A cost-sensitive deep learning-based meta-classifier for pediatric pneumonia classification using chest X-rays

Article *in* Expert Systems · March 2022

DOI: 10.1111/EXSY.12966

CITATIONS
16

READS
524

3 authors, including:

Vinayakumar Ravi
Prince Mohammad University
321 PUBLICATIONS **8,174** CITATIONS

SEE PROFILE

Tuan D Pham
Linköping University
346 PUBLICATIONS **4,667** CITATIONS

SEE PROFILE

Received: 30 June 2021 | Revised: 2 January 2022 | Accepted: 28 January 2022

DOI: 10.1111/exsy.12966

ORIGINAL ARTICLE

Expert Systems WILEY

A cost-sensitive deep learning-based meta-classifier for pediatric pneumonia classification using chest X-rays

Vinayakumar Ravi[1] | Harini Narasimhan[2] | Tuan D. Pham[1]

[1]Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

[2]Smart Materials Structures and Systems Lab, Indian Institute of Technology, Kanpur, India

Correspondence
Vinayakumar Ravi, Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia.
Email: vinayakumarr77@gmail.com; vravi@pmu.edu.sa

Funding information
None.

Abstract

Literature survey shows that convolutional neural network (CNN)-based pretrained models have been successfully employed to diagnose and detect childhood pneumonia using chest X-rays (CXR). However, most of the existing methods are prone to imbalance problems, which become even more significant in medical image classification for example most importantly childhood pneumonia classification using CXR. This is due to the fact that some classes in childhood pneumonia have a very little support in the training dataset. Additionally, though the existing methods have reported better performances for training and testing, in most of the test cases the existing models will not be effective on variants of the childhood pneumonia CXR images or CXR samples from a new pediatric patient. In addition, the models may be effective in detecting latent stage pediatric pneumonia but not show better performances for CXR samples from pediatric patients who are early stage, sick but not pneumonia, sick with other lung diseases, and so on. Generalization is an important term to be considered while designing a pneumonia classifier that can perform well on completely unseen pneumonia CXR datasets. This article presents a cost-sensitive large-scale learning with stacked ensemble meta-classifier and transfer learning-based deep feature fusion approach for pediatric pneumonia classification using CXR. With the aim to identify the importance among the classes of pneumonia, the larger cost items are introduced based on the class-imbalance degree during the back-propogation learning methodology in transfer learning models such as Xception, InceptionResNetV2, DenseNet201, and NASNetMobile. Next, the features from the penultimate layer (global average pooling) of Xception, InceptionResNetV2, DenseNet201, and NASNetMobile were extracted and dimensionality of the extracted features were reduced using kernel principal component analysis (KPCA). The reduced features were fused together and passed into a stacked ensemble meta-classifier for classifying the CXR into either pneumonia or normal. A stacked ensemble meta-classifier is a two stage approach in which the first stage employs random forest and support vector machine (SVM) for prediction and followed by logistic regression for classification. Experiments of the proposed model were done on publicly available benchmark pediatric pneumonia classification CXR dataset. In addition, the experiments for existing methods as well as various cost-insensitive models were conducted. In all the experiments, the proposed method has achieved better performances compared to the existing methods as well as various cost-insensitive models.

In particular, the proposed method showed 6% improvement in precision, 10% improvement in recall, 9% improvement in F1 score with less misclassification costs (0.0321) and accuracy (96.8%). Most importantly, the proposed method is insensitive to the imbalance data and more effective to handle variants of the childhood pneumonia CXR images. Thus, the proposed approach can be used as a tool for point-of-care diagnosis by healthcare professionals.

KEYWORDS
chest X-ray images, cost-sensitive learning, data imbalance, feature fusion, meta-classifier, pneumonia, stacked classifier, transfer learning

1 | INTRODUCTION

Pneumonia is an acute respiratory infection that affects the alveoli in lungs which are responsible for breathing. Along with air in alveoli, a person affected by pneumonia has fluid and pus. This affects the breathing process by limiting the intake of oxygen. According to the World Health Organization (WHO), Pneumonia is the major cause of death in children around the world. The latest report states that about 15% of deaths in children under 5 years is caused by pneumonia infection (https://www.who.int/news-room/fact-sheets/detail/pneumonia, n.d.). The 2019 report from CDC states the crude rate of 13.4 per 100,000 population is due to pneumonia in the United States (Centers for Disease Control and Prevention, National Center for Health Statistics, n.d.). The environmental risk factors of Pneumonia in children are majorly influenced by air pollution and crowded homes which is most common in developing countries. Children with pre-existing infections and malnutrition have high risks of developing pneumonia infection which can lead to death. Pneumonia can be a bacterial, viral, or fungal infection with different mortality rates. Most of the weakly immune persons are affected by fungal pneumonia. Bacterial pneumonia is comparatively more severe in children than viral infection (Siddiqi, 2019a). Preventing pneumonia in children being an essential component for world health; adequate nutrition and early diagnosis are some of the best preventive measures presented by WHO (https://www.who.int/news-room/fact-sheets/detail/pneumonia, n.d.). Pneumonia can be diagnosed using multiple modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and CXRs. CXRs are preferred to other imaging techniques due to its low cost and high availability in almost all countries across the world. CXRs are applied for diagnosing heart infections, lung disorders, injury or fracture and also prevents the use of high dosage of radiation (Viergever et al., 2001).

The outbreak of COVID-19 increased the mortality rate of this infection, COVID-19 is a variant of viral pneumonia. Presence of pneumonia is indicated by the lung inflammation and lesions. Radiologists and doctors look for airspace or interstitial opacities in the CXRs to detect pneumonia (Clouet et al., 2001). Early diagnosis of pneumonia is challenging as well for the medical experts on analysing the lung textures, density, and patches in the CXRs. Some of the patches in pneumonia can be similar to other infections such as tuberculosis and bronchitis (Saraiva, Ferreira, et al., 2019). Misdiagnosis is always a possibility at the early stages of the infection. To avoid such difficulties, Computer aided Diagnosis (CAD)-based tools are widely used by radiologists to confirm their diagnosed results (Litjens et al., 2017).

Radiologists and Doctors around the world prefer using CAD-based tools for diagnosis as they are less time consuming and accurate. These tools are utilized in diagnosing injuries, fractures, and infections. Recent advancements in developing CAD-based tools for diseases diagnosis is applied to identify tuberculosis (Melendez et al., 2016), to detect cancer and also to diagnose diabetic eye disease (Esteva et al., 2021). Automation is the major contribution by CAD-based tools on analysing the CXRs to detect, classify, and also examine various diseases. This automation is induced by employing machine learning algorithms which uses the extracted features from the CXRs. To detect pneumonia, features like texture, spatial relativity, and other local features are extracted from CXRs and then automated using machine learning models. Significant limitations of these feature extraction steps are the need for domain expert knowledge and consumes more time.

Deep learning-based algorithms are evolving in the field of CAD-based early disease diagnosis in order to mitigate the issue of feature extraction. Most of the recent state-of-the-art performances in image classification are introduced with the use of pretrained convolutional neural networks (CNNs) (Litjens et al., 2017). CNN-based pretrained deep learning models use convolutional filters for automatic feature extraction. This model also incorporates dimensionality reduction using pooling layers and induces feature importance using activation functions. Training such deep learning models demands large volumes of data and also a balanced data. A data imbalance is met in a classification problem when there is a significant difference with the number of data samples between the classes in the dataset. Generally, these problems affect the performance of the deep learning algorithm by inducing bias. In most of the cases, data resampling and augmentation techniques are implemented to overcome the class imbalance problem (Thabtah et al., 2020). In this article, we propose a cost-sensitive (CS) approach to handle the data imbalance problem in diagnosing pneumonia from child CXRs. We have also used various CNN-based pretrained models such as Xception, ResNet50, ResNet152, ResNet101V2, ResNet152V2, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet169, DenseNet201, NASNetMobile, and EfficientNet for feature extraction from CXRs and a feature fusion method is employed with dimensionality reduction for penultimate layer features of

cost-sensitive fine-tuned CNN-based models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. Main contributions and novelty proposed in this article are as follows.

- Pretrained CNN-based feature extraction from pediatric pneumonia CXRs.
- A cost-sensitive approach to handle disadvantages caused by an imbalance in the pediatric pneumonia CXR Images dataset.
- Meta-classifiers for reduced false predictions and to achieve generalization.
- Feature fusion for merging CNN-based pretrained models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile.
- KPCA was employed for dimensionality reduction of pediatric pneumonia CXR images database features of CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile models.
- t-Distributed Stochastic Neighbour Embedding (t-SNE) approach was employed for penultimate layer feature visualization of the proposed model.
- State-of-the-art performance among the existing techniques on the same dataset.

The remaining sections of this article are organized as follows. Section 2 includes the detailed discussion on the existing approaches for pneumonia classification using CXR images, Section 3 includes proposed architecture for Pneumonia classification using CXR images, Section 4 includes description of CXR database, Section 5 includes details of performance metrics, experiments, results and its discussions placed in Section 6, finally, conclusion and future works directions are discussed in Section 7.

2 | LITERATURE SURVEY

Improvements in image processing algorithms with the help of deep learning techniques surpasses the human eye in various biomedical applications involving pathologic, radiographic, and mammographic images (Litjens et al., 2017). Various machine learning methods are used along with the Random Forest method to analyse the CXRs on detecting pneumonia. As machine learning methods perform well with extracted features rather than the raw images, wavelet-based features from images are used to feed the models.

Currently, CNN-based deep learning method is brought into play to detect and classify pneumonia infection in CXR images. CNN architecture is customized to learn the presence of pneumonia infection in CXRs as a binary classification task (Siddiqi, 2019b). This method incorporates convolutional, pooling, dense, and dropout layers and has achieved state-of-the-art performance by offering around 94% accuracy on testing. Another approach proposed a custom CNN architecture along with preprocessing and augmentation techniques in order to improve generalization during the training process (Stephen et al., 2019). Saraiva et.al introduced a two stage network as CNN-based feature extractors and a deep learning Multilayer perceptron for classifying the CXRs as normal and pneumonia (Saraiva, Santos, et al., 2019). Furthermore, a 10 layer CNN architecture is used to train CXRs on identifying the appearance of pneumonia infection (Saraiva, Ferreira, et al., 2019). This method uses K-fold cross-validation technique to evaluate the performance of the model and also to increase randomness in the training process. To persuade the CNN algorithm to be more powerful, Digital Image Processing (DIP) is implemented as the feature extractors and then combined to feed a multichannel CNN network for diagnosing the pneumonia infection (Nahid et al., 2020). Explaining the performance of CNN-based on various visualization solutions are proposed to understand the working of CNN with respect to CXRs (Nguyen et al., 2020). This method is explained only for shallow CNN architecture using smaller CXR images of size 64×64. These methods deal with the limitations such as less abstract representation of input due to the smaller number of layers in the network. Also, training deeper networks demands large samples of data which can be addressed with the use of transfer learning approach.

Comparison between the performances of different existing CNN architectures such as AlexNet, DenseNet, SqueezeNet, and ResNet18 for pneumonia detection is implemented (Rahman et al., 2020). This method provides a faster diagnosis solution for pneumonia patients. Fine-tuning the ImageNet pretrained models has improved the performance on detecting pneumonia in CT scans and CXRs (Asnaoui et al., 2020). Almost eight different architectures are trained and validated to diagnose pneumonia variants and satisfactory performances are reported with various performance measures. Another ImageNet pretrained architecture Xception is transfer learned to identify pneumonia CXRs and the results obtained are also visualized (Luján-García et al., 2020). In this method, the class imbalance problem is also addressed with random undersampling and cost-sensitive learning. Furthermore, a suitable performance metric is also chosen to evaluate the results irrespective of the imbalance in the dataset.

Transfer learning approach is to utilize the weights learned by any model on different dataset. Mahajan et.al has experimented on different transfer learning approaches such as weights learned from natural images (ImageNet dataset) and weights learned from medical images (CheXNet dataset) (Mahajan et al., 2019). The performances on transfer learning from CheXNet pretrained model is found better than the train from scratch or other pretrained models. A deep transfer learning method is well suitable for extracting features from the images and a separate finetuning is implemented to prompt it as a classifier (Chhikara et al., 2020). This tuning after the feature extraction performs better than the stand-alone

transfer learned networks in terms of accuracy and other classification metrics. Islam et al has experimented on selecting suitable deep transfer learning models as feature extractors and suitable classifiers such as support vector machines (SVM), artificial neural networks (ANN), and so on (Islam et al., 2020). A well suitable method is selected based on performance and time complexity.

Other fusion methods to improve the performance of the transfer learned deep models on domain specific were also introduced. Approaches such as combining image processing techniques with deep learning models are performed with deeper CNN architectures like VGG-16 and VGG-19 (Hasan et al., 2019). Digital image processing methods such as Histogram equalization, Vertical cropping, and so on, enhances the domain specific features in the CXRs which then combined with deep CNN models enhances the classification performance. CXRs can be noisy due to various reasons and not all the residual and bottleneck connections help with noisy inputs. Customizing the convolutions in the existing architectures with the use of depthwise separable convolutions has performed highest in terms of test accuracy (Siddiqi, 2020). In addition to the improved performance, a depthwise separable convolution has lesser trainable parameters than the usual convolutions. The performance of a deep CNN model can also be enhanced by fusing the machine learning methods. Habib et al experimented on the pretrained CheXNet model for pneumonia detection combined with principal component analysis (PCA)-based dimensionality reduction and followed by machine learning-based logistic regression for binary classification (Habib et al., n.d.). This method is compared with other existing pretrained models and has surpassed the classification performance in terms of various metrics.

Along with Transfer learning practice as a feature learning method, hand crafted features are ensembled to boost the classification performance of VGG-19 model (Dey et al., 2021). This method has achieved better performance compared to the transfer learning from existing VGG-19 model, but the major future scope is listed as better feature fusion method as the extracted features are ensembled with serial fusion. Similar method with VGG16 extracted features and SVM model for classification is proposed to detect COVID-19 from lung CT scans (Singh et al., 2021). In this method, a pre-processing method for CT data is proposed that enhances the performance of the feature extraction. Some of the methods to improve performance includes a semi-supervised learning based deep learning model which is based on an automatic segmentation in lung CTs with fully connected layers (Konar et al., 2021). The model is light-weight and also shallow which enables it to perform with lower computational cost. However, the performance of the model with respect to dice similarity is less compared to some other published methods and the stability of the model for different patch sizes of CTs is under study. Another method that highly affects the performance of the deep learning model is the optimization algorithm to converge with the loss function. Ant colony optimization is proposed which improves the speed of the convergence and also the capacity of the search highly which in turn effectively diagnosing COVID-19 in Xrays (Liu et al., 2021). Similarly, a swarm intelligent optimization algorithm is proposed to perform better segmentation in COVID-19 CT images (Zhang, Wang, et al., 2021).

Another approach to improve the performance of the transfer learning method is to ensemble the trained models. A weighted average based ensemble technique is implemented among five different transfer learned models to achieve about 0.08 test loss (Kora Venu, 2020). Here, the classification performance for pneumonia identification is also measured using other metrics such as precision, recall, F1 score and the ensemble technique outperforms on classifying pneumonia X-rays. Various ensemble methods are introduced in which the method of fusing the trained models and it is predictions is different from each other. One such method is the ensemble feature extraction technique along with Random Forest classifier for detecting pneumonia in child CXRs (Habib et al., 2020). This method fuses the pretrained CheXNet and VGG-19 models incorporating the advantages of weights learned from both medical and natural images. Another ensemble approach for pneumonia detection is proposed which uses Multi-view ensemble CNN architecture (Ferreira et al., 2020). The complete architecture consists of the following steps as cropping chest cavity, preprocessing, ensembling Multi-view CNN and classifying with Multi-layer perceptrons. Mittal et al. proposed the use of Ensemble of Convolutions and Capsules (ECC) to perform better on classifying the CXRs into pneumonia class (Mittal et al., 2020). Variants of ECC models are experimented to obtain optimum in the ECC model and its performance is measured and compared between the experiments. All of those ensemble models were proposed to mitigate the limitation in classification accuracy due to training with single existing architectures. Single architecture for a specific classification task can be customized as proposed by Liang and Zheng (2020). This architecture is a deep CNN with residual connections designed and tested for child pneumonia classification. A transfer learning approach is also used by pretraining the network based on Chest X-ray 14 dataset and then transferring the weights for child pneumonia detection. A comparative study with other most common deep CNN architectures is also performed in terms of confusion matrix and receiver operating characteristic (ROC) curve. Though most of the existing methods perform with accuracy of about 0.95, the False positives are seen due to the presence of imbalance in the dataset. This problem is addressed with traditional methods such as sampling and augmentation in a few cases. Handling imbalance in the dataset is essential for generalization on testing unobserved data samples and reliability in real-time application. A cost-sensitive method improves classification performance by solving the data imbalance problem such as CS-ResNet by introducing weights as an adjustment layer based on the degree of balance present in the dataset along with a weighted loss function (Zhang, Jiang, et al., 2021). Similarly, an attention to important features in the data improves the performance by achieving generalization such as meta-classifiers (Ravi et al., 2021). The detailed analysis of the related works based on the algorithm and its effect with the classification performance are listed in Table 1, shows that there are few studies that reports the performance on pediatric pneumonia. Some classes in pediatric pneumonia have a very little support in the training dataset. The existing models not be effective in handling the samples of minority class of pediatric pneumonia. In addition, generalization is another important concept to be considered while designing a classifier. Thus, this work has employed the cost-sensitive approach to handle imbalanced pediatric pneumonia CXR images and ensemble meta-classifier to achieve generalization.

TABLE 1 Summary of related works on pnemonia detection and classification

References	Disease	Model	Feature fusion	Meta-classifier	Data imbalance	Learning method	Accuracy
Siddiqi (2019b)	Pneumonia	Sequential CNN	No	No	Yes	Cost-insensitive	94.39%
Stephen et al. (2019)	Pneumonia	CNN	No	No	No	Cost-insensitive	93.73%
Saraiva, Santos, et al. (2019)	Pneumonia	Two stage CNN	No	No	Yes	Cost-insensitive	94.40%
Saraiva, Ferreira, et al. (2019)	Pediatric Pneumonia	CNN	No	No	Yes	Cost-insensitive	95.30%
Nahid et al. (2020)	Pneumonia	Multi-channel CNN	Yes	No	Yes	Cost-insensitive	97.92%
Nguyen et al. (2020)	Pneumonia	Shallow CNN	No	No	Yes	Cost-insensitive	84.80%
Rahman et al. (2020)	Bacterial and Viral Pneumonia	Transfer Learning with CNN	No	No	No	Cost-insensitive	98.00%
Asnaoui et al. (2020)	Bacterial and Viral Pneumonia	Pre-trained CNN	No	No	No	Cost-insensitive	96.61%
Luján-García et al. (2020)	Pneumonia	Transfer Learning with CNN	No	No	Yes	Cost-sensitive	97.30%
Mahajan et al. (2019)	Pneumonia	Domain specific Transfer Learning with CNN	No	No	Yes	Cost-insensitive	88.78%
Chhikara et al. (2020)	Pneumonia	Transfer Learning with CNN and additional layers	No	No	Yes	Cost-insensitive	90.10%
Islam et al. (2020)	Pneumonia	Deep Transfer Learning Framework	Yes	No	Yes	Cost-insensitive	98.99%
Hasan et al. (2019)	Pneumonia	Image Processing with CNN	Yes	No	Yes	Cost-insensitive	96.20%
Siddiqi (2020)	Pediatric Pneumonia	Depthwise Seperable Convolutions	No	No	Yes	Cost-insensitive	94.87%
Habib et al. (n.d.)	Pneumonia	Fusion of CNN and Logistic regression	No	No	Yes	Cost-insensitive	92.60%
Dey et al. (2021)	Pneumonia	Customized VGG19 model	Yes	No	Yes	Cost-insensitive	97.94%
Konar et al. (2021)	COVID-19	Semi-supervised shallow	No	No	No	Cost-insensitive	98.40%
Liu et al. (2021)	COVID-19	Swarm intelligent Segmentation model	No	No	No	Cost-insensitive	NA
Zhang, Wang, et al. (2021)	COVID-19	GBSSA with Stochastic Fractal Search segmentation model	No	No	No	Cost-insensitive	NA
Kora Venu (2020)	Pneumonia	Ensembled based Deep Transfer Learning	No	No	Yes	Cost-insensitive	98.46%
Habib et al. (2020)	Pediatric Pneumonia	CheXNet and VGG-19	Yes	No	No	Cost-insensitive	98.93%
Ferreira et al. (2020)	Bacterial and Viral Pneumonia	Multi-View Ensemble Convolutional Neural Network	No	No	No	Cost-insensitive	97.30%
Mittal et al. (2020)	Pneumonia	Convolutions and Dynamic Capsule Routing	No	No	No	Cost-insensitive	96.36%
Liang and Zheng (2020)	Pediatric Pneumonia	Transfer Learning with ResNet	No	No	Yes	Cost-insensitive	90.50%
Singh et al. (2021)	COVID-19	Ensembled Transfer Learning and Machine model	Yes	No	No	Cost-insensitive	95.67%
Zhang, Jiang, and Li (2021)	PCB Cosmetic Defect	Cost-sensitive ResNets	No	No	Yes	Cost-sensitive	89.00%
Ravi et al. (2021)	Cassava leaf disease	Attention based DL classifiers	Yes	Yes	No	Cost-insensitive	87.08%
Current work (Ravi et al., 2021)	Pediatric Pneumonia	Ensemble DL classifiers	Yes	Yes	Yes	Cost-sensitive	97%

3 | PROPOSED ARCHITECTURE FOR PEDIATRIC PNEUMONIA CLASSIFICATION USING CXR IMAGES

The proposed architecture for pneumonia classification using CXR images is shown in Figure 1. The architecture employs four different cost-sensitive pretrained CNN models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. All four pretrained models were trained on an ImageNet database, which has more than 1 million images which were grouped into 1000 classes. This model has learnt rich features which represent images from different classes. To achieve better performances on the ImageNet database, researchers used different CNN architectures that have the same property with different scaling schemes. Scaling means arbitrarily increasing the CNN model depth or width or usage of larger input image resolution for training and evaluation. In this work, these models were reused as a transfer learning approach with the aim to transfer similar performances for the pneumonia classification using CXR images. This type of learning approach for pneumonia classification using CXR images can reduce the training time, faster convergence rate, and achieve optimal performances in detecting patients' data samples of CXR as either pneumonia or normal. The proposed architecture is composed of three steps that are discussed as follows.

3.1 | Preprocessing

Due to the scanning device and the operator's work habits, the collected CXR images in Guangzhou Women and Children's Medical Center have different dimensions. However, all the images were resized into 299×299, 299×299, 224×224, 224×224 for CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile, respectively. Finally, normalization was employed for the CXR database to convert the data in the range [0–1]. The final result obtained by the proposed model using fixed CXR samples as input is good and performed better than the existing methods. The CXR dimension of pediatric pneumonia patient collected in real-time will be higher and these were transformed into fixed size according to the input dimension of pretrained CNN model. There may be possibility that information loss can happen with this approach and since we did not have access to clinical expert we have not done any study. This is very important in healthcare and medical domain and this will be another direction of future work of the proposed work.

3.2 | Feature extraction

Literature survey shows that CNN-based pretrained models were successfully employed in medical image classification (Litjens et al., 2017). The detailed investigation and analysis of various CNN-based pretrained models are necessary to identify an optimal model. The performances of the pretrained models vary across different tasks. Applying various pretrained models and identifying the best model have been considered as an

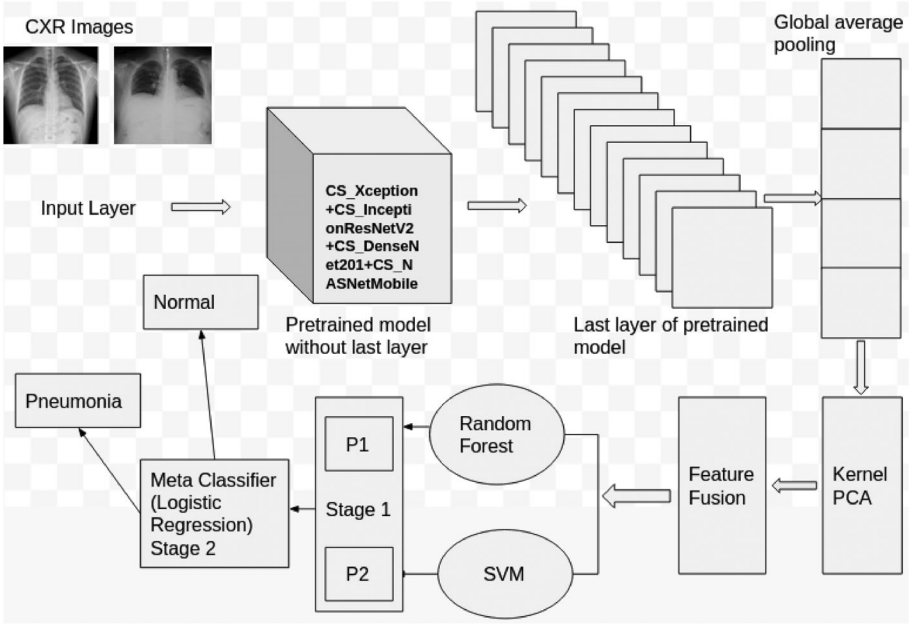

FIGURE 1 Proposed architecture for pneumonia classification using CXR images

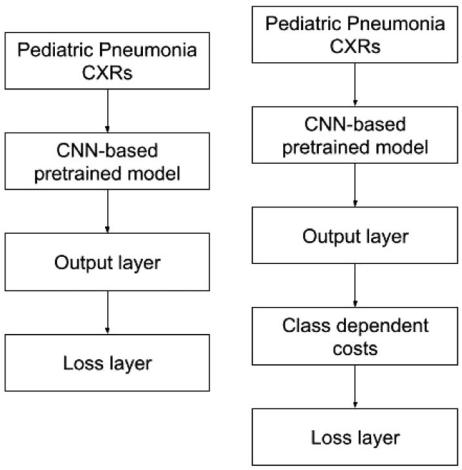

FIGURE 2 Cost-insensitive CNN-based pretrained model (left) and cost-sensitive CNN-based pretrained model

active area of research (Litjens et al., 2017). Thus, this work employs various CNN-based pretrained models such as Xception, VGG16, VGG19, ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet121, DenseNet169, DenseNet201, NASNetMobile, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7. All the pretrained models were composed of an input layer, hidden layers, and classification layer. The hidden layers contain more than one convolutional layer, pooling layer, and fully connected layer. An input layer takes an input in the form of tensor and the following convolution layer employs convolution operation to extract features. The extracted features of the convolution layer are called a feature map. To reduce the dimension of the feature map, pooling is used. Pooling can be minimum, maximum, or average. Pooling helps to control overfitting by reducing the number of training parameters. As a result, pooling with CNN will be computationally effective compared to CNN without pooling. There are many methods available in the literature that can be used instead of pooling. Some methods are principal component analysis, Autoencoder, Attention, and so on. A detailed investigation and analysis of various methods is required to identify the best method. We will consider this as future work and this has been included as one of the future work in the revised article. The major issue with pooling operation is that the method throws away a lot of information about where a feature occurs. To avoid this, capsule neural network proposed and this avoids pooling. A detailed study can be done on Capsule neural network (Sabour et al., 2017) to identify the effectiveness of the model for pediatric pneumonia classification. This will be another future directions of the proposed work. Fully connected layer serves as an output layer that contains neurons for pneumonia and normal. In a fully connected layer, each neuron has connections to every other neuron of the previous layer. All the aforementioned CNN-based pretrained models are cost-insensitive and cannot be efficiently used to handle the imbalance CXR database. Since the pneumonia CXR database of this work is highly imbalanced, the cost weights were introduced for classes such as pneumonia and normal during backpropogation. Class distribution plays an important role in classification related problems. The learning model will be biased when the dataset is imbalanced. This is mainly due to the model assigns equal importance for all the classes. However, this can be avoided by assigning larger weights to the classes that has less number of data samples. This type of approach is called as cost-sensitive learning. An overview of cost-sensitive and cost-insensitive learning model of CNN-based pretrained model is shown in Figure 2. The CNN-based pretrained model considered in this study are cost-insensitive models such as Xception, InceptionResNetV2, DenseNet201, and NASNetMobile and cost-sensitive models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. The cost-sensitive CNN-based pretrained models were CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile and its parameters 20,809,001, 54,277,729, 18,094,849, and 4,234,035 respectively. This work has followed the methodology proposed by (Vinayakumar et al., 2020) and the authors followed algorithmic approach to directly include the misclassification costs in the CNN-based pretrained model during backward pass. Each pediatric pneumonia sample P connected with a cost item $C[class(P), pred]$, where $class(p)$ and $pred$ are the actual and predicted class respectively. Here, less cost is assigned to the classes that contain more number of samples and higher cost to the classes that contains lesser number of samples. The values of cost-matrix is not known initially and there are many methods in the literature to include the initial values to the cost matrix. The loss function for cost-insensitive model is given below:

$$F = -\sum_{s \in samples} \sum_{n} t^n \log pred^n \tag{1}$$

The loss function for cost-sensitive model is given below:

$$F = -\sum_{s \in samples} \sum_{n} t^n \log pred^n C[class(s), n] \quad (2)$$

where F is a function of predicted output $pred^n$ of the nth output neuron and the target t^n, $C[class(s),n]$ is the cost with $class(s)$ is the exact value of sample s, and n is the predicted class of sample s.

3.3 | Classification

The penultimate layer (global average pooling) features of CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile models were extracted and dimension of each extracted feature was 1280. The pediatric CXR data samples were highly non-linearly separable, as shown in Figure 3. Linear classifier is not optimal on this type of non-linearly separable data points medical dataset. In this work, KPCA was employed that includes kernel-based trick to map the data into a higher dimensional space in which the decision boundary becomes linear. It is an extension of PCA with a kernel-based trick. In this work radial-basis function (RBF) was used as kernel in KPCA. There are many type of kernels available and this study chooses the default kernel of scikit-learn KPCA implementation. There may be possibility that the performance may increase with other kernel and this type of a detailed investigation and analysis of various kernel considered to be as one of the future work. More details of KPCA available in (Wang et al., 2017). Using KPCA the dimension of features of CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile are reduced into 320. Finally, the feature dimension of 320 of models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile were fused and passed into meta-classifiers. The meta-classifier was used in this work because the CXR data samples were highly non-linearly separable, as shown in Figure 3. it was generated using t-SNE. t-SNE is an important approach for feature visualization which uses PCA implicitly to reduce the pneumonia feature dimension from 1280 to two dimensions. These two dimensions are two principal components and these were plotted in the X axis and Y axis. The meta-classifier is a stacking approach which contains random forest and SVM in the first stage and logistic regression in the second stage. This type of stacking approach has the capability to show better performances than any single model and most importantly can achieve model generalization. For the random forest, we set n_components = 100, for SVM, the kernel was set to rbf, and c = 10. During training and validation of the model, various trails of experiments were run to identify the optimal parameter values for n_components and c. We run experiments with n_components as 10, 25, 50, 75, 100, 125, and c as 1, 4, 7, 10, 15. The experiments with n_components 100 and c 10 performed better than other values for validation dataset. Thus the model parameter values n_components and c were set into 100 and 10 respectively. The logistic regression uses the sigmoid activation function, which outputs 0 or 1, where 0 indicates normal and 1 indicates pneumonia.

4 | CHEST X-RAY DATABASE

To evaluate the performance of the proposed method as well as other existing methods for pneumonia classification using CXR images, this work has taken the pneumonia CXR database collected in Guangzhou Women and Children's Medical Center (Liang & Zheng, 2020). Though there are

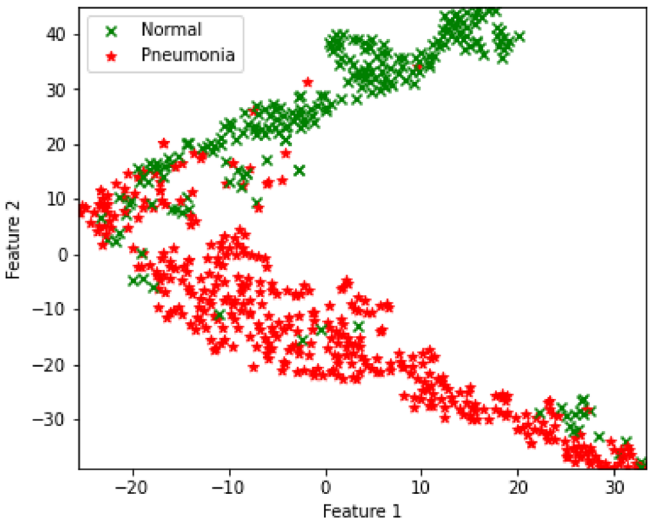

FIGURE 3 T-SNE penultimate layer feature representation of the proposed model for pediatric pneumonia classification using CXR images

TABLE 2 Details of pneumonia CXR dataset

Type	Normal	Pneumonia	Total
Training	1349	3883	5232
Testing	234	390	624
Total	1583	4273	5856

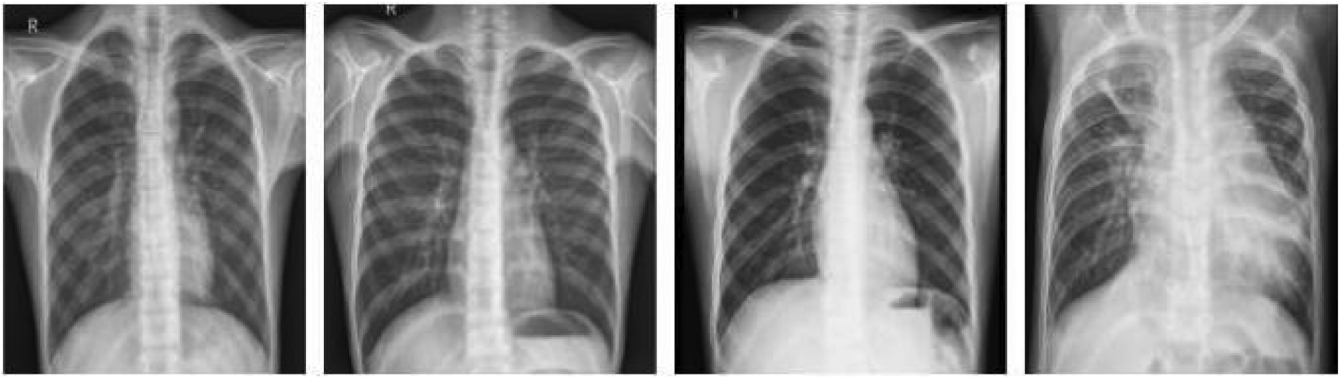

FIGURE 4 (left to right) first two images are normal patient CXR images and next two are pneumonia patient CXR images

many CXR datasets available for pneumonia classification, literature survey shows that there is only one benchmark dataset, that is, Guangzhou Women and Children's Medical Center publicly available for research for pediatric pneumonia classification. It contains CXR images of retrospective cohorts of pediatric patients of one to 5 years old. During the dataset preparation, low-quality and unreadable CXR images were removed and diagnosis was performed by two doctors and after that the validation was done by another expert to avoid misclassification. The database has train and test and both contain CXR image samples for normal and pneumonia patients. Both the train and test datasets are completely unseen and have different distributions. The detailed statistics of the database is reported in Table 2. The CXRs of normal and pneumonia samples were shown in Figure 4 (left to right) the first two images for normal patients and next two are for pneumonia patients. The fourth image can be easily distinguished since the image is more blur compared to the first two CXR images of normal patients. However the third image of pneumonia CXR sample cannot be easily distinguished from normal patients without the required domain knowledge in medical science about pneumonia. Thus, this work employs a deep learning-based approach which implicitly extracts features from CXR images and classifies them into normal and pneumonia. The CXR images in both training and testing have different dimensions and these are changed to the same dimension according to the deep learning pretrained model input data requirements.

As Table 2 shows that the pneumonia CXR dataset is highly imbalanced, most of the available cost-insensitive models cannot be directly applied on this type of dataset. Some of the most commonly used methods to handle imbalanced dataset are upsampling, undersampling, and cost-sensitive learning. Since the CXR images can be very sensitive to slight changes for various most commonly used data augmentation operations such as rotation, flipping, zooming, shearing, and so on, this work has not employed any of the undersampling or oversampling approaches. Instead, this work has employed a cost-sensitive learning approach to handle the pneumonia CXR data imbalance.

5 | PERFORMANCE METRICS

This work employs the following most commonly used metrics to evaluate the performance of the proposed model as well as the existing models for pneumonia classification using CXR dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

where TP, FP, TN, and FN are true positive, false positive, true negative and false negative respectively. In this work, TP, FP, TN, and FN are defined as follows

- TP is the number of pneumonia CXR samples predicted accurately as pneumonia.
- FN is the number of pneumonia CXR samples incorrectly predicted as normal.
- FP is the number of normal CXR samples incorrectly predicted as pneumonia.
- TN is the number normal CXR samples predicted accurately as normal.

The values for TP, FP, TN, and FN were estimated from a confusion matrix. It is a table that displays and compares actual values with the predicted values. With the aim to identify the performance of the proposed model at each class level of pneumonia detection, precision, recall, and F1 score were used along with accuracy.

Macro is considered as a good metric for precision, recall, and F1 score in the case of imbalanced datasets. Macro metric computes the precision, recall, and F1 score for each class and returns the average without considering the proportion for each class in the pneumonia CXR dataset. The weighted metric computes the precision, recall, and F1 score for each class and returns the average by considering the proportion for each class in the pneumonia CXR dataset.

6 | EXPERIMENTS, RESULTS, AND DISCUSSIONS

TensorFlow[1] with Keras[2] was used to implement all the CNN-based pretrained models. All of the machine learning algorithms were implemented using scikit-learn.[3] The experiments of all the models were run using Google Colab[4] with K80 GPU and 25 GB RAM.

The proposed method for pneumonia classification using CXR images was based on transfer learning. Various models were employed in transfer learning and the models were Xception, VGG16, VGG19, ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet121, DenseNet169, DenseNet201, NASNetMobile, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7. The input image size for Xception, VGG16, VGG19, ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNet, MobileNetV2, DenseNet121, DenseNet169, DenseNet201, NASNetMobile, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7 are 299×299, 224×224, 224×224, 224×224, 224×224, 224×224, 224×224, 224×224, 224×224, 299×299, 299×299, 200×200, 224×224, 224×224, 224×224, 224×224, 224×224, 224×224, 240×240, 260×260, 300×300, 380×380, 456×456, 528×528, and 600×600, respectively. During training, optimizer, learning rate, batch size, and epochs parameter values were set as adam, 0.0001, 64, and 8, respectively. To identify the optimal values for parameters such as optimizer, learning rate, batch size, and epochs, various trials of experiments were run and best parameters were identified. Initially, the experiments were run for adam and sgd optimizer with epochs 10 and learning rate 0.0001. The proposed method showed better performances for adam. When the same experiments were run for more epochs, the training and validation performance after epochs 8 remains same. With the aim to avoid overfitting, the epochs parameter value was set to 8. Further, experiments were run with learning rate 0.0001, 0.001, 0.01, and 0.1. The model showed better training and validation performances with lower learning rate 0.0001. To identify the optimal parameter for batch size, the same experiments were further run with 32, 64, and 128 batch size. The model with 64 performed better during training and validation compared to batch size 32 and 128. Thus the experiments were not run further and batch size was set to 64. The training data samples were shuffled for each epoch during training and the models were initialized with ImageNet pretrained model weights. The trainable parameter details for DenseNet121, DenseNet169, DenseNet201, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, EfficientNetB7, InceptionResNetV2, InceptionV3, MobileNet, MobileNetV2, NASNetMobile, ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, ResNet152V2, VGG16, VGG19, and Xception are 6,954,881, 12,486,145, 18,094,849, 4,008,829, 6,514,465, 7,702,403, 10,697,769, 17,550,409, 28,342,833, 40,738,009, 63,789,521, 54,277,729, 21,770,401, 3,208,001, 2,225,153, 4,234,035, 23,536,641, 23,521,409, 42,554,881, 42,530,945, 58,221,569, 58,189,953, 14,715,201, 20,024,897, and 20,809,001, respectively.

During training, the 20% of the train dataset was used for validation. As a result, the train dataset contains 1078 and 3107 CXR data samples for normal and pneumonia respectively. The validation had 271 and 776 CXR data samples for normal and pneumonia respectively. Both the train and validation datasets are highly imbalanced. In each epoch, the train and validation data were shuffled. The validation data helped to monitor the train accuracy across epochs and most importantly played an important role in finding an optimal value for hyperparameters such as epochs, learning rate, and batch size. The train accuracy and validation accuracy and train loss and validation loss is shown in Figures 5 and 6 respectively.

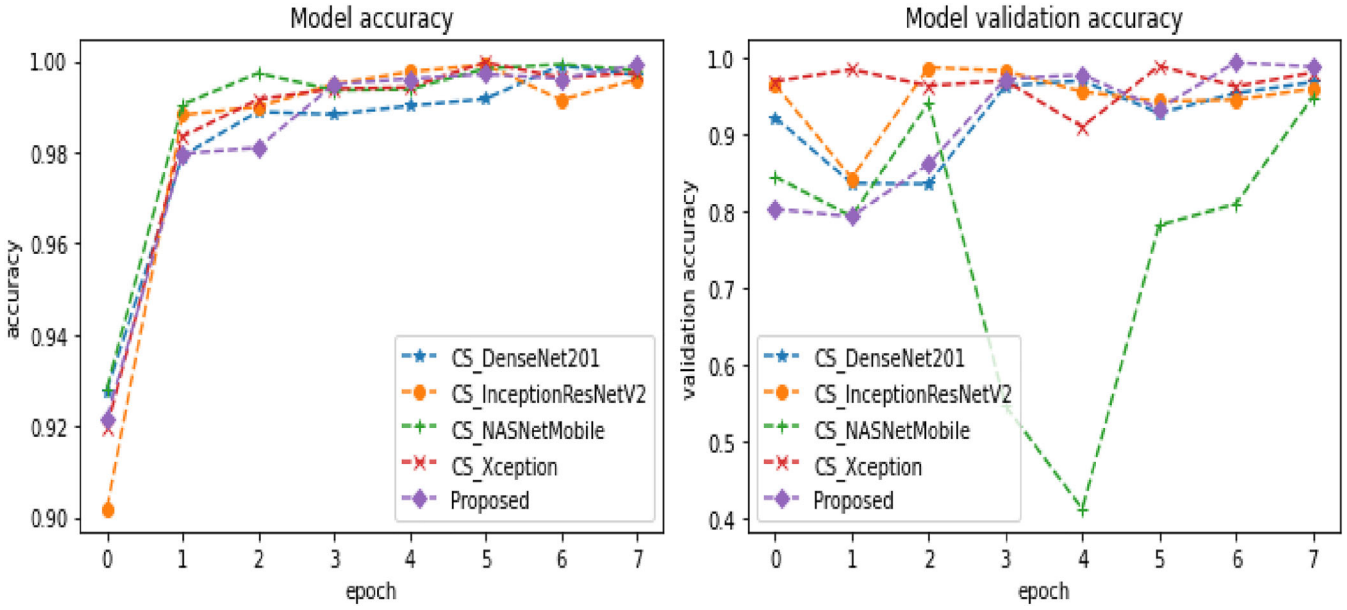

FIGURE 5 CNN-based models accuracy, training and validation (left to right)

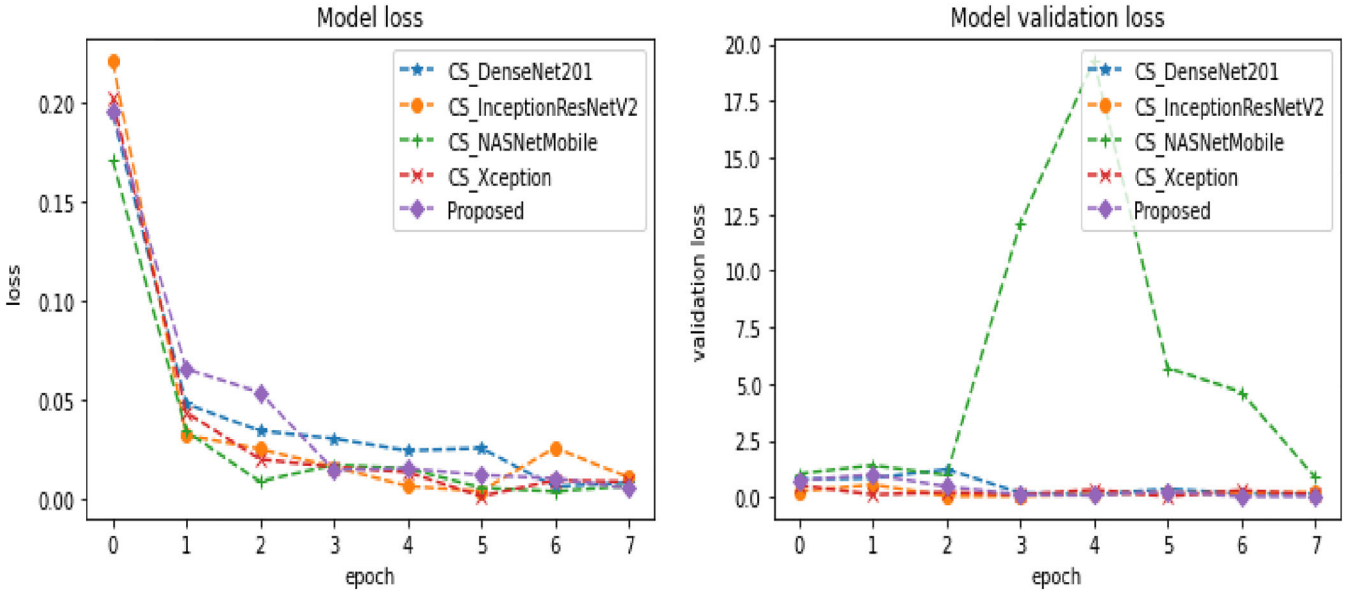

FIGURE 6 CNN-based models loss, training and validation (left to right)

The figures show a steady increase and steady decrease in train and validation accuracy and train loss and validation loss, respectively on CXR dataset. All five different CNN-based pretrained models converged within eight epochs. As Figure 5 shows, all the five models have achieved more than 95% train accuracy and 90% validation accuracy. Most importantly, the proposed method has shown slightly better performances in both the train and validation accuracy.

Next, the trained models performances were evaluated on the CXR pneumonia test dataset. This dataset has been completely unseen during the training. The model's performances were shown in Table 3. It contains results for various CNN-based pretrained models. The results for both macro and weighted precision, macro and weighted recall, macro and weighted F1 score. Since the CXR pneumonia dataset is a highly imbalanced macro with precision, recall, and F1 score considered to be effective metrics in comparison to weighted precision, recall, and F1 score. The models such as Xception, InceptionResNetV2, DenseNet201, and NASNetMobile have achieved better performances in terms of macro and weighted precision, macro and weighted recall, and macro and weighted F1 score in comparison to other models such as ResNet50, ResNet152, ResNet101V2, ResNet152V2, MobileNet, MobileNetV2, DenseNet169, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3,

TABLE 3 Results for pediatric pneumonia classification using CNN-based pretrained models on CXR images

Model	Type	Precision	Recall	F1 score
Xception	**Macro**	**0.91**	**0.84**	**0.86**
	Weighted	**0.89**	**0.88**	**0.87**
VGG16 (Liang & Zheng, 2020b)	Macro	0.31	0.50	0.38
	Weighted	0.39	0.62	0.48
VGG19 (Islam et al., 2020)	Macro	0.83	0.66	0.66
	Weighted	0.80	0.75	0.71
ResNet50	Macro	0.31	0.50	0.38
	Weighted	0.39	0.62	0.48
ResNet101 (Yu et al., 2021)	Macro	0.31	0.50	0.38
	Weighted	0.39	0.62	0.48
ResNet152	Macro	0.31	0.50	0.38
	Weighted	0.39	0.62	0.48
ResNet50V2 (Luján-García et al., 2020)	Macro	0.81	0.50	0.39
	Weighted	0.77	0.63	0.48
ResNet101V2	Macro	0.66	0.51	0.42
	Weighted	0.65	0.63	0.51
ResNet152V2	Macro	0.61	0.56	0.44
	Weighted	0.67	0.47	0.41
InceptionV3 (Chouhan et al., 2020)	Macro	0.80	0.82	0.79
	Weighted	0.84	0.79	0.79
InceptionResNetV2	**Macro**	**0.91**	**0.84**	**0.86**
	Weighted	**0.89**	**0.88**	**0.87**
MobileNet	Macro	0.84	0.61	0.58
	Weighted	0.80	0.71	0.64
MobileNetV2	Macro	0.86	0.70	0.70
	Weighted	0.83	0.77	0.74
DenseNet121 (Liang & Zheng, 2020b)	Macro	0.83	0.56	0.51
	Weighted	0.78	0.67	0.58
DenseNet169	Macro	0.81	0.50	0.39
	Weighted	0.77	0.63	0.48
DenseNet201	**Macro**	**0.91**	**0.83**	**0.85**
	Weighted	**0.89**	**0.87**	**0.86**
NASNetMobile	**Macro**	**0.92**	**0.86**	**0.88**
	Weighted	**0.90**	**0.89**	**0.89**
EfficientNetB0	Macro	0.86	0.68	0.68
	Weighted	0.83	0.76	0.72
EfficientNetB1	Macro	0.85	0.66	0.66
	Weighted	0.81	0.75	0.70
EfficientNetB2	Macro	0.76	0.76	0.71
	Weighted	0.80	0.71	0.71
EfficientNetB3	Macro	0.83	0.57	0.52
	Weighted	0.79	0.68	0.59
EfficientNetB4	Macro	0.88	0.75	0.77
	Weighted	0.85	0.81	0.80
EfficientNetB5	Macro	0.85	0.64	0.63
	Weighted	0.81	0.73	0.68
EfficientNetB6	Macro	0.85	0.68	0.68
	Weighted	0.82	0.76	0.72
EfficientNetB7	Macro	0.85	0.65	0.65
	Weighted	0.82	0.74	0.69

Note: Bold indicates the results of Proposed Approach for Pediatric Pneumonia Classification using CXR samples.

TABLE 4 Results for pediatric pneumonia classification using cost-sensitive CNN-based pretrained models on CXR images

Model	Type	Precision	Recall	F1 score
CS_Xception	Macro	0.93	0.87	0.89
	Weighted	0.92	0.90	0.90
CS_InceptionResNetV2	Macro	0.93	0.87	0.89
	Weighted	0.92	0.90	0.90
CS_DenseNet201	Macro	0.93	0.85	0.87
	Weighted	0.91	0.89	0.89
CS_NASNetMobile	Macro	0.94	0.89	0.91
	Weighted	0.93	0.92	0.92
Proposed	**Macro**	**0.98**	**0.96**	**0.97**
	Weighted	**0.97**	**0.97**	**0.97**

Note: Bold indicates the results of Proposed Approach for Pediatric Pneumonia Classification using CXR samples.

EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7. Most importantly, the models such as Xception, InceptionResNetV2, DenseNet201, and NASNetMobile have performed better than existing CNN-based pretrained models such as VGG16 (Liang & Zheng, 2020), VGG19 (Islam et al., 2020), DenseNet121 (Liang & Zheng, 2020), ResNet50V2 (Luján-García et al., 2020), ResNet101 (Yu et al., 2021), and InceptionV3 (Chouhan et al., 2020). All these models were cost-insensitive and these models will consider all data samples equally. Thus, these models were not effective for imbalanced dataset. This is the main reason behind not considering these models were not effective though the models achieved above 84% performances in both macro and weighted precision, macro and weighted recall and macro and weighted F1 score.

The proposed method i.e. cost-sensitive large-scale learning stacked meta-classifier with feature fusion of CNN-based pretrained models such CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile results was reported in Table 4. Also, the Table 4 contains results for other cost-sensitive models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. As the Table 4 shows that the cost-sensitive models were shown better performances in both the macro and weighted metrics in comparison to the other cost-insensitive models. Moreover, the models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile have used SVM for classification instead of softmax. This is due to the reason that SVM has the capability to produce good decision surfaces when applied to well-behaved feature vectors (Huang & LeCun, 2006). Mostly, CNN has the capability to extract optimal invariant feature vectors by passing the raw data into many hidden layers. As can be seen, the proposed method has achieved better performance in terms of precision, recall, and F1 score of weighted and macro compared to the single cost-sensitive classifiers such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. The main reason is that each CNN-based pretrained model has the ability to learn unique features. Thus, the features of well-known cost cost-sensitive CNN-based pretrained models were fused and with the help of meta-classifier the performances were increased for classifying the CXR data sample into either pneumonia or normal. In addition, the meta-classifier helps to achieve model generalization. Thus, this type of classifier has the capability to detect pneumonia samples from unseen CXR or variants of existing CXR data samples.

6.1 | Comparisons of the proposed model with others for pediatric pneumonia classification

The proposed approach for pneumonia classification using CXR images was compared with various existing methods such as VGG16 (Liang & Zheng, 2020), VGG19 (Islam et al., 2020), DenseNet121 (Liang & Zheng, 2020), ResNet50V2 (Luján-García et al., 2020), ResNet101 (Yu et al., 2021), and InceptionV3 (Chouhan et al., 2020) and the proposed CNN-based pretrained models such as ResNet50, ResNet152, ResNet101V2, ResNet152V2, MobileNet, MobileNetV2, DenseNet169, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7 and cost-sensitive models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. The detailed investigations and analysis of the proposed method was shown on the CXR pediatric pneumonia CXR dataset. In all the experiments, the proposed method performed better than all the models. This indicates that the proposed method is more robust and generalizable and most importantly the model has the capability to detect the unseen CXR images as well as the variants of the existing CXR images. The results were reported in Table 4 and the detailed result of the proposed method were included in Table 5.

The confusion matrix for the existing methods such as VGG16 (Liang & Zheng, 2020), VGG19 (Islam et al., 2020), DenseNet121 (Liang & Zheng, 2020), ResNet50V2 (Luján-García et al., 2020), ResNet101 (Yu et al., 2021), and InceptionV3 (Chouhan et al., 2020) are shown in Figure 7a, b, f, d, c, and e, respectively. Both VGG16 and ResNet101 have classified all the samples of normal as pneumonia with accuracy 63% and misclassification error rate 38%. ResNet50V2 has classified only one CXR sample as normal out of 244 CXR samples. Thus, VGG16, ResNet101, and ResNet50V2 were not effective for pneumonia classification using CXR images. The models DenseNet121, InceptionV3, and VGG19 showed

TABLE 5 Detailed results of the proposed model for pediatric pneumonia classification using CXR images

Class	Precision	Recall	F1 score
Normal	1.00	0.91	0.96
Pneumonia	0.95	1.00	0.97
Accuracy	0.97		

misclassification error rate 33%, 21%, and 25%, respectively. These three models were considered to be good in comparison to the other models such as VGG16, ResNet101, and ResNet50V2. Most importantly, DenseNet121 and VGG19 were more effective to classify the pneumonia CXR samples and these models have misclassified 0 and 4 CXR data samples, respectively. Both DenseNet121 and VGG19 were able to classify 29 and 79 normal CXR samples out of 234. But, InceptionV3 has misclassified only 13 normal CXR samples as pneumonia but it was able to correctly classify 272 CXR samples as Pneumonia out of 390. Thus, there may be possibility that the misclassification error rate in normal and pneumonia classes can be reduced by combining the features of DenseNet121, InceptionV3, and VGG19. This type of study can be considered as one of the significant directions towards future work. However, since the data is highly imbalanced, this type of experiment was not done for pneumonia classification using CXR images. The confusion matrix for the proposed method is shown in Figure 7g and this approach has shown 0 misclassification error rate for pneumonia and 20 CXR samples of normal patients were misclassified as pneumonia. Thus further study can be done to identify the reasons behind the misclassification and this type of study can be considered as another direction towards future work. Overall, the proposed method has achieved less misclassification error rate for normal class and 0 misclassification error rate for pneumonia. Thus, the proposed method is considered to be more robust and generalizable compared to the existing methods for pneumonia classification using CXR images.

Overall, the proposed approach based on cost-sensitive learning has showed better performances compared to cost-insensitive models for pediatric pneumonia classification. In addition, the proposed approach has performed better than six different existing methods that are based on CNN pretrained models. The existing methods are VGG16 (Liang & Zheng, 2020), VGG19 (Islam et al., 2020), ResNet101 (Yu et al., 2021), ResNet50V2 (Luján-García et al., 2020), InceptionV3 (Chouhan et al., 2020), and DenseNet121 (Liang & Zheng, 2020). All these six models are cost-insensitive. In addition, the existing methods does not employ any feature fusion or ensemble meta-classifier. The feature fusion method employed in this method relies on features extracted from four different cost-sensitive pretrained models and finally all these four feature sets are merged column-wise. The four cost-sensitive pretrained models are CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile. Also, the proposed method employs ensemble meta-classifier instead of a single fully connected layer and this enhances the results and in addition helps to achieve generalization. In the experimental analysis, various experiments were shown for the proposed method and compared with cost-insensitive models and six different CNN-based pretrained existing models for pediatric pneumonia classification.

6.2 | Visualization of penultimate layer features of the proposed model for pediatric pneumonia classification

Since deep learning models were considered to be balck-box, interpretation plays an important role in deep learning. t-SNE feature representation is one of the most commonly used approaches to visualize the penultimate layer features of deep learning models. In this work, the t-SNE feature representation for the proposed model on CXR test data samples were shown in Figure 3. t-SNE takes features of dimension 1280 and implicitly uses PCA to reduce the dimensionality of features into two dimensions. These two dimensions are two principal components which are shown in Figure 3 *X*-axis and *Y*-axis, respectively. This type of visualization approach allowed us to verify the pneumonia and normal CXR data samples in separate clusters. Though normal and pneumonia classes have formed a distinct cluster, still both showed an overlapping region between them. This indicates that further study is required to analyse the misclassified samples and a further development needed to minimize misclassified CXR samples.

6.3 | Computational analysis of models for pediatric pneumonia classification

The experiments were performed on Google cloud platform supporting GPU with the system configuration of 25 GB RAM with K80 GPU and leveraged the Google Colab notebook. We were able to access this environment for free and now the size of the RAM has been reduced from 25 GB to 12 GB. In addition, we do not have access to GPU using the same Ipython notebook kernel. Since we do not have any funding support, we did not go for the paid version of Google Colab. Computation cost highly depends on the hardware and we have included only the parameter details of the models. The parameter details of deep learning models training time is directly proportional to the number of parameters in the model during training and fine-tuning. We rerun the models on a single randomly chosen CXR sample of pediatric pneumonia to determine the

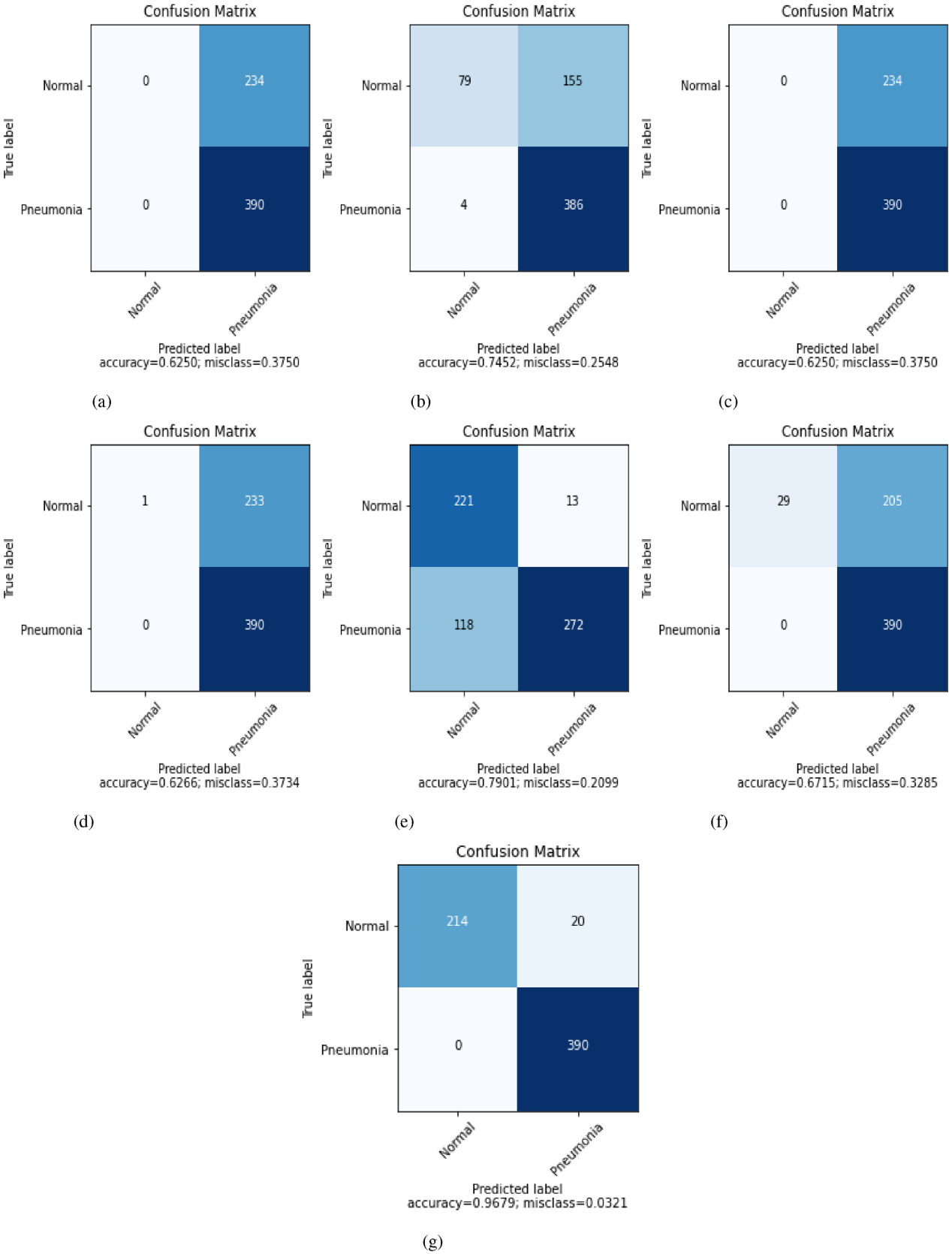

FIGURE 7 Confusion matrix obtained from the proposed approach for pediatric pneumonia classification using CXR images. (a) VGG16 (Liang & Zheng, 2020). (b) VGG19 (Islam et al., 2020). (c) ResNet101 (Yu et al., 2021). (d) ResNet50V2 (Luján-García et al., 2020). (e) InceptionV3 (Chouhan et al., 2020). (f) DenseNet121 (Liang & Zheng, 2020). (g) Proposed

testing time. The computational time on classification is: CS_Xception: 11.54 s, CS_InceptionResNetV2: 12.55 s, CS_DenseNet201: 13.64 s, and CS_NASNetMobile: 12.39 s. The proposed approach computational time on a single randomly chosen CXR sample of pediatric pneumonia classification is 5.12 s. However, the proposed approach computational time will be higher when feature extraction, dimensionality reduction, and feature fusion is included to the model. The reported time of proposed approach is for only classification.

7 | CONCLUSION AND FUTURE WORK

A cost-sensitive large-scale learning with stacked ensemble meta-classifier and transfer learning-based deep feature fusion approach for pediatric pneumonia classification using CXR has been presented. Four cost-sensitive transfer learning models such as CS_Xception, CS_InceptionResNetV2, CS_DenseNet201, and CS_NASNetMobile were employed and penultimate layer features of these four models were extracted, dimensionality of each model features were reduced using KPCA and finally, fused together. Next, the fused features were passed into a meta-classifier for classification. The meta-classifier has employed random forest and SVM in the first stage for prediction and logistic regression in the second stage for classifying the predictions into either pneumonia or normal. Additionally, various existing approaches were implemented. Both the proposed method and the existing methods performances were evaluated on the benchmark pneumonia CXR dataset. In all the test cases, the proposed method achieved better performances than the existing methods. The present results suggest that the cost-sensitive deep learning approach is more effective in handling data imbalance in pneumonia CXR database and meta-classifier approach for classification is generalizable and robust for classifying unseen and variants of existing CXR images of patients and controls.

The current study did not consider the sub-classification of pneumonia CXR samples of patients into early stage, latent stage, sick but not pneumonia due to unavailability of data labelling for CXR data samples. Identifying the CXR into either the early or latent stage helps doctors provide patients with proper treatment. In addition, the study can be extended to detect the pneumonia area in the pneumonia CXR data sample. However, simultaneous pneumonia CXR image classification and pneumonia area detection system can be a challenging task. Due to unavailability of bounding box information for pneumonia area in the CXR pneumonia images, this study did not investigate the role of object detection models for pneumonia area detection. Attention-based approaches can be leveraged to identify the infected regions of the CXR sample even without bounding box information. However, still this type of research study requires help from radiologist for validation. Integrating, an attention-based approach to the proposed study will be considered as future works.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ENDNOTES

[1] https://www.tensorflow.org

[2] https://keras.io

[3] https://scikit-learn.org/

[4] https://colab.research.google.com/

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Vinayakumar Ravi https://orcid.org/0000-0001-6873-6469

REFERENCES

Asnaoui, K. E., Chawki, Y., & Idri, A. (2020). Automated methods for detection and classification pneumonia based on X-ray images using deep learning. arXiv preprint arXiv:2003.14363.

Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999–2019 on CDC WONDER Online Database. http://wonder.cdc.gov/ucd-icd10.html.

Chhikara, P., Singh, P., Gupta, P., & Bhatia, T. (2020). Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays. In *Advances in bioinformatics, multimedia, and electronics circuits and signals* (pp. 155–168). Springer.

Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., & De Albuquerque, V. H. C. (2020). A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Applied Sciences*, *10*(2), 559.

Clouet, P. L., Berthezene, Y., Marchand, B., Nesme, P., & Mornex, J. F. (2001). Pneumonias. In *Radiologic diagnosis of chest disease* (pp. 171–179). Springer.

Dey, N., Zhang, Y. D., Rajinikanth, V., Pugalenthi, R., & Raja, N. S. M. (2021). Customized VGG19 architecture for pneumonia detection in chest X-rays. *Pattern Recognition Letters*, *143*, 67–74.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, *4*(1), 1–9. https://doi.org/10.1038/s41746-020-00376-2

Ferreira, J. R., Cardenas, D. A. C., Moreno, R. A., de Sá Rebelo, M. D. F., Krieger, J. E., & Gutierrez, M. A. (2020). Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest X-ray images. In *2020 42nd annual international conference of the IEEE engineering in Medicine & Biology Society (EMBC)* (pp. 1238–1241). IEEE.

Habib, N., Hasan, M. M., Rahman, M. M., Parvizi, O., Khani, R. T., & Aboozari, M. Fusion of deep convolutional neural network with PCA and logistic regression for diagnosis of pediatric pneumonia on chest X-rays.

Habib, N., Hasan, M. M., Reza, M. M., & Rahman, M. M. (2020). Ensemble of CheXNet and VGG-19 feature extractor with random Forest classifier for pediatric pneumonia detection. *SN Computer Science*, *1*(6), 1–9.

Hasan, M. M., Kabir, M. M. J., Haque, M. R., & Ahmed, M. (2019). A combined approach using image processing and deep learning to detect pneumonia from chest X-ray image. In *2019 3rd international conference on electrical, Computer & Telecommunication Engineering (ICECTE)* (pp. 89–92). IEEE.

https://www.who.int/news-room/fact-sheets/detail/pneumonia

Huang, F. J., & LeCun, Y. (2006). Large-scale learning with svm and convolutional netw for generic object recognition. In 2006 IEEE computer society conference on computer vision and pattern recognition (Vol. 10).

Islam, K. T., Wijewickrema, S. N., Collins, A., & O'Leary, S. J. (2020). A deep transfer learning framework for pneumonia detection from chest X-ray images. In VISIGRAPP (5: VISAPP) (pp. 286-293).

Konar, D., Panigrahi, B. K., Bhattacharyya, S., Dey, N., & Jiang, R. (2021). Auto-diagnosis of covid-19 using lung ct images with semi-supervised shallow learning network. *IEEE Access*, *9*, 28716–28728.

Kora Venu, S. (2020). An ensemble-based approach by fine-tuning the deep transfer learning models to classify pneumonia from chest X-ray images. arXiv e-prints, arXiv-2011.

Liang, G., & Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer Methods and Programs in Biomedicine*, *187*, 104964.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88.

Liu, L., Zhao, D., Yu, F., Heidari, A. A., Li, C., Ouyang, J., Chen, H., Mafarja, M., Turabieh, H., & Pan, J. (2021). Ant colony optimization with Cauchy and greedy levy mutations for multilevel COVID 19 X-ray image segmentation. *Computers in Biology and Medicine*, *136*, 104609.

Luján-García, J. E., Yáñez-Márquez, C., Villuendas-Rey, Y., & Camacho-Nieto, O. (2020). A transfer learning method for pneumonia classification and visualization. *Applied Sciences*, *10*(8), 2908.

Mahajan, S., Shah, U., Tambe, R., Agrawal, M., & Garware, B. (2019). Towards evaluating performance of domain specific transfer learning for pneumonia detection from X-ray images. In *2019 IEEE 5th international conference for convergence in technology (I2CT)* (pp. 1–6). IEEE.

Melendez, J., Sánchez, C. I., Philipsen, R. H., Maduskar, P., Dawson, R., Theron, G., Dheda, K., & Van Ginneken, B. (2016). An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Scientific Reports*, *6*, 25265.

Mittal, A., Kumar, D., Mittal, M., Saba, T., Abunadi, I., Rehman, A., & Roy, S. (2020). Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray images. *Sensors*, *20*(4), 1068.

Nahid, A. A., Sikder, N., Bairagi, A. K., Razzaque, M., Masud, M., Kouzani, A., & Mahmud, M. A. (2020). A novel method to identify pneumonia through analyzing chest radiographs employing a multichannel convolutional neural network. *Sensors*, *20*(12), 3482.

Nguyen, H., Huynh, H., Tran, T., & Huynh, H. (2020). Explanation of the convolutional neural network classifying chest X-ray images supporting pneumonia diagnosis. *EAI Endorsed Transactions on Context-aware Systems and Applications*, *7*(21), e3.

Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., & Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, *10*(9), 3233.

Ravi, V., Acharya, V., & Pham, T. D. (2021). Attention deep learning-based large-scale learning classifier for cassava leaf disease classification. *Expert Systems*, *39*, e12862.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. arXiv preprint arXiv:1710.09829.

Saraiva, A. A., Ferreira, N. M. F., de Sousa, L. L., Costa, N. J. C., Sousa, J. V. M., Santos, D. B. S., Valente, A., & Soares, S. (2019). Classification of images of childhood pneumonia using convolutional neural networks. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOIMAGING*, *112-119*, *2019*, *Prague, Czech Republic* (pp. 112–119). https://doi.org/10.5220/0007404301120119

Saraiva, A. A., Santos, D. B. S., Costa, N. J. C., Sousa, J. V. M., Ferreira, N. M. F., Valente, A., & Soares, S. (2019). Models of learning to classify X-ray images for the detection of pneumonia using neural networks. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOIMAGING*, *76-83*, *2019*, *Prague, Czech Republic* (pp. 76–83). https://doi.org/10.5220/0007346600760083

Siddiqi R. (2019a). Automated pneumonia diagnosis using a customized sequential convolutional neural network. In Proceedings of the 2019 3rd International Conference on Deep Learning Technologies.

Siddiqi, R. (2019b). Automated pneumonia diagnosis using a customized sequential convolutional neural network. In Proceedings of the 2019 3rd International Conference on Deep Learning Technologies. (pp. 64–70).

Siddiqi, R. (2020). Efficient pediatric pneumonia diagnosis using Depthwise separable convolutions. *SN Computer Science*, *1*(6), 1–15.

Singh, M., Bansal, S., Ahuja, S., Dubey, R. K., Panigrahi, B. K., & Dey, N. (2021). Transfer learning–based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data. *Medical & Biological Engineering & Computing*, *59*(4), 825–839.

Stephen, O., Sain, M., Maduh, U. J., & Jeong, D. U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. *Journal of Healthcare Engineering*, *2019*, 4180949.

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441.

Viergever, M. A., Romeny, B. H., & Van Goudoever, J. B. (2001). Computer-aided diagnosis in chest radiography: A survey. *IEEE Transactions on Medical Imaging*, *20*(12), 1228–1241.

Vinayakumar, R., Alazab, M., Srinivasan, S., Pham, Q. V., Padannayil, S. K., & Simran, K. (2020). A visualized botnet detection system based deep learning for the internet of things networks of smart cities. *IEEE Transactions on Industry Applications*, *56*(4), 4436–4456.

Wang, W., Zhang, M., Wang, D., & Jiang, Y. (2017). Kernel PCA feature extraction and the SVM classification algorithm for multiple-status, through-wall, human being detection. *EURASIP Journal on Wireless Communications and Networking*, *2017*(1), 1–7.

Yu, X., Wang, S. H., & Zhang, Y. D. (2021). CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia. *Information Processing & Management*, *58*(1), 102411.

Zhang, H., Jiang, L., & Li, C. (2021). Cs-resnet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection. *Expert Systems with Applications*, *185*, 115673.

Zhang, Q., Wang, Z., Heidari, A. A., Gui, W., Shao, Q., Chen, H., & Chen, M. (2021). Gaussian Barebone Salp swarm algorithm with stochastic fractal search for medical image segmentation: A COVID-19 case study. *Computers in Biology and Medicine*, *139*, 104941.

AUTHOR BIOGRAPHIES

Vinayakumar Ravi is an Assistant Research Professor at Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia. My previous position was a Postdoctoral research fellow in developing and implementing novel computational and machine learning algorithms and applications for big data integration and data mining with Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA from September, 2019 to September, 2020. He received the Ph.D. degree in computer science from Computational Engineering & Networking, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India. His Ph.D. work centers on Application of Machine learning (sometimes Deep learning) for Cyber Security and discusses the importance of Natural language processing, Image processing and Big data analytics for Cyber Security. His current research interests include applications of data mining, Artificial Intelligence, machine learning (including deep learning) for biomedical informatics, Cyber Security, image processing, and natural language processing. More details available at https://vinayakumarr.github.io/. He has more than 50 research publications in reputed IEEE conferences, IEEE Transactions and Journals. His publications include prestigious conferences in the area of Cyber Security, like IEEE S&P and IEEE Infocom. He has given many invited talks on deep learning applications in IEEE conferences and Industry workshops in 2018. He has got a full scholarship to attend Machine Learning Summer School (MLSS) 2019, London. Dr. Ravi has served as a Technical Program Committee (TPC) member at international conferences including SSCC Symposium, IEEE TrustCom-2020, and IEEE SmartData-2020. He is an editorial board member for Journal of the Institute of Electronics and Computer (JIEC), International Journal of Digital Crime and Forensics (IJDCF), and he has organized a shared task on detecting malicious domain names (DMD 2018) as part of SSCC'18 and ICACCI'18. He received the Chancellor's Research Excellence Award in AIRA 2021 and was included in the World's Top 2% Scientists by Stanford University published in PLoS Biology.

Harini Narasimhan is a Project Engineer at the Smart Materials Structures and Systems Lab, Indian Institute of Technology, Kanpur, India. Contact her at harininarasimhan123@gmail.com.

Tuan D. Pham currently holds positions as (full) Senior Research Professor in AI and Founding Director of the Center for Artificial Intelligence at Prince Mohammad Bin Fahd University, Saudi Arabia. His previous position was (full) Professor of Biomedical Engineering at Linkoping University, University Hospital Campus, Linkoping, Sweden. He was appointed as (full) Professor and Leader of the Aizu Research Cluster for Medical Engineering and Informatics, and the Medical Image Processing Lab, both at the University of Aizu, Japan. Before his appointment in Japan, he was appointed as Associate Professor and the Bioinformatics Research Group Leader at the University of New South Wales, Canberra, Australia. His current research focuses on AI and machine learning methods for image processing, time-series analysis, complex networks, and pattern recognition applied to medicine, biology, and mental health. He serves as an Associate/Section Editor for a number of scholarly journals, series, and conference proceedings, such as Pattern Recognition (Elsevier), Heliyon (Cell Press), IET Signal Processing, Entropy (MDPI), Frontiers in Artificial Intelligence, Frontiers in Big Data, Computer Science Advisory Board (Cambridge Scholars), Current Bioinformatics (Bentham), IEEE-EMBC (Theme 10: Biomedical & Health Informatics), ACM, and SPIE conference proceedings. Dr. Pham has been selected as an Expert in Artificial Intelligence for consultation by the U.S. Food & Drug Administration (FDA) Center for Devices and Radiological Health (CDRH) Network of Digital Health Experts Program (NoDEx), and included in the World's Top 2% Scientists by Stanford University published in PLoS Biology.

How to cite this article: Ravi, V., Narasimhan, H., & Pham, T. D. (2022). A cost-sensitive deep learning-based meta-classifier for pediatric pneumonia classification using chest X-rays. *Expert Systems*, e12966. https://doi.org/10.1111/exsy.12966