International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Aug-24)

http://dx.doi.org/10.12785/ijcds/160174

An Ensemble Neural Architecture for Lung Diseases Prediction Using Chest X-rays

Abeer Abdelhamid[1], Oluwatunmise Akinniyi[2], Gehad A. Saleh[3], Wael Deabes[4] and Fahmi Khalifa[1,2]

[1]*Electronics and Communications Engineering Department, Mansoura University, Mansoura 35516, Egypt*
[2]*Department of Electrical and Computer Engineering, Morgan State University, Baltimore MD 21251, USA*
[3]*Diagnostic and Interventional Radiology, Faculty of Medicine, Mansoura University, Mansoura 35516, Egypt*
[4]*Computational, Engineering, Mathematical Sciences Dept., Texas A&M University SA, TX 78224, USA*

Received 18 Apr. 2024, Revised 14 May 2024, Accepted 15 May 2024, Published 10 Aug. 2024

Abstract: Accurate diagnostic tools for disease control and treatment options is of immense importance, specially during pandemics, Coronavirus (or COVID) that drew global attention in late 2019. Early detection and seclusion are the cornerstone effective ways to prevent virus spread. Artificial intelligence (AI)-based diagnostic tools for COVID detection have surged dramatically using various diagnostic imaging techniques, among which Chest X-ray (CXR) have been extensively investigated due to its fast acquisition coupled with its superior results. We propose a hybrid, automated, and efficient approach to detect COVID-19 at an early stage using CXRs. One of the main advantages of the proposed analysis is the development of a learnable input scaling module, which accommodates various CXR with different sizes with the ability to keep prominent CXRs features while filtering out noise. Additionally, the suggested method ensembles several learning modules to extract more discriminative representation of texture and appearance cues of CXRs, thereby facilitating more accurate classification. Particularly, we integrated two sets of features (texture descriptors and deeper features) representing a rich accumulation of both local and global characteristics. In addition to learnable scaling and information-rich features, an ensemble classifier using various machine learning models is used for classification. Our classification module included support vector machine, XGBoost and extra trees modules. Extensive evaluation, supported by ablation and comparison studies, is conducted utilizing two benchmark datasets to assess the model's performance via a cross-validation strategy. Using various metrics, the results document the robustness of our ensemble classification system with higher accuracy of 98.20% and 97.85% for the two data sets, respectively.

Keywords: Ensemble Classifier; Autoencoder; Artificial Intelligence; Feature Fusion

1. Introduction

The Coronavirus (COVID-19) has rapidly propagated across the globe since its emergence in late 2019. The global pandemic precipitated by COVID-19 has unleashed a catastrophic assault, afflicting over 287M individuals and claiming the lives of approximately 5.4M people. Subsequently, medical experts worldwide have relentlessly dedicated their efforts to find vaccines and treatments for COVID-19 [1]. COVID-19 could be recognized in two distinct manners. The primary one is a real-time polymerase chain reaction (RT-PCR) nucleic identification testing. RT-PCR has aided in diagnostics, discharge evaluation, and recovery monitoring. Nevertheless, the RT-PCR sample sensitivity is restricted, and this may result in an increased false negatives. Medical imaging, i.e., chest X-ray (CXR) or computed tomography (CT), is the second technique for detecting COVID-19 [2], [3]. CT scans have multiple cross sections, making diagnosis time-consuming and costly. Technicians must make numerous adjustments during the process. Inadequate disinfection between CT technicians and patients can cause cross-contamination. Furthermore, in some areas, a shortage of radiologists poses a challenge [4]. Given the enormous impact of artificial intelligence (AI) methods for health-related imagery, a number of scientists have turned to these resources in the recent COVID-19 scenario more accurately, quickly, and affordably [5]. Deep learning (DL) is highly effective and produces better results for CXR image classification. It has made significant progress in feature learning and representing features [6]. In summary, early, precise, and fast COVID-19 detection is critical for timely isolation and medical intervention. Such measures are pivotal for patient diagnosis, epidemic prevention, and public preventive care.

In this work, we propose a hybrid architecture based on an ensemble classifier that integrates a scale-invariant input module and an ensemble cascaded classification for lung disease detection. Our proposed model has been

E-mail: Addresses {abeer.abdelhamid,gehadsaleh,fahmikhalifa}@mans.edu.eg; {olaki58,fahmi.khalifa}@morgan.edu, wdeabes@tamusa.edu

evaluated using two data sets and compared against of the shelf networks as well as recent COVID-19 detection methods. The key contributions of the present study are: (i) a robust ensemble design that integrate feature fusion of information-rich features for classification as compared to direct feature concatenation; (ii) prominent disease-related features are retained through the a learnable scale-invariant module compared with conventional image resizing algorithms; (iii) we developed a hybrid architecture to extract prominent features, thereby aiding in intermediate feature learning; and (iv) we demonstrated improved system classification accuracy using cross-validation and two data sets for evaluation.

The paper is structured into five sections starting with the introduction to CXRs and its utilization in modern CAD systems for COVID detection and diagnosis long with the outlined contributions of this work, is given in Section 1. The relevant review of the recent and related literature work is provided in Section 2. Full descriptions of the methodology and the details of the learnable modules and strategies are completely specified in Section 3. Data description, employed evaluation criteria, conducted experiments, and results and discussion are given in Section 4. Associated conclusions, and future work suggestions are outlined in Section 5.

2. Related Work

In applications that use image-based data, AI approaches have consistently produced accurate and dependable results. In recent years, researchers have investigated and analyzed CXR images to identify COVID-19 using DL-based techniques (e.g., [7]–[13]). A technique to automate the detection of COVID-19 was developed using extended segmentation-based fractal texture analysis and discrete wavelets [14]. Selected optimal features are combined with an entropy-controlled genetic algorithm and implemented through serial approach. To detect the chosen features, different ML classifiers were used. The Naïve Bayes classifier achieved 92.6% accuracy, compared to other ML algorithms. A hybrid shape-based (HOG-based) and convolutional neural network (CNN)-derived features was proposed by [15]. Integrated features improved overall performance by allowing classifiers to gather insights from the amalgamated data. They used three CT datasets of 328, 1,972, and 1,608 images for pneumonia, COVID-19, and normal individuals, respectively. VGG16 + HOG+SVM accomplished an accuracy of 99.4%. This suggests that the proposed combined feature can improve SVM accuracy in COVID19 diagnosis. Pre-trained CNNs were combined by [16] with a pyramid MLP-mixer module to classify 4,099 CXR images which contained 1,464 COVID-19, 1,294 pneumonia, and 1,341 normal patients. They attained a 98.3% of accuracy with their model. Singh et al. [7] utilized a customized stacked ensemble model compromising four learners based on CNN and Naïve Bayes serving as meta-learner for categorizing CXR images. Their method scored an accuracy of 98.67%. The performance of seven CXR-DL-based COVID detection modules was introduced by El Asnaoui et al. [17]. In their study, Inception-ResNetV2 achieved an accuracy of 92.18%. A DL-based ensemble module based on CNN for binary and multi-calss classification was proposed by Bhardwaj et al. [10]. Their method provided an average accuracies of 98.33% and 92.36% for binary and multiclass, respectively. A similar approach was proposed in [11], where the experimental results showed that their approach achieved an accuracy of 91.2% for the multi-class problem. Although the used dataset was large and TL concept was applied, the overall accuracy wasn't high. A modified COVIDnet model for COVID-19 diagnosis, called EDL-net, was proposed by Tang et al. [13]. The suggested framework achieved a detection accuracy of 95%. COVID-19 identification was carried out using multiclass and hierarchical classification tasks in [18]. In the classification schema, both early- and late-fusion methods were used using texture features and CNN-extracted features. They used a dataset consisted seven classes with a total 1,144 CXR images. The evaluation results showed that the suggested method was effective, resulting in a F1-score of 0.89 to acquire COVID-19 recognition.

Recently, a method employing a pre-processing step fusing 800 CXRs was developed by Waisy et al. [19]. They combined the weighted decisions of the two trained modeled suggestions and accomplished an accuracy of 98%. Balasubramaniam et al. [20] proposed a model based on ensemble learning technique for COVID-19 detection. They applied their method on a dataset consisted of 5,000 CXR images. Their approach achieved an accuracy of 92.3%. Hossain et al. [21] presented an AI-based solution for COVID-19 infection identification by integrated a weighted CNN fusion strategy with an attention module. The former combined multiple base pretrained CNN models, such as ResNet50V3, VGG-16, and InceptionV3 models. The utilized attention module identified important features while resisting overfitting and leverage an LC layer. The dataset used includes 1,848 CXR images with equal counts from the COVID-19, healthy, and pneumonia categories. Experimental outcomes revealed that the fusion model scored an accuracy of 96.75%. Additionally, Paul et al. [22] offered an ensemble approach using an inverted bell curve weighted ensemble with DenseNet161, ResNet18, and VGG-16 networks. Their binary- and three-class classifications were performed on an internal validation dataset of 1,214 images, involving 683 COVID-19 instances, with an 99.84% accuracy. Kumar et al. [23] suggested an ensemble model for detecting COVID-19 infection using multiple transfer learning models such as GoogLeNet, EfficientNet, and XceptionNet. They used a dataset contained four classes: COVID-19, pneumonia, tuberculosis, and normal. The designed ensemble model enhances the effectiveness of the classifier on multiclass and binary COVID-19 datasets. Their proposed method scored an accuracy of 99.2% on four class classification and 99.3% for the binary classification.

In sum, various studies have been conducted and demon-

strated promising results. Most of the methods employed DL-based approaches; however, the overall accuracy still need improvement. While, some methods utilize ensemble learning, some of these studies have a small number of subjects to test. Traditionally, the original image is used as the input to the CNN, that might not be sufficient to give a high accuracy score. Furthermore, various studies have integrated CXR and CT data sets; however, higher-order texture features were not deeply investigated in addition to the deep features. This work aims to present a robust pipeline based on feature fusion of higher-order texture features and CNN-derived deep features in order to get high accuracy with minimum loss. Additionally, our design is based on an ensemble classifier that integrates three ML algorithms with five fold cross-validation for evaluation. Our proposed model is hybrid in nature and has been evaluated using two data sets and compared against of the shelf networks as well as recent COVID-19 detection methods.

3. Analysis Pipeline

Figure 1 illustrates the proposed analysis pipeline, comprising multiple main analysis blocks: learnable-scale invariant module, the proposed deep feature extraction model, feature fusion, and ensemble classification. Before data analysis, a preprocessing stage is conducted, which is essential to improve the input data for subsequent processing operations. This stage is crucial as it ensures that a given model is generalizable, particularly when evaluated on data sets outside of the training cohort. Preprocessing step also minimizes data noises and/or deformations, allowing the deep architectures or hand-crafted features extraction to perform its tasks effectively and quickly. All stages are further detailed next.

A. Learnable Input-Scaling

Evaluation data sets in the filed of medical image analysis are usually collected at different sites and machines. Thus, collected images have different sizes. Also, modern CAD system for disease diagnosis that utilize pretrained networks requires input image downscaling to fit the model's input layer. This results in a huge loss of prominent features. To avoid this, in our design we have developed a learnable module that utilizes auto-encoder to resize input CXRs images without losing rich information as compared with traditional downscaling techniques [24]. Auto-encoders are well known for their ability to filter noise and irrelevant information while minimizing information loss in output reconstruction [25]. In our pipeline, the learnable input scaling module resize the CXRs to a 224×224 using an auto-encoder structure to be utilized as the input in a pre-trained CNN-based feature extraction.

The block diagram of the learnable input scaling module is depicted in Figure 2. The encoding path comprises successive convolutional and pooling layers to produce small sized feature maps, AE_m, of the desired size of 224×224, which subsequently undergoes convolution, transposed convolution, and reshaping to achieve dimensions of $224 \times 224 \times 4$. Both original and processed AE_ms are then combined via concatenation, generating high and low-resolution images. The former is generated from AE_m and is then utilized for preprocessing and feature extraction and classification. The latter aids in module training to minimize reconstruction error and learn crucial features while discarding redundancy and noise. Training employs a hybrid customized loss function that incorporates pseudo-Huber and log-cosh loss functions. In particular, Pseudo-Huber loss (see Eq. 1) is used due to its robustness against outliers in addition to its unique behaviour that resemble squared (for small errors) and absolute losses (for large errors) using one tunable hyper-parameter, λ [26].

$$_PH(\nu) = \lambda^2 \left(\sqrt{1 + \left(\frac{\nu}{\lambda}\right)^2} - 1 \right) \quad (1)$$

The Log Cosh loss function $logcosh(x) = log(cosh(\nu))$, on the other hand is similar to Huber loss, while hold the benefits of being double differentiable everywhere [27]. In the above equations, ν is the the variance between the observed and anticipated outcomes.

After resizing, multiple image preprocessing are conducted on the resized images. Image pre-processing is a vital process in the analysis of medical images tasks. This step range from simple task, such as noisy removal to more complicated tasks like, histogram equalization and data augmentation, which are for vital for enhanced and right classification. In this study, we used data normalization, in which all pixel values were rescalled to $[-1, 1]$ using $\left(\frac{1}{255}\right)$ as the pixel-wise multiplication. Given a grayscale image I with maximum (I_{max}) and minimum (I_{min}) values, the normalized image, $R = \frac{I - I_{min}}{I_{\max} - I_{\min}}$. After data normalization, we applied histogram equalization in order to increase the global contrast. This allows areas with low local contrast to gain contrast. Histogram equalization frequently produces artificial appearances in images; however, it is extremely useful for scientific images in nature such as thermal, satellite, or X-ray images [28].

B. Feature Extraction

In medical image classification problems, feature extraction and selection are critical. A number of images may be classified based on its primary distinguishing features, that are found using a suitable feature extraction technique [29]. To enhance the accuracy of CXR classification, in the proposed pipeline we integrated two types of CXR-derived information-rich features: deep CNN–derived along with texture and radiomic features. These feature-driving algorithms are well-known for their proficiency in medical image classification tasks. The extracted features are then integrated where the layers of these models are combined using the stacking method to create an ensemble classifier, resulting in a more robust classification.

Firstly, deep features are extracted from the preprocessed image using pretrained neural network architecture. The goal of this stage is to diagnose COVID-19 (i.e.,

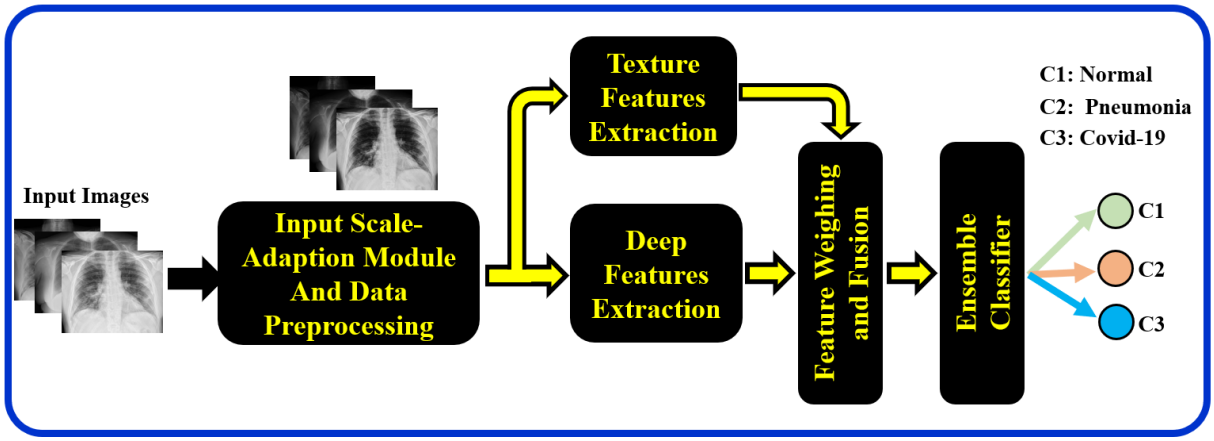


Figure 1. Schematic of the developed pipeline for lung disease detection.

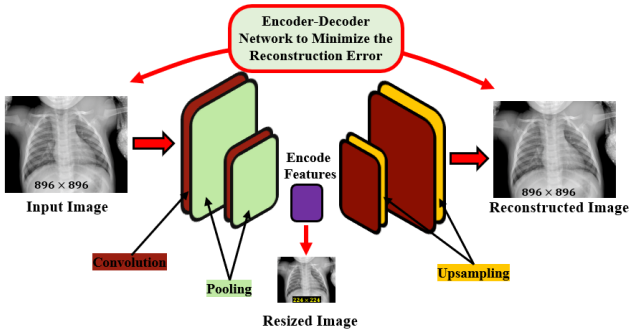


Figure 2. A schematic showing the structure of the learnable input scaling module.

categorize an input CXR image to one of the three classes). To attain the goal, we utilized ResNet50 pretrained on the ImageNet dataset [30]. ResNet is cutting-edge CNN architecture that presented an improvement over traditional CNN architectures due its ability to address the vanishing gradient problem often encountered by CNN modules, which is circumvented using the residual layers. The latter enable direct transfer of data from one layer to another, skipping some layers. By doing so, the network (with very deep architecture) to develop residual functions, that represent basically the variations between the desired output and the present result at the current layer. Thus, through intrgerating a pretrained ResNet50 into our ensemble architecture, we utilize its ability to learn detailed features and patterns from CXR images. The CXRs are resized to 224×224 using the learnable input scaling module and then we removed the final fully connected layer to get the feature vector. Obtaining these feature vectors required minimal computational capacity. We deployed pre-trained ImageNet dataset weights instead of fine-tuning ResNet50 for the data set, as it is able to offer features for most images.

Secondly, a set of texture features from the CXRs are also included to enhance disease differentiation by helping the proposed architecture to harness the spatial interactions between pixels' intensities. The features are computed using two higher-order texture feature descriptors. This set of features are derived from using GLCM based Haralick features and the gray-level run length matrix (GLRLM) features. GLCM produces a square matrix with the same dimension as the number of grey levels in the image. Each GLCM cell represents the total number of co-occurring associated grey levels in the image. The matrix must have a reasonably high occupancy level in order for the statistical estimate to be reliable. Thus, either the number of grey level values is reduced or a larger window should be used [31]. The GLCM matrix, G, can be calculated as follows [32]:

$$G_{\Delta x,\Delta y}(i,j) = \sum_{x=1}^{K}\sum_{y=1}^{M}\begin{cases}1, & R_i(x,y)=i, R_i(x+\Delta x, y+\Delta y)=j\\ 0, & \text{otherwise}\end{cases} \quad (2)$$

where $R_i(x,y)$, and $\{\Delta x, \Delta y\}$ are pre-processed data with $K \times M$ dimension at the spatial position $\{x, y\}$; and the spatial offset in the image I, respectively. The GLCM matrix's second-order statistical analysis yields various parameters that are widely used as texture features in medical data classification research [33]. In our work, we extracted the five most commonly used GLCM features from each image: energy, contrast, correlation, homogeneity and dissimilarity [34]. We statistically evaluated the GLCM features to select the most informative and distinctive ones.

Furthermore, the GLRLM texture features are utilized in the method we employ as a higher-order statistical texture feature set. GLRLM investigation, like GLCM, commonly herbal extracts the spatial plane features of pixels based on the high-order statistics of their immediate neighbors [35]. Thus, GLRLM features texture patterns assessment deliver discriminative power for image classification and supplement other features such as color, shape, and intensity for comprehensive representations and better performance. The technique generates a normalized 2D feature matrix, with each component representing the overall number of occurrences of the graylevel in the given direction [36]. Typically, GLRLM extractor captured information of pixel

pairs at angels 0, 45, 90, and 135^o. Mathematically, each element $L(i, j|\theta)$ of the run length matrix L, represents the number of runs with pixels of graylevel intensity and length of run equal to i and j,respectively along a specific orientation, $i, j \in [0, 255], \theta \in \{0, 45, 90, 135\}$. From L for an input image of size $N \times M$, many features including short/long run emphasis (SRE/LRE), greyLevel/run Length non-uniformity (GLN/RLN), run percentage (RP), low/high gray level run emphasis (LGRE/HGRE) for a given θ can be calculated using the number of greylevel (g) and number of discrete run lengths (r) of a given image as as follows:

$$SRE = \sum_{i=1}^{g}\sum_{j=1}^{r}\frac{L(i,j)}{j^2} \tag{3}$$

$$LRE = \sum_{i=1}^{g}\sum_{j=1}^{r} j^2 L(i,j) \tag{4}$$

$$GLN = \sum_{i=1}^{g}\left(\sum_{j=1}^{r} L(i,j)\right)^2 \tag{5}$$

$$RLN = \sum_{i=1}^{r}\left(\sum_{j=1}^{g} L(i,j)\right)^2 \tag{6}$$

$$RP = \frac{1}{N \times M}\sum_{i=1}^{g}\sum_{j=1}^{r} P(i,j) \tag{7}$$

$$LGRE = \sum_{i=1}^{g}\sum_{j=1}^{r}\frac{L(i,j)}{i^2} \tag{8}$$

$$HGRE = \sum_{i=1}^{g}\sum_{j=1}^{r} i^2 L(i,j) \tag{9}$$

C. Feature Fusion and Ensemble Classification

Following feature extraction, we used a feature fusion approach for combining derived CXR features for classification. This technique usually enhances the discriminative classification power of a given system by delivering comprehensive representations of information-rich features. Finally, the fused CXR features are provided as input to a ML classifier for prediction. Our design is based on an ensemble classifier that integrates three ML algorithms: SVM, XGBoost, extra trees. Also, five fold cross-validation technique was used for training the ensemble models on the CXRs datasets. **The SVM** is the first ML algorithm that is employed in our design. SVM, as a supervised ML technique, constructs a hyperplane to effectively separate two input classes while maximizing the margin. The margin indicates the difference between the SVs and the hyperplane [37]. The main advantage of SVM lies in the fact that it powerfulness ability to handle high-dimensional data and non-linear classification using kernel functions, while being effective with limited training samples.

In addition to SVM, **XGBoost** is also included in our ensemble model. It has gradient boosting at its core. The XGBoost algorithm differs from simple gradient boosting in that it uses a multi-threaded approach to add weak learners, rather than adding them sequentially. The XGBoost algorithm stands out from basic gradient boosting when it adds weak learners in a multi-threaded approach, utilizing the machine's CPU cores for faster and better performance. Sparse aware implementation includes automated missing data values handling, a block structure for parallel tree design, and ongoing training to improve an existing model on new data [38]. Finally, **extra tree**, an extremely randomized classifier, is integrated in our ensemble model [39]. It is built differently than traditional decision trees. Random splits are utilized to determine the best way to divide a node's samples into two separate sets. The best split will be determined based on randomly selected features, specifically max features. Its averages out the variance problems of a single decision tree method, making it suitable for multiple sub samples of a dataset. This improves predictive accuracy and prevents overfitting [40].

4. Experimental Results and Discussion

In recent years, the need for the development of precise and accurate computer-aided diagnostic tools for disease detection has surged dramatically. Among several diseases, lung diseases has drawn global attention for the need to develop artificial intelligence-based tools for precise, accurate and effective containment and treatment plans. The primary goal of our study is to develop and evaluate a versatile pipeline with the potential to be extended for the detection of various lung diseases.

The primary objective for developing the proposed AI-pipeline is to create a robust classification model that will perform well regardless of the domain in which it can be used. System evaluation and assessment is based on two publicly-available chest x-ray datasets with compromising a total CXRs of 20,063 images [41]–[45]. The datasets consist of three classes: normal, pneumonia, and COVID-19 and are summarized in Table I.

TABLE I. Per-class distribution for the public datasets.

	Dataset		
	First	Second	Total
Normal	4,200	1,751	5,951
Pneumonia	4,273	4,273	8,546
COVID-19	4,195	1,371	5,566
Total	12,668	7,395	20,063

For experimental and training settings, the system was built with the TensorFlow framework and Keras as the backend using the Python programming language. All experiments are conducted using a windows machine with 32GB RAM, 8GB NIVIDIA graphics card, and a 12 core i7-processor. We employed Adam optimizer [46] and set the learning rate at 10^{-4} due to its effective hyperparameters'

selection. Furthermore, the batch size was fine-tuned to 32, epochs at 50, and categorical loss function was used.

To offer comprehensive assessment of the proposed system's efficacy, quantitative evaluation is conducted using various metrics: accuracy, recall, precision, and F1-Score indexes. The accuracy (AC) represents the ratio of correctly categorized labels to the total number of tested ones. Recall (RC, specificity) is a representation of the fraction of correctly identified positive (or negative) class. Further, correctly predicted positive samples out of the total predicted patterns in the positive class is referred to as precision (PR). Finally, the F1-Score indicates the harmonic mean between RC and PR. The overall accuracy for the proposed method is summarized in Table II.

TABLE II. Multi-class accuracy of comparing the ensemble classifier against various classifiers. Here, AC: accuracy, PR: precision, RC: recall; RF: random forest; and SVM: support vector machine.

	Evaluation Metrics (%)			
Classifier	AC	PR	RC	F1-Score
XGBoost	78.23	80.64	76.51	77.15
RF	81.13	82.36	82.56	82.42
SVM	85.39	83.03	84.25	83.87
Extra Trees	88.10	87.18	86.08	86.73
Our (dataset1)	98.20	97.91	98.18	96.43
Our (dataset2)	97.85	97.33	95.67	96.67

At first, we investigated the importance of the ensemble classifier on the overall system performance. Thus, we evaluated the classification process using different single ML models and Table II represents the summary of the accuracy of the tested models. For the second evaluation phase, we investigated the importance of features fusion. Thus, we conducted an ablation study where we test the system accuracy using each type of the features separately. Table III represents the summary of the obtained accuracy. As demonstrated, the proposed model's performance is highly enhanced using the fused features compared to individual features and reached to an accuracy of 98.20% and 97.85% for the first and second data sets, respectively.

TABLE III. Overall performance for the proposed model. Here, HaC: Handcrafted Features; CNN: convolutional neural network; AC: accuracy, PR: precision, RC: recall.

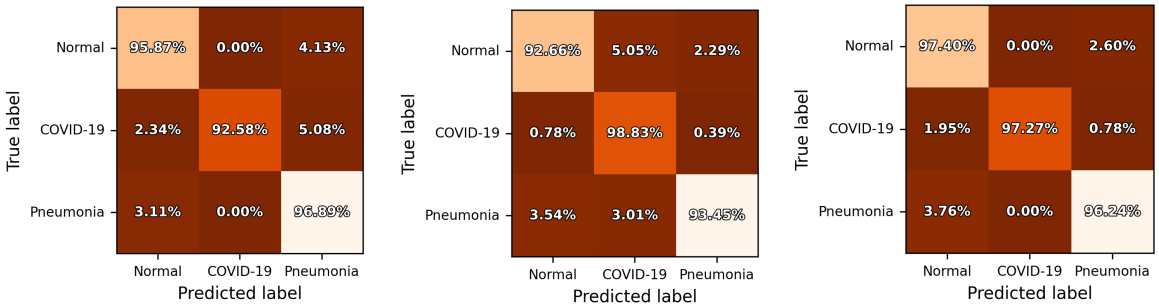
	Metrics, (%)			
Method	AC	PR	RC	F1-score
HaC Features	90.15	88.63	89.28	89.46
CNN Features	93.29	91.33	90.00	90.67
Proposed (dataset1)	98.20	97.91	98.18	96.43
Proposed (dataset2)	97.85	97.33	95.67	96.67

Particularly, the ablation and experimental studies shown in the above tables emphasize the robustness of the proposed model. First, in Table II, it observed that most of the single classifier have lower performance and the extra tress achieved the highest accuracy when tested on the extracted features. However, individual classifiers are all less than the ensemble model of the fused features. This highlights the importance of the ensemble classifier compared with single ML classification. Also, the reported results in Table III emphasis the idea that feature fusion enhances the overall performance.

Besides quantitative indexes, the architecture performance is investigated using the confusion matrices (CM) and the receiver operating characteristics (ROC) curve, which are powerful tools for the evaluation and comparison of classification models. On the first hand, CM is an extremely useful tool for determining which classes, if any, were misclassified the most. In Figure 3-c and Figure 4-a, the confusion matrices demonstrate the ability of the model correctly predicted 97%, 97%, and 96% of the normal, COVID, and pneumonia classes, respectively, for the first data set; and 95%, 96%, and 99% for the second data set.

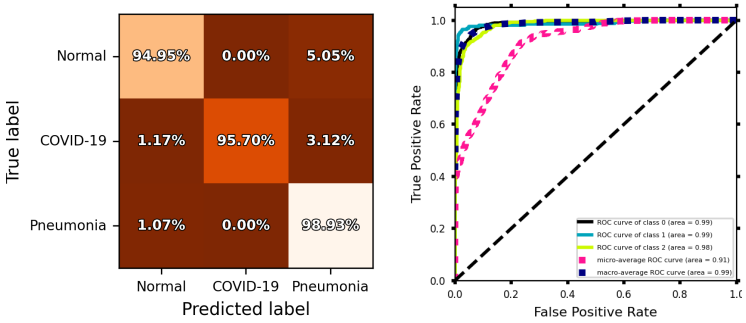
On the other hand, the ROC analysis tool is used to validate/support the reliability (and accuracy) of classification pipelines by analyzing models' output based on the relation between false and true positive rates assessed at various thresholds. The area under the curve (AUC) of an ROC can be utilized in quantitative classification models to demonstrate how well the model discriminates between classes, with "1" and "0" indicating the best and the worst performance, respectively. Figure 5-c and Figure 4-b, demonstrate the ROC curves, by the interconnected lines, for the proposed ensemble model evaluated on the first and the second data set, respectively. Notably, the ROC curve for the fused features for each class (Figure 5-c & Figure 4-b) exhibits the best performance (AUC $\approx$1.0) compared with individual feature (Figures 5-a,b). Essentially, both of the curve and its associated metric (i.e., the AUC) show that the model exhibited above-average performance in correctly classifying each class for both data sets.

Moreover, evaluation of well-known pretrained backbone CNs and recent literature work developed for COVID-19 detection are also conducted. For pretrained CNNs, we tested various architectures including VGG19, Inception, DenseNet121, and Xception and the average accuracies were 71.36%, 78.50% , 80.00%, and 86.92%, respectively and Table IV summarizes the results. Finally, the evaluation results of our method against recent literature work developed for COVID-19 detection are shown in Table V. As readily seen, the proposed model scored higher accuracy compared to other methods for COVID detection. The compared methods mainly employed on pretrained architectures ([47], [48]) and single classifier for the extracted deep features [49] or hand crafted features [50]. Our method on the other hand employed both deeper and higher order features and utilized an ensemble classifier. Additionally, for most of the compared methods the test data sets was small compared to our approach that employed a larger data set and utilized cross validation scenario for evaluation.



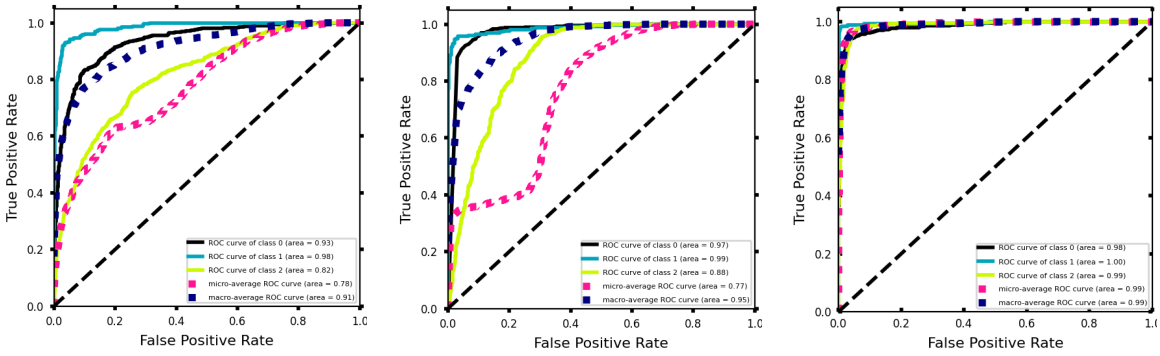
(a) (b) (c)

Figure 3. Visualization of confusion matrices of individual feature sets and the proposed method: (a) higher-order features; (b) CNN-derived features, and (c) their fusion (the proposed) on the first data set.



(a) (b)

Figure 4. Visualization of (a) confusion matrix and (b) ROC curve of our ensemble model for the second data set.



(a) (b) (c)

Figure 5. Visualization of the ROC curves of individual feature sets and the proposed method: (a) higher-order features; (b) CNN-derived features, and (c) their fusion (the proposed) on the first data set.

TABLE IV. Multi-class accuracy of various CNN architectures. Here, AC: accuracy, PR: precision, RC: recall.

	Evaluation Metrics (%)			
Model	AC	PR	RC	F1-Score
VGG16	71.36	74.47	70.11	72.22
InceptionV3	78.50	76.00	74.23	75.10
DenseNet121	80.00	82.68	85.75	84.18
Xception	86.92	89.23	87.36	88.28

TABLE V. Multi-class accuracy of recent work for COVID-19 detection. Here, AC: accuracy, PR: precision, RC: recall.

	Evaluation Metrics (%)			
Model	AC	PR	RC	F1-Score
Ismael et al. [49]	94.7	91.0	98.9	94.8
Horry et al. [47]	83.0	81.0	81.0	
Abbas et al. [48]	92.10	—	—	—
Li et al. [51]	96.4	—	92.9	—
Medeiros et al. [50]	93.05	89.55	96.28	92.80

5. Conclusions and future work

This study have introduced a hybrid pipeline incorporating ensemble feature fusion concept for distinguishing COVID-19 from normal and pneumonia cases CXRs images through multi-classification. Particularly, an AI-based combining high-level features with higher-order texture ones has been developed. Compared with literature work, our method incorporates several learning modules to extract more discriminative representation of texture and shape cues, thereby facilitating more accurate classification. Further, a cascaded machine classification module is trained to discern between classes. The potentials of the developed feature fusion architecture is supported by rigorous evaluation using two cohorts of CXR images. The experimental findings reveal in which the improved fusion approach combined with ensemble classifier outperformed the single prediction models and achieved overall accuracy of 98.20% and 97.85%. The use of hierarchical features in the design of the model enabled it to outperform traditional ML models, classical pretrained DL models, and SOTA methods.

The suggested method has a high potential for clinical application and has the possibility to reduce both pollutants and hospital burden by preventing unnecessary hospital visits. Despite the higher performance of our pipeline, our study have some limitations, such as the use of a single image modality for evaluation and the lack of challenging data sets for rigorous evaluation. Secondly, our analysis framework exploited only pre-trained CNNs as deep feature feature extraction. Thus, in future work we plan to create a multi-level classification framework that mimic physicians diagnosis of separating groups by providing global screening then performing micro classification of the potential instances. We will also investigate ensemble learning of multi-scale images. Although we have introduced a feature fusion approach, our future work will be dedicated to multi-modal features fusion. Integration between CXR-based features and CT in conjunction with using transformer-based architectures will be another research venue.

Acknowledgment

This work partially supported by the Center for Equitable Artificial Intelligence and Machine Learning Systems (CEAMLS), Morgan State University, Project #11202202; and in part funded by the National Institutes of Health (NIH) Agreement No.# 1OT2OD032581. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.

References

[1] M. Abdar, S. Salari, S. Qahremani, H.-K. Lam, F. Karray, S. Hussain, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "Uncertaintyfusenet: robust uncertainty-aware hierarchical feature fusion model with ensemble monte carlo dropout for covid-19 detection," *Information Fusion*, vol. 90, pp. 364–381, 2023.

[2] M. Nur-A-Alam, M. K. Nasir, M. Ahsan, M. A. Based, J. Haider, and M. Kowalski, "Ensemble classification of integrated ct scan datasets in detecting covid-19 using feature fusion from contourlet transform and cnn," *Scientific Reports*, vol. 13, no. 1, p. 20063, 2023.

[3] A. A. AbdElhamid, E. AbdElhalim, M. A. Mohamed, and F. Khalifa, "Multi-classification of chest x-rays for covid-19 diagnosis using deep learning algorithms," *Applied Sciences*, vol. 12, no. 4, p. 2080, 2022.

[4] C.-T. Yen, J.-X. Liao, and Y.-K. Huang, "Evaluating feature fusion techniques with deep learning models for coronavirus disease 2019 chest x-ray sensor image identification," *Sensors and Materials*, vol. 36, no. 2, pp. 683–699, 2024.

[5] M. A. Khan, M. Alhaisoni, M. Nazir, A. Alqahtani, A. Binbusayyis, S. Alsubai, Y. Nam, and B.-G. Kang, "A healthcare system for covid19 classification using multi-type classical features selection," *Comput. Mater. Contin*, vol. 74, pp. 1393–1412, 2023.

[6] J. J. M. A. Devakanth, D. A. Suresh, and R. Balasubramanian, "Feature fusion based ensemble method for multi-level lung diseases classification of x-ray images using deep learning," *International Neurourology Journal*, vol. 27, no. 4, pp. 331–349, 2023.

[7] R. K. Singh, R. Pandey, and R. N. Babu, "Covidscreen: explainable deep learning framework for differential diagnosis of covid-19 using chest x-rays," *Neural Computing and Applications*, vol. 33, no. 14, pp. 8871–8892, 2021.

[8] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, S. Mirjalili, and M. K. Khan, "Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms," in *Multimodal image exploitation and learning 2021*, vol. 11734. International Society for Optics and Photonics, 2021, p. 117340E.

[9] E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, and M. Z. Parvez, "Corodet: A deep learning based classification for covid-19 detection using chest x-ray images," *Chaos, Solitons & Fractals*, vol. 142, p. 110495, 2021.

[10] P. Bhardwaj and A. Kaur, "A novel and efficient deep learning approach for covid-19 detection using x-ray imaging modality," *International Journal of Imaging Systems and Technology*, vol. 31, no. 4, pp. 1775–1791, 2021.

[11] A. Afifi, N. E. Hafsa, M. A. Ali, A. Alhumam, and S. Alsalman, "An ensemble of global and local-attention based convolutional neural networks for covid-19 diagnosis on chest x-ray images," *Symmetry*, vol. 13, no. 1, p. 113, 2021.

[12] A. Sedik, M. Hammad, A. El-Samie, E. Fathi, B. B. Gupta, A. El-Latif, and A. Ahmed, "Efficient deep learning approach for augmented detection of coronavirus disease," *Neural Computing and Applications*, pp. 1–18, 2021.

[13] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong, and A. Barnawi, "EDL-COVID: ensemble deep learning for covid-19 case detection from chest x-ray images," *IEEE Transactions On Industrial Informatics*, vol. 17, no. 9, pp. 6539–6549, 2021.

[14] T. Akram, M. Attique, S. Gul, A. Shahzad, M. Altaf, S. S. R. Naqvi, R. Damaševičius, and R. Maskeliūnas, "A novel framework for rapid diagnosis of covid-19 on computed tomography scans," *Pattern analysis and applications*, pp. 1–14, 2021.

[15] H. Abubakar, F. Al-Turjman, Z. S. Ameen, A. S. Mubarak, and C. Alturjman, "A hybridized feature extraction for covid-19 multi-class classification on computed tomography images," *Heliyon*, 2024.

[16] Y. Liu, W. Xing, M. Zhao, and M. Lin, "A new classification method for diagnosing covid-19 pneumonia based on joint cnn features of chest x-ray images and parallel pyramid mlp-mixer module," *Neural Computing and Applications*, vol. 35, no. 23, pp. 17 187–17 199, 2023.

[17] K. El Asnaoui and Y. Chawki, "Using x-ray images and deep learning for automated detection of coronavirus disease," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 10, pp. 3615–3626, 2021.

[18] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, and Y. M. Costa, "Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios," *Computer methods and programs in biomedicine*, vol. 194, p. 105532, 2020.

[19] A. S. Al-Waisy, S. Al-Fahdawi, M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. S. Maashi, M. Arif, and B. Garcia-Zapirain, "Covid-chexnet: hybrid deep learning framework for identifying covid-19 virus in chest x-rays images," *Soft computing*, vol. 27, no. 5, pp. 2657–2672, 2023.

[20] S. Balasubramaniam and K. S. Kumar, "Optimal ensemble learning model for covid-19 detection using chest x-ray images," *Biomedical Signal Processing and Control*, vol. 81, p. 104392, 2023.

[21] M. S. Hossain and M. Shorfuzzaman, "Non-invasive covid-19 screening using deep learning-based multilevel fusion model with an attention mechanism," *IEEE Open Journal of Instrumentation and Measurement*, 2023.

[22] A. Paul, A. Basu, M. Mahmud, M. S. Kaiser, and R. Sarkar, "Inverted bell-curve-based ensemble of deep learning models for detection of covid-19 from chest x-rays," *Neural Computing and Applications*, vol. 35, no. 22, pp. 16 113–16 127, 2023.

[23] N. Kumar, M. Gupta, D. Gupta, and S. Tiwari, "Novel deep transfer learning model for covid-19 patient detection using x-ray chest images," *Journal of ambient intelligence and humanized computing*, vol. 14, no. 1, pp. 469–478, 2023.

[24] O. Akinniyi, M. M. Rahman, H. S. Sandhu, A. El-Baz, and F. Khalifa, "Multi-stage classification of retinal oct using multi-scale ensemble deep architecture," *Bioengineering*, vol. 10, no. 7, p. 823, 2023.

[25] S. Chen and W. Guo, "Auto-encoders in deep learning—a review with new perspectives," *Mathematics*, vol. 11, no. 8, p. 1777, 2023.

[26] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Transactions on image processing*, vol. 6, no. 2, pp. 298–311, 1997.

[27] X. Yu, Y. Liang, X. Lin, J. Wan, T. Wang, and H.-N. Dai, "Frequency feature pyramid network with global-local consistency loss for crowd-and-vehicle counting in congested scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9654–9664, 2022.

[28] M. Vinodhini, S. Rajkumar, M. V. K. Reddy, and V. Janesh, "Detection of post covid-pneumonia using histogram equalization, clahe deep learning techniques: Deep learning," *Inteligencia Artificial*, vol. 26, no. 72, pp. 137–145, 2023.

[29] M. A. Ozdemir, G. D. Ozdemir, and O. Guren, "Classification of covid-19 electrocardiograms by using hexaxial feature mapping and deep learning," *BMC medical informatics and decision making*, vol. 21, no. 1, pp. 1–20, 2021.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[31] T. Sathwik, R. Yasaswini, R. Venkatesh, and A. Gopal, "Classification of selected medicinal plant leaves using texture analysis," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE, 2013, pp. 1–6.

[32] W. Sun, N. Zeng, and Y. He, "Morphological arrhythmia automated diagnosis method using gray-level co-occurrence matrix enhanced convolutional neural network," *IEEE Access*, vol. 7, pp. 67 123–67 129, 2019.

[33] B. Dhruv, N. Mittal, and M. Modi, "Study of haralick's and glcm texture analysis on 3d medical images," *international journal of Neuroscience*, vol. 129, no. 4, pp. 350–362, 2019.

[34] S. Jafarpour, Z. Sedghi, and M. C. Amirani, "A robust brain MRI classification with GLCM features," *International Journal of Computer Applications*, vol. 37, no. 12, pp. 1–5, 2012.

[35] A. K. Mohanty, S. Beberta, and S. K. Lenka, "Classifying benign and malignant mass using glcm and glrlm based texture features from mammogram," *International Journal of Engineering Research and Applications*, vol. 1, no. 3, pp. 687–693, 2011.

[36] Ş. Öztürk and B. Akdemir, "Application of feature extraction and classification methods for histopathological image using glcm, lbp, lbglcm, glrlm and sfta," *Procedia computer science*, vol. 132, pp. 40–46, 2018.

[37] A. Burkov, *The hundred-page machine learning book*. Andriy Burkov Quebec City, QC, Canada, 2019, vol. 1.

[38] S. Ramraj, N. Uzir, R. Sunil, and S. Banerjee, "Experimenting xgboost algorithm for prediction and classification of different datasets," *International Journal of Control Theory and Applications*, vol. 9, no. 40, pp. 651–662, 2016.

[39] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, pp. 3–42, 2006.

[40] S. A. Manaf, N. Mustapha, M. N. Sulaiman, N. A. Husin, H. Z. M. Shafri, and M. N. Razali, "Hybridization of slic and extra tree for object based image analysis in extracting shoreline from medium resolution satellite images." *International Journal of Intelligent Engineering & Systems*, vol. 11, no. 1, 2018.

[41] Prashant Patel, "Pneumonia COVID-19 Image Dataset)," https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia, online; accessed 21 November 2023.

[42] Nabeel Sajid, "COVID-19 Patients Lungs X Ray Images," https://https://www.kaggle.com/datasets/nabeelsajid917/covid-19-x-ray-10000-images, online; accessed 21 November 2023.

[43] Tarandeep Singh, "COVID-19 & Normal Posteroanterior(PA) X-rays," https://www.kaggle.com/datasets/tarandeep97/covid19-normal-posteroanteriorpa-xrays, online; accessed 21 November 2023.

[44] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al Emadi *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.

[45] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan *et al.*, "Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images," *Computers in biology and medicine*, vol. 132, p. 104319, 2021.

[46] S. Bera and V. K. Shrivastava, "Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification," *International Journal of Remote Sensing*, vol. 41, no. 7, pp. 2664–2683, 2020.

[47] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, "X-ray image based covid-19 detection using pre-trained deep learning models," 2020.

[48] G. Maguolo and L. Nanni, "A critic evaluation of methods for covid-19 automatic detection from x-ray images," *Information Fusion*, vol. 76, pp. 1–7, 2021.

[49] A. M. Ismael and A. Şengür, "Deep learning approaches for covid-19 detection based on chest x-ray images," *Expert Systems with Applications*, vol. 164, p. 114054, 2021.

[50] E. P. Medeiros, M. R. Machado, E. D. G. de Freitas, D. S. da Silva, and R. W. R. de Souza, "Applications of machine learning algorithms to support covid-19 diagnosis using x-rays data information," *Expert Systems with Applications*, vol. 238, p. 122029, 2024.

[51] H. Li, N. Zeng, P. Wu, and K. Clawson, "Cov-net: A computer-aided diagnosis method for recognizing covid-19 from chest x-ray images via machine vision," *Expert Systems with Applications*, vol. 207, p. 118029, 2022.

Abeer AbdElhamid, M.Sc., received her M.Sc. degree in Electronics and communication Engineering, ECE department, Faculty of Engineering, Mansoura University, Egypt, in 2022. She is currently pursuing her Ph.D. studies with the ECE department, Faculty of Engineering, Mansoura University. Abeer has more than 4 years of hands on experience in the field of image/signal processing and analysis, focused on the application of AI/ML for medical data analysis for disease diagnosis.

Oluwatunmise Akinniyi, M.Sc., received her B.Sc. and M.Sc. degrees in Computer Science from The Federal University of Technology Akure, Nigeria and Morgan State University (MSU), USA in 2020 and 2023, respectively. She is currently enrolled in Computer and Electrical Systems Engineering Ph.D. program, MSU, and works with the Center for Equitable Artificial Intelligence and Machine Learning Systems. Oluwatunmise has more than 5 years of hands-on experience in the fields of data analytics, image processing, ML, medical image analysis, and computer aided diagnosis and has authored/co-authored five peer-reviewed publications.

Gehad A. Saleh, M.D., is a lecturer of radiology & intervention radiology, Faculty of medicine, Mansoura University,Egypt. Dr. Saleh received the B.Sc. and M.Sc. degrees from Mansoura University, Mansoura, Egypt, in 2010 and 2015, respectively. She also received her Ph.D. degree in 2020. Dr. Saleh's main research interests include machine learning application for medical diagnostics with application in liver, oncology and female pelvic, head and neck radiology using various imaging techniques.

Wael Deabes. Ph.D., is an Instructional Associate Professor and ESET program coordinator in the Department of Mathematical, Physical, and Engineering Sciences at the Texas A&M, San Antonio. He has his master and PhD in Electrical and Computer Engineering from Mansoura University and Tennessee Technological University in 2003 and 2010, respectively. He has more than 50 conferences and journal publications. His research interests include: tomography systems, intelligent sensing and measurement systems, IoT, smart control systems, smart cities, inverse theory, artificial intelligence and deep learning methods, intelligent transportation system, and global optimization algorithms.

Fahmi Khalifa, Ph.D., received his B.Sc. and M.Sc. degrees in Electronics and Electrical Communication Engineering from Mansoura University, Egypt in 2003 and 2007, respectively. He received his Ph.D. degree in 2014 in Electrical Engineering from ECE Department, University of Louisville, USA. Currently, he is an Associate Professor with the Electronics and Communications Engineering Department, Mansoura University, Egypt. He has more than 15 years of hands-on experience in the fields of image processing, machine learning, medical image analysis, computer-aided diagnosis, and digital and analog signal processing. Dr. Khalifa has authored/co-authored more than 180 peer-reviewed publications appearing in high-impact journals, selective peer-reviewed top-rank international conferences, and leading edited books.