Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
COVID-19 and Non-COVID-19 Classification using Multi-layers Fusion From Lung Ultrasound Images

Ghulam Muhammad\textsuperscript{a,b}, M. Shamim Hossain\textsuperscript{a,c,*}

\textsuperscript{a} Chair of Pervasive and Mobile Computing, King Saud University, Riyadh 11543, Saudi Arabia
\textsuperscript{b} Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
\textsuperscript{c} Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

ABSTRACT

COVID-19 or related viral pandemics should be detected and managed without hesitation, since the virus spreads very rapidly. Often with insufficient human and electronic resources, patients need to be checked from stable patients using vital signs, radiographic photographs, or ultrasound images. Vital signs do not often offer the right outcome, and radiographic photos have a variety of other problems. Lung ultrasound (LUS) images can provide good screening without a lot of complications. This paper suggests a model of a convolutionary neural network (CNN) that has fewer learning parameters but can achieve strong accuracy. The model has five main blocks or layers of convolution connectors. A multi-layer fusion functionality of each block is proposed to improve the efficiency of the COVID-19 screening method utilizing the proposed model. Experiments are conducted using freely accessible LUS photographs and video datasets. The proposed fusion method has 92.5% precision, 91.8% accuracy, and 93.2% retrieval using the data collection. These efficiency metric levels are considerably higher than those used in any of the state-of-the-art CNN versions.

1. Introduction

Till now, millions of people have become diagnosed by the coronavirus epidemic and mortality is rising every day. As coronavirus is extremely infective, it spreads rapidly among those linked to COVID-19 (Corona Virus Disease – 19) infected persons. The virus spills over saliva droplets or swab discharge from the nose while an affected individual with COVID-19 is tinging or sneezing. A COVID-19 afflicted individual may feel dry cough, muscle pain, headache, fever, sore throat, and gentle to moderate respiratory disease. Nevertheless, elderly adults and those with existing medical problems such as coronary disorder, asthma, chronic pulmonary disease, and cancer are most prone to experience severe illnesses.

Being an uncertain source of infection of the pneumonia form and a new virus to be created by mutation, it is very difficult for COVID-19 patients to find a remedy by vaccination or treatment. Further studies and social disparities within the community in high alerted regions in various countries impacted by the corona pandemic are also suggested as per the WHO.

The origin of the 2019 coronavirus outbreak is a novel coronavirus, SARS-CoV-2, (COVID-19). Patient assessors may wish to order X-ray images to screen COVID-19 findings, assess their severity, or to suggest additional disease etiologies. It will change the interest of the COVID-19 treaters through radiographic tests and may contribute to medical treatment or supportive therapeutic decisions, such as hospitalization, a requirement for additional supervision or expectation of disease risks.

The American College of Radiography (ACR) recommends that (i) the test for or first-line treatment of COVID-19 will not be performed using computed tomography (CT), (ii) for admitted symptomatic patients with particular clinical signs of CT, the CT will be used in a vigilant and reserved way, (iii) adequate protocols can be used for infectious management before future patients are screened, (iv) regular chest CT does not indicate that a person has no COVID-19 infection and pathological COVID-19 evaluation may not be accurate.

When used in the assessment of COVID 19 patients with symptoms of pneumonia in comparison to chest x-rays, lung ultrasound (LUS) could provide increased diagnostic accuracy. LUS is particularly sensitive and sometimes passes chest x-rays of many pulmonary infections. The detection system for pneumonia and associated lung diseases has already been developed in lung ultrasound [1]. The technique of diagnosing lung disease has been proposed as favored, particularly in resource-limited environments, such as crises or low-income countries [2] and has begun to substitute x-rays for first-line testing [3,4].

Since the outbreak of COVID-19, many researchers are trying to develop systems to detect the virus using radiographic images (CT scan or X-ray images). Very few researchers are involved in developing such
systems using LUS images, cough sound, or Electrooculogram images. A multi-source system regarding COVID-19 detection has not been developed yet because all the available related datasets are single-source. For example, there is no dataset where the CT scan images and the X-ray images are available from the same patient. As a dataset is a prerequisite to developing a system, currently, a multi-source system to combat COVID-19 is very difficult to realize. Instead, a system with different levels or layers of fusion can be a realistic one. There are some recent heterogeneous systems that deploy mass surveillance systems involving social distancing measures and body temperature measurements in a crowded place [5].

Most of the COVID-19 screening systems are developed using deep learning, especially convolutional neural networks (CNN) for images. Some systems used pretrained CNN models, while some systems used CNN models from scratch. There are two types of pretrained CNN models: (i) very deep and narrow, and (ii) moderately deep and wide. In the first type, the models consider that depth is the most important factor to increase accuracy. Some examples are AlexNet and VGG Net. In the second type, the models consider that the width of the convolutional filters is the most important factor. Some examples of the second type are wide residual network and ResNet. Both types have their advantages and disadvantages in terms of accuracy and number of parameters. To compromise between these two terms, a new term called the information density is sometimes used. The information density is calculated as a ratio between the accuracy and the number of parameters (in million) of a model. The recently proposed tree-based deep model has a very high information density [6]. In a pandemic situation such as COVID-19, due to a very large number of affected people, the screening system should be operated in less time, if not in real-time.

The fusion of deep learned features is a very effective way to enhance the accuracy of a health-related system [7,8]. Typically, the fusion can be at three levels: feature level, classification level, and decision level. In the feature level fusion, a multi-stream CNN can be used and the features from all the streams are concatenated to produce a unified feature vector. The feature vector is then fed into a single classifier. In the classification level fusion, features from each stream are fed to a different classifier. The classifiers’ outputs are then fused. In the decision level fusion, each classifier produces a decision, and then a majority-voting type technique can be used to make the final decision based on the decisions of the classifiers. One of the problems of a multi-stream CNN model is the increased computational complexity of the model. If we have three streams, the number of parameters is increased by approximately three times. Therefore, if there is a problem with the availability of a powerful computer, the multi-stream CNN model is often avoided. An alternative solution is to fuse features from different layers of a single CNN model.

In this paper, a multi-layer fusion approach of a CNN model is proposed for a COVID-19 screening system using LUS images. First, a light CNN model is proposed; the number of learnable parameters of the model is less than 0.4 million. Then features from different layers of the model are fused to produce a feature vector to be fed into a multilayer perceptron (MLP), in the form of a fully-connected layer, for the classification. Various layers of fusion are also investigated. The major contributions of the paper are as follows: (i) an efficient light CNN model is proposed for the COVID-19 screening system, (ii) a multi-layer features fusion approach is introduced in the system, and (iii) an excellent information density is achieved by the system. It can be noted that the multi-layer features fusion is not common in any recently developed COVID-19 screening systems. Also, such systems using LUS images are rare in the literature.

The rest of the paper is organized as follows. Section 2 provides a brief literature review on two points: fusion of medical signals, and systems with LUS images. Section 3 describes the proposed system using a multi-layer feature fusion of a CNN model. Section 4 presents the experiments, and Section 5 gives the conclusion of the paper.

2. Literature review

This section is divided into two subsections. In the first subsection, a brief review of the fusion of medical signals is provided. In the second subsection, some recent developments of deep learning-based systems using LUS images are discussed.

2.1. Fusion of medical signals

Steenkiste et al. [9] developed a robust model focused on sensor fusion methods to enhance sleep apnea prediction. The system was used to gather and integrate reverse shortcut connections from multi-sensor data including oxygen saturation, cardiac rhythm, thoracic breathing, and respiratory bands in the abdomen. The robustness and efficiency were tested in both CNN and LSTM prototypes.

Yang et al. [10] proposed a cross-subject emotion recognition system using multi-sensors input. The sensors’ data are vital signs such as body temperature and plethysmography, electroencephalogram (EEG) signals, electromyography (EMG) signals, and electrooculogram (EOG) signals. 10 different types of feature vectors were formed from these signals to produce a high-dimensional feature space to be fed into a support vector machine (SVM) classifier. Before feeding the features, a significant test was performed per feature to select the significant features only. Then, only the selected features were inputs to the SVM. Accuracies ranged between 72% and 89% were obtained using different datasets.

Miao et al. [11] developed a cuffless blood pressure (BP) prediction system using double-channel high-pulse waves and a single-channel electrocardiogram (ECG). 24 features from the pulse waves and 11 features from the ECG signal were extracted. Then a weakly-supervised feature selection method selected the significant features, which were fed to a multi-instance regression classifier to predict the BP. A private dataset of 85 patients having hypertension or hypotension was used to verify the system. Experimental results showed that the developed system achieved a reliable prediction accuracy.

Gu et al. [12] provided a design for the situation awareness recognition system to ensure the precision of the signal transmitting to a mine and the safety of mine employees. In the system, multi-sensor data were analyzed and knowledge entropy theory was used to give various weights to different parameters of the data. A random forest (RF)-SVM-based classifier was used to predict the situation of the mine.

Muzammal et al. [13] provided coherent activity data that could be used for therapeutic research by introducing a multi-sensor data fusion system. The data collected from the body sensor networks (BSNs) were combined and incorporated into an ensemble classifier, which was installed in cloud computing. The evidenced-based analysis found that a novel kernel random forest tree was the best performer in the system.

In [14], an audiovisua emotion recognition system was developed using CNNs and sparse autoencoder (SAE). Mel-cepstrum features with their temporal derivatives were extracted from audio signals and fed into a 2D CNN. Video signals were fed into a 3D CNN. The CNN features were fused using the SAE, followed by an SVM classifier. The system achieved good accuracy in various publicly available emotion datasets.

Lin et al. [15] developed a hybrid multi-sensor fusion BSN framework to enable smart healthcare. The goal was to address the lack of conventional multi-sensor fusion approaches in clinical uses. This research discussed in depth the basic roles of different layers of the framework. The first layer was a data preprocessing layer, where data were preprocessed using data cleaning, data conversion, data integration, and data compaction. In the fusion layer, an explainable neural network and AI-based technique were used to fuse data from cross-domain and multi-sensor.

Amin et al. [16] proposed an EEG-based motor imagery classification system by fusing different layers of a CNN model. In the system, the features from different layers were concatenated before being fed into a fully-connected layer. The authors proved that fusing features from
various layers achieved better results than without fusing.

A system of medical image fusion relying on a Rolling Guidance Filtering (RGF) was introduced in [17]. The system used the RGF to transform the source artifacts into either geometric or detailed elements. The types of fusion were used. A linear prediction-based fusion was used for the geometric component, while a sum-modified Laplacian was used for the system had its own fusion role; the LP-based fusion rule was used for the structural components and the Sum-Modified-Laplacian was used for the detailed components. The system had good accuracy compared with other relevant systems.

Simjanoska et al. [18] proposed a multi-level fusion system to predict BP from the ECG sensor data. In this system, the process was summarized into five separate levels. In level 1, data were combined from various ECG sensors, and in level 2, data generated features were derived using separate approaches. The combining of feedback knowledge from seven separate classifiers was included in level 3. With each BP type, knowledge from multi-target regression models was integrated into level 4, and then a single prediction was obtained in level 5 using the probability outputs of the models.

The challenge of obtaining structural and functional knowledge from both MRI and PET images by the usage of the same decomposition technique is present in multi-scale transform fusion approaches. To address the challenge, Du et al. [19] introduced a method called the intrinsic image decomposition. The method decomposed MRI and PET images into two-scales in the spatial domain. Then the original image coefficients were integrated with these two-scales elements. Experiments were performed on both gray and pseudo-color images, and the obtained accuracies were good.

Gravina et al. [20] offered a detailed and comprehensive review of new multi-sensor fusion technologies for BSNs by offering a formal categorization of the fusion in the BSN domain. This also provided an in-depth study and evaluation of the data-fusion physical activity recognition. The study described general characteristics and variables that have an effect on the option of fusion architecture at the data-level, feature-level and decision-level of image fusion, as well as certain development decisions and fusion approaches for the general implementation of health and emotional identification.

The authors in [21] proposed a voice pathology detection system by fusing multi-source signals. The signals that were used are voice signals and electroglottograph (EGG) signals. The signals were processed and modeled separately, and then fused using a score fusion method. Experimental results proved that the detection accuracy was superior to that using a single source signal. Some other types of fusion is available in [33,34]. Security concern in the healthcare system was discussed in [35].

In order to improve heart rate regulation with speckle noise and wearable sensors, Nathan and Jafari [22] proposed a way of utilizing particulate filters. In this strategy, the heart rate was formulated as the only condition to be projected apart from different signal characteristics. It led to the fuse of knowledge from various sensors and signal modes to improve the precision of the monitoring. The efficiency of this approach was tested on actual motion objects caused by ECG and PPG data with corresponding accelerometer readings and showed encouraging average error levels of less than 2 beats/min.

2.2. Deep learning-based systems using lus images

LUS images are typically low-resolution monochrome images. Roy et al. proposed a network using deep space transformer, which simultaneously forecasts the magnitude of the disease correlated with the input LUS images and provide a weakly regulated localization of pathological artifacts for COVID-19 screening [23]. Besides, the authors proposed a way of aggregating frame-level scores to predict video-level ultrasound inputs. The network gave 60% precision and 70% recall in the case of video-level inputs.

Born et al. recently developed a deep CNN named POCOVID-Net to screen COVID-19 using LUS images [24]. The authors first developed an annotated LUS images dataset by collecting images from different sources. The named the dataset as the POCUS (point of care ultrasound) dataset. The dataset is available online to use publicly. POCOVID-Net
used an extension of VGG-16 Net and achieved 86% precision and 78% recall on the dataset.

Cristina et al. developed a CNN model to quantify the assessment of B-lines in LUS images [25]. B-lines corresponding to alveolar interstitial syndrome may appear in COVID-19 or pneumonitis diseases. The model consisted of eight intermediate layers having 3D filters followed by fully-connected layers. Severity measurement was also performed using the model.

A CNN model to classify five major lung features linked to abnormal lung conditions: B-lines, converged B-lines, lack of pulmonary sliding, and pleural effusion was proposed in [26]. The model was trained with closely measured lung conditions of pig’s LUS videos. Pleuropneumothorax (absence of pulmonary sliding) was observed with the proposed CNN model is shown in Table 1. As can be seen from the table, the total number of learnable parameters is around 0.4 M (million).

| Blocks   | Layers                        | Learnable parameters |
|----------|-------------------------------|----------------------|
| BLOCK 1  | Conv; BN                      | 448                  |
| Conv; BN | 4704                          |
| Conv; BN | 9312                          |
| Skip Conv| 544                           |
| BLOCK 2  | Conv; BN                      | 13,968               |
| Conv; BN | 20,880                        |
| Skip Conv| 1584                          |
| BLOCK 3  | Conv; BN                      | 27,840               |
| Conv; BN | 37,056                        |
| Skip Conv| 3136                          |
| BLOCK 4  | Conv; BN                      | 46,320               |
| Conv; BN | 57,840                        |
| Skip Conv| 5200                          |
| BLOCK 5  | Conv; BN                      | 69,409               |
| Conv; BN | 83,232                        |
| Skip Conv| 7776                          |
| FC       | 291                           |
| Total    | 389,540 ≈ 0.4M                |

3. Proposed cnn model with multi-layer features fusion

There are many pretrained CNN models in the literature. Most of these models have high computational complexity but have very high accuracy in many applications. In this paper, an efficient CNN model is proposed for a COVID-19 screening system using LUS images.

3.1. The proposed cnn model

The architecture of the proposed CNN model is shown in Fig. 1. The input LUS image size is 512 × 512. A fundamental convolutional module called ResF is the building module of the model. The ResF consists of convolutional operations (Conv.), batch normalization (BN), and a rectifier linear unit (ReLU). The structure of the ResF is shown in the upper part of Fig. 1.

After the first ResF module, there are five blocks (denoted as BLOCK 1 to BLOCK 5 in Fig. 1), where each block has two successive ResF modules and a connector ResF module, followed by an average pooling. The connector ResF module has a convolutional layer having 1 × 1 filters. The addition operator adds point by point values. The number of filters, the filter size, and the stride are shown in each ResF in Fig. 1. For example, the first ResF (before the blocks) has 16 filters, each filter has size 5 × 5, the stride is 2. The stride of 2 for filters reduces the size of the input by half in one direction. Initial convolutional layers generally extract gross features such as edges and stripes; hence, the stride of 2 in the initial convolutional layers will not significantly affect the accuracy.

On the other hand, this stride will reduce the computation complexity of these layers, and we can use more computation in later layers (using a stride of 1), which are more complex and denser than the initial layers. Before each convolution, the input matrix to that convolution is zero-padded to maintain the matrix dimension. The BN is used to speed up the training and to overcome the overfitting. All the average pooling operations have 2 × 2 window size and stride 2.

At the end of the fifth block (BLOCK 5), there is one global average pooling (GAP), a fully-connected (FC) layer, and a softmax layer having three outputs. The purpose of the GAP is to compress each feature map to its average value. The number of learnable parameters of the proposed CNN model is shown in Table 1. As can be seen from the table, the total number of learnable parameters is around 0.4 M (million).

The input LUS images are preprocessed as follows. First, the images are resized to 512 × 512 pixels. Then, the mean value and the standard deviation of all the pixels in the dataset are calculated. After that, the pixel values are mean normalized (subtracted from the mean value). Finally, the values are divided by the standard deviation.

During training the model, the images are augmented by applying several operations. They are randomly reflected around x- and y- axes, randomly rotated by [-20° to +20°] angles, and randomly scaled along x- and y- directions by [0.8 to 1.2] scale factor. This augmentation is natural for any LUS images.

**Fig. 2.** Block diagram of the proposed multi-layer features fusion. BLOCK 1 to BLOCK 5 are taken from Fig. 1.
The initial learning rate is set to 5. The gradient decay factor parameters. The properties of the adam optimizer were set as follows: class-wise error function. The adam optimizer is used to optimize the randomize the samples and decrease overfitting. The cost function is a constructing a batch, the samples are shuffled in every epoch to result (we investigated 1 layer, two layers, and three layers are investigated, where two layers had parameters of the FC layers. It is noted that in the experiments FC one system.

After developing and training the proposed model, features from different layers of the model are fused to enhance the performance of the COVID-19 screening system. Fig. 2 shows the block diagram of the proposed multi-layer features fusion. The output size of each block is given in Table 2. In the table, for example, 64 × 64 × 32 refers to the dimension of the matrix 64 × 64 and 32 filters. As we can see from the table, the numbers of features (output size) from different blocks are completely unbalanced, which may affect the fusion. Therefore, max pooling is applied to the outputs of BLOCK 1 and BLOCK 2. Max pooling of size 4 × 4 with a stride of 4 is used at the output of BLOCK 1, and a max pooling of size 2 × 2 with a stride of 2 is applied to the output of BLOCK 2. With these max pooling, the output size of BLOCK 1 and BLOCK 2 becomes 16 × 16 × 32 and 16 × 16 × 48, respectively. After this, the outputs are flattened and concatenated sequentially to produce the fused feature vector. The vector is fed to FC layers (2 layers), followed by the softmax layer. The main purpose of feature fusion is to utilize features from different layers because various layers encode different types of information. Therefore, fusing features encoding different types of information may enhance the accuracy of the system.

After fusing the features, the model is trained only to update the parameters of the FC layers. It is noted that in the experiments FC one layer, two layers, and three layers are investigated, where two layers had the best result. The number of neurons per layer is also varied.

### Table 2

| Block# | BLOCK 1 | BLOCK 2 | BLOCK 3 | BLOCK 4 | BLOCK 5 |
|--------|---------|---------|---------|---------|---------|
| Size   | 64 × 64 | 32 × 32 | 16 × 16 | 8 × 8   | 4 × 4   |
|        | 32 × 48 | 64      | 80      | 96      |

The mini-batch size is set to 5 and the maximum epoch is 150. Before constructing a batch, the samples are shuffled in every epoch to randomize the samples and decrease overfitting. The cost function is a class-wise error function. The adam optimizer is used to optimize the parameters. The properties of the adam optimizer were set as follows: gradient decay factor = 0.9, and squared gradient decay factor = 0.999. The initial learning rate is set to 5 × 10⁻⁴, because it gave the optimum result (we investigated 1 × 10⁻⁴, 5 × 10⁻⁴, 1 × 10⁻⁵, and 5 × 10⁻⁵). Initial parameters are randomly assigned with zero mean and one standard deviation normalization.

#### 3.2. Multi-layer features fusion

In the experiments, a five-fold cross-validation approach was used. Three metrics, accuracy, precision, and recall were used to evaluate the performance of the system. Also, receiver operating characteristics (ROC) curve and area under the ROC curve (AUC) were used to compare the performance. The three metrics are reported here as average accuracy, average precision, and average recall of five folds. We compared four systems, which are the proposed model without fusion, the proposed model with fusion, ResNet50 [30], and SqueezeNet [31]. The ResNet50 gives very good accuracy in many image processing applications and the SqueezeNet is a light model with moderate accuracy is many applications.

MATLAB (R2020a) Deep Learning toolbox was used for the coding. The computer configuration was as follows: two CPUs of 2.80 GHz and 6 cores each, 64.0 GB RAM, NVIDIA GPU of GeForce GTX 1080 (8 GB).

#### 4.3. Results and discussion

First, we present the confusion matrix of the system. Fig. 3-6 show the confusion matrices of the proposed system without fusion, the proposed system with fusion, the ResNet50, and the SqueezeNet, respectively. The left-side matrix was produced by using the number of samples during testing (five-folds), the middle one is for the precision of various applications.
and the right-side matrix is for the recall. The precision and the recall matrices were obtained from the left-side matrix. From these matrices, we find that the proposed system with fusion increased the precision and the recall for all the three classes (COVID-19, pneumonia, and healthy) compared to that without fusion. For example, for COVID-19 class, the precision increased from 0.898 to 0.952, and the recall increased from 0.855 to 0.902. The proposed system with fusion had the best precision and recall values among the compared systems for all three classes.

Table 3 shows the average accuracy, precision, and recall (%) of the systems.

| Systems               | Accuracy (%) | Precision (%) | Recall (%) |
|-----------------------|--------------|---------------|------------|
| SqueezeNet            | 84.4         | 83.4          | 84.3       |
| ResNet50              | 90.0         | 89.1          | 90.8       |
| Proposed without fusion | 86.6         | 85.7          | 86.8       |
| Proposed with fusion  | 92.5         | 91.8          | 93.2       |

Fig. 4. Confusion matrix of the proposed system with fusion.

Fig. 5. Confusion matrix of the ResNet50.

Fig. 6. Confusion matrix of the SqueezeNet.
had the highest accuracy, precision, and recall values. The proposed system without fusion had better accuracy than that using the SqueezeNet, but worse accuracy than that using the ResNet50. In the proposed system, the fusion strategy increased the accuracy by 5.9%, the precision by 6.1%, and the recall by 6.4%. This increase suggests that fusing features from different layers of a CNN model significantly improves the performance of a system.

Different blocks’ fusion was also investigated. In the experiments, we fused BLOCK 3, BLOCK 4, and BLOCK 5. The accuracy while fusing these three blocks was 90.1%. This accuracy is inferior to the accuracy obtained by the proposed system with five blocks’ (all blocks) fusion. Therefore, all blocks’ fusion is better than the fusion of a subset of blocks.

The number of learnable parameters of a CNN model plays a great role in terms of computation. If the number is high, the computation will be high. Therefore, the information density value is required to see how a model performs in terms of both the accuracy and the number of parameters. The information density is calculated as a ratio between the accuracy (normalized to the range between 0 and 1) and the number of parameters in million. The more the information density is, the better the model is. Table 4 shows the information density of the four systems mentioned above. From the table, we find that the information densities of the proposed systems are very high compared to the SqueezeNet and the ResNet50.

Fig. 7 shows the ROC curves of the four systems. The ROC AUC of the systems is 0.98824, 0.99802, 0.99387, and 0.99929 for SqueezeNet, ResNet50, proposed without fusion, and proposed without fusion, respectively. Based on the AUC, we find that the proposed system with fusion had the best performance.

Fig. 8 shows the learning curve of the proposed CNN model. Fig. 8(a) shows the accuracy curve and Fig. 8(b) shows the loss curve. From the curves, we see that the proposed model was well trained.

Fig. 9 shows the clustering capability of the proposed CNN model. The upper part of the figure shows the distribution of the samples of three classes after the first pooling activations. The figure was drawn using t-SNE [32]. t-Distributed Stochastic Neighbor Embedding (t-SNE) is an algorithm that maps a high-dimensional feature space to a 2D feature space for the visualization purpose. The lower part of the figure shows the clustering of samples after the softmax activations. We can see that the samples are clearly clustered into three separate regions. This visual representation illustrates the efficiency of the proposed model to cluster the samples according to their classes.

Fig. 10 shows gradient-weighted class activation mapping (Grad-CAM) of the systems for two samples, one each for COVID-19 class and pneumonia class. The left-side figures are the actual samples and the right-side figures have the activation shown in different colors, where the yellow represents the highest activation and the blue represents the lowest activation. The figures show that the proposed system can also be used for visualization of the affected area of LUS images that can be beneficial to the doctors.

| Systems                  | Accuracy [0–1] | # of parameters (M) | Information density |
|--------------------------|----------------|---------------------|---------------------|
| SqueezeNet               | 0.844          | 1.24                | 0.681               |
| ResNet50                 | 0.900          | 25.6                | 0.035               |
| Proposed without fusion  | 0.866          | 0.40                | 2.165               |
| Proposed with fusion     | 0.925          | 0.42                | 2.202               |

Table 4 The information density of the systems.
5. Conclusion

An efficient CNN model with multi-layer features fusion for the COVID-19 screening system using LUS images was proposed. The model had five blocks of convolution connectors, and the features from each block were fused to enhance the accuracy of the system. Experiments were performed using the POCUS dataset having LUS images and videos. The results showed that the proposed system with fusion significantly improved the performance from that of the system without fusion. The proposed model also had a high information density, which means it can give a high accuracy with a few numbers of learnable parameters. The proposed system also outperformed systems with the ResNet50 and the SqueezeNet models in terms of accuracy, precision, recall, and AUC. It can be noted that during the pandemic, all the preventive and hygiene protocols should be well maintained. For example, proper disinfection measures should be taken while taking the LUS images.

In a future study, a multi-source system using the proposed CNN model with fusion will be developed when different modalities of input will be available from the same person infected by COVID-19 or similar pandemic. In another direction, we can utilize beyond 5 G or 6 G communication technology to boost the real-time experience of the smart healthcare.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are grateful to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, for funding this work through the Vice Deanship of Scientific Research Chairs: Chair of Pervasive and Mobile Computing.

References

[1] A. Pagano, et al., Lung ultrasound for diagnosis of pneumonia in emergency department, Intern Emerg Med 10 (7) (2015) 851–854.
[2] Y. Amatya, et al., Diagnostic use of lung ultrasound compared to chest radiograph for suspected pneumonia in a resource limited setting, Int J Emerg Med 11 (1) (2018) 8.
[3] M. Shorfuzzaman, M.S. Hossain, MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients, Pattern Recognition, Pattern Recognition 113 (2021) 107700, pages.
[4] J.-E. Bourcier, et al., Performance comparison of lung ultrasound and chest x-ray for the diagnosis of pneumonia in the ED., Am J Emerg Med 32 (2) (2014) 115–118.
(3) March (2020) 2022–2028.

[17] J. Chen, L. Zhang, L. Lu, Q. Li, M. Hu, X. Yang, A novel medical image fusion method based on Rolling Guidance Filtering, Internet of Things (Feb. 2020), 100172, https://doi.org/10.1016/j.iot.2020.100172.

[18] M. Simjanoska, S. Kochev, J. Tanevski, A.M. Bogdanova, G. Papa, T. Eftimov, Multi-level information fusion for learning a blood pressure predictive model using sensor data, Inf. Fusion 58 (Jun. 2020) 24–39.

[19] J. Du, W. Li, H. Tan, Intrinsic Image Decomposition-Based Grey and Pseudo-Color Medical Image Fusion, IEEE Access 7 (2019) 56443–56456.

[20] R. Gravina, P. Alinia, H. Ghazemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (May 2017) 68–80.

[21] M. Alhussein, G. Muhammad, Voice Pathology Detection Using Deep Learning on Mobile Healthcare Framework, IEEE ACCESS 6 (December 2018) 41034–41041.

[22] V. Nathan, R. Jafari, Particle Filtering and Sensor Fusion for Robust Heart Rate Monitoring Using Wearable Sensors, IEEE J. Biomed. Health Inform. 22 (6) (Nov. 2018) 1834–1846.

[23] S. Roy, et al., Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound, IEEE Trans Med Imaging 39 (8) (Aug. 2020) 2676–2687.

[24] J. Born, et al., POCOID-Net: automatic detection of COVID-19 From a New Lung Ultrasound Imaging Dataset (POCUS),” arXiv:2004.12084v3 [eess.IV], 2020.

[25] C. Balouescu, et al., Automated Lung Ultrasound B-Line Assessment Using a Deep Learning Algorithm, IEEE Trans Ultrason Ferroelectr Freq Control 67 (11) (Nov. 2020) 2312–2320.

[26] Kulhare S. et al., Ultrasound-Based Detection of Lung Abnormalities Using Single Shot Detection Convolutional Neural Networks,” In: D. Stoyanov et al. (eds) Simulation, Image Processing, and Ultrasound Systems For Assisted Diagnosis and Navigation. POCUS 2018, BIVPCS 2018, CuRIous 2018, CPM 2018. Lecture Notes in Computer Science, vol 11042. Springer, Cham.

[27] G. Muhammad, M.S. Hossain, A Deep-Learning-Based Edge-Centric COVID-19-Like Pandemic Screening and Diagnosis System within a 5G Framework Using Blockchain, IEEE Netw 35 (2) (2021) 1–8, https://doi.org/10.1109/MNET.011.2000326.

[28] D. Buonsenso, D. Pata, A. Chiaretti, COVID-19 outbreak: less stethoscope, more ultrasound, The Lancet, Respiratory Medicine 8 (5) (May 2020) E27.

[29] L. Demi, T. Egan, M. Muller, Lung Ultrasound Imaging, a Technical Review, Applied Science 10 (2) (2020) 462.

[30] K. He, Y. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770–778.

[31] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, “SqueezeNet: alexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” arXiv:1602.07360v4 [cs.CV], 2016.

[32] L. Maaten, G. Hinton, Visualizing Data using t-SNE, Journal of Machine Learning Research 9 (2008) 2579–2605.

[33] M.S. Hossain, G. Muhammad, A. Alamri, Smart healthcare monitoring: a voice pathology detection paradigm for smart cities, Multimedia Systems 25 (5) (2019) 565–675.

[34] Z. Ali, G. Muhammad, M.F. Alhamid, An Automatic Health Monitoring System for Patients Suffering from Voice Complications in Smart Cities, IEEE Access 5 (1) (2017) 3900–3908.

[35] A. Ghoneim, et al., Medical image forgery detection for smart healthcare, IEEE Communications Magazine 56 (4) (April 2018) 33–37.