Volume 13, Issue 1, Page 54-61, 2024 https://doi.org/10.46810/tdfd.1363218 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Detection of Pneumonia Using A Hybrid Approach Consisting of MobileNetV2 and Squeeze-and-Excitation Network

Hüseyin FIRAT[1], Hüseyin ÜZEN[2]

[1] Dicle University, Faculty of Engineering, Department of Computer Engineering, Diyarbakır, Türkiye
[2] Bingol University, Faculty of Engineering-Architecture, Department of Computer Engineering, Bingöl, Türkiye

Hüseyin FIRAT ORCID No: 0000-0002-1257-8518
Hüseyin ÜZEN ORCID No: 0000-0002-0998-2130

**Corresponding author: huseyin.firat@dicle.edu.tr*

(Received: 19.09.2023, Accepted: 10.02.2024, Online Publication: 26.03.2024)

Keywords
Deep Learning, Chest X-Ray Images, Pneumonia Detection, MobileNet V2, Squeeze-and-Excitation Network

Abstract: Pneumonia is a global health concern, responsible for a significant number of deaths. Its diagnostic challenge arises from visual similarities it shares with various respiratory diseases, such as tuberculosis, complicating accurate identification. Furthermore, the variability in acquiring and processing chest X-ray (CXR) images can impact image quality, posing a hurdle for dependable algorithm development. To address this, resilient data-centric algorithms, trained on comprehensive datasets and validated through diverse imaging methods and radiology expertise, are imperative. This study presents a deep learning approach designed to distinguish between normal and pneumonia cases. The model, a hybrid of MobileNetV2 and the Squeeze-and-Excitation (SE) block, aims to reduce learnable parameters while enhancing feature extraction and classification. Integration of the SE block enhances classification performance, despite a slight parameter increase. The model was trained and tested on a dataset of 5856 CXR images from Kaggle's medical imaging challenge. Results demonstrated the model's exceptional performance, achieving an accuracy of 98.81%, precision of 98.79%, recall rate of 98.24%, and F1-score of 98.51%. Comparative analysis with various Convolutional neural network-based pre-trained models and recent literature studies confirmed its superiority, solidifying its potential as a robust tool for pneumonia detection, thus addressing a critical healthcare need.

MobileNetV2 ve Sıkma-Uyarma Ağından Oluşan Hibrit Bir Yaklaşım Kullanılarak Pnömoni Tespiti

Anahtar Kelimeler
Derin Öğrenme, Göğüs Röntgeni Görüntüleri, Pnömoni Tespiti, MobileNet V2, Sıkma ve Uyarma Ağı

Öz: Pnömoni, önemli sayıda ölümden sorumlu olan küresel bir sağlık sorunudur. Teşhis zorluğu, tüberküloz gibi çeşitli solunum yolu hastalıklarıyla paylaştığı ve doğru tanımlamayı zorlaştıran görsel benzerliklerden kaynaklanmaktadır. Ayrıca, göğüs röntgeni görüntülerinin elde edilmesi ve işlenmesindeki değişkenlik, görüntü kalitesini etkileyerek güvenilir algoritma geliştirmenin önünde bir engel oluşturabilir. Bu sorunu çözmek için, kapsamlı veri kümeleri üzerinde eğitilen ve çeşitli görüntüleme yöntemleri ve radyoloji uzmanlığı ile doğrulanan esnek veri merkezli algoritmalar zorunludur. Bu çalışma, normal ve pnömoni vakalarını ayırt etmek için tasarlanmış bir derin öğrenme yaklaşımı sunmaktadır. MobileNetV2 ve Sıkıştırma-Uyarma (SU) bloğunun bir hibriti olan model, özellik çıkarma ve sınıflandırmayı geliştirirken öğrenilebilir parametreleri azaltmayı amaçlamaktadır. SU bloğunun entegrasyonu, hafif bir parametre artışına rağmen sınıflandırma performansını arttırmaktadır. Model, Kaggle'ın tıbbi görüntüleme yarışmasından alınan 5856 göğüs röntgeni görüntüsünden oluşan bir veri kümesi üzerinde eğitilmiş ve test edilmiştir. Sonuçlar, %98.81 doğruluk, %98.79 kesinlik, %98.24 geri çağırma oranı ve %98.51 F1 puanı elde ederek modelin olağanüstü performansını göstermiştir. Çeşitli Evrişimsel sinir ağı tabanlı önceden eğitilmiş modellerle ve son literatür çalışmalarıyla yapılan karşılaştırmalı analiz, modelin üstünlüğünü doğrulayarak pnömoni tespiti için sağlam bir araç olarak potansiyelini sağlamlaştırdı ve böylece kritik bir sağlık hizmeti ihtiyacını karşıladı.

1. INTRODUCTION

Pneumonia stands as a respiratory ailment provoking inflammation in either one or both lungs, leading to manifestations like coughing, fever, and respiratory distress. The timely identification of pneumonia proves crucial in ensuring effective therapy and enhanced patient recovery. Regrettably, pneumonia constitutes merely one among numerous pulmonary ailments, and as a result, radiographic findings don't consistently verify a pneumonia diagnosis. Consequently, given the present technology available, it remains unfeasible to definitively differentiate pneumonia from alternative lung disorders based on radiological criteria [1,2].

Developing accurate algorithms to detect pneumonia requires large amounts of meticulously annotated data, a task that can be particularly demanding. The complexity escalates when tackling pneumonia, as it necessitates skilled radiologists for data annotation, and the availability of labeled images is limited. Deep learning (DL), a subset of artificial intelligence, has emerged as a formidable tool for pneumonia identification and diagnosis through medical imaging, such as chest X-rays (CXR) [3,4]. DL algorithms possess the capability to undergo training on vast archives of CXR images, enabling them to discern unique features and attributes indicative of pneumonia's presence. This procedure involves utilizing convolutional neural networks (CNNs), a specialized category of DL frameworks exceptionally skilled in tasks related to image recognition. By scrutinizing the texture, shape, and pixel intensity within CXR images, CNNs can develop the ability to precisely identify areas within the image that correspond to regions of the lungs affected by infection or inflammation [5,6].

Following their training, DL models possess the capability to categorize novel CXR images, determining whether they exhibit indications of pneumonia or not. This process can be executed instantly, rendering it a potentially valuable resource for healthcare practitioners in the prompt diagnosis and treatment of pneumonia in patients. Furthermore, DL models can provide support to radiologists in the interpretation of CXR, thereby diminishing the chances of incorrect diagnoses and enhancing patient results [7-9]. Considering these advantages offered by DL techniques, they are commonly employed in scientific studies to diagnose pneumonia through the analysis of CXR images. Some of these studies are as follows.

Reshan et al. [2] introduced an advanced DL model for distinguishing between severe and normal cases of pneumonia. They harnessed eight pretrained models: ResNet50, ResNet152V2, DenseNet121, DenseNet201, Xception, VGG16, EfficientNet. To assess the efficiency of their model, they utilized the identical Kaggle dataset comprising 5856 CXR images. Their findings demonstrated that the MobileNet model yielded the highest accuracy, achieving an impressive 94.23%. In this research, Singh et al. [4] suggested a quaternion-based residual network to classify pneumonia cases in CXR images. They extended the use of the residual network into the quaternion realm to differentiate between CXR images as either indicative of pneumonia or normal. This approach yielded an accuracy of 93.75% and an F1-score of 0.94, outperforming the traditional real-numbered residual network in terms of performance. Szepesi et al. [10] presented a CNN framework crafted to ensure an accurate and efficient method for detecting pneumonia through the analysis of CXR images. A notable innovation in their approach was the integration of a dropout layer positioned amidst the convolutional layers within the network. Rather than relying on pre-existing networks, they constructed a CNN model from the ground up, incorporating models of transfer learning. The method under consideration underwent training and testing using a collection of 5856 annotated CXR images provided within the context of a medical imaging competition hosted on Kaggle's platform. As a consequence of the experimental investigations, accuracy level of 97.2% was achieved. Ahmad et al. [11] devised an approach to autonomously identify pneumonia by combining pre-trained ResNet and DenseNet169 models. They assessed the model's effectiveness using a set of 5856 CXR images from the well-balanced Kaggle platform. This method attained a 90% accuracy rate in its performance evaluation. Shah et al. [12] suggested an CNN method based on the VGG16 model. They tested the proposed method using 5856 CXR images. They achieved 96.6% accuracy with their proposed method. Stephen et al. [13] presented a CNN model specifically designed to classify and detect the presence of pneumonia in a dataset containing CXR images. They assessed the effectiveness of this innovative approach by employing a dataset consisting of 5856 CXR images, ultimately achieving a classification accuracy of 93.73%. Rajaman et al. [14] introduced a CNN-driven decision support platform for the rapid and precise identification of pneumonia in pediatric CXR. They employed innovative and cutting-edge visualization techniques to elucidate model predictions, which holds great importance in guiding clinical decision-making. Furthermore, the encouraging results exhibited by the tailored VGG16 model, trained on the present tasks, indicate its ability to effectively acquire knowledge from a limited set of intricate data, resulting in reduced bias and enhanced generalization capabilities. They tested the proposed method using 5856 CXR images. They achieved 96.2% accuracy with their proposed method. Toğaçar et al. [15] amalgamated profound attributes extracted from multiple deep models to create an effective deep attribute collection. Subsequently, they employed various algorithms like decision trees and k-nearest neighbors for categorizing this deep attribute ensemble. Additionally, they utilized methods such as linear discriminant analysis and k-nearest neighbors in the realm of pneumonia detection. They tested the proposed method using 5849 CXR images. They achieved 96.84% accuracy with their proposed method. Chouhan et al. [16] introduced a model structured around an ensemble architecture, integrating results from multiple pre-existing models to aid in the diagnosis of pneumonia. They achieved an

accuracy of 96.39% as a result of experimental studies using 5232 CXR images.

In this study, a hybrid MSENet model consisting of pre-trained MobileNetV2 and Squeeze-and-Excitation (SE) block is proposed to detect normal and pneumonia disease conditions. The purpose of the hybrid MSENet model is to improve the classification accuracy by decreasing the number of trainable parameters. The number of trainable parameters is reduced with the pre-trained MobileNetV2. Given that the SE block is tailored to enhance the representational quality of the CNN, it is incorporated into CNNs within this investigation to enhance both feature extraction and classification performance. The integration of SE block into the proposed method is driven by their ability to bolster classification performance with only minimal augmentation to the overall parameter count. To assess the performance of the MSENet model put forward, an examination was conducted using a dataset consisting of 5856 publicly accessible CXR images within the Kaggle platform. While MobileNetV2 achieved 97.78% accuracy, MSENet, which is a combination of MobileNetV2 and SE block, achieved 98.81% accuracy. MobileNetV2 has the fewest parameters compared to other pre-trained network architectures compared with 2.228.994 trainable parameters. MSENet has 2.228.995 trainable parameters. Compared to MobileNetV2, the number of trainable parameters increased by 1, while the classification accuracy increased by 1.03%. Besides, the MSENet model was compared with the pre-trained models and it was observed that it gave the best classification accuracy value. In addition, it was compared with methods from the literature using the same dataset and it was observed that the MSENet model gave the most successful results.

The remainder of this manuscript comprises three distinct sections. In Section 2, we introduce the CXR image dataset employed in this research. Also, within this section, we delve into the description of our MSENet approach, along with the theoretical foundation of MobileNetV2 and the SE block. Section 3 outlines the experiments carried out in conjunction with our proposed method. Section 4, Conclusions, provides a comprehensive overview of the entire paper.

2. MATERIAL AND METHOD

2.1. Dataset

The dataset employed in this research was sourced from one of Kaggle's DL competitions [17]. This dataset comprises lung CXR images of children aged one to five years old, sourced from Guangzhou Women and Children's Medical Center. These lung CXR images underwent validation by healthcare experts as part of standard patient care procedures. The dataset consists of a total of 5856 labeled images, with 4273 indicating cases of pneumonia and the remaining 1583 classified as negative cases (normal). All scans were represented as grayscale images, with dimensions varying from 1346x1044 to 2090x1858 pixels. To match the anticipated input format of most CNN network architectures, all images were resized to dimensions of 224x224x3. Figure 1 displays some sample images from the dataset.

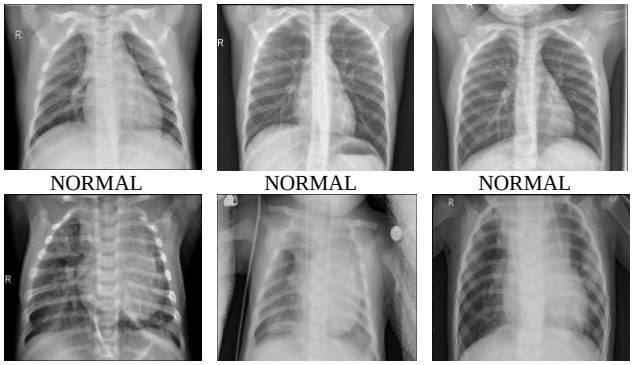

Figure 1. Samples of CXR images

2.2. MobileNetV2

MobileNetV2 stands out as a compact deep neural network, characterized not only by its reduced size but also its computational efficiency, all while delivering strong performance [18]. MobileNetV2 represents an enhanced iteration of MobileNetV1 [19], incorporating a linear bottleneck layer and the inverted residual block as additional components. The network architecture's specifications are detailed in Table 1. Within the inverted residual block (IRB), the initial step involves expanding the input channels via a 1x1 convolution to acquire additional features. Subsequently, feature extraction occurs through a 3x3 convolution, followed by channel reduction via 1x1 pointwise convolution. This entire sequence of operations follows the "expansion-convolution-compression" scheme, which proves to be more efficient and less reliant on parameters compared to the direct utilization of a 3x3 convolutional network.

The linear bottleneck (LB) layer substitutes the ReLU6 activation function within the second-to-last layer with a linear function. This change eliminates the ReLU6 activation between the high-dimensional and low-dimensional sections, effectively addressing the issue of ReLU6 discarding low-latitude information following the IRB. The core structure of MobileNetV2, known as the bottleneck residual block, is created by combining the IRB with the LB layer. When stride (s) equals 1, the shortcut is employed; when s equals 2, the shortcut is omitted, as depicted in Figure 2.

The model took as input the preprocessed chest X-ray images resized to 224x224x3. To start, a convolution operation with 32 channels and a stride of 2 was executed, resulting in a feature layer sized 112x112x32. Subsequently, after applying 7 bottleneck residual blocks, the feature layer expanded to 7x7x320. Following this, a 1x1 convolution operation was applied with 1280 channels and a stride of 1, resulting in a feature layer of 7x7x1280. Ultimately, a global average pooling operation was incorporated to diminish the feature dimensions from 7x7x1280 to a mere 1280, by

implementing a downsizing procedure following this feature map.

Table 1. Parameters defining the structural configuration of the MobileNetV2 feature extraction network

Input	Operator	t	c	n	s (stride)
224x224x3	Conv2D 3x3	-	32	1	2
112x112x32	Bottleneck	1	16	1	1
112x112x16	Bottleneck	6	24	2	2
56x56x24	Bottleneck	6	32	3	2
28x28x32	Bottleneck	6	64	4	2
14x14x64	Bottleneck	6	96	3	1
14x14x96	Bottleneck	6	160	3	2
7x7x160	Bottleneck	6	320	1	1
7x7x320	Conv2D 1x1	-	1280	1	1
7x7x1280	avgpool 7x7	-	-	1	-

Within Table 1, "t" represents the factor by which the initial 1x1 convolutional expansion channel is multiplied within each bottleneck residual block. "c" denotes the count of output feature layer channels, "n" signifies the frequency at which the present convolutional block is iterated, and "s" indicates the convolutional step size utilized within the current convolutional block.

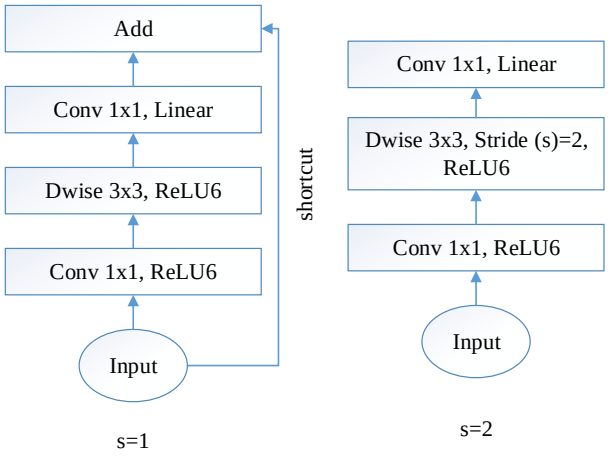

Figure 2. Bottleneck residual block in MobileNetV2

2.3. Squeeze-and-Excitation (SE) Block

The SE block offers a framework for CNNs that enhances channel relationships without substantially increasing computational overhead. This block effectively enhances vital feature information by recalibrating the input features it receives. Utilizing the SE block contributes to the enhancement of interdependencies among channels, thereby augmenting the relevance of feature information for chest X-ray image classification tasks. Figure 3 illustrates the structural outline of the SE block, which functions as follows: Initially, the SE block takes in a feature map along with its current channel count. Subsequently, employing Global Average Pooling (GAP), each channel is transformed into a single numerical value (referred to as "squeeze"). In this process, the feature maps associated with each channel are condensed into 1x1 feature maps using a channel descriptor technique like GAP. This stage produces a scalar value encapsulating general channel information. The primary objective of the "squeeze" operation is to establish a global receptive field, facilitating the utilization of global information even by the lower network layers. Following the "squeeze" operation, the "excitation" process comes into play, generating weights for individual feature channels based on specified parameters. These parameters are trained explicitly to capture the correlations existing among feature channels. The architecture incorporates two fully connected layers (FC layers) to manage the method's complexity, facilitate generalization, and introduce a bottleneck structure for modeling channel correlations. The initial FC layer serves to reduce the feature dimensions, which are then expanded back to their original size in the subsequent FC layer. Between these two FC layers, the ReLU activation function is employed to infuse non-linearity into the network, allowing it to better adapt to intricate channel correlations. Contrasted with a direct FC layer, the utilization of two FC layers enhances the model's non-linear characteristics, enabling improved accommodation of complex channel relationships while also reducing computational load and parameter count. After the FC-ReLU-FC sequence, the sigmoid function is invoked to compute normalized weights ranging between 0 and 1. Subsequently, a scaling procedure is applied to adjust the normalized weights in accordance with the characteristics of each channel. Importantly, these operations within the SE block incur minimal additional computational overhead. This block can be effortlessly integrated into various methods due to its capability to enhance classification performance [20-22].

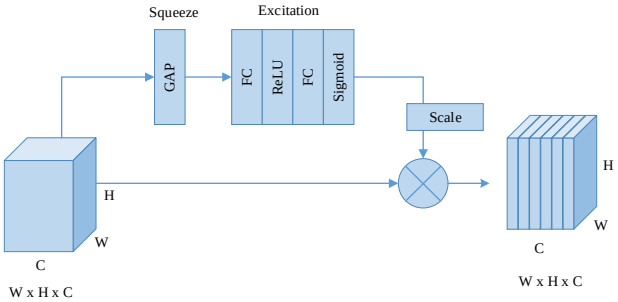

Figure 3. Squeeze-and-Excitation (SE) Block

2.4. Proposed MSENet Method

The proposed MSENet model is a hybrid model consisting of MobileNetV2 and SE block as shown in Figure 4. MobileNetV2 aims to decrease the number of trainable parameters. As the SE block is intended to enhance the representational quality of a CNN, this research combines it with MobileNetV2 to enhance both feature extraction and classification performance. Besides, the SE block is integrated into the proposed model as it increases the classification accuracy by minimizing the total number of parameters.

The input image from the proposed MSENet model is 224x224x3.The MobileNetV2 model was first applied to this image. In the MobileNetV2 model, a convolution process is first applied with a kernel size of 3x3 and a stride value of 2. After this convolution, a feature map of size 112x112x32 is obtained. Then, after applying bottleneck residual blocks seven times in a row, a feature map of size 7x7x320 is obtained. The image size given to the input of each bottleneck residual blocks and the

image size obtained at the output are given in Table 1. After the bottleneck residual blocks, a convolution process of 1x1 size and 1280 filters is implemented to the 7x7x320 feature map. The size of the feature map obtained after this process is 7x7x1280. A global mean pooling process is then applied to this feature map followed by a downscaling procedure to reduce the feature sizes from 7x7x1280 to only 1280. After the operations performed in the MobileNetV2 feature extraction block, the output is given to the input of the SE block. The SE block consists of Squeeze, Excitation and Scaling stages as shown in Figure 3. After these operations are performed in the SE block, GAP and then batch normalization (BN) operations are performed. GAP is a pooling technique designed to substitute the FC layers typically found in standard CNNs. Using GAP, a separate feature map is generated for each category in the classification task's final layer. Instead of introducing FC layers atop these feature maps, each map undergoes averaging, resulting in a vector that is directly passed to the softmax layer. Furthermore, the incorporation of GAP in the proposed method offers the advantage of not having any parameters to optimize in the GAP layer, effectively preventing overfitting. The output of the GAP layer is fed into a softmax function for feature classification and the prediction of CXR images. However, prior to softmax, BN is applied to streamline and accelerate the training process. Finally, Softmax assigns probabilities to each class and the sum of these probabilities equals one.

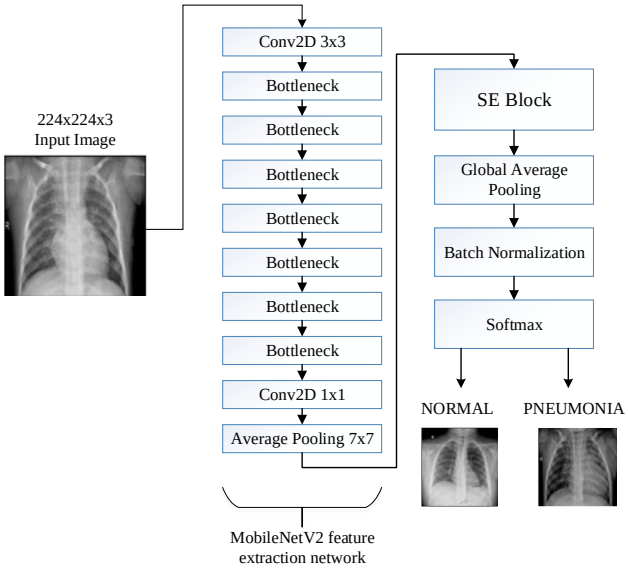

Figure 4. Proposed MSENet model

3. RESULTS OF EXPERIMENTS

In this section, the hyperparameters used in the experimental studies, the evaluation criteria used, the classification results obtained from the experimental studies and the discussion of these results are presented.

3.1. Parameter Setting

A specific set of hyperparameters is used to train the model using the Keras and TensorFlow library in the kaggle environment. Experimental studies for the CXR images dataset were performed with TPU VM v3-8, a hardware accelerator in the kaggle environment. Overall, we expect that the use of hyperparameters, optimizers and callbacks in combination with Keras-TensorFlow and the Kaggle environment will allow us to achieve state-of-the-art results with our proposed method architecture. The hyperparameters used include batch size, image size, training-test-validation separation, number of epochs. To train the proposed method, a batch size of 128, an image size of 224x224x3 and a training-test-validation separation of 80%-10%-10% are used. Out of 5856 CXR images, 4684 images were used for training, 586 for validation and 586 for testing. In addition, the model is trained for 100 epochs. Adam optimizer is used to minimize the loss function and optimize the model. In addition to the hyperparameters, two special callbacks are used to optimize the training process. The first callback is the ReduceLROnPlateau callback, which is used to reduce the learning rate when the loss of validation stops the recovery. This helps to stabilize the training process and avoid overlearning. In this callback, 0.000001 is taken as the lower bound of the learning rate (min_learning_rate). Also, the factor value to reduce the learning rate is 0.3. The second callback is the ModelCheckpoint callback, which is used to record model weights at regular intervals during training. This allows us to save the best model based on validation accuracy and load it for future use.

3.2. Performance Metrics

The effectiveness of the proposed approach is gauged by employing performance metrics like precision, F1-score, recall, and classification accuracy. Performance metrics provide a quantitative and objective measure of the effectiveness of a model's predictions. They are also necessary to evaluate classification performance. These metrics provide different perspectives on the performance of the model, each with its own strengths and limitations. A detailed description of these metrics is as follows. Accuracy, a key performance evaluation metric, measures the percentage of correct estimations produced by the method. It is determined as the number of correct estimations divided by the total number of estimations made. The computation of the accuracy value is done as detailed in Equation (1). Precision, a measure of the proportion of true positives in all positive estimations made by the proposed method, is determined as the number of true positives divided by the sum of true positives and false positives. The calculation of the precision value is as in Equation (2). Recall, a metric that measures the proportion of true positives among all true positive samples in the dataset, is determined as the number of true positives divided by the sum of true positives and false negatives. Recall is formulated as in Equation (3). The F1 score, which is the harmonic mean of recall and precision, is an indispensable metric for balancing recall and precision, especially when classes are unbalanced. It ensures a single score that captures both precision and recall, making it a powerful measure for overall model performance evaluation [5,9,23]. The F1 score is computed as outlined in Equation (4) .

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

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

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

$$F1 - score\ (F1S) = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

All these performance metrics are deduced from the confusion matrix. An illustrative representation of a confusion matrix can be observed in Figure 5. Within a standard confusion matrix, there exist four primary elements: True Positives (TP): These denote cases where the genuine category is affirmative (for example, class 1), and the model accurately forecasts it as affirmative. True Negatives (TN): These represent instances in which the true category is negative (e.g., class 0), and the model correctly anticipates it as negative. False Positives (FP): These instances arise when the true category is negative, but the model incorrectly anticipates it as affirmative. False Negatives (FN): These arise from situations where the authentic category is affirmative, yet the model erroneously predicts it as negative.

		Actual	
		Positive (P)	Negative (N)
Prediction	Positive (P)	TP	FP
	Negative (N)	FN	TN

Figure 5. Confusion matrix

3.3. Experimental Studies and Results

In order to analyze the classification performance of the proposed MSENet model, extensive experimental studies were performed on a 2-class (Normal and Pneumonia) CXR images dataset. The confusion matrix of the MSENet model is given in Figure 6. In the test dataset, 159 out of 164 Normal CXR images and 420 out of 422 Pneumonia CXR images were correctly predicted. It is seen that a total of 579 images are correctly predicted in all classes from a total of 586 test datasets. Based on these results in the confusion matrix, the classification accuracy of the proposed MSENet model was obtained as 98.81%.

The MobileNetV2 model used in the proposed MSENet model is a pre-trained model. Accordingly, the MSENet model is compared with different pre-trained models and the results are given in Table 2. Table 2 shows that 98.81% accuracy, 98.79% precision, 98.24% recall and 98.51% F1-score were obtained with the proposed MSENet model. The closest accuracy, precision and F1-score values to the proposed method are obtained with the Xception model with 98.12%, 98.05% and 97.49%, while the closest recall value is obtained with the MobileNetV2 model with 97.22%. The lowest accuracy value was obtained in VGG16 with 96.42%, the lowest precision value was obtained in EfficientNetB0 with 95.94%, the lowest recall value was obtained in VGG16 with 94.32% and the lowest F1-score value was obtained in VGG16 with 95.61%. Considering all the results in Table 2, it is seen that the proposed MSENet model is more successful than all methods. Considering the parameter numbers of each model, it is seen that the lowest number of parameters is obtained with MobileNetV2. With MobileNetV2, 97.78% accuracy, 97.02% precision, 97.22% recall and 97.12% F1 score values were found. In the proposed MSENet model, the SE block was added to the MobileNetV2 model. SE block increased the number of parameters by 1. However, it increased the accuracy by 1.03%, precision by 1.77%, recall by 1.02% and F1 score by 1.39%.

The proposed MSENet model was compared with different studies from the literature using the same dataset and the results are given in Table 3. The proposed MSENet model achieved 98.81% classification accuracy. The MobileNet model developed by Reshan et al. [2] achieved an accuracy of 94.23%, while the Quaternion CNN model developed by Singh et al. [4] reached 93.75%. Szepesi et al. [10] achieved a 97.2% accuracy using CNN with modified dropout. Ahmad et al. [11] attained 90% accuracy with the Pre-trained ResNet and DenseNet169 models, whereas Shah et al. [12] obtained 96.6% accuracy with the Efficient VGG16 model. Stephen et al. [13] achieved an accuracy of 93.73% using their CNN model when compared to other models. It is evident that the MSENet model introduced in this study outperforms all other models in terms of success.

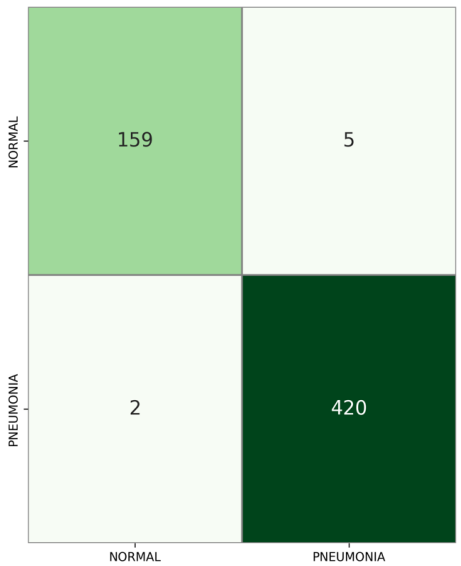

Figure 6. Confusion matrix of the proposed MSENet model

Table 2. Comparison of different pre-trained models (%)

Model	Acc	P	R	F1S	#Params
VGG19	96.42	96.94	94.32	95.61	20.026.434
Xception	98.12	98.05	96.94	97.49	20.815.146
ResNet50	97.10	96.43	95.78	96.10	23.542.786
DenseNet121	97.10	96.34	96.14	96.24	6.957.954
NasNetMobile	97.61	96.85	96.85	96.85	4.237.204
EfficientNetB0	96.93	95.94	96.3	96.12	4.012.670
MobileNet	97.44	97.23	96.29	96.76	3.211.074
MobileNetV2	97.78	97.02	97.22	97.12	2.228.994
MSENet	**98.81**	**98.79**	**98.24**	**98.51**	**2.228.995**

Table 3. Comparison results with studies in the literature using the same dataset (%)

Study in the Literature	Model	Acc
Reshan et al. [2]	MobileNet	94.23
Singh et al. [4]	Quaternion CNN	93.75
Szepesi et al. [10]	CNN + modified dropout	97.2
Ahmad et al. [11]	Pre-trained ResNet and DenseNet169	90
Shah et al. [12]	Efficient VGG16	96.6
Stephen et al. [13]	CNN	93.73
Proposed MSENet Model	**MobileNetV2 and SE block**	**98.81**

4. CONCLUSION

Within the scope of this study, a new DL-based hybrid method for pneumonia detection using CXR images is proposed. The proposed method consists of a combination of pre-trained MobileNetV2 and SE block (MSENet). The primary objective of this proposed MSENet model is to enhance classification accuracy while simultaneously reducing the number of learnable parameters. The utilization of pre-trained MobileNetV2 effectively reduces the count of learnable parameters, a signifcant aspect of our approach. The SE block is specifically tailored to augment the quality of representation within CNNs, and it is seamlessly integrated into our investigation to amplify both feature extraction and classification performance. The integration of the SE block into our proposed MSENet model is driven by its capability to enhance classification performance with only marginal adjustments to the overall parameter count. To assess the performance of the MSENet model proposed in this study, we employed a dataset consisting of 5856 publicly available CXR images within the Kaggle environment. MobileNetV2 achieved an accuracy of 97.78%, whereas the combined MSENet, comprising MobileNetV2 and the SE block, achieved an accuracy of 98.81%. It's worth noting that MobileNetV2 boasts the lowest parameter count compared to other pre-trained network architectures, with a mere 2.228.994 trainable parameters. In contrast, the MSENet incorporates 2.228.995 trainable parameters. This results in a minor increase of just 1 parameter compared to MobileNetV2, while concurrently achieving a notable increase in classification accuracy by 1.03%. Furthermore, when compared to various pre-trained models, our proposed MSENet demonstrated the highest classification accuracy. Additionally, in a comparative analysis with existing literature methods on the same dataset, the MSENet approach exhibited the most remarkable results.

REFERENCES

[1] Hu Z, Yang Z, Lafata KJ, et al. A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images. Med Phys. 2022; 49: 3213–3222.

[2] Reshan MS Al, Gill KS, Anand V, et al. Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model. Healthc; 11. Epub ahead of print 2023. DOI: 10.3390/healthcare11111561.

[3] Jaiswal AK, Tiwari P, Kumar S, et al. Identifying pneumonia in chest X-rays: A deep learning approach. Meas J Int Meas Confed. 2019; 145: 511–518.

[4] Singh S, Tripathi BK. Pneumonia classification using quaternion deep learning. Multimed Tools Appl. 2022; 81: 1743–1764.

[5] Zhang D, Ren F, Li Y, et al. Pneumonia detection from chest x-ray images based on convolutional neural network. Electron; 10. Epub ahead of print 2021. DOI: 10.3390/electronics10131512.

[6] Kundu R, Das R, Geem ZW, et al. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PLoS One; 16. Epub ahead of print 2021. DOI: 10.1371/journal.pone.0256630.

[7] Mercaldo F, Belfiore MP, Reginelli A, et al. Coronavirus covid-19 detection by means of explainable deep learning. Sci Rep. 2023; 13: 1–11.

[8] Ayan E, Ünver HM. Diagnosis of pneumonia from chest X-ray images using deep learning. 2019 Sci Meet Electr Biomed Eng Comput Sci EBBT 2019. İstanbul: IEEE; 2019; p. 0–4.

[9] Sharma S, Guleria K. A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks. Procedia Comput Sci. 2023; 218: 357–366.

[10] Szepesi P, Szilágyi L. Detection of pneumonia using convolutional neural networks and deep learning. Biocybern Biomed Eng. 2022; 42: 1012–1022.

[11] Al-Taani AT, Al-Dagamseh IT. Automatic Detection of Pneumonia Using Concatenated Convolutional Neural Network. Jordanian J Comput Inf Technol. 2023; 9: 118–136.

[12] Shah U, Abd-Alrazeq A, Alam T, et al. An efficient method to predict pneumonia from chest X-rays using deep learning approach. Stud Health Technol Inform 2020; 272: 457–460.

[13] Stephen O, Sain M, Maduh UJ, et al. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. J Healthc Eng; 2019. Epub ahead of print 2019. DOI: 10.1155/2019/4180949.

[14] Rajaraman S, Candemir S, Kim I, et al. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. Appl Sci; 8. Epub ahead of print 2018. DOI: 10.3390/app8101715.

[15] Toğaçar M, Ergen B, Cömert Z, et al. A Deep Feature Learning Model for Pneumonia Detection

Applying a Combination of mRMR Feature Selection and Machine Learning Models. Irbm. 2020; 41: 212–222.
[16] Chouhan V, Singh SK, Khamparia A, et al. A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci; 10. Epub ahead of print 2020. DOI: 10.3390/app10020559.
[17] Mooney P. Chest X-Ray Images (Pneumonia). Kaggle, https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia (accessed 10 September 2023).
[18] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2018. Salt Lake City: IEEE; 2018. p. 4510–4520.
[19] Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, http://arxiv.org/abs/1704.04861 (2017).
[20] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2018. Salt Lake City: IEEE; 2018. p.7132–7141.
[21] Fırat H. Sıkma - Uyarma Artık Ağı kullanılarak Beyaz Kan Hücrelerinin Sınıflandırılması. Bilişim Teknol Derg. 2023; 16: 189–205.
[22] Asker ME. Hyperspectral image classification method based on squeeze-and-excitation networks, depthwise separable convolution and multibranch feature fusion. Earth Sci Informatics. 2023; 1427–1448.
[23] Dayı B, Üzen H, Çiçek İB, et al. A Novel Deep Learning-Based Approach for Segmentation of Different Type Caries Lesions on Panoramic Radiographs. Diagnostics. 2023; 13: 202.