Neural Computing and Applications (2023) 35:8259–8279
https://doi.org/10.1007/s00521-022-08099-z

ORIGINAL ARTICLE

Stacked ensemble learning based on deep convolutional neural networks for pediatric pneumonia diagnosis using chest X-ray images

J. Arun Prakash[1] · Vinayakumar Ravi[2] · V. Sowmya[1] · K. P. Soman[1]

Received: 23 April 2022 / Accepted: 22 November 2022 / Published online: 7 December 2022

Abstract

Pneumonia is an acute respiratory infection caused by bacteria, viruses, or fungi and has become very common in children ranging from 1 to 5 years of age. Common symptoms of pneumonia include difficulty breathing due to inflamed or pus and fluid-filled alveoli. The United Nations Children's Fund reports nearly 800,000 deaths in children due to pneumonia. Delayed diagnosis and overpriced tests are the prime reason for the high mortality rate, especially in underdeveloped countries. A time and cost-efficient diagnosis tool: Chest X-rays, was thus accepted as the standard diagnostic test for pediatric pneumonia. However, the lower radiation levels for diagnosis in children make the task much more onerous and time-consuming. The mentioned challenges initiate the need for a computer-aided detection model that is instantaneous and accurate. Our work proposes a stacked ensemble learning of deep learning-based features for pediatric pneumonia classification. The extracted features from the global average pooling layer of the fine-tuned Xception model pretrained on ImageNet weights are sent to the Kernel Principal Component Analysis for dimensionality reduction. The dimensionally reduced features are further trained and validated on the stacking classifier. The stacking classifier consists of two stages; the first stage uses the Random-Forest classifier, K-Nearest Neighbors, Logistic Regression, XGB classifier, Support Vector Classifier (SVC), Nu-SVC, and MLP classifier. The second stage operates on Logistic Regression using the first stage predictions for the final classification with Stratified K-fold cross-validation to prevent overfitting. The model was tested on the publicly available pediatric pneumonia dataset, achieving an accuracy of 98.3%, precision of 99.29%, recall of 98.36%, F1-score of 98.83%, and an AUC score of 98.24%. The performance shows its reliability for real-time deployment in assisting radiologists and physicians.

Keywords Pneumonia · Chest X-rays · Computer-aided diagnosis · Deep learning · Transfer learning · Stacking classifier · Principal component analysis · Stratified K-fold

1 Introduction

Over the years, the number of respiratory diseases and infections has increased drastically. Degradation in the air quality has paved the way to numerous lung-related contaminations [1]. Pneumonia is one such acute lower respiratory infection that fills the alveoli with pus and fluid, leading to reduced oxygen holding capacity in the lungs. Lack of oxygen directly impacts the standard functioning of the body. Fatigue and lethargy are a few of many symptoms caused by inadequate oxygen levels. In severe cases, it can deter the brain and heart. Symptoms of pneumonia include fever, shallow breathing, and coughing. In extreme cases, it causes sharp chest pains when breathing and coughing. Sepsis, one of the many

✉ Vinayakumar Ravi
vravi@pmu.edu.sa

J. Arun Prakash
arun.jayakanthan@gmail.com

V. Sowmya
v_sowmya@cb.amrita.edu

K. P. Soman
kp_soman@amrita.edu

[1] Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

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

complications of pneumonia, can lead to tissue damage, organ failure, and even death if left untreated. Studies show that people with a weaker immune system are highly susceptible to pneumonia [2]. This acute respiratory infection poses a much bigger problem in children predominantly between 1 and 5 years of age whose immunity system is in its embryonic stages of development [3]. Symptoms of severe pneumonia in children include vomiting, severe malnutrition, and the inability to consume food and water [4].

Pediatric pneumonia accounts for nearly 800,000 deaths of young children, as reported by the United Nations Children's Fund (UNICEF). Based on factors like age group and other medical conditions, there are several diagnostic tests for pneumonia. The most widely used diagnostic tests in children include pulse oximetry to check the oxygen levels, complete blood count (CBC) to check the activity of the immune system when there is an infection, sputum test, and chest X-rays to look for inflammation in the lungs. Abnormal CBC can be due to a variety of medical conditions. The count may decrease or increase even with mild infections. Thus, CBC is not guaranteed to confirm the presence of pneumonia. Children below the age of 10 have reduced sputum production. This reduced sample quantity restricts conducting various tests and eliminates its possibility as a confirmatory diagnostic test. Though pulse oximetry might seem like the best alternative, it cannot assure the presence of pneumonia as there may be other lung contaminations causing the low oxygen levels in the body. In addition to the limitations of these tests, they are time and cost-inefficient. These two factors are critical in saving lives. Cost in specific is a prime challenge in underdeveloped countries where people scarcely avail such diagnosis measures due to its high costs. An affordable and rapid standardized test was adopted considering these issues: Chest X-rays.

Chest X-rays being time and cost-effective are the most common modality of pneumonia diagnosis. Doctors and radiologists with years of expertise examine the X-rays to detect the presence of pneumonia. Radiation level for chest X-rays in children is lower in contrast to the radiation levels used in adults to eliminate the risk of developing cancer. Low radiation levels in X-rays lead to loss of important information, making the task of pediatric pneumonia detection much more laborious and strenuous. With the ongoing COVID-19 virus advancing into other variants, several doctors across the globe are being transferred to emergency wards. The current situation might place children with pneumonia in jeopardy of not getting the required medical attention and thus, motivates the need for a computer-aided diagnostic model that is accurate and immediate.

Several Computer-Aided Diagnosis (CAD) methods are currently in use for various biomedical applications, such as breast cancer detection [5], heart disease detection [6], tuberculosis detection [7], Alzheimer disease detection [8], diabetes-related retinal disease detection [9], and pneumonia detection. Literature survey shows that machine learning-based pneumonia diagnosis from chest X-rays using several feature extraction techniques helped physicians automate the process of diagnosis. However, this feature extraction process requires the usage of handcrafted filters. Feature engineering in biomedical tasks requires tremendous proficiency and relies laboriously on experts, hence hindering the widescale development of CADs.

The applications of computer-aided diagnosis are now limitless with the advent of deep learning. Deep learning has been rooted down firmly in different domains owing to the availability of enormous data and ample computational resources. Deep Convolutional Neural Networks (CNN) has gained lots of attention in recent years leading to state-of-the-art performances in various image classification problems. The advantage of automatic feature extraction and engineering in deep learning, which was not previously possible with machine learning, has propelled the surge in computer-aided diagnosis-based systems. Transfer learning, a splendid breakthrough in artificial intelligence, has helped researchers overcome the disadvantage of the inadequate dataset that arises due to privacy concerns. Most of the deep learning architectures used for pediatric pneumonia diagnosis perform well nonetheless, their performance is limited. The cause for this is in the learning of a neural network: huge parameters in neural networks tend to overfit and thereby limit their performance on the test data. Most of the models proposed in the literature are not generalizable and robust as their performance has not been validated on similar datasets belonging to the same disease. Possible reasons for the limited performance include poor outlier handling, training on high class imbalance datasets, models with convoluted structure, and overfitting. The existing models are not guaranteed to perform well on unseen data. Robustness and generalizability are the key factors to be considered before real-time deployment. Therefore, it is of utmost importance to validate the performance on datasets of similar lung diseases or the same disease. Accounting to the above-mentioned concerns, the major contributions of the proposed work are summarized as follows:

- A stacked classifier-based learning approach, leveraging the strengths of machine learning classifiers and neural networks for pediatric pneumonia detection.
- A comparative performance analysis of the various pretrained deep CNN architectures with the proposed method for the task at hand.
- Class activation maps (CAM) to visualize the area of interest pertinent to the classification of normal and pneumonia X-rays.

- t-distributed stochastic neighbor embedding (t-SNE) based feature visualization for layman interpretability of the features predicted by the deep CNN architecture.
- Investigation on the effect of the kernel PCA on the performance of the classification model.
- An up-to-date comparison with other recent works tested on the publicly available Kermany et al. [10] dataset.
- A detailed investigation on the advantages and limitations of the proposed architecture on the pediatric pneumonia dataset.
- Performance analysis of the proposed method on similar pneumonia datasets to prove its generalizability and robustness.

The contributions made in the field of pediatric pneumonia diagnosis that motivated us to advance with the idea of stacked ensemble learning are as follows:

- The unprecedented research on the diagnosis of pediatric pneumonia with transfer learning, unfurled the possibilities of research along with the open-source availability of the dataset [10].
- The current impediment in the performance of deep convolution layers was solved using dilated convolutions, residual structures, and transfer learning [11].
- A fusion technique involving a deep CNN model with PCA and logistic regression [12].
- A weighted average ensemble of deep CNN models incorporating deep transfer learning [13].
- A majority voting ensemble of the predictions from deep CNN models [14].
- CheXNet [15], a DenseNet121 model trained on the ChestX-ray14 dataset whose performance exceeded that of the average radiologist.

The rest of this article is organized as follows: Sect. 2 describes the literature survey and discusses the existing gap in the literature and how our approach completes it. The proposed approach is discussed in Sect. 3. Section 4 contains the description of the dataset. Section 5 details the performance metrics used in this study. The experimental results are analyzed and discussed with plots in Sect. 6. In Sect. 7, we conclude our work; summarizing the problem and the limitations of our approach, along with the possible future works.

2 Literature survey

Convolution Neural Network (CNN) gets its name from the mathematical operation called convolution. CNN is widely used for feature extraction and consists of three types of layers: convolutional layer, pooling layer, and fully connected layer. The first study on pediatric pneumonia detection using deep learning facilitated the onset of pediatric pneumonia-based diagnosis research [10]. The dataset was made public, and researchers began experimenting with different neural network approaches. Multilayer Perceptron (MLP) and CNN-based approaches were proposed in [16]. As a continuation of his previous work, Saraiva et al. [17] used CNN for feature extraction, followed by cross-validation for extensive learning from the limited dataset. Several state-of-the-art deep learning models were fine-tuned for pediatric pneumonia detection on the Kermany et al. [10] dataset with competing performances. However, the performances of these models were limited. The current limitation of deep CNN architectures is the degradation of spatial information with increasing layers. In classification tasks pertinent to medical imaging, spatial information is of acute necessity. Gaobo Liang et al. [11] proposed an elegant solution to this shortcoming. They presented a deep learning framework based on dilated convolution to preserve spatial information alongside residual structures to prevent over-fitting. In addition to dilated convolutions, their study emphasizes using transfer learning for better training on the small-scale dataset.

CheXNet [15], a deep CNN model built by Stanford's researchers trained on the ChestX-ray14 dataset, achieves a diagnosis capability better than the average radiologist. Additionally, they executed a secondary check on the given dataset for proper classification. In transfer learning, predefined weights are a key factor in determining the performance of a model. The knowledge of CheXNet weights was transferred to the task of pediatric pneumonia diagnosis in several studies. It is highly favorable if the weights chosen belong to the same field. The differences in performance when using CheXNet weights, ImageNet weights, and random weights are detailed in [29]. Stephen et al. [30] investigated the performance of simple CNN architecture in the absence of transfer learning.

Several studies focus on existing deep CNN architectures, such as MobileNets, VGGs, DensNets, and ResNets. Rahman et al. [21] studied the performance of AlexNet, ResNet18, DenseNet201, and SqueezeNet using transfer learning for normal vs. pneumonia, bacterial vs. viral pneumonia, and normal, bacterial, and viral pneumonia classification. Novel architectures were proposed as a solution to the existing limitations in these deep CNN architectures. Deep sequential CNNs for pediatric pneumonia detection are introduced in [19]. In [20], the authors exemplify the use of depthwise separable convolutions for the task of pediatric pneumonia diagnosis. A hybrid system consisting of adaptive median filter Convolutional Neural Network (CNN) recognition model based on Random Forest (RF) for detecting pneumonia from chest X-Ray images was introduced in [35]. In addition to different

architectures, several feature extraction techniques were also employed. Wavelet transform is another technique for feature extraction based on a set of predefined filters. Akgundogdu et al. [18] analyzed the performance of 2D discrete wavelet transform for feature extraction with random forest for classification.

Image enhancement techniques have become a topic of interest to improve the quality of the image and highlight essential features in an image. The effect of HE, CLAHE, image complement, gamma correction, and balance contrast enhancement techniques for chest X-rays are described in Tawsifur et al. [23]. Rubini et al. [24] compared two prominent spatial processing techniques- Adaptive histogram equalization (AHE) and Contrast Adaptive histogram equalization (CLAHE) for enhancing MRI images. El Asnaoui et al. [22] compares fine-tuned deep-learning architectures' performances for binary classification in pediatric chest X-rays. Their work details the advantage of using Contrast Limited Adaptive Histogram Equalization (CLAHE) as an image enhancement technique for better learning.

The class imbalance problem is a necessity that needs to be addressed in machine learning. Machine learning is heavily dependent on a balanced dataset for unbiased training. Sampling is an important solution to deal with class imbalance problems. Habib, Nahida, et al. [25] proposed the use of Random Under Sampling, Random Over Sampling, and SMOTE on ensembled features from VGG-19 and CheXNet. Luján-García et al. [26] explored random undersampling (RUS) for unbiased training and used a cost-sensitive learning approach for the Xception network. However, such approaches' performance was limited because the data generated from SMOTE was unable to capture the required features for pediatric CXRs, and no new data was generated to improve learning in RUS.

The performance of a model can be increased using several techniques. Increasing the feature set is one way to improve the performance of the model. This idea applied to pneumonia diagnosis was introduced by Nahid et al. [27] where they proposed a novel two-channel CNN architecture for pneumonia diagnosis. Predictions using feature concatenations from SqueezeNet and InceptionV3 along with ANNs are detailed by Islam et al. [28]. Their work entails retraining with modified parameters in addition to redistributing the existing dataset for unbiased training. Hyperparameters are a major contributing factor to the performance of a model. The right choice of optimizers is crucial to get the best results. While most of the recent related research focused on Adam optimizer, the effect of Stochastic Gradient Descent (SGD) optimizer was explained in [31].

Ensemble approaches are another important technique to improve the predicting accuracy of a model. Chouhan et al. [14] studied the performance of a majority voting ensemble combining the predictions from AlexNet, DenseNet121, Inception V3, GoogLeNet, and ResNet18. Sagar Kora Venu [13] proposed a weighted average ensemble of these deep CNN models—MobileNetV2, Xception, DenseNet201, ResNet152V2, and InceptionResNet. Nahida et al. [12] proposed a combination of a deep convolutional neural network for feature extraction, Principal Component Analysis (PCA) for dimensionality reduction, and logistic regression for classification. Improved feature representation may increase the performance of a classification model. A graph knowledge embedded convolutional network called CGNet was proposed by Yu et al. [33]. They used the transfer learning technique for feature extraction followed by graph-based feature reconstruction for classification. Mittal et al. [34] proposed a CapsNet architecture for classifying normal and pneumonia images.

The main impeding factor for the complete transition to artificial intelligence (AI) is the lack of transparency. A promising field of research called explainable AI (XAI) has been gaining momentum lately. A unique approach in integrating explainability for pneumonia detection was introduced by Nguyen, Hai, et al. [32]. They proposed a combination of custom CNN architecture and Grad-CAM for pneumonia detection. An abundance of research has been done in this field. However, there exist limitations which are discussed below:

1. Most studies propose data augmentation techniques to increase the number of samples for training to ensure improved performance. Artificially increasing the dataset is time and space inefficient.
2. Studies emphasize the use of CheXNet weights for custom CNN training which is a challenging task.
3. Lot of research proposes the use of custom complex architectures that are not easily replicable and hampers the reproducibility of the work.
4. The absence in the exploration of ensemble approaches pertinent to pediatric pneumonia diagnosis was observed. The same was witnessed concerning the use of machine learning classifiers.
5. The pressing need for dimensionality reduction using PCA has not been stated firmly.
6. Data sampling methods like RUS, ROS, and SMOTE lead to longer training times and over-fitting.
7. Most of the above-mentioned studies failed to cover the aspect of feature visualization. This is very important to ensure the learned features are meaningful for predictions.

Our work proposes a detection pipeline to bridge the gap in the existing literature. The dataset has been redistributed for unbiased training instead of using data sampling methods. The proposed methodology is based on the

Xception architecture pretrained on the commonly available ImageNet weights for feature extraction. The extracted feature maps from the global average pooling layer are passed to the t-SNE for feature visualization. Kernel PCA is then used for dimensionality reduction. Stacking ensemble classifier approach with KNN, SVC, RandomForest classifier, Nu-SVC, MLP classifier, and Logistic Regression was used along with Stratified K-Fold cross-validation to overcome overfitting. All additional details are discussed in the forthcoming sections.

3 Proposed approach

This section details the workflow of the proposed pediatric pneumonia detection model, from the collection of data to the final classification as illustrated in Fig. 1. The dataset contains images of varying sizes. In this study, we reshape the images according to the requirement of different deep CNN models. Each image is normalized to bring the pixel values between the range 0–1 using the Keras image generator in addition to the introduction of sheer, zoom, and flip augmentations as shown in Table 1. Image augmentations are a necessary part of modeling to prevent overfitting. These augmentations are generated on the fly in concurrence with the training.

The proposed architecture is trained on a two-step process. The first step was to use train deep CNN architecture for feature extraction. The Xception network was selected among all the other existing deep CNN architectures based on its performance for this task. A global average pooling layer was added to obtain feature maps. To prevent overfitting, a dropout rate of 0.4 was used and the Xception network was trained using binary cross-entropy loss. The ImageNet weights were used for transfer learning from second half of the layers in the Xception network. This resulted in better feature extraction from the CXRs. With the model now being able to extract the required features, the second step was to train the stacking classifier using the extracted features. The extracted features from the fine-tuned Xception network are sent through Kernel PCA for dimensionality reduction. The reduced features are trained on the stacking classifier with Nu-SVC, XGB classifier, Logistic Regression, K-Nearest classifier, Support Vector classifier, Randomforest classifier and MLP classifier for the first stage. The predictions from the base estimators (first stage classifiers) are trained on a meta classifier (logistic regression) for the final binary NORMAL and PNEUMONIA classification.

3.1 Transfer Learning

The performance of any deep learning model relies on the amount of data available. Accessibility to large datasets is guaranteed to increase the performance of deep learning models and make them more robust. Large datasets allow the model to learn much more intrinsic patterns. However, this is not always the case in medical imaging pertinent to pediatric pneumonia due to concerns, such as patient privacy and the time-consuming task of inspecting and labeling the data. Transfer learning [36] serves as a solution to this problem. In transfer learning, we use the existing knowledge gained when trained on a similar task and apply it to our detection of pediatric pneumonia. In our study, we fine-tune models pre-trained on ImageNet weights (trained on more than 14 million images ranging across 1000 classes).

3.2 Deep learning models

The literature survey concludes on the observation that competing performances were obtained when using pretrained deep CNN models. A detailed investigation on existing pretrained CNN architectures was performed to find the architecture best suited to the task at hand. These models pretrained on ImageNet weights were trained and tested on the Kermany dataset [10] to understand its advantages and limitations for pediatric pneumonia diagnosis. The initial half of layers were frozen while the second half of the models were fine-tuned. Pre-trained deep CNN models, such as VGG16 [37], VGG19 [37], MobileNet [38], MobileNetV2 [38], MobileNetV3Large [50], MobileNetV3Small [50], InceptionResNetV2 [39], DenseNet121 [40], DenseNet169 [40], DenseNet201 [40], InceptionV3 [41], ResNet50 [42], ResNet101 [42], ResNet152 [42], ResNet50V2 [43], ResNet101V2 [43], ResNet152V2 [43], EfficientNetB0 [51] and Xception [44] are trained on the Kermany dataset [10] to find the best performing model. The features are then extracted using the best performing model. The extracted features are passed through the global average pooling layer to extract one feature map from each image. These features are used for further processing. Details on the parameters used to conduct all the experimentations are explained in results and discussions.

3.2.1 Xception

CNNs rely on the gradients in the image for feature retrieval. Increasing convolution layers introduces the vanishing gradient problem, hence explaining the staggering performance with the increasing number of

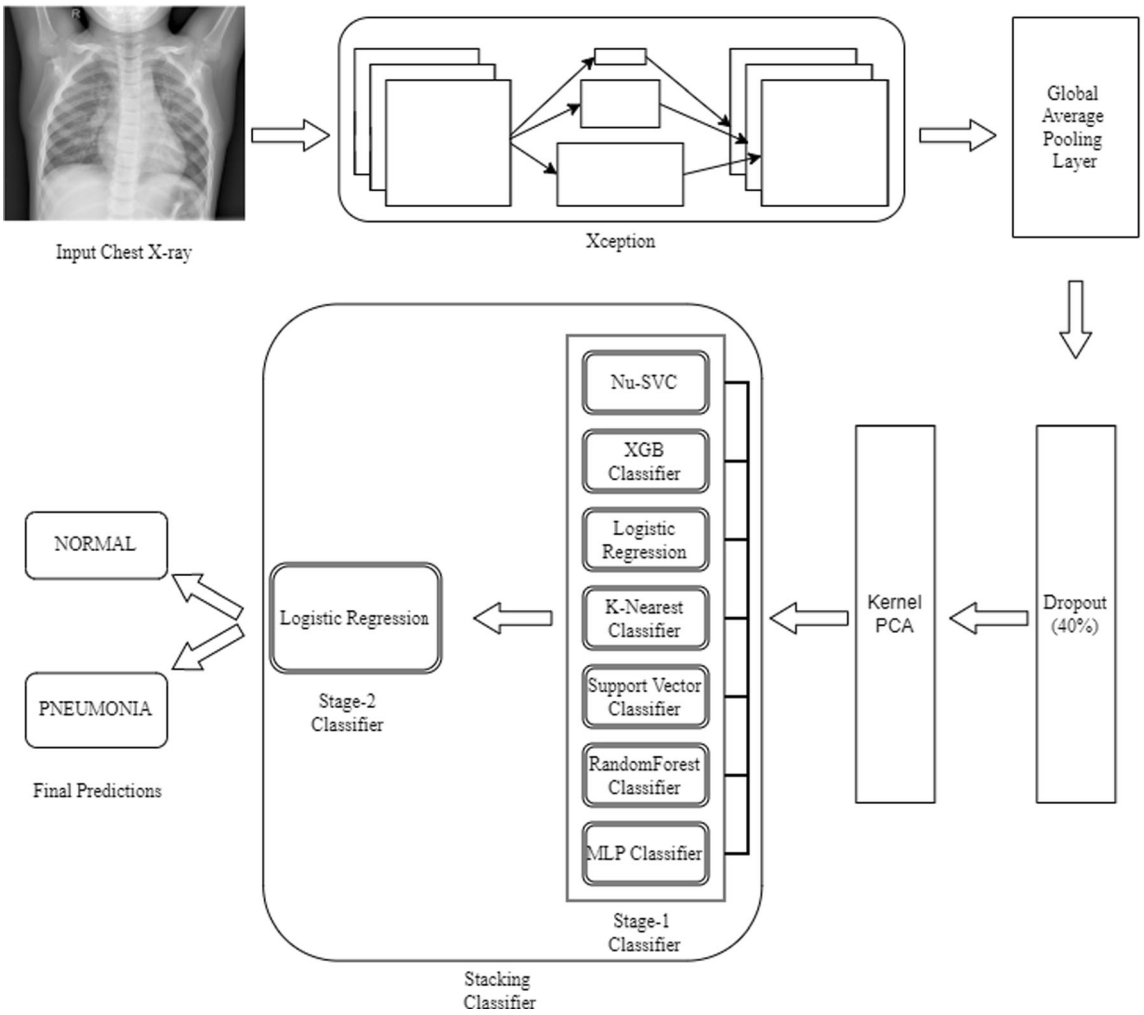

Fig. 1 Proposed architecture for pediatric pneumonia classification

Table 1 Augmentations used in our study and their corresponding values

Methods	Corresponding parameters
Rescale	255
Shear	0.2
Zoom	0.2
Horizontal Flip	True

convolutional layers. Residual connections were introduced as a solution to the vanishing gradient problem. In time, researchers began incorporating residual structure in deep CNN models. Inception made its way into the research community along with its successors- InceptionV3 and InceptionResNets. Inception was built on the hypothesis that the spatial and cross-channel correlations in feature maps can be decoupled. Xception, leveraging this hypothesis pushed it to the extreme, thereby getting the name Xception, the extreme version of Inception.

The Xception architecture is a stack of 14 modules (36 convolutional layers) with linear residual connections except for the first and the last modules. The entry flow initiates the flow of data and is followed by the middle flow where the set of operations is repeated 8 times. The architecture incorporates residual structure to tackle the vanishing gradient problem. The exit flow terminates the order of convolutions. The detailed architecture flow diagram is shown in Fig. 2. Each convolution and separable convolution layer is succeeded by batch normalization. In contrast to depthwise separable convolution where depthwise convolution is followed by pointwise convolution as shown in Fig. 3, Xception follows the reverse. The process starts with pointwise convolution followed by depthwise convolution.

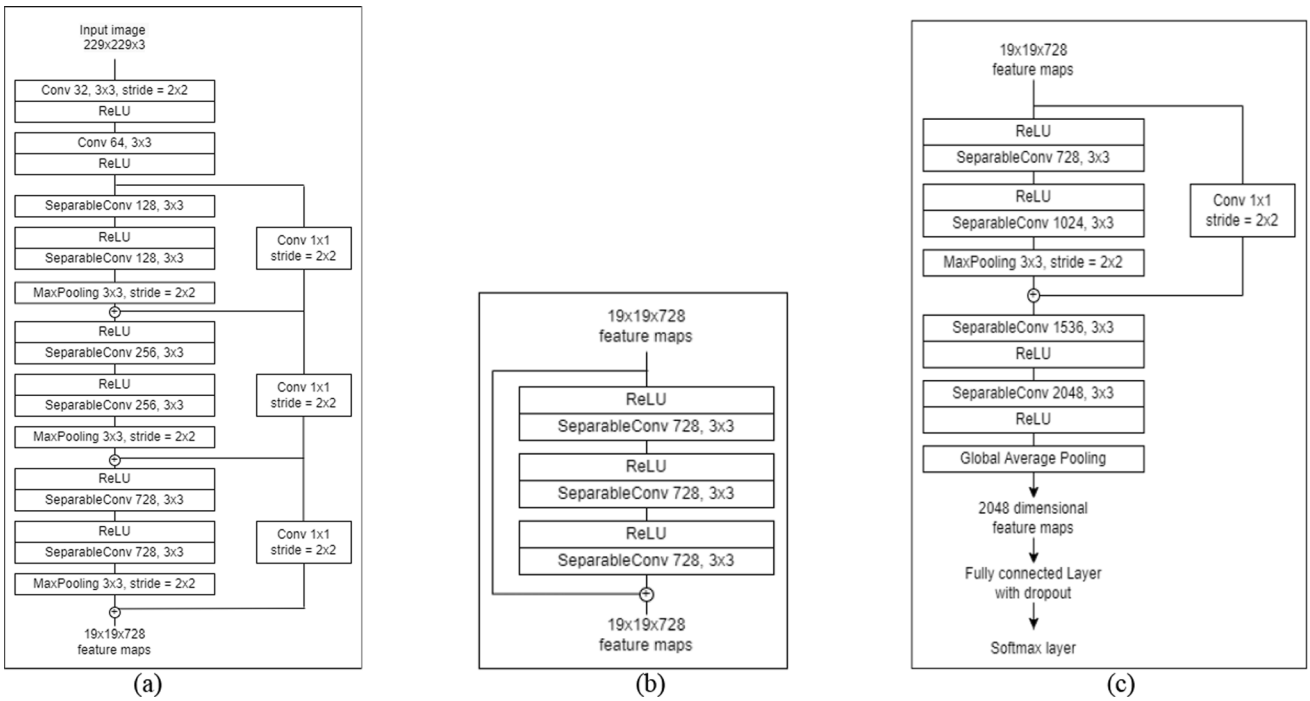

Fig. 2 The architecture for xception deconstructed as (a) Entry flow, (b) Middle flow and (c) Exit flow

3.2.2 Hyperparameters

To improve the feature extraction capability of the models, the best performing deep CNN architecture was first selected among existing deep CNN architectures. This selection was based on training all the architectures with a learning rate of 0.001 with adam as the optimizer and selecting the best performing model. The sigmoid activation function was used for this binary classification task. Hyperparameter tuning is a crucial step to boost the performance of a model. In our study, we fine-tuned the models based on different combinations of optimizers and learning rates as shown in Table 2, to select the perfect composition that results in the highest validation accuracy.

Binary cross-entropy is used as the loss function which calculates the difference between the expected and actual output. The value for the loss function ranges between 0 and 1 and is given by Eq. 1.

$$\text{Loss} = \sum_{i=0}^{\text{output size}} y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i) \qquad (1)$$

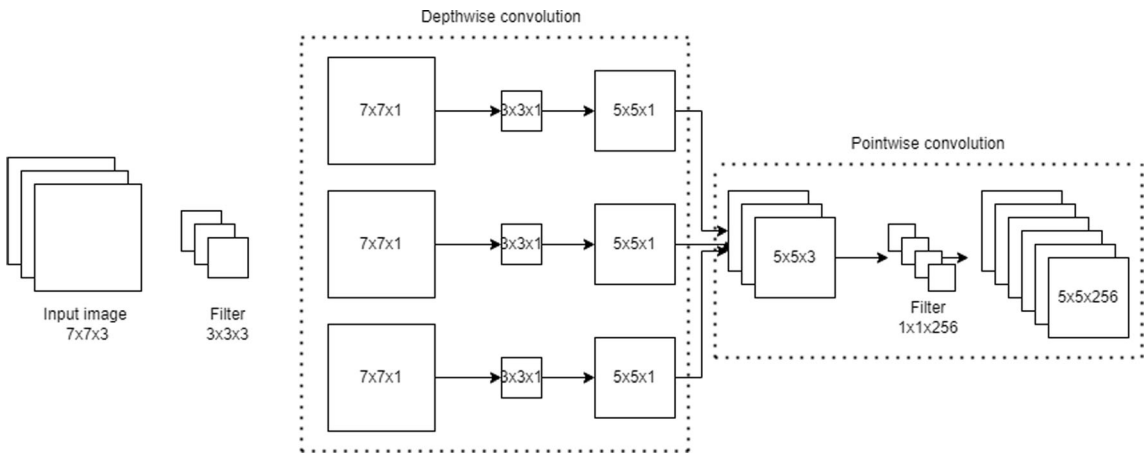

Fig. 3 Illustration of the working of a depthwise separable convolution network

Table 2 List of hyperparameters and their values used in our study to finalize the perfect combination for the task at hand

Hyper parameter	Corresponding values
Optimizer	Adam, SGD, Nadam, RMSprop, Adamax, Adagrad
Learning rate	0.01, 0.001, 0.0001, decay rate from 0.001 to 0.000001
Batch size	32

3.3 Principal component analysis

Large datasets often have redundant features which makes them difficult to interpret. No additional benefit is gained by learning the redundant feature rather it burdens the time taken to train a model. An immediate solution to this complication is Principal component analysis [45]. PCA is a famous statistical method used for dimensionality reduction. The PCA algorithm reduces the dimensionality of the given input in such a way that it minimizes the loss in reduced data. The objective of the PCA algorithm is to maximize the variance by creating new uncorrelated variables. The first set of uncorrelated variables forms the first principal axes. Similarly, subsequent sets of variables form their corresponding principal axes. The first principal axes capture the maximum variance and subsequent axes that are orthogonal to the previous axes capture decreasing variances in order. PCA performs well for linearly separable data however, this is never the case in real-world data. Kernel PCA [46] was developed as a solution to deal with nonlinear dimensionality reduction. It captures much more intrinsic correlations between the given high-dimensional features.

3.4 Stacking classifiers

Ensemble learning has attracted considerable attention in the past years. Studies emphasize it as a promising way to improve the performance of a model. Stacking Classifier is one such ensemble learning technique. As the name suggests, it stacks the predictions from individual classifiers (base classifiers) and uses them as features. These features are trained on a final classifier called the meta classifier. Stacking exploits the strengths of individual predictions, making predictions much richer and accurate**.**

3.5 Stratified K-fold cross-validation

Cross-validation is the most widely employed technique to estimate the model's performance on unseen data. The performance on unseen data is of utmost importance for real-world deployment. The cross-validation technique facilitates the model to learn the most out of the provided data and prevent over-fitting. The Stratified K-Fold is an extension of the K-Fold cross-validation technique developed for the purpose of dealing with imbalanced class distributions. It ensures that each fold has same class distribution as in the original dataset. The dataset used has a higher number of pneumonia CXRs than normal CXRs for training. The class distribution in the training set was preserved in each of the folds. In this study we used Stratified K-Fold cross-validation with n_splits = 10.

4 Dataset description

The Kermany et al. [10] dataset was used for all the experiments in this comparative study. The dataset comprises 5856 Chest X-Ray images belonging to two categories- Normal (1538 X-rays) and Pneumonia (4273 X-rays). The dataset was split on an 80–10-10 (train-test-validation) split ratio after recombining the train and test data of the Kermany et al. [10] dataset. The data distribution used in this study is shown in Table 3. These chest X-ray images are from routine screening in Pediatric patients between 1 – 5 years of age from the Guangzhou Women and Children's Medical Centre. The faint white occlusions, present in the X-rays in the second row in Fig. 4 are due to the occupancy of pus and fluids in the alveoli.

5 Performance metrics

Performance metrics are imperative to distinguish the performance of classification models. The metrics used in this study are accuracy, precision, recall, F1-score, and the AUC value. The Confusion matrix counts the distribution of predictions across the actual labels as shown in Fig. 5. Accuracy, Precision, Recall, and F1-score are derived from the confusion matrix.

The accuracy of a model is calculated as the ratio between correct predictions and total predictions as shown

Table 3 Distribution of the dataset for our study

Category	Train	Test	Validation
Normal	1266	159	158
Pneumonia	3418	427	428
Total	4684	586	586

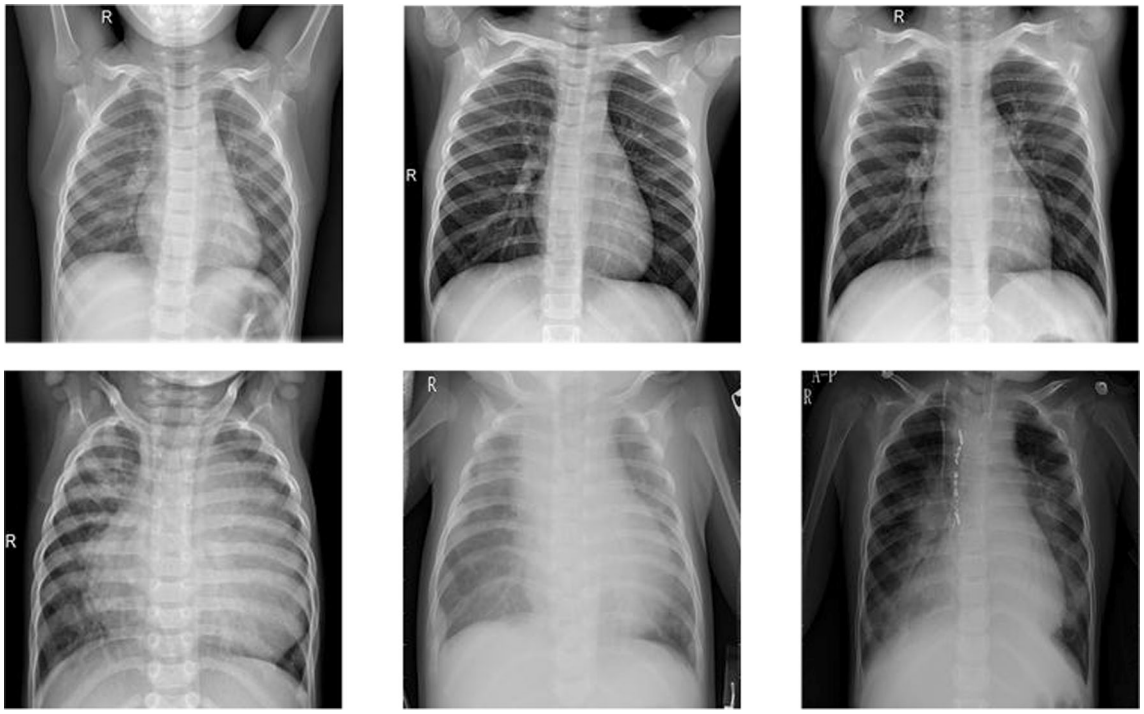

Fig. 4 Samples of Normal x-rays and Pneumonia x-rays from the dataset in the first row and second row, respectively

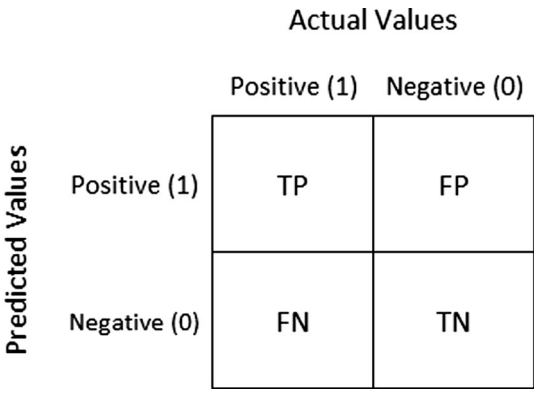

Fig. 5 Confusion matrix. True Positive (TP)—number of pneumonia x-rays correctly predicted as pneumonia. False Negative (FN)—number of pneumonia x-rays wrongly predicted as normal. True Negative (TN)—number of normal x-rays correctly predicted as normal. False Positive (FP)—number of normal x-rays predicted wrongly as pneumonia

in Eq. 2. The precision of a model is calculated as the ratio between true positives and total positives as shown in Eq. 3. It summarizes the quality of positive predictions made by the model. For a good classifier, precision is close to 1. Recall of a model is calculated using Eq. 4, which shows how well the predictions are classified as actual positive. F1-score is the harmonic mean of precision and recall as shown in Eq. 5. The Area Under Curve (AUC) score is the area under the receiver operating characteristics (ROC) curve. It defines the ability of the model to distinguish between patients with and without pneumonia. For a good classification model, the AUC score must be close to 1.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F1 score} = 2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right) \tag{5}$$

6 Results and discussion

Several deep-CNN models were trained on the 4684 x-ray images for 30 epochs and evaluated on the test data consisting of 586 images to determine the model best suited for the task at hand. Google Colab resourced with K80 GPU and 12 GB RAM was used to conduct all the forementioned experiments in this study. Tensorflow2 and Keras2 were used to build and evaluate the models.

The following models are compared and the best performing model is used for feature extraction: VGG16, VGG19, MobileNet, MobileNetV2, MobileNetV3Small, MobileNetV3Large, InceptionResNetV2, DenseNet121, DenseNet169, DenseNet201, InceptionV3, ResNet50,

ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2, Xception and EfficientNetB0. Each of the models above were pre-trained on ImageNet weights with the corresponding input image of size 224 × 224 for all architectures except for InceptionV3, ResNet152V2 and Xception with an input image of size 299 × 299. All deep CNN models were trained with a constant learning rate of 0.001 and Adam as the optimizer. The initial layers of all the deep CNN models were frozen during training. Table 4 describes the layer count at which fine-tuning commenced for each deep CNN model with the corresponding count of trainable parameters. Table 5 illustrates the performance of the existing deep CNN architectures for the binary classification of no-pneumonia vs pneumonia detection along with the proposed method.

When noticed, the family of DenseNet models performs consistently well. The reflection of collective knowledge in DenseNets enabled it to achieve an accuracy of 0.96. InceptionResNetV2, ResNet152V2, and Xception are the best performing architectures with the highest accuracy compared to the rest of the models for the task of pediatric pneumonia detection. The residual connections are a key factor that has suppressed over-fitting and thus enabled the above models to perform well on the test data. Though ResNet152V2 and InceptionResNetV2 achieve the same accuracy of 0.97 and an AUC of 0.98 similar to that of Xception, the latter has a higher recall of 0.97 compared to the former architectures. The recall of a model is of utmost importance as we do not want X-rays with pneumonia to be classified as normal. The confusion matrix for the test data predictions from the Xception architecture is shown in Fig. 6. From the confusion matrix, we conclude that the Xception in itself is unable to deal with false positives and false negatives. Figure 7 shows the ROC curve for the test data predictions. Xception proves to be a good feature extractor with an AUC of 0.97 still, its performance can be improved by looking at the feature representations.

The training and validation plots are shown in Fig. 8. Though the loss initially peaks at irregular intervals, it substantially decreases. It can also be inferred that the validation loss and accuracy are constrained to certain bounds from 1 to 0 and 0.75 to 1, respectively. The validation data of 586 images were used for hyperparameter tuning. The Xception model was first fine-tuned on different optimizers to find the best fit for the task at hand. The Adam optimizer performs best as seen in Fig. 9. This combination was further tested on different learning rates. Figure 10 illustrates the competing performances of these learning rates when set to a static and a continuously regressing value. Based on Figs. 9 and 10, the optimal hyperparameters with the adam optimizer and a constant learning rate of 0.001 were chosen for feature extraction.

Inspection of the features learned by deep learning models is crucial especially in the biomedical domain for its adaptability as a life-saving resource. This inspection was made possible with class activation maps [57] giving an overall vision of what the Xception model has learned. Figures 11 and 12 show the pixels that contributed the most while looking at pediatric pneumonia diagnosis for misclassified and correctly classified samples, respectively.

Table 4 Fine-tuning information and the number of trainable parameters associated with each model used in our study

Deep CNN model	Fine-tuned from	Total number of trainable parameters
VGG16	9	13,569,793
VGG19	11	17,699,329
MobileNet	50	2,665,473
MobileNetV2	77	2,064,769
MobileNetV3Small	117	1,371,849
MobileNetV3Large	134	4,028,273
ResNet50	87	21,364,225
ResNet50V2	95	21,352,449
ResNet101	172	30,640,129
ResNet101V2	188	30,625,793
ResNet152	257	39,855,617
ResNet152V2	282	39,836,673
InceptionV3	155	16,791,489
Xception	66	14,860,313
InceptionResNetV2	390	41,922,529
DenseNet121	213	4,632,897
DenseNet169	297	8,544,833
DenseNet201	353	12,741,185
EfficientNetB0	118	3,700,169

Table 5 Performance chart of deep learning models with values rounded off to the nearest two decimal positions

Model	Accuracy	Precision	Recall	F1-score	AUC
VGG16	0.73	0.73	1.00	0.84	0.5
VGG19	0.73	0.73	1.00	0.84	0.5
MOBILENET	0.96	1.00	0.96	0.97	0.97
MOBILENETV2	0.94	0.99	0.91	0.95	0.95
MOBILENETV3SMALL	0.27	0.00	0.00	0.00	0.5
MOBILENETV3LARGE	0.89	0.89	0.97	0.93	0.82
RESNET50	0.69	1.00	0.57	0.73	0.79
RESNET50V2	0.96	0.99	0.96	0.98	0.97
RESNET101	0.23	0.46	0.32	0.37	0.16
RESNET101V2	0.96	1.00	0.94	0.97	0.97
RESNET152	0.75	0.75	1.00	0.86	0.54
RESNET152V2	0.97	1.00	0.96	0.98	0.98
DENSENET121	0.96	1.00	0.95	0.97	0.97
DENSENET169	0.96	1.00	0.94	0.97	0.97
DENSENET201	0.96	1.00	0.95	0.97	0.97
INCEPTIONV3	0.92	1.00	0.89	0.94	0.95
XCEPTION	0.97	0.99	0.97	0.98	0.98
EFFICIENTNETB0	0.312	1	0.05	0.11	0.53
INCEPTIONRESNETV2	0.97	1.00	0.96	0.98	0.98
PROPOSED METHOD	0.98	0.99	0.98	0.99	0.98

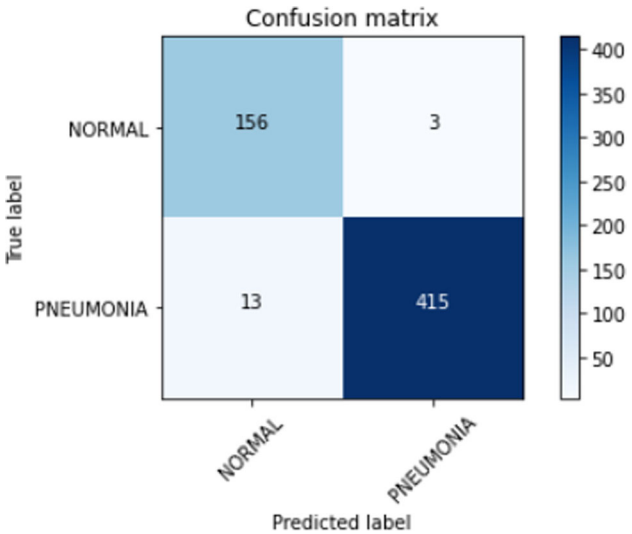

Fig. 6 Confusion matrix for xception predictions on the test data

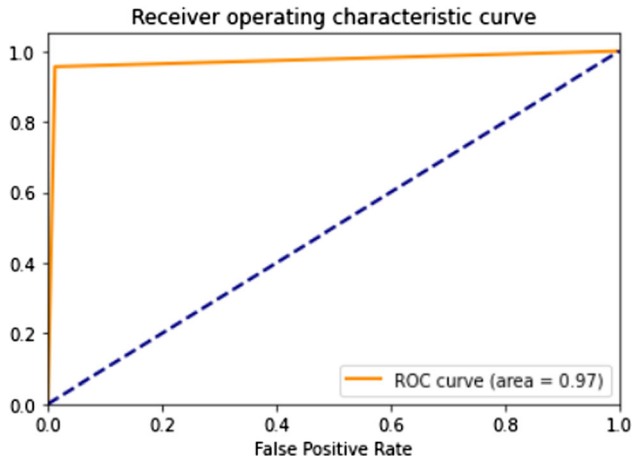

Fig. 7 ROC curve for test data predictions made by the fine-tuned xception model

Our method uses Xception for feature extraction with adam as the optimizer, the learning rate set to a constant value of 0.001 throughout the experiment, and a batch size of 32. The extracted features are visualized using the t-SNE [58] feature representation for the layman interpretability of the features predicted by the model. The t-SNE is a nonlinear dimensionality reduction technique that tries to preserve the local structure of the data. The feature maps of the test data are visualized using the t-SNE feature representation. The two dimensions (x and y-axes) shown in Fig. 13 are the first two principal components of the test data. This approach allowed us to visualize the normal and pneumonia samples in separate clusters. The cluster formation gives an idea of how well the predictions are made. In addition, the visualization element gives an insight into the possible classifiers that can be used for the classification task.

The parameter values used for visualization are n_components = 2, perplexity = 40, and n_iter = 300**.** The t-SNE plot of the extracted feature maps from the Xception architecture is shown in Fig. 13. Looking at the cluster

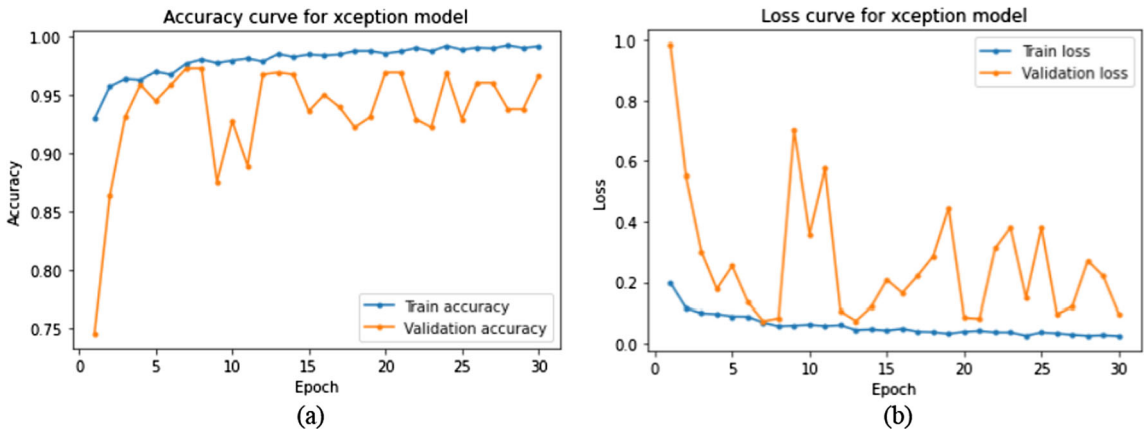

Fig. 8 Training and validation accuracy-loss history of the fine-tuned xception model

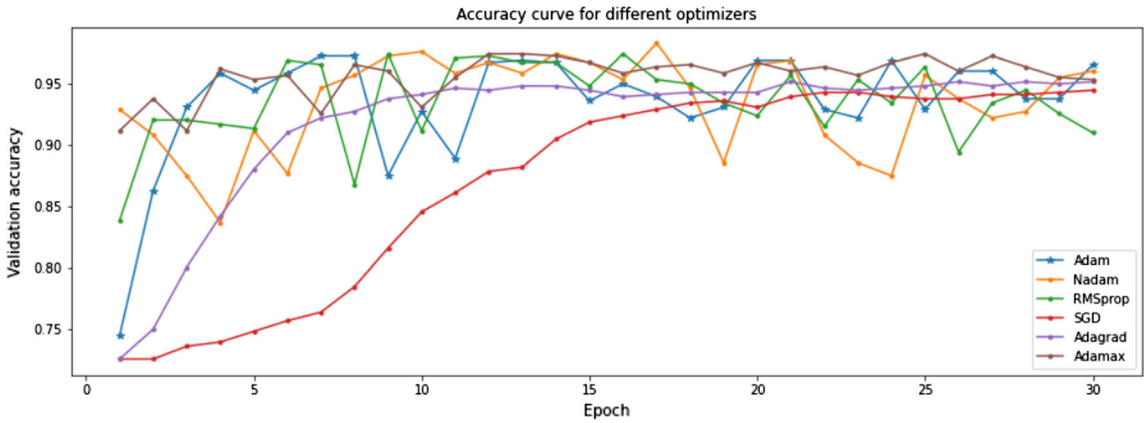

Fig. 9 Xception model performance on the validation set using different optimizers

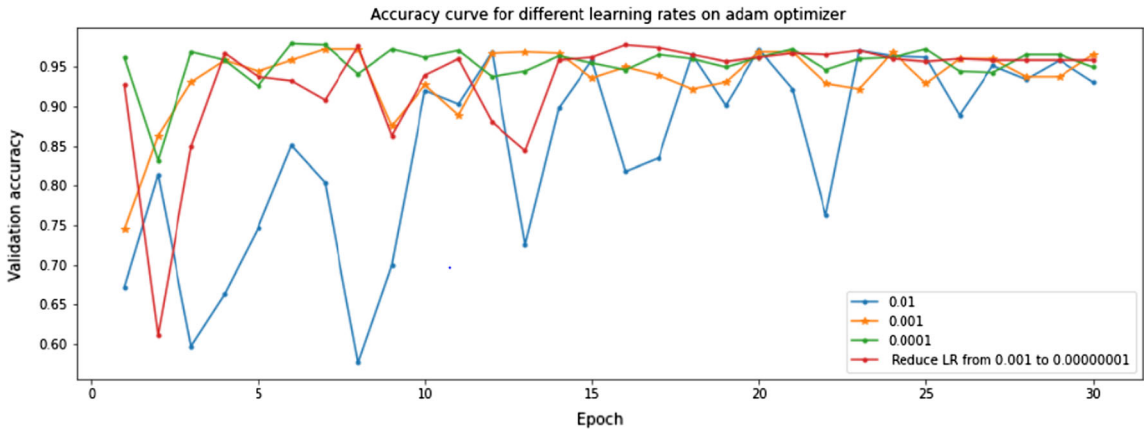

Fig. 10 Xception model performance on the validation set using different learning rates with adam as the optimizer

formations, we conclude that the test samples are nonlinearly separable with minor overlaps between the predictions and that we need a classifier that is able to deal with such complexity. This study proposes the use of the stacking classifier to deal with the nonlinearly separable classification.

Thus, finalizing Xception as the feature extractor, the next step is dimensionality reduction using PCA (Principal Component Analysis). Dimensionality reduction is an important step to prevent the model from learning redundant features. In our study, we use the RBF (radial basis function) kernel with the number of resulting components

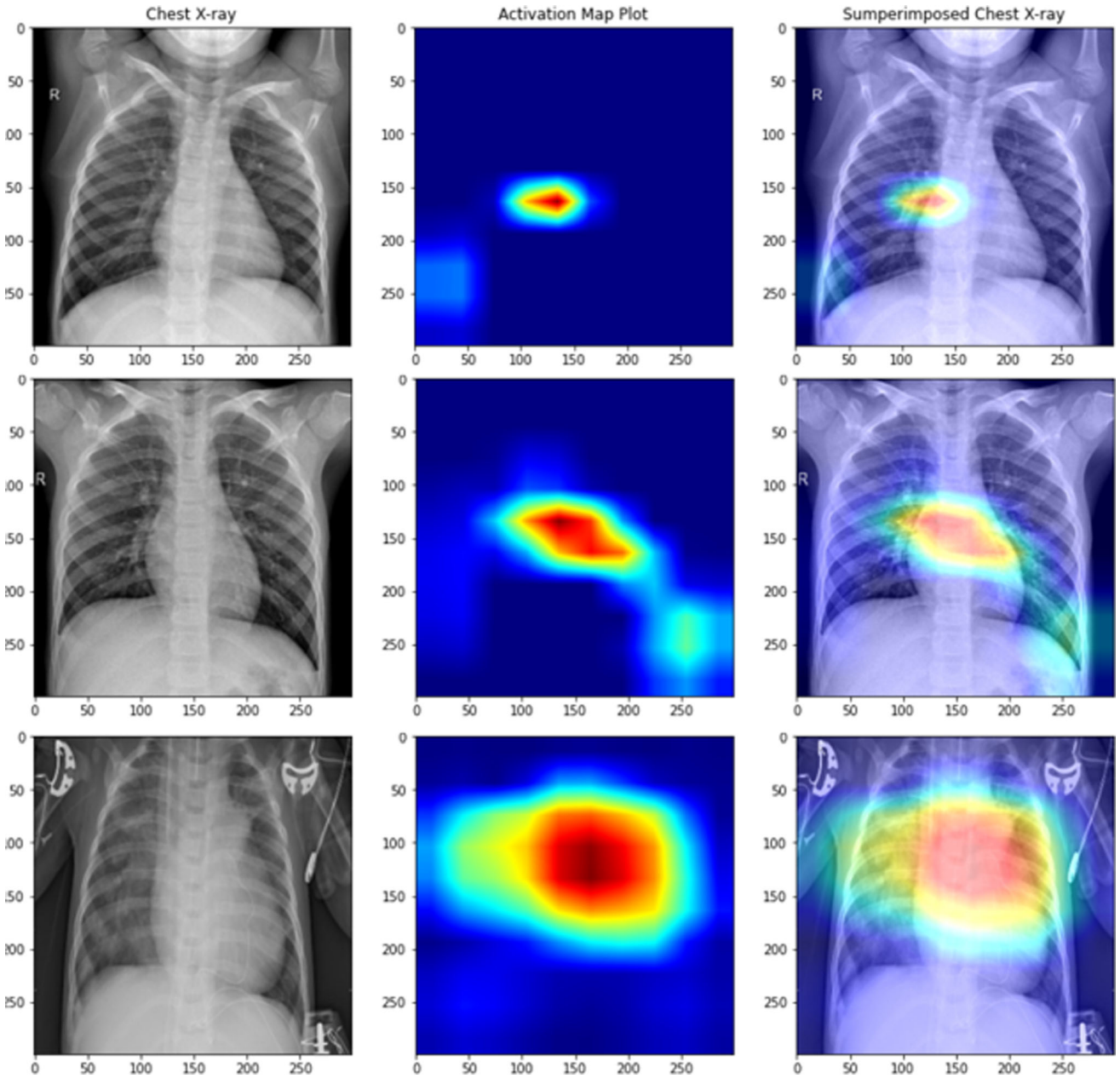

Fig. 11 Class activation maps of misclassified X-rays (row 1: normal classified as pneumonia, row 2: normal classified as pneumonia, row 3: pneumonia classified as normal)

as 200. This number has been chosen based on careful examination of the cumulative variance plot with a 95% cut-off threshold, shown in Fig. 14. Several machine learning classifiers were trained on the dimensionally reduced features and validated against the stacking classifier for the binary classification of normal and pneumonia CXRs. Table 6 concludes that the stacking classifier outperforms all machine learning classifiers by leveraging the strength of individual estimators.

Redundant features are detrimental to the performance of a classification model. The existing correlations between the important and redundant features are the key explanation for the hampering performance. The beneficial effect of removing redundant features in the task pertinent to pediatric pneumonia diagnosis is illustrated in Table 7 (Normal vs Pneumonia classification). The cumulative variance plot describes the percentage of the total variance captured by the first n components from the entire data. Higher variance indicates better preservation of important information from the data. The cumulative variance plot, Fig. 14 shows that the first 200 components capture most of the variance and that all additional principal components

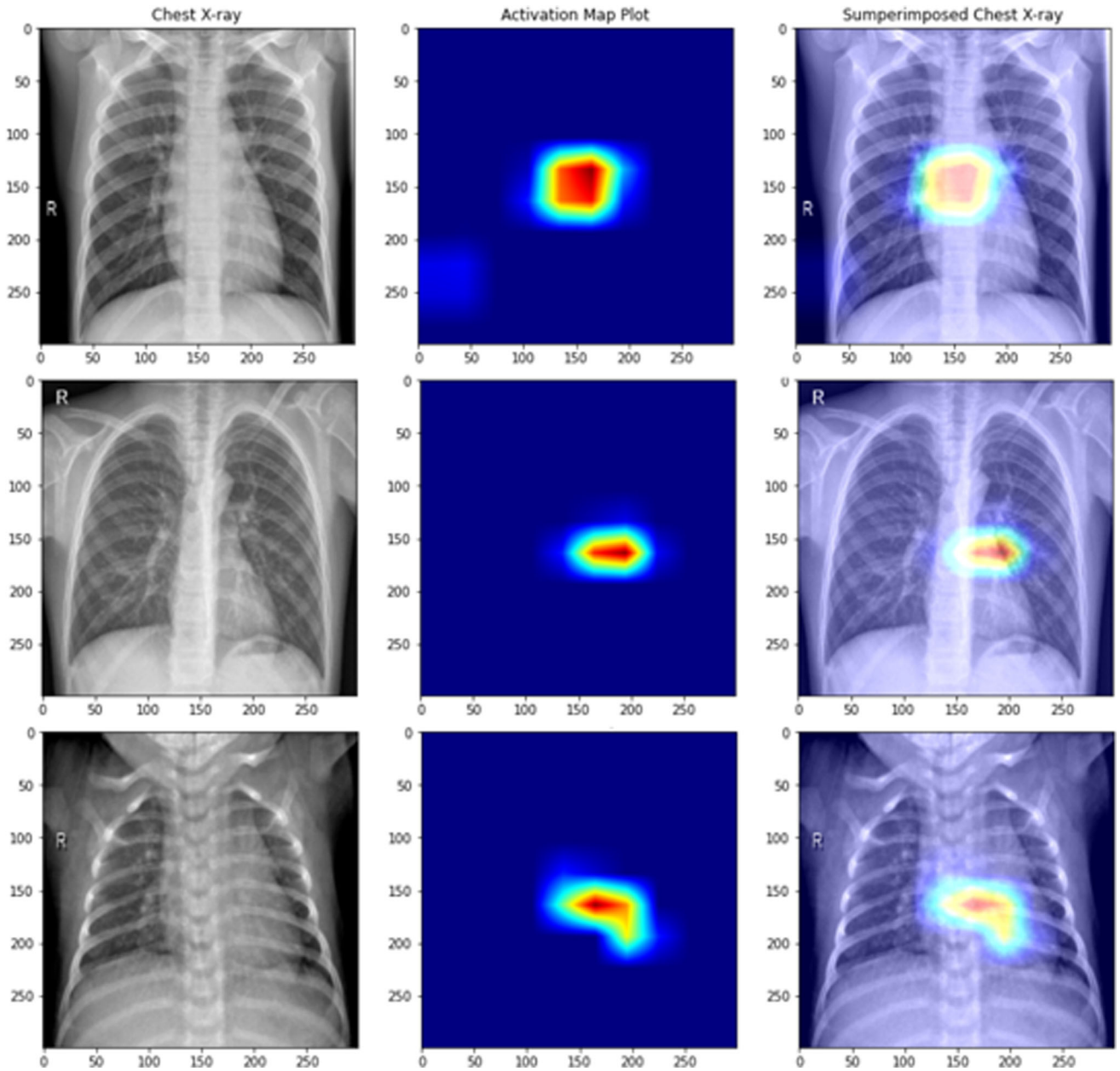

Fig. 12 Class activation maps of correctly classified X-rays (row 1: normal classified as normal, row 2: normal classified as normal, row 3: normal classified as normal)

henceforth are redundant. The 200-dimensional output is passed to the two-stage stacking classifier.

The first stage in the stacking classifier leverages the RandomForestClassifier, Support Vector Classifier, KNeighborsClassifier, XGBClassifier, LogisticRegression, Nu-Support Vector Classifier, and MLPClassifier. The hyperparameters for each of these classifiers were selected using GridsearchCV and are detailed in Table 8. Individual predictions from each of the five classifiers are sent to the meta-classifier for the final classification. The meta classifier uses LogisticRegression with penalty = l2, tol = 1e-4, C = 1.0, solver = 'lbfgs' and max_iter = 100. Stratified K-Fold cross-validation with n_splits = 10 was employed to help the model learn the most from the existing limited dataset and prevent over-fitting.

The confusion matrix for Stratified K-Fold cross-validation stacking classifier predictions on the test set is shown in Fig. 15. Lesser false-positive predictions from the stacking classifier are observed compared to the raw predictions made by the Xception architecture due to the strengths of individual classifiers. Thus, the strength of a stacking classifier solely relies on the individual strengths

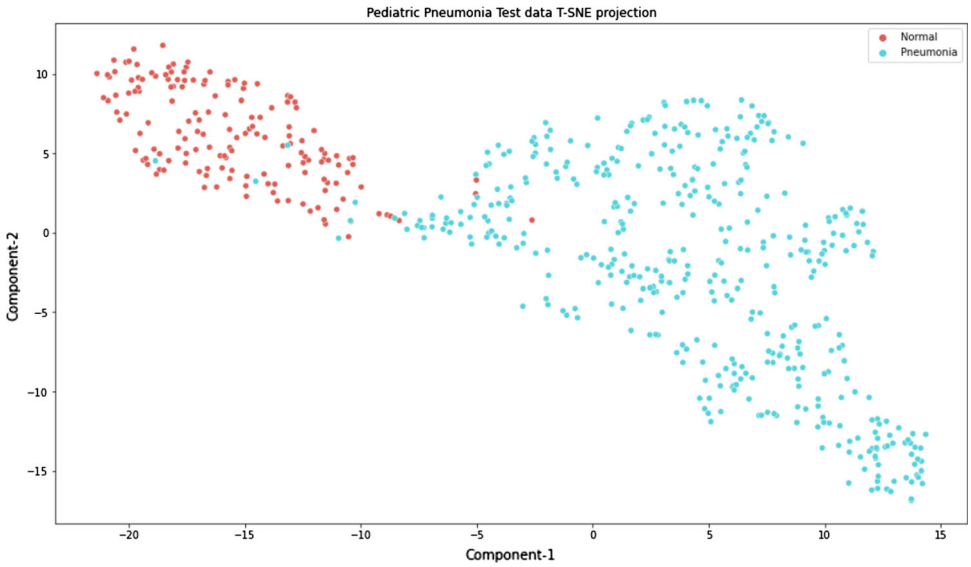

Fig. 13 t-SNE feature representation of the test data extracted from the xception model

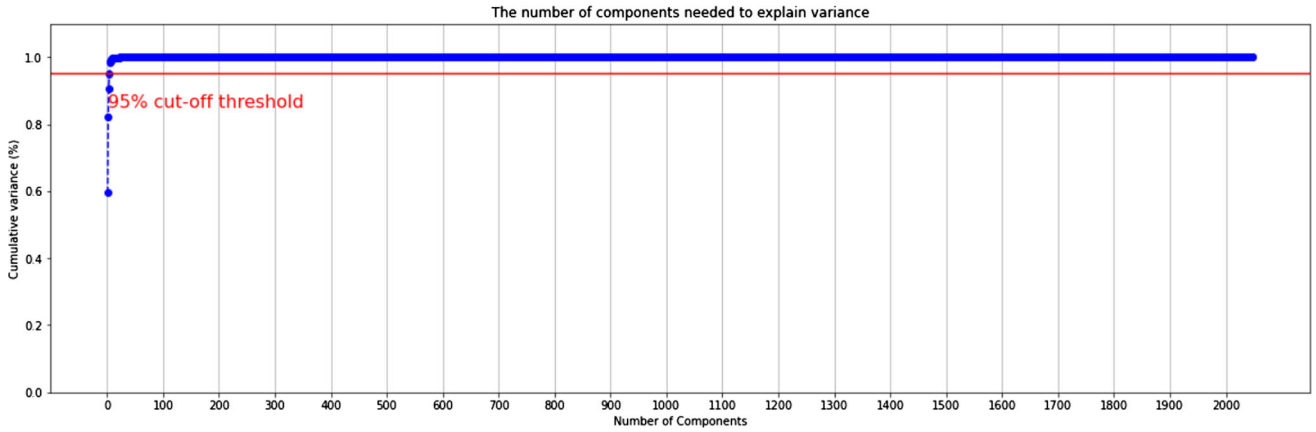

Fig. 14 Cumulative variance plot of the extracted xception features

of the predictors. Principal component analysis has facilitated in lowering the number of false positives and false negatives which can be seen as a comparison between Figs. 15 and 16. The ROC curve, shown in Fig. 17 has an AUC value of 0.98. The AUC value from the stacking classifier has a 1% increase from the previously obtained AUC value. Looking at the confusion matrix, the loss of 1.7% in the accuracy of the model might favorably be due to the imbalanced dataset or insufficient training samples for training. The proposed method achieves a much higher accuracy of **98.30%**.

Table 9 compares the performance, technique, and classification classes of our proposed approach with other recent works. The proposed work exhibits competing performances with other literary works for the binary classification of normal and pneumonia CXRs. All the works mentioned in the Table validated their results tested on the Kermany et al. [10] dataset. Since the Xception model used as the feature extractor is based on the commonly available ImageNet weights, reproducibility is easier. In addition to stacking various machine learning classifiers for rich predictions, the proposed method was tested on unseen pneumonia datasets for model generalization and robustness which was previously absent in recent works. The limitation of the proposed model is in its heavy reliance on the correct combination of base classifiers for accurate classification. The comparison hints at a possible future direction for using feature concatenations (Islam et al. [28]) followed by a stacking classifier for better results.

7 Robustness and generalization of the proposed approach for lung disease classification

The generalization of a proposed approach is essential to validate its performance. The proposed stacking classifier trained on the Kermany et al. [10] pediatric pneumonia

Table 6 Performance comparison of different machine learning classifiers with the stacking classifier with values rounded off to the nearest two decimal positions

Classifier	Accuracy	Precision	Recall	F1-score	AUC
Logistic regression	98.13	98.83	98.60	98.71	97.73
Support vector classifier	98.13	99.29	98.13	98.71	98.12
Nu- Support vector classifier	97.44	97.47	99.07	98.26	96.07
K-Nearest classifier	98.13	99.29	98.13	98.71	98.12
MLP classifier	97.10	98.81	97.20	98.00	97.03
Gaussian naïve bayes	95.91	96.12	98.36	97.23	93.84
Bernoulli NB	94.89	98.77	94.16	96.41	95.50
Gradient boosting classifier	94.72	97.37	95.33	96.34	94.20
XGB classifier	96.59	99.28	96.03	97.62	97.07
Decision Tree classifier	94.72	97.37	95.33	96.34	94.20
Random forest classifier	96.08	96.13	98.60	97.35	93.95
Extra Trees classifier	96.76	97.01	98.60	98.80	95.21
Bagging classifier	98.13	98.60	98.83	98.72	97.53
AdaBoost classifier	95.06	97.84	95.33	96.57	94.83
LGB classifier	97.10	98.58	97.43	98.00	96.83
CatBoost classifier	97.96	99.29	97.90	98.59	98.00
HistGradient boosting classifier	96.08	99.27	95.33	97.26	96.72
Proposed method	98.30	99.29	98.36	98.83	98.24

Table 7 Performance comparison with and without PCA with values rounded off to the nearest two decimal positions

Method	Accuracy	Precision	Recall	F1-score	AUC
Stacking classifier in the absence of PCA	97.79	99.05	97.90	98.47	97.69
Stacking classifier with PCA	98.30	99.29	98.36	98.83	98.24

Table 8 Fine-tuning information and the number of trainable parameters associated with each model used in our study

Classifier	Hyperparameters
RandomForest	n_estimators = 100, criterion = 'gini', min_samples_split = 2, min_samples_leaf = 1, min_weight_fraction_leaf = 0.0, min_impurity_decrease = 0.0, ccp_alpha = 0.0
Support vector	C = 1.0, kernel = 'poly', degree = 3, gamma = 'scale', coef0 = 0.0, tol = 1e-3
Nu-Support vector	kernel = 'rbf', degree = 1, gamma = 'scale', probability = True, nu = 0.25, tol = 1e-3
K-Neighbors	n_neighbors = 5, weights = 'uniform', leaf_size = 30, p = 2
XGB	loss = 'deviance', learning_rate = 0.1, n_estimators = 100, subsample = 1.0, criterion = 'friedman_mse', min_samples_split = 2, min_samples_leaf = 1, max_depth = 3, min_weight_fraction_leaf = 0.0
Logistic regression	penalty = 'l2', tol = 1e-4, C = 1.0, solver = 'lbfg', max_iter = 100
MLP	Hidden_layer_sizes = (50,10,10,10), activation = 'tanh', solver = 'adam'

dataset was tested on other pneumonia datasets [55, 56]. The confusion matrix of the predictions made on the test data on the two pneumonia datasets is shown in Figs. 18 and 19, respectively. The misclassifications in the first [55] and second [56] datasets are 25 and 31 false positives (normal predicted as pneumonia), respectively. The proposed method shows null false negatives in both unseen datasets. Tables 10 and 11 discuss the classification report for the corresponding datasets [55, 56]. The proposed method achieves an accuracy of 88% on the unseen test dataset [55] with 100 images belonging to normal and pneumonia classes each as shown in Table 10. The model's reliability is supported by the precision of 100%, recall of 75% for the normal class, and precision of 80%, and recall

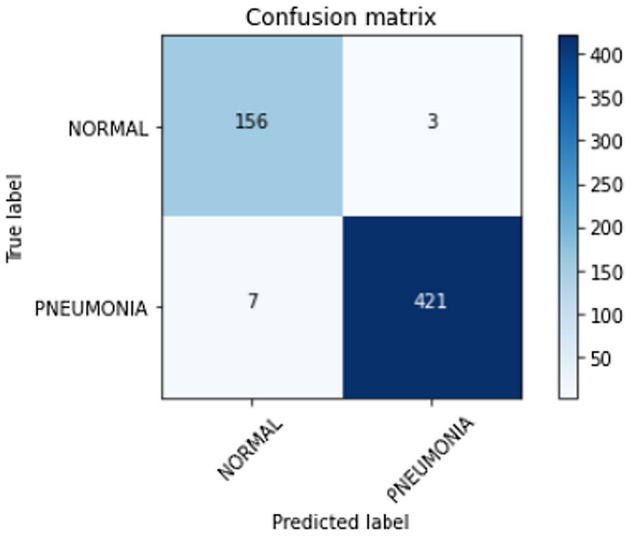

Fig. 15 Confusion matrix for predictions made on the test dataset using the stacked classifier with kernel PCA

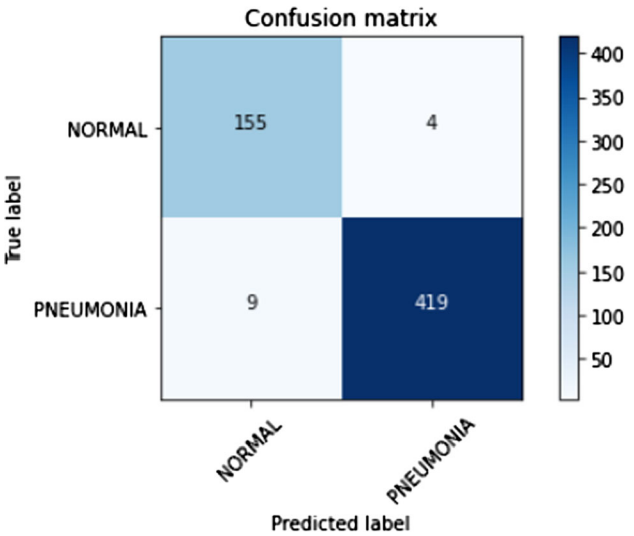

Fig. 16 Confusion matrix for predictions made on the test dataset using the stacked classifier without kernel PCA

of 100% correct prediction for the pneumonia class. In unseen dataset [56], the proposed method achieves an accuracy of 95% supported by 234 X-rays belonging to class normal and 390 X-rays belonging to class pneumonia as shown in Table 11. The model's reliability is supported by the precision of 100%, recall of 87% for the normal class, and precision of 93%, and recall of 100% correct prediction for the pneumonia class. The weighted and macro averages differ by a small margin because of the class imbalance but are limited within 93−96%.

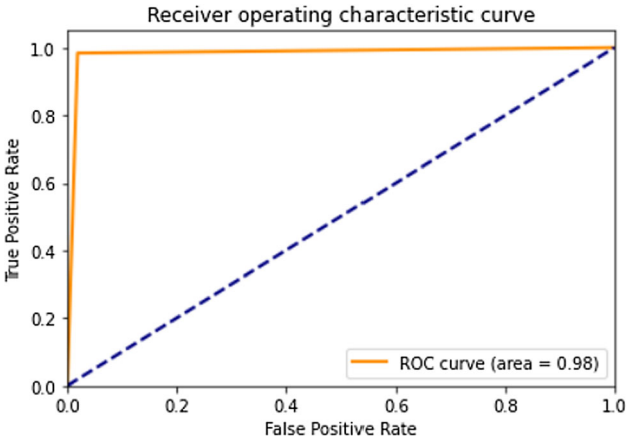

Fig. 17 ROC curve for predictions made on the test dataset using the stacked classifier

The results conclude that though the challenges of pediatric pneumonia diagnosis are characteristically different from adult pneumonia, the proposed method can be extended to aid with the diagnosis of adult pneumonia.

8 Conclusion and future work

In this work, we propose a computer-aided diagnosis tool for pneumonia detection in infants using chest X-rays. Pediatric pneumonia is one of the substantial causes of the increasing death toll among children. Lower radiation levels in chest X-rays for children make detection a cumbersome and time-consuming task. Other works in the same field include using novel architectures and an ensemble of deep CNN models with the added advantage of using an augmented dataset to increase the number of samples in each category. Our work uses the existing deep CNN models for feature extraction; visualized using t-SNE feature representations and class activation maps, followed by Kernel PCA for dimensionality reduction. The reduced features advance into the stacking classifier for the final normal or pneumonia classification. Redistribution of the dataset instead of added augmentations to ensure unbiased training was the initial dominant factor for reliable performance. Our work uses transfer learning on pre-trained models to compensate for the availability of a limited dataset and introduces data augmentations to prevent overfitting. The Xception model achieves the highest accuracy and is used as the feature extractor. The advantage of Xception for this task in specific has been studied in detail along with the addition of PCA on the performance of the classification model. Dimensionality reduction is used to eliminate the redundant features. A stacking

Table 9 Performance of other recent works on the Kermany et al. [10] dataset with values rounded off to the nearest two decimal positions

Authors	Classes	Technique	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)
Kermany et al. [10]	Normal and Pneumonia	Inception V3 pretrained CNN model	92.8	90.1	93.2	–
Nahida et al. [27]	Normal and Pneumonia	Two-channel CNN model	97.92	98.38	97.47	97.97
Stephen et al. [30]	Normal and Pneumonia	Custom CNN model without Transfer Learning	93.73	–	–	–
Chouhan et al. [14]	Normal and Pneumonia	Majority voting ensemble model	96.39	93.28	99.62	99.34
Rajaraman et al. [47]	Normal and Pneumonia	Custom VGG-16 model	96.2	97.0	99.5	99.0
Siddiqi et al. [19]	Normal and Pneumonia	Deep sequential CNN model	94.39	92.0	99.0	–
Hashmi et al. [48]	Normal and Pneumonia	Weighted classifier	98.43	–	–	99.76
Yu Xiang et al. [33]	Normal and Pneumonia	CGNET	98.72	97.48	99.15	–
El Asnaoui et al. [22]	Normal and Pneumonia	Deep CNN model	96.27	98.06	94.61	–
Saraiva et al. [16]	Normal and Pneumonia	MLP and NN approach	92.16	–	–	–
Saraiva et al. [17]	Normal and Pneumonia	Custom CNN	95.30	–	–	–
Mittal et al. [34]	Normal and Pneumonia	CapsNet architecture	96.36	–	–	–
Rahman et al. [21]	Normal and Pneumonia	Deep CNN model	98.0	97.0	99.0	98.0
Sagar Kora Venu et al. [5]	Normal and Pneumonia	Weighted average ensemble model	98.46	98.38	99.53	99.60
Toğaçar et al. [49]	Normal and Pneumonia	Deep CNN model	96.84	96.88	96.83	96.80
Nahida et al. [25]	Normal and Pneumonia	SMOTE on ensembled features from VGG-19 and CheXNet	98.90	–	–	99.00
Islam et al. [28]	Normal and Pneumonia	Feature concatenations with ANN	98.99	99.18	98.90	–
Proposed Work	Normal and Pneumonia	Stacking classifier based on features extracted from Xception	98.3	99.29	98.36	98.24

classifier covering nearly all machine learning models and neural networks was employed. Stacking classifier with Stratified K-Fold cross-validation results in an accuracy of 98.3%. The proposed approach was tested other pneumonia datasets to validate the performance across unseen data for generalization.

As for future work, we would like to explore the effects of spatial domain data pre-processing techniques like Histogram Equalization (HE), Local Histogram Equalization (LHE), and Contrast Limited Adaptive Histogram Equalization (CLAHE) for the task of pediatric pneumonia detection. Reinforcement Learning-based hyperparameter tuning is another potential area of research. In the t-SNE plot (Fig. 13), we notice a few outliers and feature overlap between normal and pneumonia chest X-rays. This visualization pinpoints a potentially better model for better classification results. Custom CNN architectures with fewer parameters specific to occlusion-based categorization can be employed. The introduction of augmentations for training might help the model perform much better and reduce the current misclassification rate of 1.7%. In addition to that, we would like to explore simple yet powerful

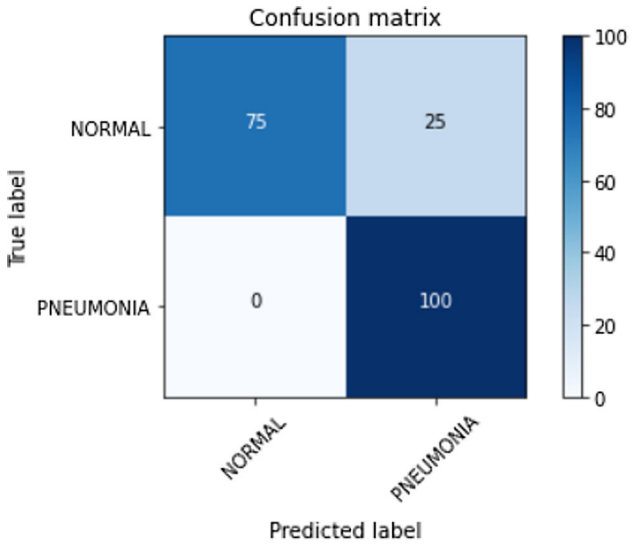

Fig. 18 Confusion matrix for predictions made on the test dataset of normal vs pneumonia classification dataset [55]

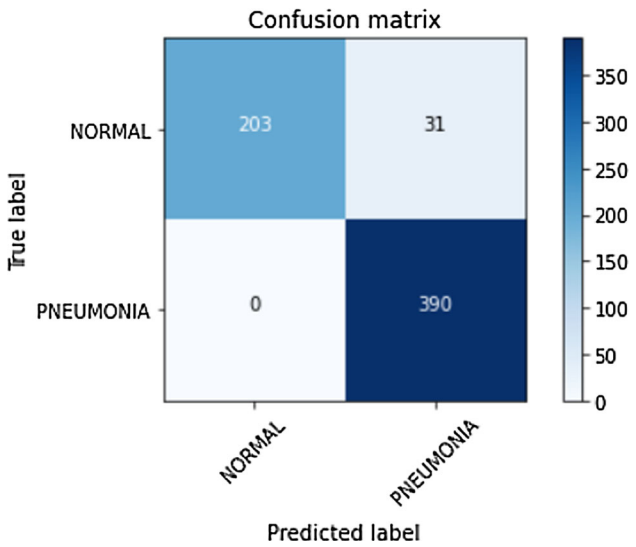

Fig. 19 Confusion matrix for predictions made on the test dataset of normal vs pneumonia classification dataset [56]

Table 10 Classification report on the test data of normal vs pneumonia classification dataset [55]

	Precision	Recall	F1-score	Support
Normal	1.00	0.75	0.86	100
Pneumonia	0.80	1.00	0.89	100
Accuracy			0.88	200
Macro avg	0.90	0.88	0.87	200
Weighted avg	0.90	0.88	0.87	200

feature extraction models. CheXNet [15] was set as a benchmark for this study for it has reached the diagnostic level of human radiologists. With our work performing better CheXNet [15], it will be of immense help to all physicians and radiologists for accurate diagnosing in a matter of seconds. This early detection will help reduce the mortality rate of children suffering from pneumonia.

Table 11 Classification report on the test data of normal vs pneumonia classification dataset [56]

	Precision	Recall	F1-score	Support
Normal	1.00	0.87	0.93	234
Pneumonia	0.93	1.00	0.96	390
Accuracy			0.95	624
Macro avg	0.96	0.93	0.95	624
Weighted avg	0.95	0.95	0.95	624

Funding Not applicable.

Availability of data and material The data that support the findings of this study are available from the first author upon reasonable request.

Code availability The code is available from the first author upon reasonable request.

Declarations

Conflicts of interest The authors declare no conflict of interest.

Compliance with ethical standards None.

Informed consent None.

References

1. Neupane B et al. (2010) Long-term exposure to ambient air pollution and risk of hospitalization with community-acquired pneumonia in older adults.“ American journal of respiratory and critical care medicine 181(1):47–53
2. Ramezani M, Aemmi SZ, Moghadam ZE (2015) Factors affecting the rate of pediatric pneumonia in developing countries: a review and literature study. Int J Pediatrics 3(6.2):1173–1181
3. Lee GE et al. (2010) National hospitalization trends for pediatric pneumonia and associated complications. Pediatrics 126(2):204–213
4. Dean P, Florin TA (2018) Factors associated with pneumonia severity in children: a systematic review. J Pediatric Infect Dis Soc 7(4):323–334
5. Rahman MM et al (2021) Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features. Irbm 42(4):215–226
6. Cherradi B et al. (2021) Computer-aided diagnosis system for early prediction of atherosclerosis using machine learning and K-fold cross-validation. In: 2021 International congress of advanced technology and engineering (ICOTEN). IEEE
7. Qin ZZ et al. (2021) Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of

five artificial intelligence algorithms. Lancet Digital Health 3(9):e543-e554
8. Kundaram SS, Ketki CP (2021) Deep learning-based alzheimer disease detection. In: Proceedings of the fourth international conference on microelectronics, computing and communication systems. Springer, Singapore
9. Perdomo O et al. (2019) Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography. Comput Methods Prog Biomed 178: 181–189
10. Kermany DS et al. (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122–1131
11. Liang G, Zheng L (2020) A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Comput Methods Prog Biomed 187:104964
12. Habib N, Hasan MM, Rahman MM (2020) Fusion of deep convolutional neural network with PCA and logistic regression for diagnosis of pediatric pneumonia on chest X-Rays. Network Biol 76
13. 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 (2020): arXiv-2011
14. Chouhan V et al. (2020) A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci 10(2):559
15. Rajpurkar P et al. (2017) Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint* arXiv:1711.05225
16. Saraiva AA et al. (2019) Models of learning to classify x-ray Images for the detection of pneumonia using neural networks. Bioimaging
17. Saraiva AA et al. (2019) Classification of images of childhood pneumonia using convolutional neural networks.Bioimaging
18. Akgundogdu A (2021) Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest. Int J Imaging Syst Technol 31(1):82–93
19. Siddiqi R (2019) Automated pneumonia diagnosis using a customized sequential convolutional neural network. In: Proceedings of the 2019 3rd international conference on deep learning technologies
20. Siddiqi R (2020) Efficient pediatric pneumonia diagnosis using depthwise separable convolutions. SN Comput Sci 1(6):1–15
21. Rahman T et al. (2020) Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Appl Sci 10(9):3233
22. El Asnaoui K, Chawki Y, Idri A (2021) Automated methods for detection and classification pneumonia based on x-ray images using deep learning. Artificial intelligence and blockchain for future cybersecurity applications. Springer, Cham, pp 257–284
23. Rahman T et al. (2021) Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput Biol Medicine 132:104319
24. Rubini C, Pavithra N (2019) Contrast enhancement of MRI images using AHE and CLAHE techniques. Int J Innov Technol Explor Eng 9(2):2442–2445
25. Habib N et al. (2020) Ensemble of CheXNet and VGG-19 feature extractor with random forest classifier for pediatric pneumonia detection. SN Comput Sci 1(6):1–9
26. Luján-García JE et al. (2020) A transfer learning method for pneumonia classification and visualization. Appl Sci 10(8):2908
27. Nahid A et al. (2020) A novel method to identify pneumonia through analyzing chest radiographs employing a multichannel convolutional neural network. Sensors 20(12):3482
28. Islam KT et al. (2020) A deep transfer learning framework for pneumonia detection from chest X-ray images. VISIGRAPP (5: VISAPP)
29. Mahajan S et al. (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). IEEE
30. Stephen O et al. (2019) An efficient deep learning approach to pneumonia classification in healthcare. J Healthcare Eng
31. Manickam A et al. (2021) Automated pneumonia detection on chest X-ray images: a deep learning approach with different optimizers and transfer learning architectures. Measurement 184:109953
32. Nguyen H et al. (2020) Explanation of the convolutional neural network classifying chest X-ray images supporting pneumonia diagnosis. EAI Endors Trans Context Aware Syst Appl 7(21)
33. Yu X, Wang S-H, Zhang Y-D (2021) CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia. Inf Process Manage 58(1):102411
34. Mittal A et al. (2020) Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray images. Sensors 20(4):1068
35. Wu H et al. (2020) Predict pneumonia with chest X-ray images based on convolutional deep neural learning networks. J Intell Fuzzy Syst 39(3):2893–2907
36. Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
37. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
38. Howard AG et al. (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861
39. Szegedy C et al. (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence
40. Huang G et al. (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
41. Szegedy C et al. (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition
42. He K et al. (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition
43. He K et al. (2016) Identity mappings in deep residual networks. In: European conference on computer vision. Springer, Cham
44. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition
45. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 374(2065):20150202
46. Ezukwoke K, Zareian SJ (2019) Kernel methods for principal component analysis (PCA) A comparative study of classical and kernel PCA. A preprint
47. Rajaraman S et al. (2018) Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. Appl Sci 8(10):1715
48. Hashmi MF et al. (2020) Efficient pneumonia detection in chest xray images using deep transfer learning. Diagnostics 10(6):417
49. Toğaçar M et al (2020) A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. Irbm 41(4):212–222
50. Howard A et al. (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision

51. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR
52. Chowdhury MEH et al (2020) Can AI help in screening viral and COVID-19 pneumonia? IEEE Access 8:132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287
53. Yang X et al. (2020) COVID-CT-dataset: a CT scan dataset about COVID-19. arXiv preprint arXiv:2003.13865
54. Nafi'iyah N, Setyati E (2021) Lung X-ray image enhancement to identify pneumonia with CNN. In: 2021 3rd East Indonesia conference on computer and information technology (EIConCIT). IEEE
55. https://www.kaggle.com/c/detecting-pneumonia-using-cnn-in-pytorch/data
56. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
57. Zhou B et al. (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition
58. Van der Maaten L, Hinton, G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(11)