

# An advanced approach for accurate pneumonia detection using combined deep convolutional neural networks

Ola M. El Zein<sup>1</sup>, Naglaa E. Ghannam<sup>1,2</sup>

<sup>1</sup>Department of Mathematics, Faculty of Science, Al-Azhar University (Girls' Branch), Cairo, Egypt

<sup>2</sup>Department of Computer Engineering and Information, College of Engineering in Wadi Alldawasir, Prince Sattam Bin Abdulaziz University, Wadi Alldawasir, Saudi Arabia

## Article Info

### Article history:

Received Jan 20, 2024

Revised Feb 15, 2024

Accepted Mar 5, 2024

### Keywords:

Chest X-rays

Concatenation technique

Deep feature extraction

Pediatric pneumonia

Transfer learning

## ABSTRACT

Pneumonia, a lung infection caused by viral or bacterial agents, poses a significant health risk by affecting one or both lungs in humans. Accurate diagnosis, particularly in pediatric cases, is crucial for timely intervention. Chest X-rays (CXRs) are a common and non-invasive diagnostic tool to detect pneumonia-related abnormalities. Nonetheless, the minimal radiation exposure suitable for pediatric diagnosis poses a challenge in accurately detecting pneumonia in children. This work proposes a concatenation model that combines two pre-trained convolutional neural networks (CNNs) depending on the transfer learning (TL) technique and optimizes the training parameters to build a highly accurate model for detecting pediatric pneumonia from CXR images. The concatenated extracted features from the two pre-trained CNNs are passed through a convolutional layer to select more valuable semantic features to reduce the extracted features, which helps reduce the model parameters and execution time. Experimental results demonstrate that the feature concatenation technique, along with optimization of training parameters, surpasses the performance of individual CNNs and several state-of-the-art methods. The proposed method achieves a classification accuracy of 98.5%, precision of 99.5%, sensitivity of 98.4%, and F1 score of 99.1%. The primary objective of the proposed approach is to aid radiologists in achieving accurate pneumonia diagnosis in real-time.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Ola M. El Zein

Department of Mathematics, Faculty of Science, Al-Azhar University (Girls' Branch)

Cairo, Egypt

Email: olaelzin@azhar.edu.eg

## 1. INTRODUCTION

In recent years, the worldwide spread of various infections and diseases has become a significant concern. Pneumonia is a respiratory infection resulting from viruses or bacteria affecting the lungs [1], [2]. The World Health Organization (WHO) has recognized pneumonia as the primary cause of death worldwide among children under the age of five, constituting approximately 12.8% of annual child deaths [3]. Tragically, in 2016, over 800,000 children succumbed to pneumonia, with the majority being under two years old. This death toll surpassed the combined fatalities from malaria, acquired immunodeficiency syndrome (AIDS), and measles [4], [5]. The gravity of the situation persisted in 2019, with pneumonia claiming the lives of 740,180 children, accounting for 14% of all deaths in child under 5 years old and 22% in child deaths aged 1-5 years [6]. These alarming statistics underscore the urgent need for improved strategies in pneumonia detection, treatment, and prevention to mitigate its devastating impact on global child mortality rates.

The chest X-ray (CXR) proves to be a time and cost-efficient method for pneumonia diagnosis

compared to alternative diagnostic tests, playing a crucial role in clinical nursing and specialized medical analysis [7], [8]. However, a challenge arises as some CXR images exhibit significant similarities. Physicians find themselves compelled to make decisions based on factors such as lung texture, patchy shadow, shadow density, inflammation site, and more. Even experienced doctors are susceptible to errors in judgment and misdiagnosis. Detecting pediatric pneumonia poses an additional challenge due to the varying radiation levels in X-rays for adults and children. Young children are subjected to reduced radiation levels in X-rays, complicating the diagnostic process due to information loss. Consequently, identifying pediatric pneumonia in CXR images becomes a complex task heavily reliant on the individual expertise of medical professionals [9]. Hence, there is a pressing need for an accurate computer-aided diagnostic (CAD) model to assist radiologists in pneumonia diagnosis based on CXR images [10], [11]. Such a model would enhance the precision and efficiency of the diagnostic process, addressing the challenges posed by image similarity and variations in radiation levels, ultimately improving the overall accuracy of pediatric pneumonia detection. Deep learning networks, like convolutional neural networks (CNNs), have streamlined the development of CAD models through automation of the feature engineering process and learning abstract features, leading to enhanced performance [12]. Attention-based CNNs, in particular, have demonstrated improved capabilities by incorporating attention modules that focus on crucial input elements. The integration of residual-based attention helps prevent gradient loss and enhances error propagation. CNNs, known for their effectiveness, are widely employed in the realm of medical image analysis and classification [13]. This approach has garnered significant interest from researchers involved in pneumonia detection within the field of medical imaging. The utilization of attention mechanisms and residual-based strategies showcases the continuous innovation in deep learning techniques, contributing to the advancement of accurate and efficient pneumonia detection models [14], [15].

The detection and diagnosis of pediatric pneumonia through CXR images have garnered significant attention from researchers. Several studies have been conducted where researchers have sought to detect pediatric pneumonia by utilizing deep CNN models. Kermany *et al.* [16] introduced a method for detecting and treating medical diseases by utilizing image-based deep learning models. The approach classified various medical datasets, containing the dataset employed in this paper. The performance of their proposed method was noteworthy, achieving accuracy comparable to that of human specialists at 92.8%. This paper enhanced the performance, but the drawback was it depended on lower-level features. Stephen *et al.* [17] presented an effective deep-learning methodology for detecting and classifying pneumonia in healthcare, utilizing the same dataset employed in this paper. Their approach involved a deep learning model comprising four convolutional and two dense layers. Additionally, traditional image augmentation techniques were applied in their work, resulting in a testing accuracy of 93.7%. In contrast, this paper focuses on enhancing accuracy using classical data augmentation across various data sizes. El Asnaoui *et al.* [18] proposed automated methods for detecting and binary classifying pneumonia images, employing various CNN architectures, including Visual Geometry Group 16 (VGG16), VGG19, densely connected convolutional network with 201 layers (DenseNet201), Inception residual network version 2 (ResNetV2), InceptionV3, residual network 50 (ResNet50), mobile network version 2 (MobileNetV2), and extreme inception (Xception). Notably, models such as Inception\_ResNet\_V2, Inception\_V3, ResNet50, DenseNet201, and MobileNetV2 demonstrated highly satisfactory performance, outperforming baseline CNNs like Xception, VGG16, and VGG19, which exhibited lower performance. This paper highlights the noteworthy performance of Resnet50, MobileNet\_V2, and Inception\_Resnet\_V2, achieving accuracy levels exceeding 96%. These models utilized the use of a deeper convolutional layer with a large number of parameters. Liang and Zheng, as presented in [19], introduced a transfer learning (TL) approach employing a deep residual network for pediatric pneumonia diagnosis. The deep learning model utilized in their work comprised 49 convolutional layers and two dense layers for diagnosing and detecting childhood pneumonia, achieving a testing accuracy of 90.05%. However, a limitation of this technique was the extended execution time due to the utilization of a substantial number of convolutional layers. El Zein *et al.* [20] proposed a hybrid model of EfficientNetB0 as a TL-based model with support vector machine (SVM) hinge loss. The classification process is performed utilizing the hinge loss function as SVM loss instead of sigmoid loss. The proposed hybrid method achieved an impressive classification accuracy of 97%. Güler *et al.* [21] introduced TL methods to classify CXR images into two classes: healthy and pneumonia. Different image preprocessing techniques are applied to CXR images, and the classification process employs TL models including InceptionV3, DenseNet169, ResNet50, ResNet101, Xception, MobileNetV2, VGG16, and DenseNet121 from the Keras library. The Xception model achieved the highest performance, demonstrating a validation accuracy of 96.16% and a test accuracy of 95.73%.

Comparing the existing methods, it becomes evident that many of the proposed models face challenges related to feature extraction and data limitations. Consequently, these issues have an impact on the outcomes of these studies. This paper introduces a novel and accurate concatenated model for addressing these issues and detecting pediatric pneumonia based on CXR images. The innovation presented lies in the

application of the concatenation technique, where two pre-trained models (ResNet50V2 and MobileNetV2) are combined based on TL. ResNet50V2 and MobileNetV2 were chosen due to their capacity to achieve high performance with fewer parameters and faster execution times. A concatenated neural network is created through the integration of the deep extracted features from (ResNet50V2 and MobileNetV2). The deep extracted features are fed into a convolutional layer to extract more meaningful semantic features from the spatial points across all channels. This process aids the network in learning more effectively from the concatenated features derived from ResNet50V2 and MobileNetV2. Also, the convolutional layer helps to reduce the extracted features which helps reduce the model's parameters and execution time. Also, we focus on optimizing training hyperparameters and propose augmentation techniques to overcome data limitations and imbalanced datasets. The proposed model demonstrates an overall increase in accuracy and achieves its peak performance by leveraging multiple deep features extracted from two robust networks rather than relying solely on the features of each network individually. Also, optimization of training hyperparameters helps to improve the accuracy of the concatenated proposed model from 95.7% to 98.5%. This comprehensive strategy aims to address the challenges in feature extraction and data constraints, ultimately contributing to the advancement of pneumonia detection accuracy. The influential contributions of this work are outlined as follows: i) Developing and implementing an accurate medical diagnosis system designed to detect pediatric pneumonia in CXR images utilizing a concatenation technique based on TL models; ii) Utilize efficient TL and fine-tuning methods instead of training a CNN from scratch to address the constrained performance of deep convolution layers; iii) Applying a parallel deep feature extraction process utilizing TL models during the feature extraction stage of the concatenation-based models proposed in this study; and iv) Exploring the effect of optimizing the training parameters on the performance of the classification model.

The structure of the rest of this paper is as follows: section 2 provides an outline of the proposed method architecture, the dataset utilized, preprocessed data, and the training techniques applied. The exploratory results, accompanied by a comparative examination against state-of-the-art models, are displayed in section 3. Finally, section 4 introduces conclusions based on the study findings and outlines probable areas for future research.

## 2. METHOD

This paper proposes a concatenation model that combines two pre-trained CNN models based on the TL technique and optimizes training parameters for detecting pediatric pneumonia from CXR images. First, the collected CXR images are preprocessed and sent as inputs to the two pre-trained CNN models (ResNet50V2 and MobileNetV2). Afterward, the deep features extracted from the two pre-trained CNNs are concatenated to derive combined deep features, aiding the network in better learning from the combined features extracted. These deep features are connected to a convolutional layer to select more valuable semantic features to reduce the extracted features, which helps reduce the model's parameters and execution time, which is, in turn, linked to the classifier for the classification tasks. Also, exploring the effect of optimizing the training hyperparameters during the training of the proposed model. The following subsections give a point-by-point portrayal of each stage of the proposed architecture. The architectural representation of the proposed concatenation model is shown in Figure 1.

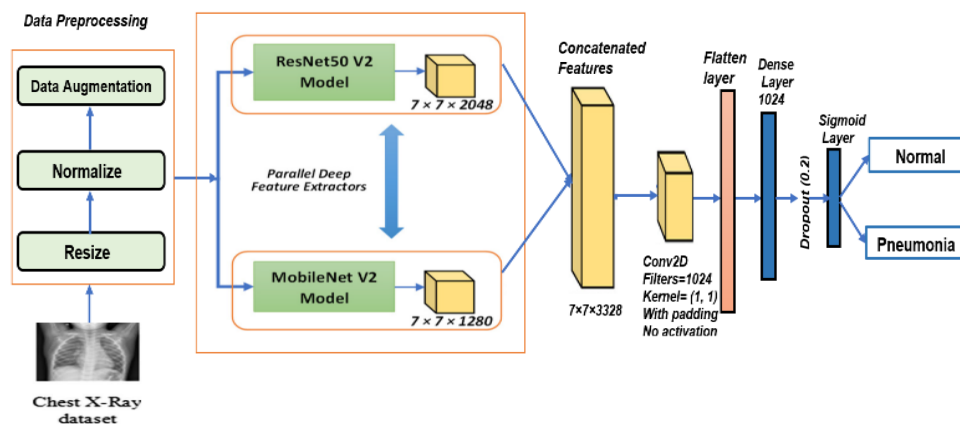


Figure 1. The structure of the proposed concatenation model

## 2.1. Dataset description

The dataset employed in this research consists of 5,856 CXR images [22], distributed across two categories: 4,273 pneumonia cases and 1,583 normal cases. The benchmark dataset is freely accessible on Kaggle and includes three folders: training, validation, and test. However, the distribution of data across these folders was highly imbalanced. Therefore, the dataset was consolidated and divided into training, validation, and testing sets, as outlined in Table 1. Ultimately, the training set contained 4,185 images, the validation set contained 624, and the test set contained 1,047 images. Figure 2 depicts several CXR image samples from the dataset.

Category	Training	Validation	Testing
Normal	1,070	234	279
Pneumonia	3,115	390	768
Total	4,185	624	1,047

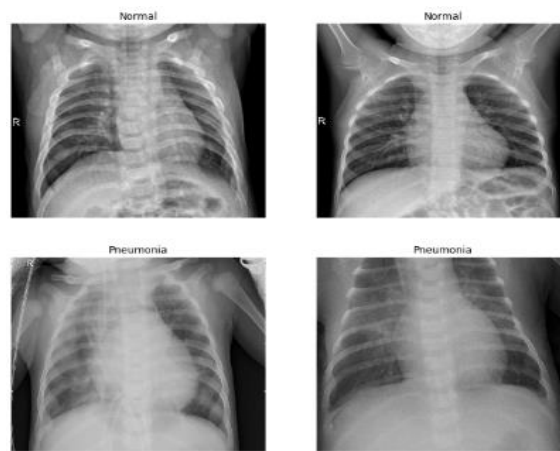


Figure 2. Examples of radiographs of the CXR

## 2.2. Image preprocessing

Preprocessing is a vital phase in the data preparation pipeline for machine learning models, requiring numerous processes to improve the quality and usability of input data. Resizing, normalization, and data augmentation are among the most important of these procedures, as each plays a crucial role in structuring the data for the best model performance. The preprocessing steps' detailed methods are elaborated on in this subsection.

### 2.2.1. Resize images

As evident, all input X-ray images vary in size and shape, posing a challenge for effective classification. To address this, image preprocessing is conducted. In the proposed approach, all input images are resized to dimensions of  $224 \times 224 \times 3$  pixels to ensure suitability for classification tasks in the ResNet50V2 and MobileNetV2 models. Smaller images require fewer computations during both training and inference, leading to faster processing times.

### 2.2.2. Normalization

The normalization of data is a crucial step, often employed to maintain numerical stability in CNN architectures. Failure to normalize data can result in increased network complexity and a reduction in the learning rate. Normalization plays an essential part in enhancing the speed and stability of gradient descent during training in CNN models. Additionally, it contributes to improving the model's performance and its ability to generalize effectively. In this study, the pixel values of input images, post-resize, are normalized within the range of 0 to 1. The dataset comprises of grayscale images, and during the rescaling process, each pixel value is multiplied by  $1/255$  for normalization. This normalization procedure ensures that the CNN model can effectively learn and optimize its parameters during the training process, promoting stability and efficiency in the gradient descent.

### 2.2.3. Data augmentation

Deep learning relies heavily on extensive datasets to yield reliable results. However, acquiring a substantial amount of data, especially in medical domains, proves challenging due to cost and time constraints. Data augmentation techniques involve synthetically expanding the dataset size by applying diverse conversions to existing data, which helps the model learn to recognize patterns and features under different conditions. These are implemented in the training process after dataset preprocessing and splitting. These techniques not only address the challenge of limited data but also mitigate the risk of overfitting, contributing to improved model performance. Table 2 describes the values of parameter for various data augmentation techniques, including shifts, rescaling, flipping, rotation, zooms, and shear. The visual representation of images post different augmentation techniques is illustrated in Figure 3. In scenarios of class imbalance, data augmentation plays a pivotal role in image processing. It not only increases the instances of smaller classes but also enhances model accuracy, introducing variability into the data and contributing to a more robust training process.

Table 2. The augmentation techniques used in this work and their corresponding values

Parameter	Values
Horizontal shift range	0.2
Vertical shift range	0.2
Shear range	0.2
Rotation degree	20
Zoom	0.2
Horizontal flip	True

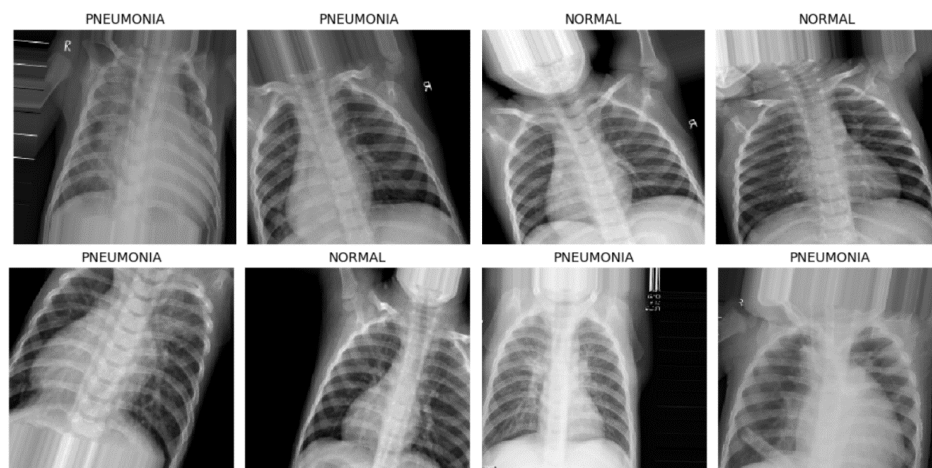


Figure 3. Some samples of CXR images by utilizing different augmentation techniques

### 2.3. The concatenated pre-trained CNN model

CNN models have shown exceptional performance in tasks of processing medical images. Nevertheless, training these models from the ground up for pneumonia detection poses challenges due to the restricted availability of CXR samples. In such scenarios, employing pre-trained models utilizing TL proves beneficial. TL involves utilizing a pre-trained model's structure from a relevant database, either from the same or a different domain, leveraging the acquired knowledge to efficiently and effectively address new problems [23]. This technique is particularly valuable in accelerating the resolution of new tasks, such as predicting pneumonia, by building upon existing knowledge. TL reduces the training time and improves the performance of model, especially when the new dataset is small [24]. An essential aspect of the TL process concerns choosing a pre-trained model. This model serves as the initial point for specific tasks, avoiding the lengthy process of training from scratch with haphazardly initialized weights. This approach significantly conserves the computational resources required for developing neural network models to address these challenges. In this paper, the pre-trained CNN models (ResNet50V2 and MobileNetV2) were chosen due to their capacity to achieve high performance with fewer parameters and faster execution times. These models are employed as a foundation, harnessing their learned features to enhance the efficiency and effectiveness of pneumonia detection in CXR images.

ResNetV2, an evolution of the 2016 ResNetV1 by Microsoft Research Asia, offers three models (ResNet152V2, ResNet101V2, and ResNet50V2). Notably, each layer within these models leverages distinct optimizers, contributing to their superior accuracy compared to the original architecture. This research employed ResNet50V2 [25] for image classification, achieving 93% accuracy on the ImageNet dataset, a widely used benchmark for computer vision tasks. By outperforming other ResNet models like ResNet50 and ResNet101, ResNet50V2 demonstrated its effectiveness for our specific application. This superior performance stems from modifications to the way information flows between different parts of the network, enabling more efficient learning. Choosing ResNet50V2 allowed us to leverage its proven accuracy and enhanced capabilities, contributing significantly to the success of our research for pneumonia detection based on CXR images.

The MobileNetV2 model [26] is a CNN consisting of 53 layers and 88 depths. It is specifically designed for mobile and edge devices. It is an efficient model that aims to provide high accuracy while minimizing computational and memory requirements. MobileNetV2 is an evolution of the original MobileNet architecture, and it introduces improvements in terms of both accuracy and efficiency. In 2018, a million images from the ImageNet database were utilized for training. This model acquired a 90.1% accuracy on this dataset.

The preprocessed input images in our dataset have dimensions of  $224 \times 224 \times 3$  pixels. From the input image, the ResNet50V2 model generates a  $7 \times 7 \times 2048$  feature map on its final feature extractor layer, while the MobileNetV2 model produces a  $7 \times 7 \times 1280$  feature map on its last layer. By concatenating the output features from both pre-trained models, we aim to enhance the quality of the semantic features generated. This concatenation forms a neural network where the concatenated features of size  $7 \times 7 \times 3328$  are connected to a convolutional layer with a  $1 \times 1$  kernel size and 1,024 filters, without an activation function. This strategic addition is intended to extract more valuable semantic features from the spatial relationship between all channels, treating each channel as an individual feature map. By learning from the concatenated features extracted from ResNet50V2 and MobileNetV2, the convolutional layer enhances the network's ability and performance in identifying meaningful patterns. Moreover, it assists in reducing the dimensionality of the extracted features, thereby decreasing the model's parameters and execution time.

#### 2.4. Classification

After automatically extracting features using two pre-trained models, these features are combined via concatenation and passed through a convolutional layer. The resulting output from this convolutional layer is then directed to the classifier. Positioned at the end of the proposed CNN model, the classifier is essentially an artificial neural network (ANN) known as a dense layer. This classifier requires individual features (vectors) for computations, similar to other classifiers. Consequently, the output of concatenated features is flattened into a 1D feature vector for the classifier's use. This process, referred to as flattening, involves transforming the output of the concatenation operation into a single extended feature vector of size 50,176. This vector is then utilized by the dense layer for the final classification task. The classification layer comprises a flattened layer, two dense layers with sizes of 1024 and 1 respectively, a dropout layer with a rate of 0.2 between them, a rectified linear unit (ReLU) activation function between the dense layers, and a sigmoid activation function for performing the tasks of classification. The sigmoid function outputs a value within the range of 0 to 1, making it easy to predict the result as 1 if the value exceeds 0.5 and 0 otherwise. This activation function is commonly employed when dealing with binary classification problems, defined as such [27]:

$$f(x_i) = \frac{1}{(1+e^{-x_i})} \quad (1)$$

where  $x_i$  represents the input vector.

During model training, the binary cross-entropy loss function is utilized. This function assesses the performance of a classification model and is computed as (2) [28], [29].

$$L(y, p) = -y \log(p) + (1 - y) \log(1 - p) \quad (2)$$

where  $y$  denotes the actual value, and  $p$  represents the predicted value.

ResNet50V2 [25], MobileNetV2 [26], and a concatenated proposed network consisting of ResNet50V2 and MobileNetV2, based on the described method, undergo training. Remarkably, the concatenated neural network demonstrates superior accuracy compared to the individual models. Several networks are tested in our project, the ResNet50V2 [25] and MobileNetV2 [26] networks were selected due to their capacity to achieve high performance with fewer parameters and faster execution times. Concatenating the output features from both networks proved to enhance the network's performance and classification capability

by leveraging information from both feature vectors, leading to improved accuracy. Details of the training hyperparameters can be found in Table 3. As indicated in Table 3, the networks were trained utilizing the binary cross-entropy loss function and the Adam optimizer, with learning rates of 0.00001 and 0.0001, respectively. The networks are trained for 50 epochs. For ResNet50V2, MobileNetV2, and the concatenated network, the selected batch size was set to 32, 64 for train, and 32 for test and validation. The optimization of training hyperparameters helps improve the accuracy and performance of the proposed model.

Table 3. Details of the training hyperparameters for the proposed model

Hyperparameters of proposed concatenated model			
Learning rate	0.00001		0.0001
Optimizer	Adam		Adam
Batch Size	32 for train, test and validation	64 for train	32 for test and validation
Max Epochs	50		50

### 3. RESULTS AND DISCUSSION

This section showcases the outcomes derived from a series of experiments, delivering a detailed analysis of the diagnostic performance in pediatric pneumonia detection from CXR images. The evaluation involves employing both a pre-trained CNN model and a concatenation model that combines ResNet50V2 and MobileNetV2. A comparative assessment is carried out among these models, with their results compared against recent approaches, to determine the foremost compelling performing model.

The proposed model underwent training on Google Colaboratory (Colab), utilizing a dedicated Google Cloud project to access free GPU resources for deep learning projects and research. Users are currently allocated 12 GB of RAM, with the potential for expansion up to 25 GB. Google Colab provides access to a single 12 GB NVIDIA Tesla K80 GPU, available for continuous utilization for up to 12 hours.

#### 3.1. Evaluation metrics of model performance

The concatenation proposed model's performance is assessed using various evaluation metrics such as testing accuracy, precision, recall (sensitivity), F1 score, specificity, and area under the curve (AUC) [29], [30]. AUC is a crucial evaluation metric for classification model performance, indicating the model's ability to distinguish between classes by measuring the degree of separability. When the AUC is equal to 1, the classifier is capable of accurately distinguishing between children with and without pneumonia.

#### 3.2. Confusion matrix and testing accuracy

The confusion matrix acts as a tabular representation to visually assess the performance of a prediction model. It contains counts of predictions made by the model, differentiating between correct and incorrect classifications in each cell. In this work, we explore the impact of different hyperparameters. Figure 4 showcases the confusion matrices of ResNet50V2, MobileNetV2, and the concatenated proposed model utilized in this study, with batch size 32 and learning rate 0.00001 in Figures 4(a), 4(b), and 4(c), respectively. Similarly, Figure 5 shows the confusion matrices for the same models but with a batch size of 64 and a learning rate of 0.0001 in Figures 5(a), 5(b), and 5(c), respectively. These matrices show how model performance differs with various parameter configurations.

In Figure 4, the confusion matrix indicates that the ResNet50V2 model, see Figure 4(a), accurately predicted 276 images as "Normal". For the "Pneumonia" class, it correctly identified 726 images and misclassified 42. Similarly, the MobileNetV2 model, see Figure 4(b), correctly identified 730 "Pneumonia" images and misclassified 38. In the "Normal" class, 274 images were correctly classified, while 5 images were incorrectly labeled as "Pneumonia." Furthermore, the concatenated model 1, see Figure 4(c), displayed robust performance by correctly identifying 748 "Pneumonia" images with only 20 misclassifications. In the "normal" class, 274 images were correctly classified, while 5 were erroneously labeled as "Pneumonia".

In Figure 5, the confusion matrix reveals that the ResNet50V2 model, see Figure 5(a), correctly predicted 276 images as "Normal". For the "Pneumonia" class, it identified 740 images correctly and misclassified 28. Conversely, the MobileNetV2 model, see Figure 5(b) correctly identified 759 "Pneumonia" images and misclassified 9. In the "Normal" class, 256 images were correctly classified, with 23 misclassified as "Pneumonia". Additionally, the concatenated model 2, see Figure 5(c) demonstrated robust performance by correctly identifying 756 "Pneumonia" images with only 12 misclassifications. In the "Normal" class, 275 images were correctly classified, while 4 were erroneously labeled as "Pneumonia." These findings underscore the effectiveness of the models in distinguishing between "Normal" and "Pneumonia" instances, with the concatenated model exhibiting particularly strong accuracy and minimal misclassifications. Moreover, concatenation model 2, utilizing two pre-trained models (ResNet50V2 and

MobileNetV2) with a batch size of 64 and a learning rate of 0.0001, outperforms concatenation model 1, which employs two modified models (ResNet50V2 and MobileNetV2) with a batch size of 32 and a learning rate of 0.00001, respectively. Concatenation Model 2 achieves successful classification for all cases in both classes with minimal misclassifications. These findings underscore the significance of optimizing training hyperparameters for improving model performance.

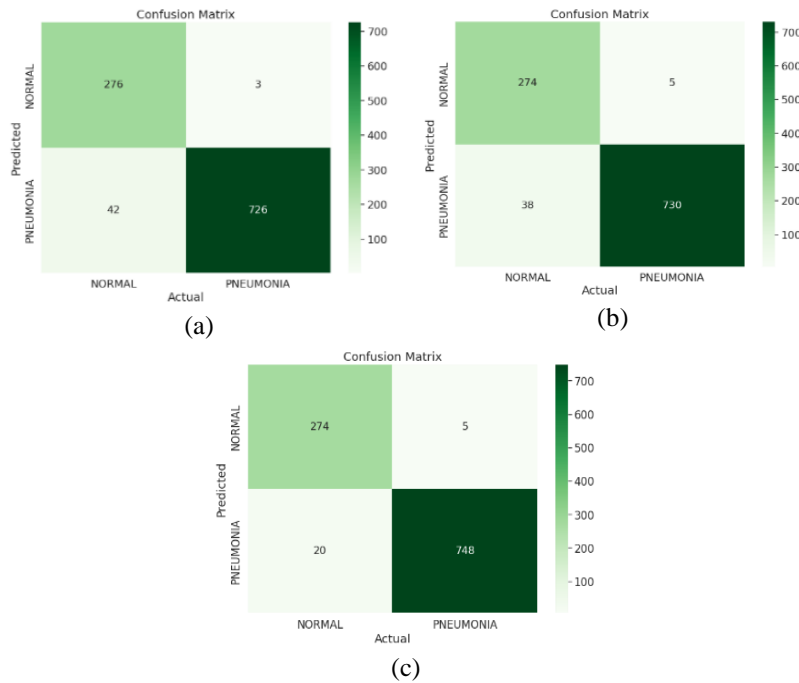


Figure 4. The confusion matrix for (a) ResNet50V2, (b) MobileNetV2, and (c) concatenated model 1 with a batch size of 32 and a learning rate of 0.00001

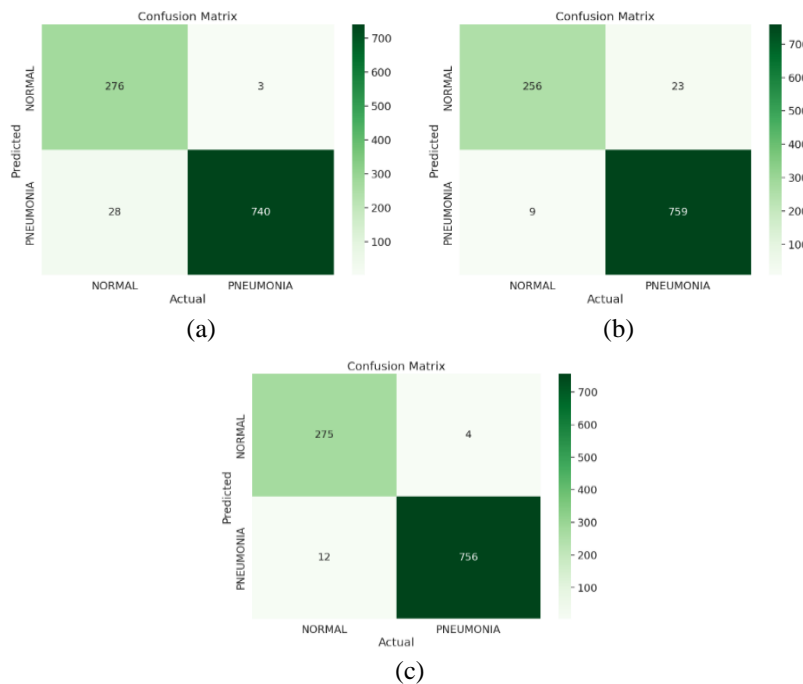


Figure 5. The confusion matrix for (a) ResNet50V2, (b) MobileNetV2, and (c) concatenated model 2 with a batch size of 64 and a learning rate of 0.0001



Table 4 and Table 5 present the performance metrics for ResNet50V2, MobileNetV2, and the concatenation models when trained with batch sizes and learning rates set to 32, 64, and 0.00001, 0.0001, respectively. According to Table 4, the concatenation model 1 outperformed ResNet50V2 and MobileNetV2 in most metrics, achieving the highest percentages for sensitivity (97.4%), F1 score (98.4%), and AUC (97.8%) along with the highest accuracy of 97.6%. In contrast, Table 5 illustrates the superior performance in precision, sensitivity, F1 score, specificity, and AUC when compared to the results in Table 4. Concatenation Model 2, which combines two adapted pre-trained models (ResNet50V2 and MobileNetV2), attained exceptional results with percentages of 99.5%, 98.4%, 98.6%, 99.1%, and 98.5% for precision, sensitivity, F1 score, specificity, and AUC, respectively.

The testing accuracy, a crucial evaluation metric, for ResNet50V2, MobileNetV2, and concatenation model 1 is 95.7%, 95.9%, and 97.6%, respectively, with a batch size of 32 and a learning rate of 0.00001. When the batch size was set to 64 for training and 32 for validation and testing, with a learning rate of 0.0001, ResNet50V2, MobileNetV2, and concatenation model 2 achieved the highest accuracy, reaching 97%, 97%, and 98.5%, respectively. Remarkably, both concatenation model 1 and concatenation model 2 demonstrate superior accuracy compared to ResNet50V2 and MobileNetV2 models. Figure 6 depicts the comparative performance of Precision, Recall, F1 score, Specificity, Sensitivity, AUC, and testing accuracy for all proposed models. Based on the results, it can be deduced that concatenation model 2 displayed a higher level of performance, achieving an overall accuracy rate of 98.5% compared to concatenation model 1 and all individual CNN models. The findings demonstrate that the overall accuracy of the proposed model increases significantly, reaching its peak accuracy when leveraging multiple features extracted from two robust networks, rather than relying solely on the features of each network individually. Furthermore, the optimization of training hyperparameters enhances the accuracy of the concatenated proposed model, improving it from 95.7% to 98.5%. In general, these architectures used within the scope of the study exhibited strong performance on the provided dataset, with some potentially showing slightly lower performance.

Table 4. Performance metrics for the proposed model with a batch size of 32 and a learning rate of 0.00001

Models\Metrics	ResNet50V2 model 1	MobileNetV2 model 1	Concatenation model 1
Precision	99.6%	99.3%	99.3%
Sensitivity	94.5%	95.1%	97.4%
Specificity	98.9%	98.2%	98.2%
F1 score	95.6%	97.1%	98.4%
AUC	96.7%	96.6%	97.8%
Accuracy	95.7%	95.9%	97.6%

Table 5. Performance metrics for the proposed model with a batch size of 64 and a learning rate of 0.0001

Models\Metrics	ResNet50V2 model 2	MobileNetV2 model 2	Concatenation model 2
Precision	99.6%	97.1%	99.5%
Sensitivity	96.4%	98.8%	98.4%
Specificity	98.9%	91.8%	98.6%
F1 score	97.9%	97.9%	99.1%
AUC	97.6%	95.3%	98.5%
Accuracy	97%	97%	98.5%

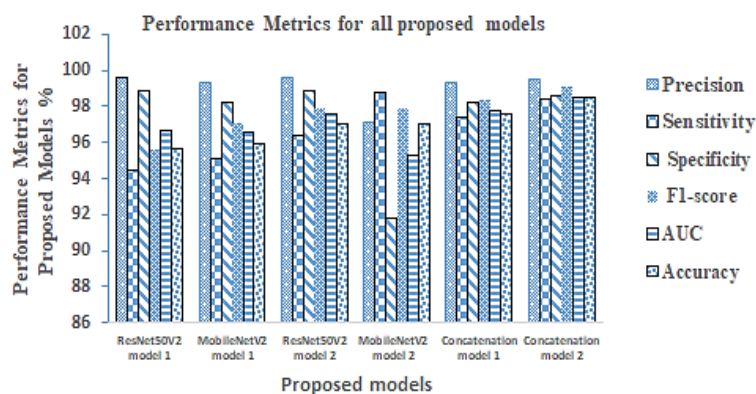


Figure 6. Comparison of performance metrics for all proposed models when trained with batch sizes and learning rates set to 32, 64, and 0.00001, 0.0001, respectively

### 3.3. Comparative analysis

In this section, we assess the efficacy of the proposed concatenation model for pediatric pneumonia detection and conduct a comparative assessment with state-of-the-art models. The concatenation model, which combines two pre-trained models (ResNet50V2 and MobileNetV2), is compared with previous related work conducted on the same dataset, as elaborated in Section 1. Table 6 and Figure 7 illustrate a comparison between the proposed concatenation model and recent existing models conducted on the [22] dataset. Notably, the concatenation model 2 demonstrated a remarkable testing accuracy of 98.5%, surpassing the concatenation model 1 and the state-of-the-art models. The optimal experimental outcome is attained by training the model with the Adam optimizer, utilizing a learning rate of 0.0001 and a batch size of 64. This configuration yields an accuracy of 98.5% on the test data, demonstrating the superior performance of the proposed concatenation model. Furthermore, the proposed concatenation model 2 demonstrates competitive performance with outstanding precision, sensitivity, specificity, F1 score, and an AUC curve of 99.5%, 98.4%, 98.6%, 99.1%, and 98.5%, respectively. These findings highlight the efficacy of the proposed approach in accurately detecting pediatric pneumonia from CXR images. The outcomes depicted in Table 6 and Figure 7 suggest that utilizing the concatenation of two pre-trained models (ResNet50V2 and MobileNetV2) through TL and optimized hyperparameters notably enhances the accuracy for pneumonia detection. ResNet50V2 and MobileNetV2 networks were selected in this study due to their capability to achieve high performance with fewer parameters and faster execution time. So, the concatenation of output features from both networks proves to be an effective strategy, improving the network's capability to classify pneumonia by leveraging information from diverse feature vectors and obtaining deep features, ultimately resulting in enhanced accuracy.

Although the proposed model for detecting pediatric pneumonia disease has been successfully implemented and analyzed, achieving highly satisfactory performance, it is essential to note that the evaluation is limited to this specific scenario. Therefore, the effectiveness of this method may not generalize well to other datasets. However, if other datasets share the same distribution characteristics and samples are obtained independently and identically, then the model is likely to perform effectively across those datasets as well.

Table 6. Comparative results for the obtained model with existing models

Models	Precision%	Sensitivity %	Specificity%	F1 score%	AUC%	Accuracy%
Stephen <i>et al.</i> [17]	-	-	-	-	-	93.7
El Asnaoui <i>et al.</i> [18]	98.5	94.9	98.4	96.7	-	96.6
Liang and Zheng [19]	89.1	96.7	-	92.7	-	90.5
El Zein <i>et al.</i> [20]	100	95.8	100	97.9	98	97.0
Güler <i>et al.</i> [21]	-	-	-	-	-	95.73
Concatenation model 1	99.3	97.4	98.2	98.4	97.8	97.6
Concatenation model 2	99.5	98.4	98.6	99.1	98.5	98.5

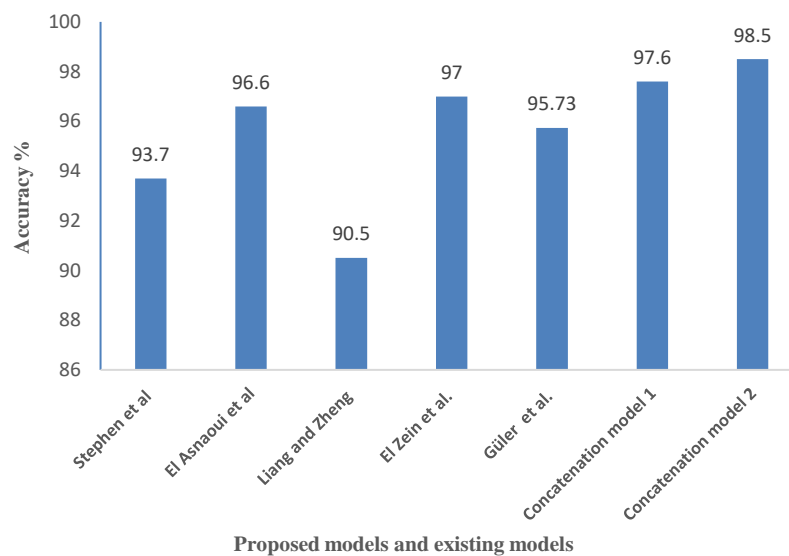


Figure 7. Comparison of the accuracy of the proposed concatenation model with existing models

#### 4. CONCLUSION

This research proposes an accurate concatenation model based on TL and optimizing the training hyperparameters that leverage the strengths of ResNet50V2 and MobileNetV2 architectures to detect pediatric pneumonia. The combined features extracted from ResNet50V2 and MobileNetV2 significantly enhance the network's performance to detect and classify pediatric pneumonia by leveraging deep information from both feature vectors. Also, different augmentation techniques are proposed for solving imbalanced and limited data. This study proved the importance and benefits of concatenating the output features from both networks with optimizing training parameters (such as batch size and learning rate). In this paper, we not only introduced a model for pediatric pneumonia detection but also explored, evaluated, and compared various architectural approaches to determine the most effective one. Our investigation yielded a highly satisfactory detection outcome. Comparing our model with others and state-of-the-art approaches, the proposed concatenation model 2 demonstrated superior performance and accuracy in pneumonia detection. The best experimental result was achieved by training the model using the Adam optimizer, with a learning rate of 0.0001 and a batch size of 64, resulting in an accuracy of 98.5% on the test data. This highlights the outstanding performance of our proposed Concatenation Model.

Although our study has demonstrated promising results in accurately detecting pediatric pneumonia diseases through the proposed concatenation method and optimized training parameters, several questions remain unanswered. One key inquiry pertains to the model's ability to achieve high accuracy on other medical disease datasets. Additionally, future research should focus on evaluating the model's performance on pneumonia and other real-world medical disease datasets to assess its adaptability to diverse medical conditions. These assessments will offer deeper insights into the potential and limitations of the proposed concatenation model, enabling further enhancements and advancements in medical disease detection.

#### ACKNOWLEDGEMENT

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1445).




#### REFERENCES

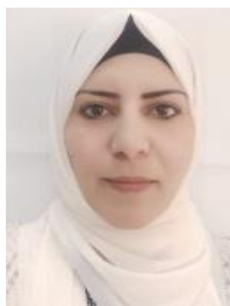
- [1] Z. Gilani *et al.*, "A literature review and survey of childhood Pneumonia etiology studies: 2000–2010," *Clinical Infectious Diseases*, vol. 54, no. suppl\_2, pp. S102–S108, Apr. 2012, doi: 10.1093/cid/cir1053.
- [2] V. Rajnikanth, S. Kady, R. Damasevicius, C. Pandeewaran, M. Abed Mohammed, and G. Glan Devdhas, "Pneumonia detection in chest x-ray using InceptionV3 and multi-class classification," in *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, Aug. 2022, pp. 972–976, doi: 10.1109/ICICICT54557.2022.9917698.
- [3] A. Du Toit, "Causes of severe pneumonia in children," *Nature Reviews Microbiology*, vol. 17, no. 9, Art. no. 529, Sep. 2019, doi: 10.1038/s41579-019-0245-y.
- [4] Y. Tian *et al.*, "The impact of ambient ozone pollution on pneumonia: a nationwide time-series analysis," *Environment International*, vol. 136, Mar. 2020, doi: 10.1016/j.envint.2020.105498.
- [5] A. P. J. *et al.*, "Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep CNN architectures," *Engineering Applications of Artificial Intelligence*, vol. 123, Aug. 2023, doi: 10.1016/j.engappai.2023.106416.
- [6] WHO, "Pneumonia in children," World Health Organization, 2022. <https://www.who.int/news-room/fact-sheets/detail/pneumonia> (accessed Nov. 11, 2022).
- [7] T. Cherian *et al.*, "Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies," *Bulletin of the World Health Organization*, vol. 83, pp. 353–359, 2005, doi: 10.1590/S0042-96862005000500011.
- [8] D. Poap, M. Wozniak, R. Damasevicius, and W. Wei, "Chest radiographs segmentation by the use of nature-inspired algorithm for lung disease detection," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Nov. 2018, pp. 2298–2303, doi: 10.1109/SSCI.2018.8628869.
- [9] M. I. Neuman *et al.*, "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," *Journal of Hospital Medicine*, vol. 7, no. 3, pp. 294–298, 2012, doi: 10.1002/jhm.955.
- [10] E. Ayan and H. M. Ünver, "Diagnosis of Pneumonia from chest x-ray images using deep learning," in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, Apr. 2019, pp. 1–5, doi: 10.1109/EBBT.2019.8741582.
- [11] P. S. Mann, S. D. Panchal, S. Singh, G. S. Saggi, and K. Gupta, "A hybrid deep convolutional neural network model for improved diagnosis of pneumonia," *Neural Computing and Applications*, vol. 36, no. 4, pp. 1791–1804, Feb. 2024, doi: 10.1007/s00521-023-09147-y.
- [12] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9.
- [13] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, Feb. 2020, doi: 10.1016/j.neucom.2019.10.008.
- [14] E. Ayan, B. Karabulut, and H. M. Ünver, "Diagnosis of pediatric Pneumonia with ensemble of deep convolutional neural networks in chest x-ray images," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 2123–2139, Feb. 2022, doi: 10.1007/s13369-021-06127-z.
- [15] K. Kassylkassova, B. Omarov, G. Kazbekova, Z. Kozhamkulova, M. Maikotov, and Z. Bidakhmet, "Automated Pneumonia diagnosis using a 2D deep convolutional neural network with chest x-ray images," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023, doi: 10.14569/IJACSA.2023.0140281.




- [16] D. S. Kermany *et al.*, “Identifying medical diagnoses and treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
- [17] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, “An efficient deep learning approach to Pneumonia classification in healthcare,” *Journal of Healthcare Engineering*, vol. 2019, pp. 1–7, Mar. 2019, doi: 10.1155/2019/4180949.
- [18] K. El Asnaoui, Y. Chawki, and A. Idri, “Automated methods for detection and classification Pneumonia based on x-ray images using deep learning,” *Electrical Engineering and Systems Science, Image and Video Processing*, pp. 257–284, 2021, doi: 10.1007/978-3-030-74575-2\_14.
- [19] G. Liang and L. Zheng, “A transfer learning method with deep residual network for pediatric pneumonia diagnosis,” *Computer Methods and Programs in Biomedicine*, vol. 187, Apr. 2020, doi: 10.1016/j.cmpb.2019.06.023.
- [20] O. Zein, M. Soliman, A. Elkholy, and N. Ghali, “Transfer learning based model for pneumonia detection in chest x-ray images,” *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 5, pp. 56–66, Oct. 2021, doi: 10.22266/ijies2021.1031.06.
- [21] O. Güler and K. Polat, “Classification performance of deep transfer learning methods for Pneumonia detection from chest X-Ray images,” *Journal of Artificial Intelligence and Systems*, vol. 4, no. 1, pp. 107–126, 2022.
- [22] D. Kermany, K. Zhang, and M. Goldbaum, “Labeled optical coherence tomography (oct) and chest x-ray images for classification,” *Mendeley data*, vol. 2, no. 2, 2018.
- [23] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Journal of Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/s40537-016-0043-6.
- [24] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, “Automated deep transfer learning-based approach for detection of COVID-19 infection in chest x-rays,” *IRBM*, vol. 43, no. 2, pp. 114–119, Apr. 2022, doi: 10.1016/j.irbm.2020.07.001.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in *European conference on computer vision*, 2016, pp. 630–645.
- [26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: inverted residuals and linear bottlenecks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
- [27] C. K. Dewa and Afiahayati, “Suitable CNN weight initialization and activation function for Javanese vowels classification,” *Procedia Computer Science*, vol. 144, pp. 124–132, 2018, doi: 10.1016/j.procs.2018.10.512.
- [28] R. Wali, “Xtreme Margin: a tunable loss function for binary classification problems,” *arXiv preprint arXiv:2211.00176*, 2022.
- [29] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Morgan Kaufmann, 2012.
- [30] Y. Kim, “Convolutional neural networks for sentence classification,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751, doi: 10.3115/v1/D14-1181.

## BIOGRAPHIES OF AUTHORS



**Ola M. El Zein**    received the B.Sc. degree in pure math and computer science, in 2010, and the M.Sc. and Ph.D. degrees in computer science from the Faculty of Science, Al-Azhar University, Cairo, Egypt, in 2018 and 2022, respectively. She is currently a lecturer at the Mathematics Department, Faculty of Science, Al-Azhar University. She is a supervisor of some master’s theses. Her research activities include data hiding and artificial intelligence, machine learning, deep learning, and image processing. She can be contacted at email: olaelzin@azhar.edu.eg.



**Naglaa E. Ghannam**    received the B.Sc. degree in science, in 2002, and the M.Sc. and Ph.D. degrees in computer science, in 2017 and 2020, respectively. She is currently a lecturer in computer science with the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. She has published several research papers in the field of machine learning, software engineering, software quality, and security. She is also a supervisor of some master’s theses. She can be contacted at email: naglaasaed@azhar.edu.eg.