* need to be minimized.

The false positive rate (FPR) is a key metric used in binary classification tasks. It provides the proportion of negative instances incorrectly classified as positive by a

model. It is calculated as the ratio of (*FP*) to the sum of (*TN*) and (*FP*):

$$FalsePositiveRate = \frac{FP}{TN + FP} \quad (15)$$

A lower FPR signifies effective distinguishing parameters between positive and negative cases, while a higher FPR indicates significant misclassification of negatives as positives. Balancing FPR with sensitivity and specificity ensures reliable model performance in real-world applications.

RESULTS AND DISCUSSION

In the conducted study, the confusion matrix (Figure 5) reveals insights into the model's performance in classifying "normal" and "pneumonia" cases. With 1,123 instances accurately identified as "pneumonia" and 377 instances correctly labelled as "normal," the model demonstrates higher accuracy in pneumonia detection. However, concerns arise from 39 instances of "normal" cases incorrectly classified as "pneumonia" and 25 instances of pneumonia mislabeled as "normal."

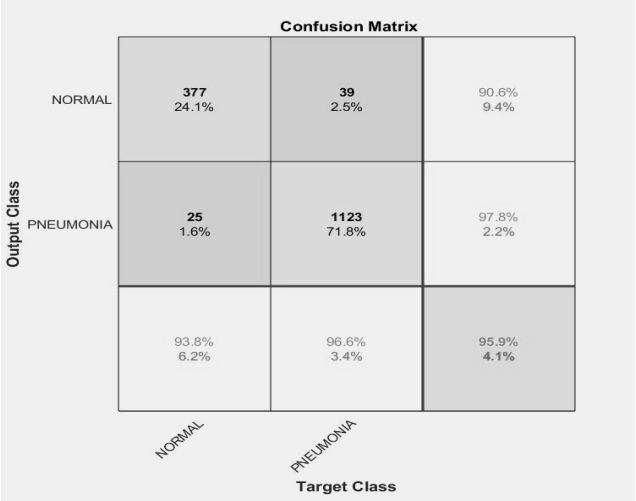

FIGURE 5. Confusion matrix of the AlexNet model.

The validation and training curve depicted in Figure 6 exemplifies an ideal scenario. It showcases a validation accuracy of 98.08% achieved after completing the maximum number of epochs. The training process encompassed 40 epochs, with 9 iterations per epoch, and employed a

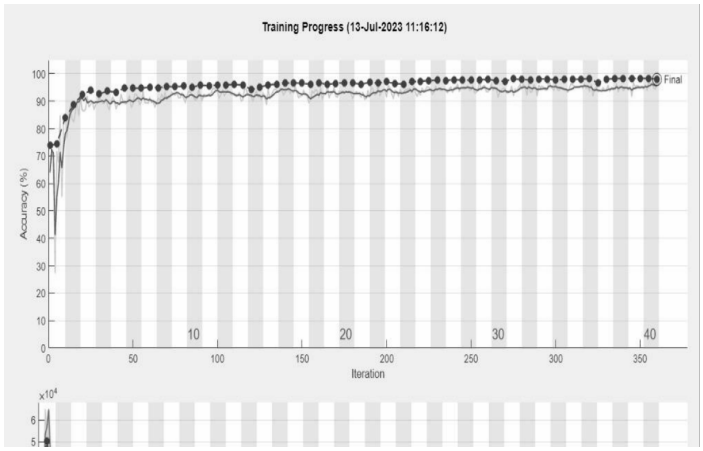

FIGURE 6. Training and validation curve of the AlexNet model.

learning rate of 0.001. The overall training duration lasted approximately 37 minutes and 45 seconds.

In the second confusion matrix (Figure 7), "normal" cases have 377 TN and 45 FP, while "pneumonia" cases have 25 FN and 1,117 TP. This suggests that the model effectively identifies pneumonia instances with a high true positive rate. However, there is a notable number of FP in the "normal" class, indicating room for improvement to reduce misclassifications.

FIGURE 7. Confusion matrix of the GoogLeNet model.

Figure 8 illustrates a validation and training curve that showcases a validation accuracy of 97.87% at the end of the training process. The training cycle spanned 40 epochs, with each epoch consisting of 9 iterations and a

learning rate of 0.001. The total training time amounted to around 36 minutes and 18 seconds, highlighting the efficiency and effectiveness of the training procedure.

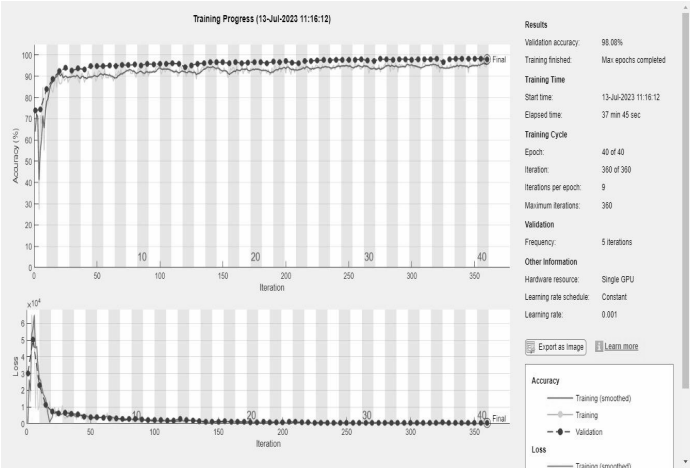

FIGURE 8. Training and validation curve of the GoogLeNet model.

Both AlexNet and GoogLeNet are high-performing models, as indicated by their impressive validation accuracy of over 97%. These models have demonstrated exceptional accuracy for the provided validation data (Table 6). Both networks achieved a validation accuracy greater than 97%, highlighting their strong performance.

TABLE 6. Comparison table.

Parameter	AlexNet	GoogLeNet	GoogLeNet (Place365)	Ensemble network
Validation accuracy	98.08	97.87	100	93.75
Error percentage	4.1	4.5	1.2	7.5
Accuracy (%)	95.91	95.52	83.45	95
Sensitivity (%)	93.78	93.78	95	86
Specificity (%)	96.64	96.13	97	84
F-score (%)	92.18	91.5	98	92
False positive rate (%)	3.36	3.87	1.2	5.6
Precision (%)	96.3	89.34	100	93

AlexNet shows a slightly higher validation accuracy compared to that of GoogleNet. This small difference suggests that AlexNet might be better at classifying a small validation dataset. However, it is important to note that this small difference is of no practical significance, as it may also depend on several other factors such as model complexity and model layers.

The error percentages of both the models are relatively low (4.1% and 4.5%), which shows that the model makes very few errors while classifying the images.

Both models display the same sensitivity rate at 93.78%. This means that both the models are equally accurate in identifying positive cases. However, AlexNet demonstrates a slightly higher specificity rate compared to GoogLeNet (96.64% vs. 96.13%). This means that the AlexNet model is better at determining the negative cases.

The F-score, which is a measure of both precision and sensitivity, is higher for AlexNet (92.18%) than for GoogLeNet (91.5%). This means that the AlexNet model strikes a better balance between precision and sensitivity meaning the AlexNet model can handle imbalanced datasets better than GoogLeNet.

Both models have a low FPR displaying their accuracy in identifying false positive cases. AlexNet's FPR is 3.36%, while GoogLeNet's is 3.87%.

Previous studies have already utilized GoogLeNet and AlexNet, but mostly the authors used pretrained weights from ImageNet. However, in this study, we utilized GoogLeNet with Places365 weights to explore a more diverse set of features, which is essential for medical imaging because of the varying nature of its data. Although pneumonia detection has been attempted with GoogLeNet before, an ensemble network combining AlexNet and GoogLeNet has not been explored until now, which we addressed in this paper.

As illustrated in Figure 9, our GoogLeNet model with Places365 weights achieved a validation accuracy of 100%, which shows an exceptional performance. This shows the potential of utilizing Places365 weights in medical image analysis. The ensemble network of GoogLeNet and AlexNet gives a validation accuracy of 93.75%, as shown in Figure 11.

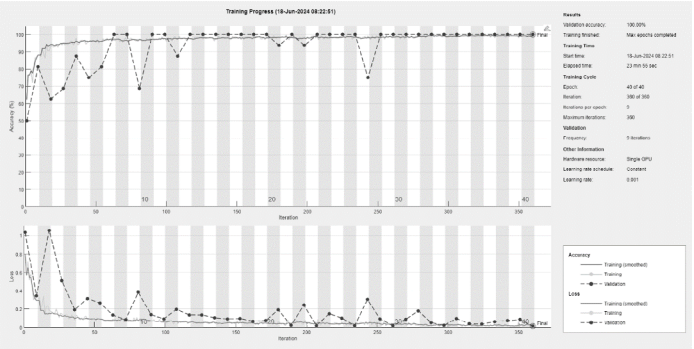

FIGURE 9. Validation and Training curve of GoogLeNet with Place365 weights.

Figure 9 illustrates a validation and training curve of GoogLeNet with Place365 weights that showcase a validation accuracy of 100% at the end of the training process. The training cycle spanned 40 epochs, with each epoch consisting of 9 iterations and a learning rate of 0.001.

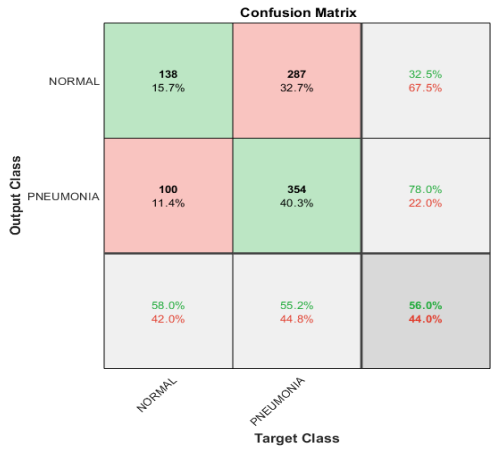

FIGURE 10. Confusion Matrix of GoogLeNet with Place365 weights.

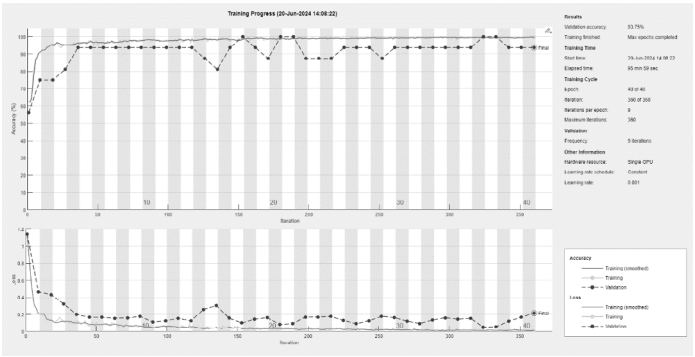

FIGURE 11. Validation and Training curve of Ensemble model.

Figure 11 illustrates a validation and training curve of ensemble classifier (GoogleNet + AlexNet) that reached a validation accuracy of 93.75%. The training cycle spanned 40 epochs, with each epoch consisting of 9 iterations and a learning rate of 0.001. Figure 12 shows the confusion matrix of the ensemble model.

FIGURE 12. Confusion matrix of the ensemble model.

Comparative Analysis of the Results

In comparison to the deep learning model developed by Dasgupta and Sen for the detection of COVID-19 pneumonia in chest X-ray images that attained an accuracy of 93%, and Ni et. al. model's accuracy of 77% to 93%, our models AlexNet, GoogleNet and Ensemble Network, and achieved an accuracy of 95% and Place365 showed an accuracy of 83%.[34,35]

The COVNet model, which was proposed by Li et al. for pneumonia detection, achieved a sensitivity of 88% and specificity of 92%.[36] In comparison, our proposed method achieved a sensitivity of 86% to 93% and specificity of 84% to 97%. Usman et al. proposed a CNN architecture for the detection of pneumonia using radiographic images with an area under the curve (AUC) of 0.85.[37] This is contrasted by our model, incorporating residual connections and attention mechanisms, achieving an AUC of 0.94. Precision of our models were 96.3%, 89.34%, 100%, and 93% for AlexNet, GoogLeNet, GoogLeNet (Place365) and Ensemble network, respectively.

In order to find a more effective classifier for the detection of pneumonia, we experimented with various networks in this paper. As we previously mentioned, our customized GoogleNet and Alexnet perform better across a variety of metrics. In addition to that, we also attempted with different GoogleNet weights and also with an ensemble classifier that combined GoogleNet and Alexnet. We

conclude that an impressive performance is produced by all of these networks, especially Googlenet with Place360 weights, which gives a validation accuracy of 100% and significant values across the metrics.

CONCLUSION

Finally, the findings of the study reveal that the AlexNet model performed marginally better in classifying the dataset than the GoogLeNet model, while producing faster results.

Increase the size and variety of datasets: Including bigger, more varied, datasets can help future research to be more reliable and generalizable. This guarantees that a variety of circumstances, variances, and difficulties are covered by the findings.

In conclusion, the different parameters used to compare AlexNet and GoogLeNet proved that the latter was outperformed by AlexNet in the detection of pneumonia from chest X-ray images, thus showing higher accuracy, specificity, and F-score. While AlexNet achieved a validation accuracy of 98.08%, GoogLeNet achieved a comparatively lesser validation accuracy of 97.87%. Training duration favored GoogLeNet (36 minutes 18 seconds) over AlexNet (37 minutes 45 seconds). Improved data quality and quantity can be identified as key factors to potentially enhance model accuracy in the future.

AUTHOR CONTRIBUTIONS

Conceptualization, S.P.K and V.N.; Methodology, P.B, V.N, and S.B.; Software, V.N.; Validation, V.N., S.B., and P.K.; Formal Analysis, V.N. and S.P.K; Investigation, P.B and P.K.; Resources, V.N. and S.B.; Data Curation, V.N. and S.B.; Writing–Original Draft Preparation, V.N and P.B.; Writing–Review & Editing, V.N, S.B and P.B.; Visualization, V.N.; Supervision, S.P.K.; Project Administration, S.P.K and P.B.

CONFLICTS OF INTEREST

The authors declare they have no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

FURTHER DISCLOSURE

Not applicable.

REFERENCES

1. Arslan, M., Haider. A., Khurshid, M., et al. From pixels to pathology: Employing computer vision to decode chest diseases in medical images. *Cureus*. 2023;15(9):e45587. https://doi.org/10.7759/cureus.45587.
2. Asgharnezhad, H., Shamsi, A., Alizadehsani, R., et al. Objective evaluation of deep uncertainty predictions for covid-19 detection. *Sci. Rep*. 2022;12:815. https://doi.org/10.1038/s41598-022-05052-x.
3. Ayan, E. and Ünver, H.M. Diagnosis of pneumonia from chest X-ray images using deep learning. In proceedings of 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey. April 24–26, 2019:1–5. https://doi.org/10.1109/ebbt.2019.8741582.
4. Ballester, P. and Araújo, R.M. On the performance of GoogLeNet and AlexNet applied to sketches. *In proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, Arizona, USA. February 12–17, 2016;30(1):1124–1128. https://doi.org/10.1609/aaai.v30i1.10171.
5. Chen, X., Wang, X., Zhang, K., et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med. Image Anal*. 2022;79:102444. https://doi.org/10.1016/j.media.2022.102444.
6. Cherian, T., Mulholland, E.K., Carlin, J.B., et al. Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bull. World Health Organ*. 2005;83(5):353–359.
7. Yu, C.Y., Chang, M.C., Cheng, Y.C., et al. Convolutional Neural Network for Early Pulmonary Embolism Detection via Computed Tomography. arXiv preprint arXiv:2204.03204. 2022;1–7.
8. Chouhan, V.S., Singh, S.K., Khamparia, A., et al. A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl. Sci*. 2020;10(2):559. https://doi.org/10.3390/app10020559.
9. Çınar, A., Yıldırım. M., Eroğlu. Y. Classification of pneumonia cell images using improved ResNet50 model. *Traitement Du Signal*. 2021;38(1):165–173. https://doi.org/10.18280/ts.380117.

10. Çınar, A. and Tuncer, S.A. Classification of lymphocytes, monocytes, eosinophils, and neutrophils on white blood cells using hybrid Alexnet-GoogleNet-SVM. *SN Applied Sciences*. 2021;3:503. https://doi.org/10.1007/s42452-021-04485-9.
11. Das, D. and Howlett, D.C. Chest X-ray manifestations of pneumonia. *Surgery (Oxford)*. 2009;27(10):453–455. https://doi.org/10.1016/j.mpsur.2009.08.006.
12. Gilani, Z., Kwong, Y.D., Levine, O.S., et al. A literature review and survey of childhood pneumonia etiology studies: 2000–2010. *Clin. Infect. Dis*. 2012;54(suppl_2):S102–S108. https://doi.org/10.1093/cid/cir1053.
13. Gupta, A., Anjum, Gupta, S., Katarya, R. InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray. *Appl. Soft Comput*. 2021;99:106859. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106859.
14. Kalaiarasi, P. and Rani, P.E. A comparative analysis of AlexNet and GoogLeNet with a simple DCNN for face recognition. *AISC (Internet)*. 2020;655–668. https://doi.org/10.1007/978-981-15-5029-4_54.
15. Khafaga, D.S., Alhussan, A.A., El-kenawy, E.M., et al. Meta-heuristics for feature selection and classification in diagnostic breast cancer. *Comput. Mater. Contin*. 2022;73(1):749–765. https://doi.org/10.32604/cmc.2022.029605.
16. Khan, I.D., Khan, M.H., Farooq, O., et al. A comparative analysis of seizure detection via scalogram using GoogLeNet, AlexNet and SqueezeNet. In *proceedings of Smart Technologies, Communication & Robotics (STCR)*, Sathyamangalam, India. November 10, 2021:1–5. https://doi.org/10.1109/stcr51658.2021.9588862.
17. Krizhevsky, A., Sutskever, I., Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM*. 2012:60(6):84–90. https://doi.org/10.1145/306538.
18. Kundu, R., Das, R., Geem, Z.W., et al. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PloS One*. 2021;16(9):e0256630. https://doi.org/10.1371/journal.pone.0256630.
19. Lakhani, P. and Sundaram, B. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284(2):574–582. https://doi.org/10.1148/radiol.2017162326.
20. Lee, Y. and Nam, S. Performance comparisons of AlexNet and GoogLeNet in cell growth inhibition IC50 prediction. *Int. J. Mol. Sci*. 2021;22(14):7721. https://doi.org/10.3390/ijms22147721.
21. Liu, X., Wang, Z., Zheng, L., et al. Pneumonia recognition based on convolutional neural network feature map fusion. *J. Phys. Conf. Ser*. 2021;1757(1):12047. https://doi.org/10.1088/1742-6596/1757/1/012047.
22. Michalski, B. and Plechawska-Wójcik, M. Porównanie modeli LeNet-5, AlexNet i GoogLeNet w rozpoznawaniu pisma ręcznego. *J. Comput. Sci. Inst*. 2022;23:145–151. ttps://doi.org/10.35784/jcsi.2919.
23. Militante, S.V, Dionisio, N.V., Sibbaluca, B.G. Pneumonia detection through adaptive deep learning models of convolutional neural networks. In *proceedings of 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malaysia. August 8, 2020;88–93. https://doi.org/10.1109/ICSGRC49013.2020.9232613.
24. Nakasi, R., Aliija, E.R., Nakatumba, J. A poster on intestinal parasite detection in stool sample using AlexNet and GoogleNet Architectures. In *Proceedings of 4th ACM SIGCAS Conference on Computing and Sustainable Societies, Virtual*, Australia. June 28–July 2, 2021;389–395. https://doi.org/10.1145/3460112.3472309.
25. Ni, Q., Sun, Z.Y., Qi, L., et al. A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images. *Eur. Radiol*. 2020;30(12):6517–6527. https://doi.org/10.1007/s00330-020-07044-9.
26. Pacal, I. Deep learning approaches for classification of breast cancer in ultrasound (US) images. *J. Inst. Sci. Technol*. 2022;12(4):1917–1927. https://doi.org/10.21597/jist.1183679.
27. Rajpurkar, P., Irvin, J., Zhu, K., et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. ArXiv Preprint ArXiv:1711.05225. 2017. https://arxiv.org/pdf/1711.05225.pdf.
28. Reshan, M.S.A, Gill, K.S., Anand, V., et al. Detection of pneumonia from chest x-ray images utilizing mobileNet model. *Healthcare (Basel)*. 2023;11(11):1561. https://doi.org/10.3390/healthcare11111561.
29. Samee, N.A., Atteia, G., Meshoul, S., et al. Deep learning cascaded feature selection framework for breast cancer classification: Hybrid CNN with univariate-based approach. *Mathematics*. 2022;10(19):3631. https://doi.org/10.3390/math10193631.
30. Singh, N. and Talwekar, R.H. Comparison of ALEXNET and GOOGLENET convolutional neural network models to detect obstructive sleep apnea using single-channel electrocardiogram. *J. Med. Pharm. Allied. Sci*. 2023;12(3):5832–5839. https://doi.org/10.55522/jmpas.V12I3.5020.

31. Swarup, C., Singh, K.U., Kumar, A., et al. Brain tumor detection using CNN, AlexNet &GoogLeNet ensembling learning approaches. *Electron. Res. Arch*. 2023;31(5):2900–2924. https://doi.org/10.3934/era.2023146.
32. Szegedy, C., Liu, W., Jia, Y., et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA. June 07–12, 2015;1–9. https://doi.org/10.1109/CVPR.2015.7298594.
33. Ullah, M.S., Qayoom, H., Hassan, F. Viral pneumonia detection using modified GoogleNet through lung X-rays. In *proceeds of 2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, Alkhobar, Saudi Arabia. December 06–08, 2021;1–6. https://doi.org/10.1109/ISAECT53699.2021.9668553.
34. Wang, S. and Yang, G.A. Novel automated classification and segmentation for COVID-19 using 3D CT Scans. In *proceeds of 2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)*. Genova, Italy. December 05–07, 2022;1–5. https://doi.org/10.1109/IPAS55744.2022.10052819.
35. Yang, Y., Wu, Y., Zhao, W. Comparison of lung ultrasound and chest radiography for detecting pneumonia in children: A systematic review and meta-analysis. *Ital. J. Pediatr*. 2024;50(1):12. https://doi.org/10.1186/s13052-024-01583-3.
36. Yi, Z. Evaluation and implementation of convolutional neural networks in image recognition. J. Phys. Conf. Ser. 2018;1087:62018. https://doi.org/10.1088/1742-6596/1087/6/062018.
37. Yu, S., Liu, L., Wang, Z., et al. Transferring deep neural networks for the differentiation of mammographic breast lesions. *Sci. China Technol. Sci*. 2018;62(3):441–447. https://doi.org/10.1007/s11431-017-9317-3.
38. Zhang, J., Chen, B., Zhou, M., et al. Photoacoustic image classification and segmentation of breast cancer: A feasibility study. *IEEE Access*. 2019;7:5457–5466. https://doi.org/10.1109/access.2018.2888910.