Neural Computing and Applications (2024) 36:5529–5544
https://doi.org/10.1007/s00521-023-09365-4

ORIGINAL ARTICLE

Comprehensive comparison of modified deep convolutional neural networks for automated detection of external and middle ear conditions

Kemal Akyol[1]

Received: 9 March 2023 / Accepted: 7 December 2023 / Published online: 10 January 2024

Abstract
Otitis media disease, a frequent childhood ailment, could have severe repercussions, including mortality. This disease induces permanent hearing loss, commonly seen in developing countries with limited medical resources. It is estimated that approximately 21,000 people worldwide die from reasons related to this disease each year. The main aim of this study is to develop a model capable of detecting external and middle ear conditions. Experiments were conducted to find the most successful model among the modified deep convolutional neural networks within two scenarios. According to the results, the modified EfficientNetB7 model could detect normal, chronic otitis media, earwax, myringosclerosis cases with high accuracy in Scenario 2. This model offers average values of 99.94% accuracy, 99.86% sensitivity, 99.95% specificity, and 99.86% precision. An expert system based on this model is expected to provide a second opinion to doctors in detecting external and middle ear conditions, particularly in primary healthcare institutions and hospitals lacking field specialists.

Keywords Ear conditions · Modified deep convolutional neural networks · Modified EfficientNetB7

1 Introduction

Otitis media, a common [1, 2] and contagious disease [2] in childhood, frequently results in conductive hearing loss, which can impair speech, language, and cognitive development [3, 4]. The prevalence of this disease, which causes permanent hearing loss, is increased by a lack of medical facilities, particularly in developing countries [5]. It is estimated that each year, approximately 21,000 people die due to complications associated with otitis media, with the highest mortality rate occurring in the first year of life [6]. A field specialist diagnoses otitis media by examining the eardrum with an otoscope device [7, 8]. Otoscopic examination is a valuable and necessary test that can accurately and efficiently distinguish tympanic membrane conditions. Endoscopy has been used in auto-neurology clinics and minor ear surgery since the 1990s [9]. The tympanic membrane separates the external auditory canal and middle ear, collects sound from the auricle and external auditory canal, and transmits it to the ossicles and cochlea via mechanical vibration [10]. One of the most frequent issues addressed by doctors delivering primary care to children and adolescents is middle ear illnesses, whose diagnosis is frequently delayed or misdiagnosed [11]. Due to the device's limited availability and high cost, expert decision support systems are needed for diagnosing otitis media, particularly in developing countries [12]. Furthermore, different subjective interpretations may result from the field expert's visual examinations [8]. False diagnosis usually occurs when the doctor is unpracticed in examining otoscopy or auto-endoscopy, leading to treatment delays or complications [13]. If used correctly in healthcare, artificial intelligence can reduce the strain on healthcare personnel while improving the quality of work by eliminating errors and boosting precision [14]. Recently, machine learning and its subfield, deep learning, have been prominent in many fields, such as speech recognition [15–17], computer vision [18–20], emotion analysis [21–23], and next-cart recommendation [24–26]. Experiments for solving many problems revealed that traditional handcrafted feature

✉ Kemal Akyol
kakyol@kastamonu.edu.tr

[1] Department of Computer Engineering, Kastamonu University, Kastamonu, Turkey

extraction approaches are less effective and time-consuming than deep learning-based approaches. In addition, handcrafted feature extraction is described as a low-level method in the literature [27]. Expert knowledge of the handcrafted techniques for obtaining the features required for solving a machine learning problem and a detailed understanding of the task definition is needed [28]. Furthermore, within-class variations and class similarities make achieving high classification accuracy difficult [29]. Deep learning models trained on large image datasets could detect diseases accurately and quickly. Especially considering that many image samples are given to a few field experts, deep learning presented promising solutions for detecting and treating diseases in the medical field. Computer-assisted tympanic membrane disease diagnosis systems are required to eliminate adverse effects and high costs caused by otolaryngologists' subjective decisions during visual examinations. A diagnostic support system would be invaluable for a general practitioner or pediatrician to accurately detect eardrum abnormalities, provide correct treatment and appropriate care, and avoid unnecessary antibiotic usage [30]. With image analysis, artificial intelligence has the potential to help clinicians identify and classify middle ear illnesses [11].

Many deep learning approaches are available to detect medical diseases. This suggests convolutional neural networks (CNN)-derived features offer more meaningful insights than handcrafted features. CNN are actively used for image-based subjects, particularly disease classification from medical images and diseased region segmentation. Some deep learning-based studies that tackled the tympanic membrane conditions are as follows: Khan et al. used CNN models to classify tympanic membrane and middle ear infections and achieved a 95% accuracy rate [10]. Chen et al. introduced a deep learning model that includes image preprocessing and data augmentation on a dataset composed by otolaryngologists at a hospital in Taiwan to detect and classify middle ear diseases. The authors used a class activation map to reveal key features in the images [11]. Myburgh et al. proposed a neural network for automatically diagnosing middle ear pathology or otitis media with smartphones and achieved an average classification accuracy of 86.84% [12]. Cha et al. presented an ensemble classifier with the best-performing Inception-V3 and ResNet101 CNN models to detect six ear diseases automatically and achieved 93.67% accuracy [31]. Başaran et al. detected the tympanic membrane regions on the augmented images dataset using the faster regional convolutional neural network. Then, the authors examined the performances of the pre-trained deep CNN models for the tympanic membrane conditions on the original and patch images they obtained. According to their results, the VGG-16 model presented an accuracy of 90.48% on the patch images set [32]. Lee et al. presented a CNN-based model to classify tympanic membrane conditions accurately. Their model showed 97.9% accuracy in detecting the tympanic membrane conditions and 91.0% in detecting the presence of perforation [33]. Table 1 summarizes related works in the literature.

The main motivation of this study is to determine a highly accurate model that can be used in high-capacity decision-support systems. With this motivation, the current study comprehensively compares the efficiencies of deep learning strategies for detecting middle ear disease. This paper focuses on the modified fully connected layers of pre-trained models according to two scenarios to detect middle ear disease with low misclassification rates. This research study presents the automatic classification of chronic otitis media, earwax, myringosclerosis, and normal cases with this perspective. Experimental studies include the performances of modified cutting-edge pre-trained deep CNN models. The visual feature extraction and classification abilities of modified cutting-edge deep learning models for external and middle ear conditions are examined and compared. Cutting-edge deep learning models, previously trained on large-scale data sets, are used to solve different problems with the transfer learning approach. Transfer learning is used in many problems, such as electromyographic hand gesture signal classification [34], COVID-19 pandemic and the Zika epidemic predictions [35], vehicle classification [36], and regression problems [37]. Since the modified EfficientNetB7 model offered significantly accurate classification in Scenario 2, this model could be used to detect E, COM, and M diseases in the real world. In this way, the workload of healthcare workers can be alleviated, and healthcare equipment can be optimized. Overall, the contributions of this study are as follows:

- A reliable method is deployed to classify external and middle ear conditions.
- The efficiencies of modifications in the classification part of the pre-trained deep CNN models are analyzed and compared.
- Empirical evidence shows that modified deep CNN models in Scenario 2 reduce the number of false positives and false negatives.

The remainder of this paper is organized as follows. Section 2 describes the dataset and gives information about the methodology used in this study. Section 3 presents the experimental studies in detail. Section 4 discusses the results, and Sect. 5 concludes this study with final remarks.

Table 1 Some related works for ear disease classification

Study	Method	Acc (%)
Khan et al. [10]	CNN model	95
Chen et al. [11]	Class activation map + deep learning	97.6
Myburgh et al. [12]	Neural network	86.84
Cha et al. [31]	Ensemble classifier (Inception-V3 + ResNet101)	93.67
Başaran et al. [32]	Region detection and VGG-16	90.48
Lee et al. [33]	CNN-based model	97.9

2 Material and methods

This study classified external and middle ear conditions with modified deep CNN models. Experiments were conducted to examine the effects of modifications performed on the fully connected layers of the cutting-edge pre-trained deep CNN architectures such as ResNet50, DenseNet121, and EfficientNetB7. Among these models, the original DenseNet extracts feature maps with a 16-block convolution process followed by batch normalization and the ReLU activation function. Original ResNet50 implements a 3-block convolution process and then performs an activation process. Original EfficientNetB7 extracts feature maps with a 7-block convolution process followed by batch normalization and an activation function. The convolution blocks of these architectures are different from each other. The classification part of the original EfficientNetB7 has global average pooling and dropout. The classification part of the original DenseNet121 and ResNet50 has global average pooling. The classification parts of these models were replaced according to Scenario 1 and Scenario 2. Accordingly, experimental studies were carried out by modifying all models under equal rules within the framework of two scenarios. The classification parts of these models constitute one flattened layer and one output layer, respectively, in Scenario 1. On the other hand, the classification parts of these models constitute one global maximum pooling layer, one hidden layer, one dropout layer, and one output layer sequentially in Scenario 2. Figure 1a, b presents these modifications applied to the pre-trained CNN models in two scenarios. The abbreviations N, COM, E, and M in the output layer of this figure refer to normal, chronic otitis media cases, earwax, and myringosclerosis cases, respectively. Each model was trained on tympanic membrane images captured with digital video-otoscopes. Then, the performances of these models in classifying the test images into normal class, chronic otitis media, earwax, and myringosclerosis cases were measured.

The following subsections provide detailed information about the dataset, pre-trained deep CNN models, and metrics used to evaluate model performance.

2.1 Dataset

The Ear Imagery dataset [38] contains 224 × 224 × 3 middle ear images in RGB color space belonging to 180 patients. The publicly available dataset includes four conditions: earwax plug, myringosclerosis, chronic otitis media, and normal otoscopy. The publishers split the dataset into training-validation and testing subsets. The dataset is balanced with 180 and 40 samples for each category in the training and testing sets, respectively. The testing set, including each class, represents 20% of the dataset. Also, 80% of the samples in the training set were reserved for training the models and the rest for validating the models. Hence, the number of samples for each class in the validation set is 36. Table 2 summarizes the number of samples of the training, validation, and testing sets, and Fig. 2 shows sample images representing each class in the dataset.

2.2 Convolutional neural networks

Deep learning, one of the most important advances in artificial intelligence, is especially successful in image processing. Using new concepts in the design of CNN architecture has increased deep learning applications in medical image processing, image identification, and data classification tasks [39]. By eliminating the drawbacks of handcrafted features used in machine learning, CNN architecture gave rise to tremendous improvements in computer vision [40]. This architecture comprises convolution filters, pooling layers, and fully connected layers [41]. Convolution and pooling processes are implemented in hidden layers of the CNN model, which is commonly employed for image-processing tasks [42]. The general structure of a CNN model is as follows: A feature map (visual features) is a result image representing features convoluted by kernels in the CNN to which an image is sent as input. Feature maps extracted from convolution and other sequential layers are used in the classification stage of a CNN model [43]. In brief, the images are sent to a CNN in the first phase, and visual features are automatically extracted in the second phase. Then, the feature vectors are given as input to the fully connected layers, and the

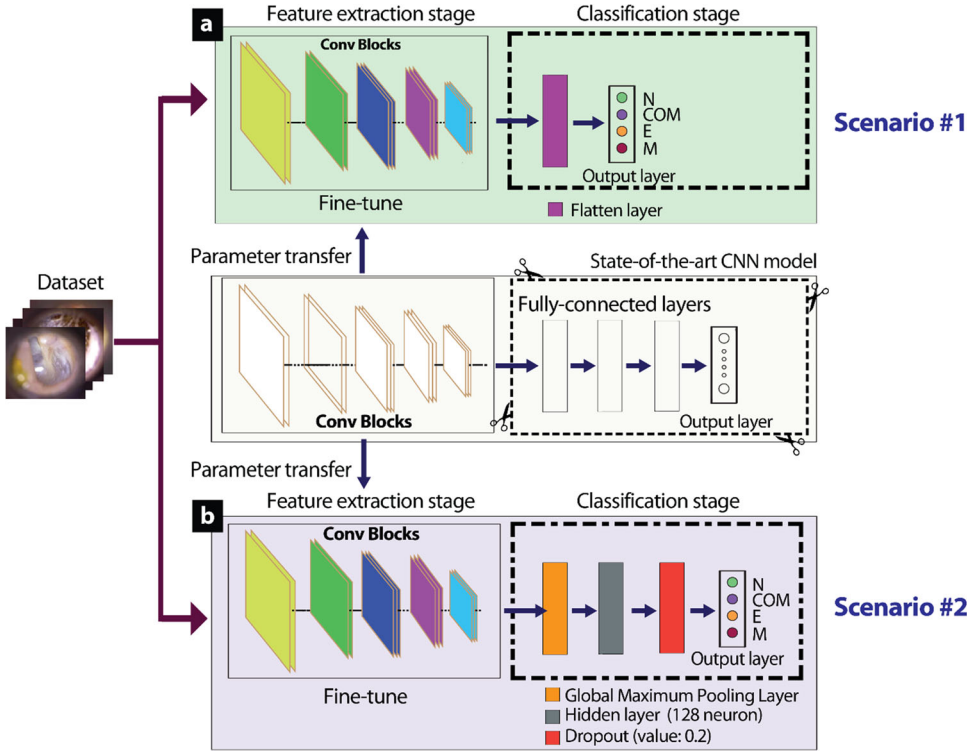

Fig. 1 Graphical representations of the modifications in the classifier part of a pre-trained CNN model

Table 2 Information about training, validation, and testing sets used in experiments

		N	COM	E	M	Total
Training-validation	Training set	144	144	144	144	576
	Validation set	36	36	36	36	144
Testing		40	40	40	40	160
	Total	220	220	220	220	880

classification process is carried out. This architecture is widely used in image classification, face detection, object detection, image noise removal, and other fields [41].

2.3 Transfer learning

Transfer learning is a machine learning method that reuses a previously trained model to perform a new task. By nature, a person who knows how to ride a bicycle will adapt to riding a motorcycle faster and easier than someone who has never ridden a bike before. It is possible to give similar samples. A person uses the information he previously learned in another process. Inspired by this, the model information obtained from the first task is transferred to the second model that focuses on the new task with the transfer learning approach. The basic idea behind transfer learning is to apply knowledge gained from previous experiences to new situations [44]. The performance and efficiency of machine learning can be considerably increased by transferring data from one task or area to another. The number of neurons of the output layer is set to the number of target classes in the dataset, while other layers of the pre-trained deep CNN model are preserved [36]. With this approach, the CNN model that was trained before on the ImageNet dataset is used with its current best weights for another problem. When there are not enough samples for the target class, transfer learning is applied that uses the knowledge obtained from the source dataset to improve learning efficiency [45]. As a result, a small dataset allows us to see the benefit of pre-trained deep CNN. The pre-trained deep CNN models presented high success on different-size new datasets in [37, 44, 46–48]. Numerous pre-trained deep CNN architectures are available, including ResNet50, DenseNet121, and EfficientNetB7.

2.3.1 Residual neural network-50 (ResNet50)

ResNet50 is one of ResNet architectures with 50 deep layers, consisting of convolutional block arrays and average pooling. Softmax function is used in the output layer for classification. Most ResNet architectures employ nonlinear (ReLU) and batch normalization with double or triple-layer skips [49]. ResNet50 first performs convolution of input data, then four residual blocks, and finally,

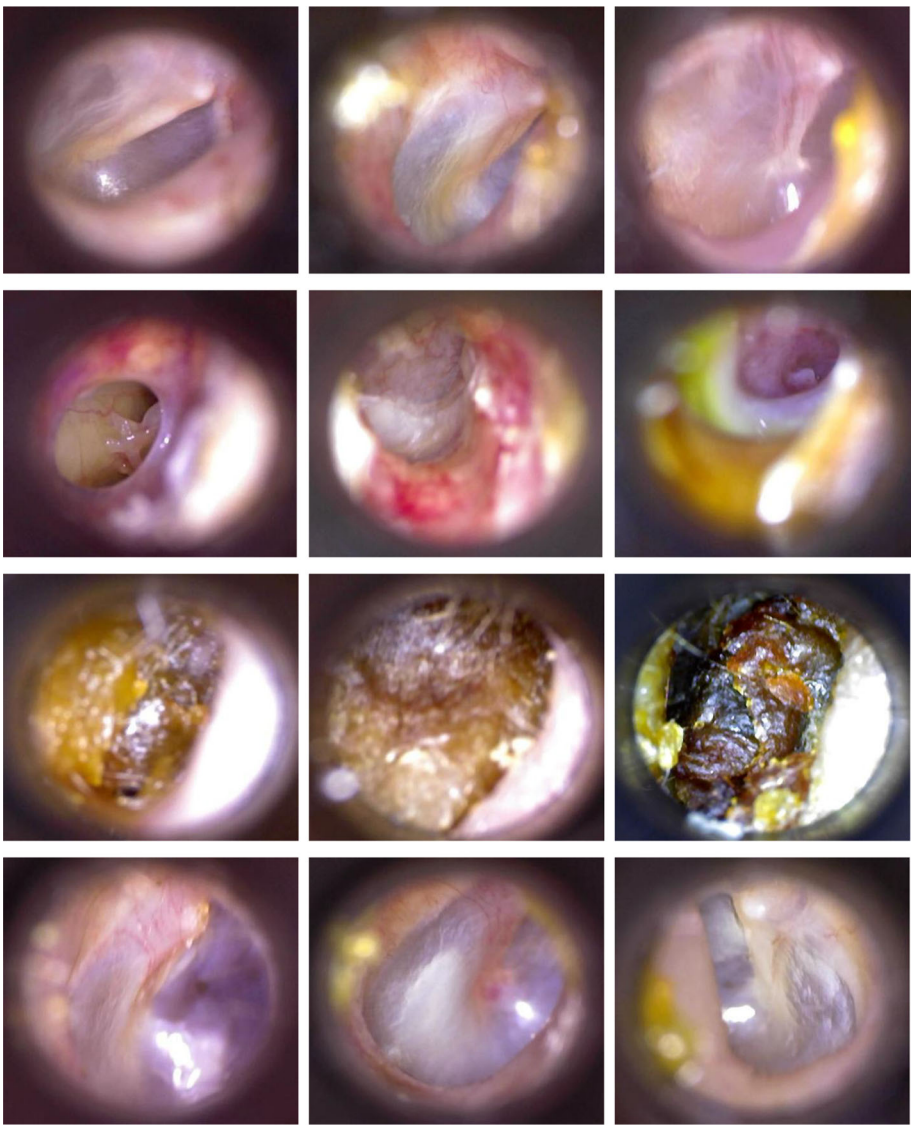

Fig. 2 Sample images for each class; From top to below rows; normal, chronic otitis media, earwax, myringosclerosis

classification with fully connected layers [50]. The input image is processed by a convolutional layer with 64 *filters* and 7 × 7 kernel sizes, followed by a maximum pooling layer. The layers of this architecture are then grouped in pairs [49]. Figure 3 depicts the network architecture of ResNet50.

2.3.2 Densely connected convolutional networks -121 (DenseNet121)

Huang et al. proposed the dense convolutional network architecture, which requires fewer parameters and computation for better performance [52]. The layers in this architecture are connected in a feed-forward manner, which

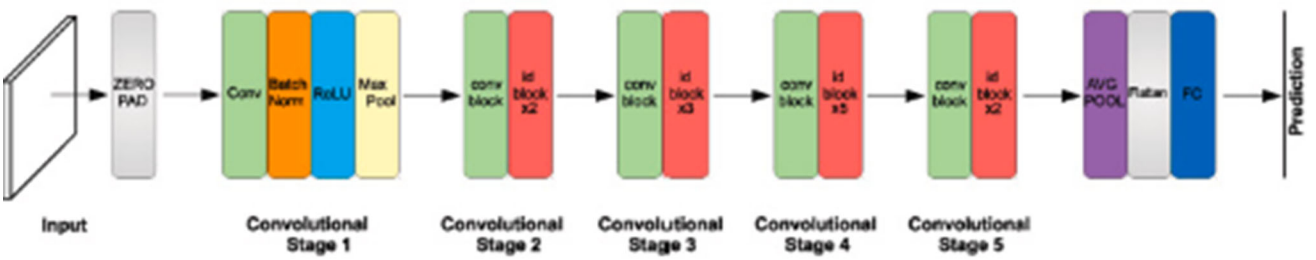

Fig. 3 The ResNet50 architecture [51]

means that each layer is fed with additional feature maps from previous layers and transmits its features to subsequent layers. These features are aggregated [53]. Figure 4 shows the DenseNet121 architecture with layers and dense blocks.

2.3.3 EfficientNetB7

EfficientNet is a model scaling approach proposed by Tan et al. [55] that uses a simple composite coefficient to scale networks in a more organized manner. In CNN, composite scaling uses a fixed set of coefficients to evenly scale width, depth, and resolution [56, 57]. To save time and computational power, the EfficientNet architecture employs transfer learning. As a result, this architecture has higher accuracy values than competing known models. This is due to the intelligently use of scaling in depth, width, and resolution [58]. Figure 5 shows the EfficientNetB7 architecture.

2.4 Dropout

Two or more fully connected layers follow convolutional and pooling layers in a CNN model. Fully connected layers of the CNN have the most significant number of parameters. These layers may prevent the model's generalization ability and cause overfitting [59]. Fully connected layers consist of neurons, weights, and biases. These layers connect neurons in one layer to neurons in another [60]. Overfitting is explained as achieving high success for the training set but not for the validation set in neural networks [61, 62]. Srivastava et al. introduced the dropout technique that increases the performance of neural networks in supervised learning tasks in computational biology, with promising results on various benchmark datasets [63]. The dropout layer prevents overfitting in networks by randomly ignoring nodes to ensure accuracy [61, 62] and increase stability [62]. The dropout is used to reduce the number of intermediate features to improve the orthogonality between the features of each layer [64] and is also utilized to avoid longer computation times [60]. There are studies that focus on these layers available in the literature. For example, Ha et al. used the dropout technique to eliminate the overfitting of probabilistic subject models in short and noisy texts [65]. Thanapol et al. used the dropout technique to reduce overfitting and improve generalization in CNN training [66]. Yang and Yang proposed a modified CNN model based on dropout and stochastic gradient descent optimizer to solve the overfitting problem. The authors designed an improved activation function to increase the convergence rate by adding a dropout layer between the fully connected and output layers [67]. Park and Kwak proposed stochastic dropout, whose drop ratio varies for each iteration to provide robustness for image variations [68]. Huynh and Nguyen used an additional dropout layer between the convolutional blocks of the Wide ResNet model to avoid overfitting problem in their study of joint age estimation and gender classification of Asian faces [69].

2.5 Global maximum pooling

Wang et al. introduced global maximum pooling (GMP) that focuses on the local salience of each channel [71]. GMP, which detects unique features [72], can increase the translation invariance of the network. Thus, the model presents a better prediction ability [73]. GMP is used to simplify the feature extraction process and improve the learning efficiency of the model [74]. GMP is also conducted to reduce feature size [74, 75]. Global pooling layers have no learnable parameters. Therefore, the layers may be less prone to overfitting and can reduce the network size [75]. Moreover, GMP could reduce the bias of the estimated value resulting from the convolution parameter error [76] and preserve the texture features well [76, 77]. GMP given in Eq. 1 [78] calculates the maximum value [74, 78] and can maintain the top feature scores in each channel [79].

$$O_{\mathrm{GMP}_i} = x_i, i = \mathrm{argmax}(\overrightarrow{x}) \tag{1}$$

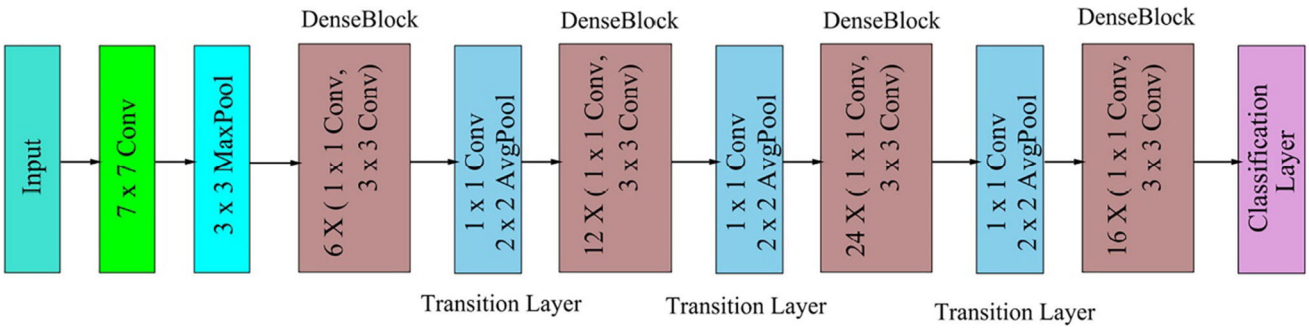

Fig. 4 The DenseNet121 architecture [54]

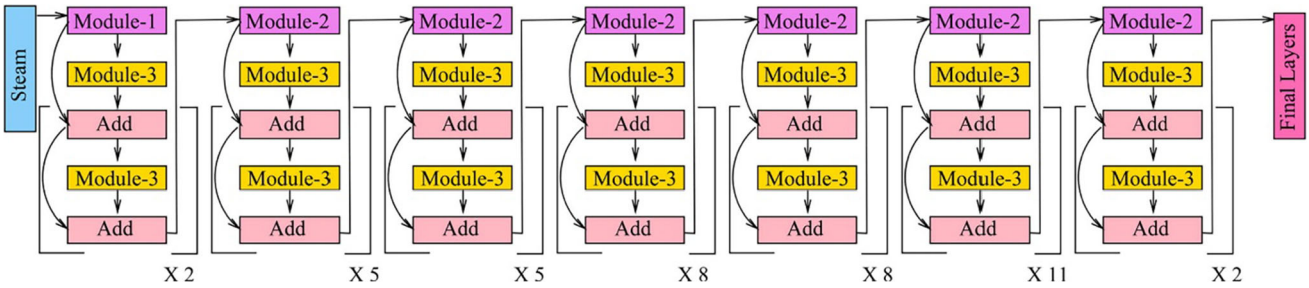

Fig. 5 The EfficientNetB7 architecture [54]

Here, $\overrightarrow{x}$ indicates the flattened vector form of feature map.

There are many studies in the literature addressing the importance of this layer. For example, Zhu et al. used the GMP layer to improve the network's generalization ability [76]. Ma et al. employed the GMP layer instead of traditional fully connected layers in the classification phase of the CNN to further increase the robustness and efficiency of the classifier [80]. Gao et al. used the GMP to increase the accuracy of lane line segmentation required in autonomous driving [81]. Hu et al. showed that the GMP contributes more than global average pooling in their proposed method based on multi-scale feature enhancement and aggregation to obtain blur-free images [82]. Chen et al. detected important points by utilizing the GMP for fine-grained image recognition [83]. Pan et al. used GMP to prevent distortions caused by some very small outliers and to increase the weight of the selected channel in their study of hand pose estimation [84].

2.6 Performance metrics

The accuracy, sensitivity, specificity, precision metrics given between Eqs. 2 and 5 are used to acquire class-wise results of the models. Accuracy is the ratio of a model's correct classified ones to all predictions. Sensitivity is a metric that shows how many of the samples that should be predicted as positive are correctly predicted. Specificity is a metric that shows how many of the samples that should be predicted as negative are correctly predicted. Precision is a metric that shows how many of the samples predicted as positive are actually positive. The true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values are utilized to calculate these metrics. TP and TN stand for the number of positive and negative images correctly identified, respectively, while FN and FP stand for the number of positive and negative images incorrectly classified. In multi-class problems, the considered class is evaluated as positive, while the others are negative. Accordingly, the averages of the class-wise results are calculated using the formulas given in Eqs. 6 and 9.

For each class, c;

$$\text{Sensitivity(Sen)} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity(Spe)} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{Accuracy(Acc)} = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

$$\text{Precision(Pre)} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{AverageSen} = \frac{1}{\text{classes}} \sum_{k=1}^{\text{classes}} \text{Sen}(c) \tag{6}$$

$$\text{AverageSpe} = \frac{1}{\text{classes}} \sum_{k=1}^{\text{classes}} \text{Spe}(c) \tag{7}$$

$$\text{AverageAcc} = \frac{1}{\text{classes}} \sum_{k=1}^{\text{classes}} \text{Acc}(c) \tag{8}$$

$$\text{AveragePre} = \frac{1}{\text{classes}} \sum_{k=1}^{\text{classes}} \text{Pre}(c) \tag{9}$$

The Area Under the Curve of the Receiver Operating Characteristic (AUC-ROC), which represent a probability curve, presents visually the performances of machine learning models [42]. The ROC curve depicts the relationship between true positive and false positive rates. The true positive rate is the ratio of correctly predicted positives to all positive data. The false positive rate is the ratio of incorrectly predicted negatives to all negative data. The AUC value close to one indicates that the model has correctly classified the data. On the other hand, the AUC value close to 0 means that the model has misclassified the samples [86].

3 Experiments

3.1 Experimental setup

Deep learning-based models were employed with Python programming language. The models were trained and tested on the Google Colaboratory environment with a Tesla P100-PCIE-16 GB GPU, an Intel(R) Xeon(R) 2.30 GHz CPU, and a 25 GB RAM configuration. Table 3 presents the tasks carried out in the experiments (Figure 6).

3.2 Model training

This section addresses the training processes of modified cutting-edge ResNet, DenseNet, and EfficientNetB7 deep CNN models. The fully connected layers of these models were removed by setting the *'Include top'* parameter to 'False.' The importance of GMP and dropout is focused on in experiments. As shown in Fig. 1, only fully connected layers that represent the classification stage were redesigned by two scenarios without making any changes to the feature extraction layers of each pre-trained model. Scenario 1 includes one flattened layer and one output layer after feature extraction layers for all models. Scenario 2 consists of one GMP layer, one hidden layer with 128 neurons, and one dropout layer (dropout rate = 0.2), respectively, after feature extraction layers. Images were resized to 224 × 224x3 in experiments. The default input size for ResNet50 and DenseNet121 pre-trained models is already 224 × 224x3. There is no problem for the EfficientNetB7 pre-trained model because 'Include_top = False' indicates that the input size can be set arbitrarily. The feature maps are passed to the classification stages of CNN models. In both scenarios, the output layer follows these layers. The activation function is ReLU in the hidden layer. All models were employed with 50 epochs. The value of the mini-batch size indicates the quantity of data used to update the network's weights during training. The value of this parameter was set to be 8 for all models. The Adam optimizer [87] with a 0.001 learning rate value was used. In addition, the *Model Checkpoint* method was used to detect the model with the best weights during the training phase. Besides, the model trainable property was set to 'True' of all models to update the Imagenet weights for the current task. The target class labels were converted to categorical values. Then, the categorical values were subjected to a one-hot encoding. The number of neurons in the output layer is set to four. SoftMax was used as the activation function and categorical cross-entropy for the loss function in this layer. The SoftMax activation function increases the flexibility of a neural network by increasing the degree of fit to the training set and transforming data from linear to nonlinear [64]. Figure 2 presents a block diagram of the model training and testing processes. Table 4 summarizes the parameters information of the models. To provide more readability to this paper, the M1 abbreviation refers to the modified classification part according to Scenario 1. The M2 abbreviation refers to the modified classification part according to Scenario 2. Accordingly, the M1 and CNN name pair (e.g., M1-EfficientNetB7) indicates the modified pre-trained deep CNN

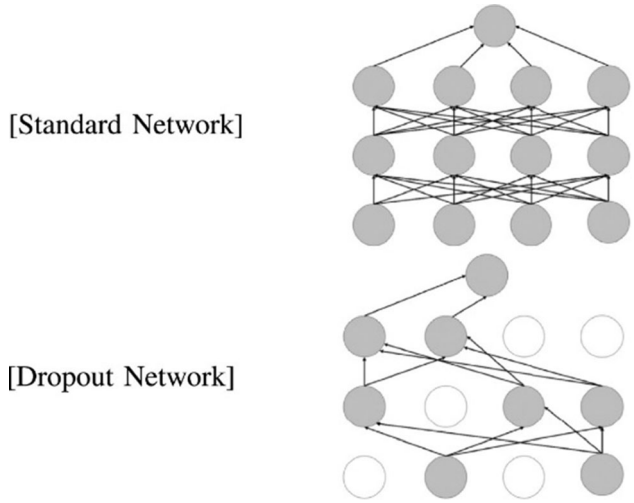

Fig. 6 Network structure [70]

Table 3 Tasks performed in this study

Task	Implement
Coding	Python *3.6*
Image processing	*Opencv, skimage, numpy*
Feature maps and modeling	*Tensorflow, Keras*
Plotting	*Matplotlib*

Table 4 Training parameters of the deep learning models

Parameter	Value
learning rate	*0.001*
Activation function in output layer	*SoftMax*
Batch size	*8*
Optimization algorithm	*RMSprop (lr = 0.001)*
Epoch	*50*
Layer trainable	*True*
Loss	*categorical_crossentropy*
Model checkpoint	*save best only = 'True'*
	monitor = 'validation accuracy'

model according to Scenario 1 throughout the paper. Similarly, the M2 and CNN name (e.g., M2-EfficientNetB7) indicates the modified pre-trained deep CNN model according to Scenario 2.

The hold-out validation technique was used to compare the performances of all deep learning models, and train subset was used for models' training and validation subset for evaluating models' performances. During the training process of a deep learning model, it is often necessary to verify the model's performance on a dataset different from the training set. In other words, it is checked whether there is a significant difference between the model's performances on the training and validating sets. Figure 7 presents accuracy curves that represent the accuracy values of the M2-EfficientNetB7 model, which offers the best performance in experiments, on the training and validation sets. The training accuracy and validation accuracy curves summarize the training process of any model and are important in detecting overfitting and underfitting issues. The accuracy curves of a model on the training and validation subsets are similar, and their accuracy values are close to one another, which is interpreted as the model's training is quite good. Also, the M2-EfficientNetB7's training history report, this model was saved with best weights in the 31st epoch via Model Checkpoint callback.

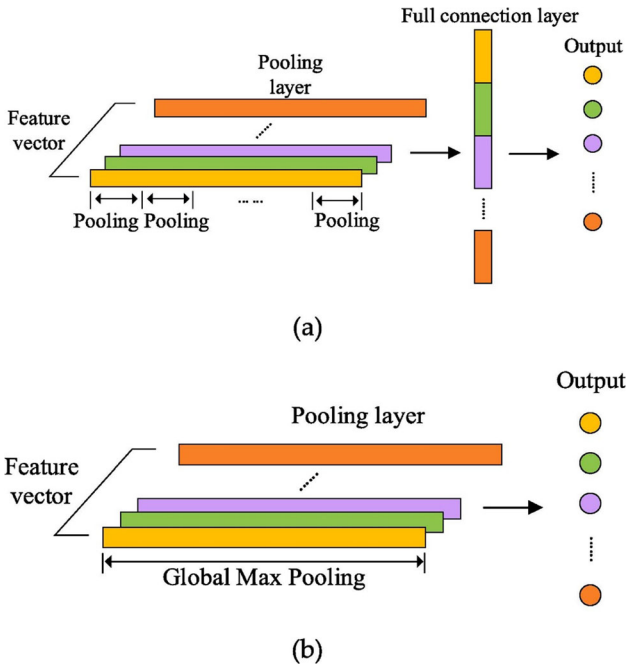

Fig. 7 Feature extraction: a) Fully connected layer extraction b) Global maximum pooling extraction [85]

4 Results and discussion

This section presents experimental results in detail. Multiple experiments were conducted to develop a robust model using cutting-edge ResNet50, DenseNet121, and EfficientNetB7 networks. Figures 7, 8 and 9 show the confusion matrices of the deep learning models performed on the testing set. Values outside the diagonal in confusion matrices that summarize the performances of the models built indicate the number of incorrectly classified samples. Compared to M1-based pre-trained models, the M2-based ones offered better performance, as can be seen from the confusion matrices. The M2-EfficientNetB7 model has less misclassification. This model misclassified only one sample image, whereas the M1-EfficientNetB7 model misclassified nine. The M1-DenseNet121 and M2-DenseNet121 models misclassified 50 and 5 sample images, respectively. Lastly, the M1-ResNet50 and M2-ResNet50 models misclassified 32 and 22 sample images, respectively. Figure 10 also shows the number of images incorrectly classified by the models.

Table 5 gives Acc, Sen, Spe and Pre measurements representing the models' performances for tympanic membrane conditions. This table shows the class-wise results of the testing set presented by each model. The rows below class-wise results show the average results. The M2-based deep CNN models performed better than the M1-based ones. Wildly, the M2-EfficientNetB7 outperformed the M1-EfficientNetB7 with an average accuracy of 99.94%. The M2-EfficientNetB7 model presents 1 FN, and the M1-EfficientNetB7 model offers 9 FNs. This shows that the M2-EfficientNetB7 model is also better in Sen, Spe, and Pre measures. Similar results are available for other models as well. The M2-DenseNet121 model outperformed the M1-DenseNet121 model by 3.11%. The M2-ResNet50 model outperformed the M1-ResNet5 model by 0.71%. Overall, the modifications in Scenario 2 are more successful than Scenario 1. Separate sentences are not given here for each one to avoid repetition.

The ROC curves with AUC values were also presented to validate the robustness of the M2-EfficientNetB7 model, which provides the highest accuracy. Figures 11 and 12 show the class-wise ROC curves of each model. According to these results, the macro-average AUC values of the M1-EfficientNetB7, M1-DenseNet121, and M1-ResNet50 models are 0.992, 0.954, and 0.970, respectively. The M2-EfficientNetB7, M2-DenseNet121, and M2-ResNet50 models have macro-average AUC values of 0.999, 0.995, and 0.980, respectively. The M2-EfficientNetB7 model outperformed other deep learning models with high AUC values for each class (Figs. 13, 14 and 15).

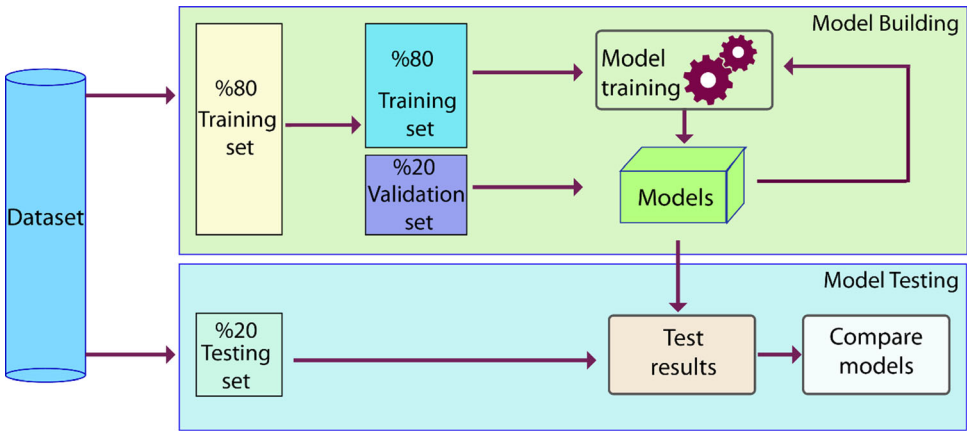

Fig. 8 A model's performance evaluation workflow

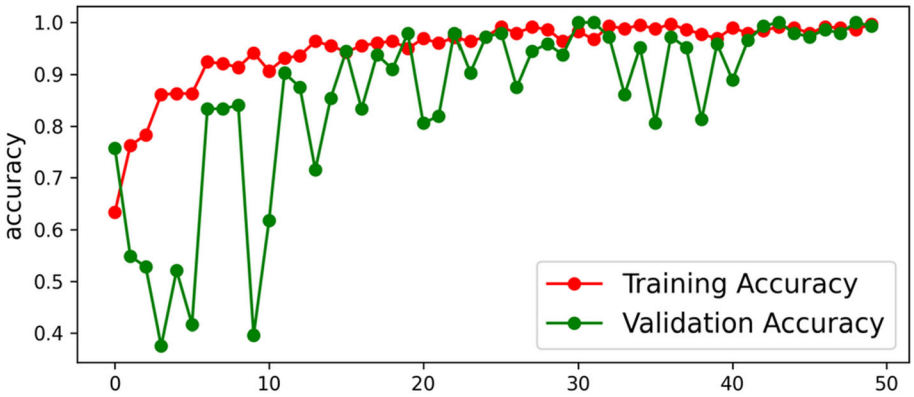

Fig. 9 Training and validation accuracy curves of the M2-EfficientNetB7 model

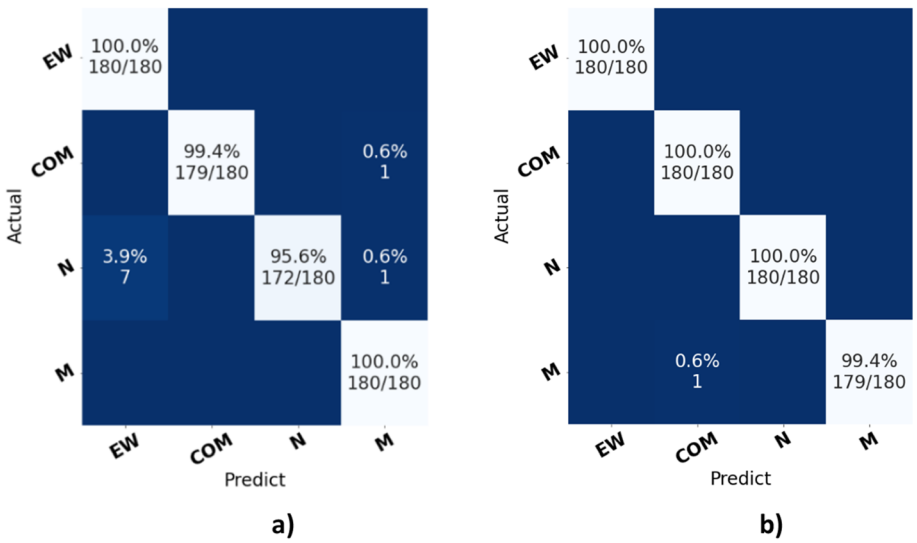

Fig. 10 Confusion matrices; a) M1-EfficientNetB7, b) M2-EfficientNetB7

In the literature, cutting-edge deep learning models gave high classification accuracies for various ear diseases and conditions. However, the datasets' sample size and class distribution are a noteworthy issue. Multiple approaches were addressed to deal with unbalanced dataset problems, small sample sizes, and a limited number of classes. For example, Cha et al. proposed an ensemble model with two different deep-learning models offering the best performance. The authors in their study combined the relatively slight samples according to features such as common properties, common pathogenesis, and physical findings to balance the sample size for each class [31]. Chen et al. introduced a deep learning model trained on a dataset that they applied preprocessing and augmentation techniques.

Table 5 Hold-out validation results of pre-trained CNN models

CNN	Modification	Class	TP	FP	FN	TN	Sen	Spe	Pre	Acc
EfficientNetB7	**M1**	EW	180	7	0	533	100	98.7	96.26	99.03
		COM	179	0	1	540	99.44	100	100	99.86
		N	172	0	8	540	95.56	100	100	98.89
		M	180	2	0	716	100	99.72	98.9	99.78
	Average						**98.75**	**99.61**	**98.79**	**99.39**
	M2	EW	180	0	0	540	100	100	100	100
		COM	180	1	0	539	100	99.81	99.45	99.86
		N	179	0	0	540	100	100	100	100
		M	179	0	1	719	99.44	100	100	99.89
	Average						**99.86**	**99.95**	**99.86**	**99.94**
DenseNet121	**M1**	EW	138	0	42	540	76.67	100	100	94.17
		COM	179	45	1	495	99.44	91.67	79.91	93.61
		N	173	1	7	539	96.11	99.81	99.43	98.89
		M	180	4	0	712	100	99.44	97.83	99.55
	Average						**93.06**	**97.73**	**94.29**	**96.56**
	M2	EW	179	0	1	540	99.44	100	100	99.86
		COM	180	4	0	536	100	99.26	97.83	99.44
		N	177	0	3	540	98.33	100	100	99.58
		M	179	1	1	717	99.44	99.86	99.44	99.78
	Average						**99.30**	**99.78**	**99.32**	**99.67**
ResNet50	**M1**	EW	179	11	1	529	99.44	97.96	94.21	98.33
		COM	158	1	22	539	87.78	99.81	99.37	96.81
		N	175	8	5	532	97.22	98.52	95.63	98.19
		M	176	12	4	692	97.78	98.3	93.62	98.19
	Average						**95.56**	**98.65**	**95.71**	**97.88**
	M2	EW	180	1	0	539	100	99.81	99.45	99.86
		COM	176	0	4	540	97.78	100	100	99.44
		N	162	3	18	537	90	99.44	98.18	97.08
		M	180	18	0	684	100	97.44	90.91	97.96
	Average						**96.95**	**99.17**	**97.14**	**98.59**

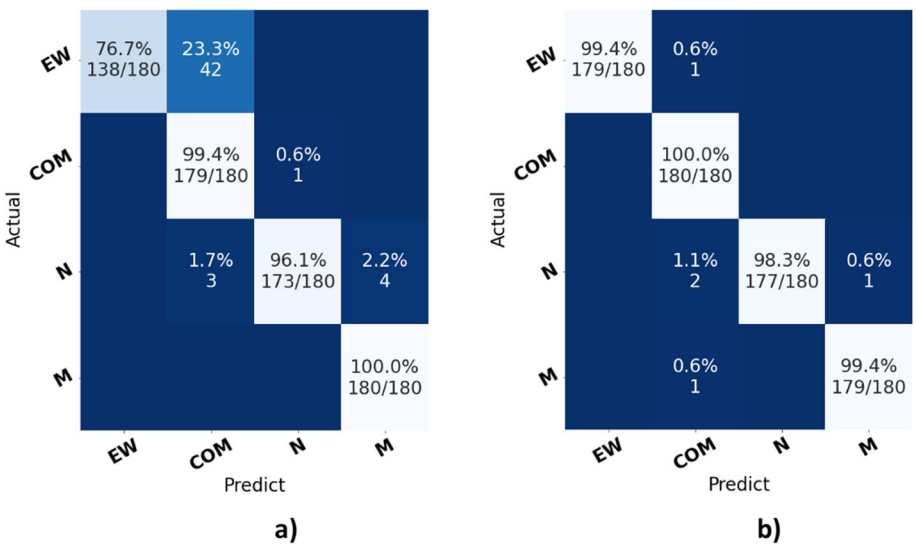

Fig. 11 Confusion matrices; a) M1-DenseNet121, b) M2-DenseNet121

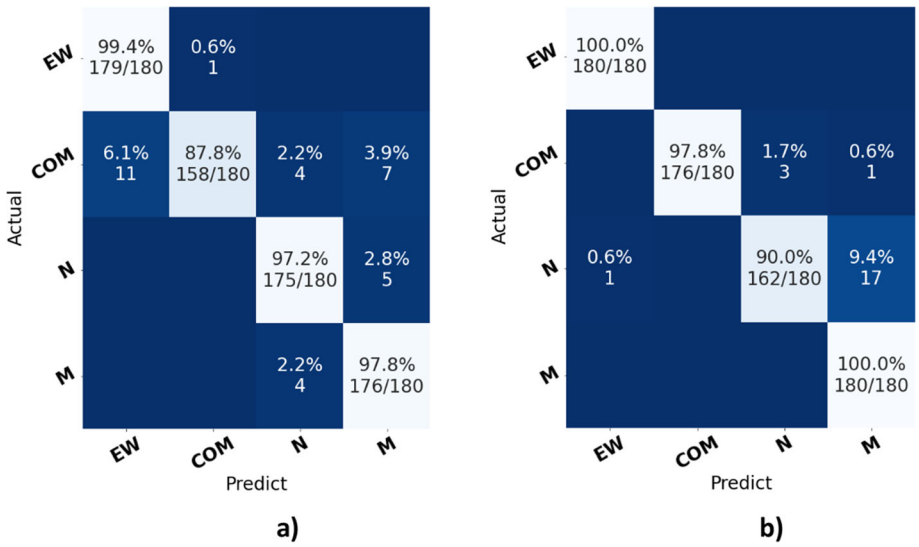

Fig. 12 Confusion matrices; a) M1-ResNet50, b) M2-ResNet50

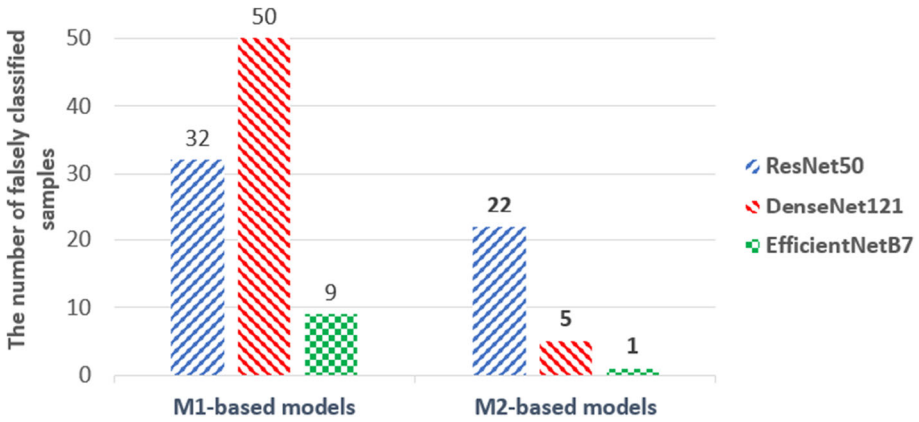

Fig. 13 Graphical representation of the number of misclassified samples

Also, the authors collected all images of these classes in this category since there were not enough samples for each class in the abnormal category. They carried out a binary classification as normal and abnormal [11]. In another study, Başaran et al. applied data augmentation to all digital otoscope images [32]. It should be noted that the success of a CNN model is measured by its ability to identify previously unseen images accurately. The last two studies above have limitations because some images produced using the data augmentation techniques from the original images may be in the training set, while others may be in the testing set [33]. Khan et al. overcame the small dataset limitation by increasing the number of training examples and developed a CNN-based deep learning model for three classes [10]. Lastly, Myburgh et al. used handcrafted features such as tympanic membrane shape, malleus bone visibility, and tympanic membrane perforation to diagnose middle ear pathology or otitis media automatically in their study [12]. The deep learning that is cutting-edge in computer vision eliminates the limitations of handcrafted features in machine learning. In this paper, the M2-EfficientNetB7 deep learning model acquired better performance than others. However, this model only produces prediction for one of the four classes, even if an image has one of the other tympanic membrane cases. This is a limitation of this model. Actually, this limitation is a common and basic problem encountered, especially in medical datasets. Khan et al. [10] and Chen et al. [11] stated also this subject as the limitation of their studies. In this context, if the datasets contain all eardrum diseases and tympanic membrane cases and enough images are collected with the approval of the relevant health committees, there will be an opportunity for successful artificial intelligence modeling ready for use in real-world clinical environments.

5 Conclusion

This study examined the classification performances of modified deep CNN models for automated detection of tympanic membrane conditions. In this context, cutting-

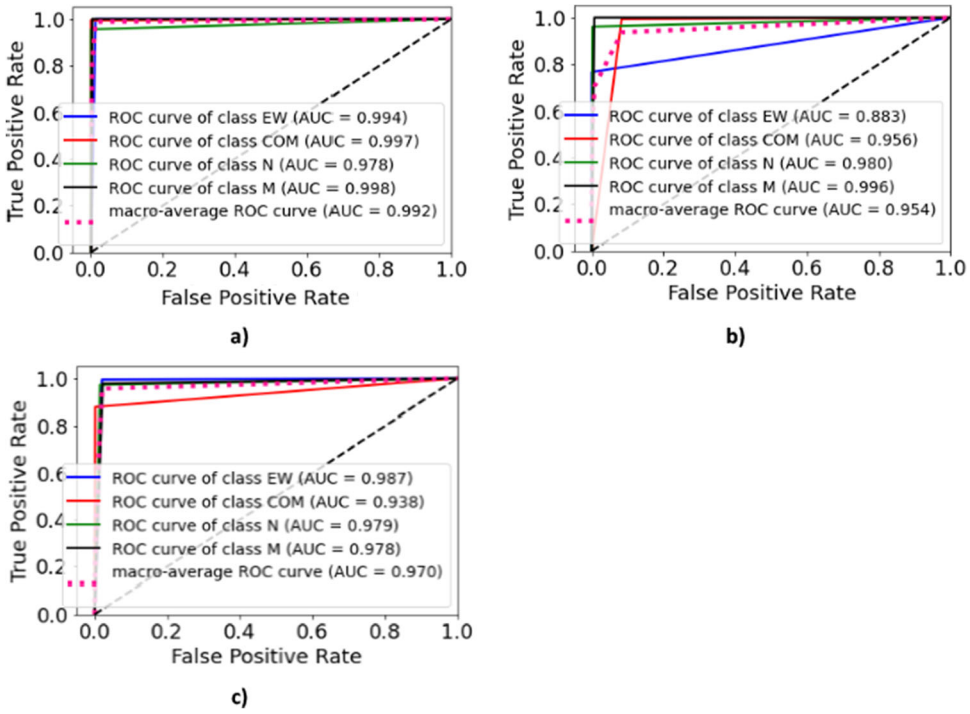

Fig. 14 ROC curves of the M1-based deep CNN models; a) EfficientNetB7 b) DenseNet121 c) ResNet50

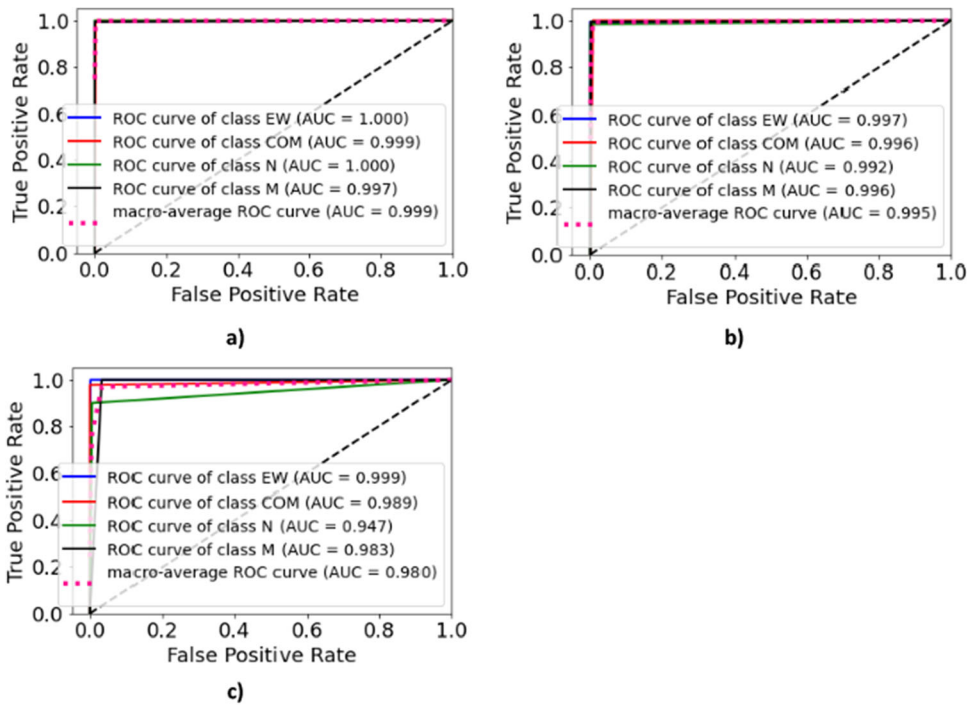

Fig. 15 ROC curves of the M2-based CNN models; a) EfficientNetB7 b) DenseNet121 c) ResNet50

edge DenseNet121, ResNet50, and EfficientNetB7 deep learning models were modified with the same rules. According to the experimental results, the M2-EfficientNetB7 model outperformed the M2-ResNet50 and M2-DenseNet121 models with average values of 99.94% Acc, 99.86% Sen, 99.95% Spe, and 99.86% Pre. An expert system including the M2-EfficientNetB7 deep learning-based artificial intelligence model will help diagnose middle ear disease, particularly in primary care institutions and hospitals that lack field specialists. Furthermore, this model will contribute to more accurately determining tympanic membrane cases by minimizing the subjective opinion mistakes of field experts. As future research objectives, it is planned to develop highly generalizable

models that can successfully classify the conditions that examined in this study and others, such as acute otitis media and cholesteatoma.

Acknowledgements The author would like to thank Viscaino et al. for providing the public ear image dataset [38].

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). Not applicable.

Data availability This study uses a public dataset available at: https://figshare.com/ndownloader/files/21793293

Declarations

Conflicts of interest The author declares that he has no conflicts of interest.

References

1. Eriksson PO, Mattsson C, Hellström S (2003) First forty-eight hours of developing otitis media: an experimental study. Annals Otol, Rhinol Laryngol 112:558–566. https://doi.org/10.1177/000348940311200614
2. Rovers MM (2008) The burden of otitis media. Vaccine 26:G2–G4. https://doi.org/10.1016/J.VACCINE.2008.11.005
3. Williams CJ, Coates HL, Pascoe EM et al (2009) Middle ear disease in Aboriginal children in Perth: analysis of hearing screening data, 1998–2004. Med J Aust 190:598–600. https://doi.org/10.5694/J.1326-5377.2009.TB02576.X
4. Williams CJ, Jacobs AM (2009) The impact of otitis media on cognitive and educational outcomes. Med J Australia. https://doi.org/10.5694/J.1326-5377.2009.TB02931.X
5. Ibekwe TS, Nwaorgu OGB (2011) Classification and management challenges of otitis media in a resource-poor country. Niger J Clin Pract 14:262. https://doi.org/10.4103/1119-3077.86764
6. Monasta L, Ronfani L, Marchetti F et al (2012) Burden of disease caused by otitis media: systematic review and global estimates. PLoS ONE 7:e36226. https://doi.org/10.1371/JOURNAL.PONE.0036226
7. Jose A (2004) Chronic suppurative otitis media: burden of illness and management options. In: World Health Organization. https://apps.who.int/iris/handle/10665/42941. Accessed 24 Jul 2022
8. Pichichero ME (2003) Diagnostic accuracy of otitis media and tympanocentesis skills assessment among pediatricians. Eur J Clin Microbiol Infect Dis 22:519–524. https://doi.org/10.1007/s10096-003-0981-8
9. Thomassin JM, Duchon-Doris JM, Emram B et al (1990) Endoscopy surgery of the ear. First assessment. Annales d'Oto-Laryngologie et de Chirurgie Cervico-Faciale 107:564–570
10. Khan MA, Kwon S, Choo J et al (2020) Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks. Neural Netw 126:384–394. https://doi.org/10.1016/J.NEUNET.2020.03.023
11. Chen Y-C, Chu Y-C, Huang C-Y et al (2022) Smartphone-based artificial intelligence using a transfer learning algorithm for the detection and diagnosis of middle ear diseases: a retrospective deep learning study. E Clin Med 51:101543. https://doi.org/10.1016/J.ECLINM.2022.101543
12. Myburgh HC, Jose S, Swanepoel DW, Laurent C (2018) Towards low cost automated smartphone- and cloud-based otitis media diagnosis. Biomed Signal Process Control 39:34–52. https://doi.org/10.1016/J.BSPC.2017.07.015
13. Cai Y, Zeng J, Lan L et al (2022) Expert recommendations on collection and annotation of otoscopy images for intelligent medicine. Int Med. https://doi.org/10.1016/J.IMED.2022.01.001
14. Aung YYM, Wong DCS, Ting DSW (2021) The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull 139:4–15. https://doi.org/10.1093/BMB/LDAB016
15. Nedjah N, Bonilla AD, de Macedo ML (2023) Automatic speech recognition of Portuguese phonemes using neural networks ensemble. Expert Syst Appl 229:120378. https://doi.org/10.1016/J.ESWA.2023.120378
16. Zhao J, Chen D, Zhao L et al (2022) Self-powered speech recognition system for deaf users. Cell Rep Phys Sci 3:101168. https://doi.org/10.1016/J.XCRP.2022.101168
17. Kheddar H, Himeur Y, Al-Maadeed S et al (2023) Deep transfer learning for automatic speech recognition: Towards better generalization. Knowl Based Syst 277:110851. https://doi.org/10.1016/J.KNOSYS.2023.110851
18. Ramírez I, Cuesta-Infante A, Pantrigo JJ et al (2020) Convolutional neural networks for computer vision-based detection and recognition of dumpsters. Neural Comput Appl 32:13203–13211. https://doi.org/10.1007/S00521-018-3390-8/FIGURES/5
19. Issac A, Dutta MK, Travieso CM (2020) Automatic computer vision-based detection and quantitative analysis of indicative parameters for grading of diabetic retinopathy. Neural Comput Appl 32:15687–15697. https://doi.org/10.1007/S00521-018-3443-Z/TABLES/3
20. Ding Y, Hua L, Li S (2022) Research on computer vision enhancement in intelligent robot based on machine learning and deep learning. Neural Comput Appl 34:2623–2635. https://doi.org/10.1007/S00521-021-05898-8/FIGURES/13
21. Han B, Yoo C-H, Kim H-W et al (2023) Deep emotion change detection via facial expression analysis. Neurocomputing 549:126439. https://doi.org/10.1016/J.NEUCOM.2023.126439
22. Rooj S, Routray A, Mandal MK (2023) Feature based analysis of thermal images for emotion recognition. Eng Appl Artif Intell 120:105809. https://doi.org/10.1016/J.ENGAPPAI.2022.105809
23. Almanza-Conejo O, Almanza-Ojeda DL, Contreras-Hernandez JL, Ibarra-Manzano MA (2023) Emotion recognition in EEG signals using the continuous wavelet transform and CNNs. Neural Comput Appl 35:1409–1422. https://doi.org/10.1007/S00521-022-07843-9/TABLES/4
24. Sanjeev D, Singh K, Craciun EM et al (2023) Next-cart recommendation by utilizing personalized ıtem frequency ınformation in online web portals. Neural Process Lett. https://doi.org/10.1007/S11063-023-11207-2/FIGURES/8
25. van Maasakkers L, Fok D, Donkers B (2023) Next-basket prediction in a high-dimensional setting using gated recurrent units. Expert Syst Appl 212:118795. https://doi.org/10.1016/J.ESWA.2022.118795

26. Arthur JK, Zhou C, Osei-Kwakye J et al (2022) A heterogeneous couplings and persuasive user/item information model for next basket recommendation. Eng Appl Artif Intell 114:105132. https://doi.org/10.1016/J.ENGAPPAI.2022.105132
27. Zhen X, Shao L, Maybank SJ, Chellappa R (2016) Handcrafted vs. learned representations for human action recognition. Image Vis Comput 55:39–41. https://doi.org/10.1016/J.IMAVIS.2016.10.002
28. Shoeibi A, Ghassemi N, Alizadehsani R et al (2021) A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals. Expert Syst Appl 163:113788. https://doi.org/10.1016/J.ESWA.2020.113788
29. Ghazal S, Qureshi WS, Khan US et al (2021) Analysis of visual features and classifiers for Fruit classification problem. Comput Electron Agric 187:106267. https://doi.org/10.1016/J.COMPAG.2021.106267
30. Sundgaard JV, Harte J, Bray P et al (2021) Deep metric learning for otitis media classification. Med Image Anal 71:102034. https://doi.org/10.1016/j.media.2021.102034
31. Cha D, Pae C, Seong SB et al (2019) Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. EBioMedicine 45:606–614. https://doi.org/10.1016/J.EBIOM.2019.06.050
32. Başaran E, Cömert Z, Çelik Y (2020) Convolutional neural network approach for automatic tympanic membrane detection and classification. Biomed Signal Process Control 56:101734. https://doi.org/10.1016/J.BSPC.2019.101734
33. Lee JY, Choi SH, Chung JW (2019) Automated classification of the tympanic membrane using a convolutional neural network. Appl Sci 9:1827. https://doi.org/10.3390/APP9091827
34. Côté-Allard U, Fall CL, Drouin A et al (2019) Deep learning for electromyographic hand gesture signal classification using transfer learning. IEEE Trans Neural Syst Rehabil Eng 27:760–771. https://doi.org/10.1109/TNSRE.2019.2896269
35. Roster K, Connaughton C, Rodrigues FA (2022) Forecasting new diseases in low-data settings using transfer learning. Chaos Solitons Fractals 161:112306. https://doi.org/10.1016/J.CHAOS.2022.112306
36. Liu F, Ye Z, Wang L (2022) Deep transfer learning-based vehicle classification by asphalt pavement vibration. Constr Build Mater 342:127997. https://doi.org/10.1016/J.CONBUILDMAT.2022.127997
37. Yang K, Lu J, Wan W et al (2022) Transfer learning based on sparse Gaussian process for regression. Inf Sci (N Y) 605:286–300. https://doi.org/10.1016/J.INS.2022.05.028
38. Viscaino M, Maass JC, Delano PH et al (2020) Computer-aided diagnosis of external and middle ear conditions: a machine learning approach. PLoS ONE 15:1–18. https://doi.org/10.1371/journal.pone.0229226
39. Rafique Q, Rehman A, Afghan MS et al (2023) Reviewing methods of deep learning for diagnosing COVID-19, its variants and synergistic medicine combinations. Comput Biol Med 163:107191. https://doi.org/10.1016/J.COMPBIOMED.2023.107191
40. Razavian AS, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: An astounding baseline for recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 512–519. https://doi.org/10.1109/CVPRW.2014.131
41. Paul E, Sabeenian RS (2022) Modified convolutional neural network with pseudo-CNN for removing nonlinear noise in digital images. Displays 74:102258. https://doi.org/10.1016/J.DISPLA.2022.102258
42. Oğuz A, Faruk Ö, Ertuğul, (2023) A survey on applications of machine learning algorithms in water quality assessment and water supply and management. Water Supply 23:895–922. https://doi.org/10.2166/WS.2023.033
43. Suha SA, Sanam TF (2022) A deep convolutional neural network-based approach for detecting burn severity from skin burn images. Mach Learn Appl 9:100371. https://doi.org/10.1016/J.MLWA.2022.100371
44. Wang H, Wang L, Zhang L (2022) Transfer learning improves landslide susceptibility assessment. Gondwana Res. https://doi.org/10.1016/J.GR.2022.07.008
45. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22:1345–1359. https://doi.org/10.1109/TKDE.2009.191
46. Li Z, Kristoffersen E, Li J (2022) Deep transfer learning for failure prediction across failure types. Comput Ind Eng 172:108521. https://doi.org/10.1016/J.CIE.2022.108521
47. Luo S, Huang X, Wang Y et al (2022) Transfer learning based on improved stacked autoencoder for bearing fault diagnosis. Knowl Based Syst 256:109846. https://doi.org/10.1016/J.KNOSYS.2022.109846
48. Bierbrauer DA, De Lucia MJ, Reddy K et al (2023) Transfer learning for raw network traffic detection. Expert Syst Appl 211:118641. https://doi.org/10.1016/J.ESWA.2022.118641
49. Victor Ikechukwu A, Murali S, Deepu R, Shivamurthy RC (2021) ResNet-50 vs VGG-19 vs training from scratch: a comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. Global Trans Proc 2:375–381. https://doi.org/10.1016/J.GLTP.2021.08.027
50. Li B, Lima D (2021) Facial expression recognition via ResNet-50. Int J Cognitive Comput Eng 2:57–64. https://doi.org/10.1016/J.IJCCE.2021.02.002
51. de Souza LA, Mendel R, Strasser S et al (2021) Convolutional neural networks for the evaluation of cancer in barrett's esophagus: explainable AI to lighten up the black-box. Comput Biol Med 135:104578. https://doi.org/10.1016/J.COMPBIOMED.2021.104578
52. Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely Connected Convolutional Networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
53. Kumar N, Sharma M, Singh VP et al (2022) An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. Biomed Signal Process Control 75:103596. https://doi.org/10.1016/J.BSPC.2022.103596
54. Hazarika RA, Kandar D, Maji AK (2021) An experimental analysis of different deep learning based models for alzheimer's disease classification using brain magnetic resonance images. J King Saud Univ Comput Inf Sci. https://doi.org/10.1016/J.JKSUCI.2021.09.003
55. Tan M, Le Q V. (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: 36th International Conference on Machine Learning, ICML 2019 2019-June:10691–10700
56. Ali K, Shaikh ZA, Khan AA, Laghari AA (2022) Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer. Neurosci Inf 2:100034. https://doi.org/10.1016/J.NEURI.2021.100034
57. Zhou A, Ma Y, Ji W et al (2022) Multi-head attention-based two-stream EfficientNet for action recognition. Multimed Syst 1:1–12. https://doi.org/10.1007/S00530-022-00961-3
58. Marques G, Agarwal D, de la Torre DI (2020) Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. Appl Soft Comput 96:106691. https://doi.org/10.1016/J.ASOC.2020.106691

59. Yadav P, Menon N, Ravi V et al (2022) EfficientNet convolutional neural networks-based android malware detection. Comput Secur 115:102622. https://doi.org/10.1016/J.COSE.2022.102622
60. Mishra RK, Urolagin S, Arul Jothi JA, Gaur P (2022) Deep hybrid learning for facial expression binary classifications and predictions. Image Vis Comput 128:104573. https://doi.org/10.1016/J.IMAVIS.2022.104573
61. Sun X, Liu Z, Wang X, Chen X (2022) Determination of ductile fracture properties of 16MND5 steels under varying constraint levels using machine learning methods. Int J Mech Sci 224:107331. https://doi.org/10.1016/J.IJMECSCI.2022.107331
62. Karthik R, Menaka R, Kathiresan GS et al (2022) Gaussian dropout based stacked ensemble CNN for classification of breast tumor in ultrasound images. IRBM 43:715–733. https://doi.org/10.1016/J.IRBM.2021.10.002
63. Srivastava N, G. H, A. K, et al (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. In: 15(56). https://jmlr.org/papers/v15/srivastava14a.html. Accessed 25 Jul 2022
64. Cui X, Chen N, Zhao C et al (2023) An adaptive weighted attention-enhanced deep convolutional neural network for classification of MRI images of Parkinson's disease. J Neurosci Methods 394:109884. https://doi.org/10.1016/J.JNEUMETH.2023.109884
65. Ha C, Tran VD, Van Ngo L, Than K (2019) Eliminating overfitting of probabilistic topic models on short and noisy text: the role of dropout. Int J Approx Reasoning 112:85–104. https://doi.org/10.1016/J.IJAR.2019.05.010
66. Thanapol P, Lavangnananda K, Bouvry P, et al (2020) Reducing Overfitting and Improving Generalization in Training Convolutional Neural Network (CNN) under Limited Sample Sizes in Image Recognition. In: CIT 2020 5th International Conference on Information Technology 300–305. https://doi.org/10.1109/INCIT50588.2020.9310787
67. Yang J, Yang G (2018) Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer. Algorithms 11:28. https://doi.org/10.3390/A11030028
68. Park S, Kwak N (2017) Analysis on the dropout effect in convolutional neural networks. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10112 LNCS:189–204. https://doi.org/10.1007/978-3-319-54184-6_12/COVER
69. Huynh HT, Nguyen H (2020) Joint age estimation and gender classification of asian faces using wide ResNet. SN Comput Sci 1:1–9. https://doi.org/10.1007/S42979-020-00294-W/FIGURES/10
70. Cao Z, Huang J, He X, Zong Z (2022) BND-VGG-19: a deep learning algorithm for COVID-19 identification utilizing X-ray images. Knowl Based Syst 258:110040. https://doi.org/10.1016/J.KNOSYS.2022.110040
71. Wang Y, Ren Y, Kang S et al (2024) Identification of tea quality at different picking periods: a hyperspectral system coupled with a multibranch kernel attention network. Food Chem 433:137307. https://doi.org/10.1016/J.FOODCHEM.2023.137307
72. Li H, Gan Y, Wu Y, Guo L (2022) EAGNet: A method for automatic extraction of agricultural greenhouses from high spatial resolution remote sensing images based on hybrid multi-attention. Comput Electron Agric 202:107431. https://doi.org/10.1016/J.COMPAG.2022.107431
73. Yu J, Liu K, He M, Qin L (2021) Insulator defect detection: a detection method of target search and cascade recognition. Energy Rep 7:750–759. https://doi.org/10.1016/J.EGYR.2021.09.197
74. Zhao Z, Lv N, Xiao R et al (2023) Recognition of penetration states based on arc sound of interest using VGG-SE network during pulsed GTAW process. J Manuf Process 87:81–96. https://doi.org/10.1016/J.JMAPRO.2022.12.034
75. He Z, He L, Xu H et al (2023) A bilateral attention based generative adversarial network for DIBR 3D image watermarking. J Vis Commun Image Represent 92:103794. https://doi.org/10.1016/J.JVCIR.2023.103794
76. Zhu Y, JiaYI H, Li Y, Li W (2022) Image identification of cashmere and wool fibers based on the improved Xception network. J King Saud Univ-Comput Inf Sci 34:9301–9310. https://doi.org/10.1016/J.JKSUCI.2022.09.009
77. Ou G, Yu G, Domeniconi C et al (2020) Multi-label zero-shot learning with graph convolutional networks. Neural Netw 132:333–341. https://doi.org/10.1016/J.NEUNET.2020.09.010
78. Halder A, Dey D (2023) MorphAttnNet: an attention-based morphology framework for lung cancer subtype classification. Biomed Signal Process Control 86:105149. https://doi.org/10.1016/J.BSPC.2023.105149
79. Dun Y, Da Z, Yang S et al (2021) Kernel-attended residual network for single image super-resolution. Knowl Based Syst 213:106663. https://doi.org/10.1016/J.KNOSYS.2020.106663
80. Ma M, Wang QF, Huang S et al (2021) Residual attention-based multi-scale script identification in scene text images. Neurocomputing 421:222–233. https://doi.org/10.1016/J.NEUCOM.2020.09.015
81. Gao X, Bai H, Xiong Y et al (2023) Robust lane line segmentation based on group feature enhancement. Eng Appl Artif Intell 117:105568. https://doi.org/10.1016/J.ENGAPPAI.2022.105568
82. Hu B, Wang S, Gao X et al (2023) Reduced-reference image deblurring quality assessment based on multi-scale feature enhancement and aggregation. Neurocomputing 547:126378. https://doi.org/10.1016/J.NEUCOM.2023.126378
83. Chen J, Hu J, Li S (2021) Learning to locate for fine-grained image recognition. Comput Vis Image Underst 206:103184. https://doi.org/10.1016/J.CVIU.2021.103184
84. Pan T, Wang Z, Fan Y (2022) Optimized convolutional pose machine for 2D hand pose estimation. J Vis Commun Image Represent 83:103461. https://doi.org/10.1016/J.JVCIR.2022.103461
85. Yao D, Li B, Liu H et al (2021) Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. Measurement 175:109166. https://doi.org/10.1016/J.MEASUREMENT.2021.109166
86. Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874. https://doi.org/10.1016/J.PATREC.2005.10.010
87. Kingma DP, Lei Ba J (2015) ADAM: A method for stochastic optimization. arXiv:14126980. https://doi.org/10.48550/arXiv.1412.6980