

Ensemble of Deep Capsule Neural Networks: An Application to Pneumonia Prediction

Jyostna Bodapati (✉ jyostna.bodapati82@gmail.com)

Vignan University <https://orcid.org/0000-0002-5185-882X>

Rohith V N

Vignan's Foundation for Science Technology and Research

Venkatesulu Dondeti

Vignan's Foundation for Science Technology and Research

Research Article

Keywords: Capsule layers, Capsule networks, Dynamic routing, Transfer Learning, Pneumonia Prediction

Posted Date: January 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1038081/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Ensemble of Deep Capsule Neural Networks: An Application to Pneumonia Prediction

Jyostna Devi Bodapati · VN Rohith · Venkatesulu Dondeti

Received: date / Accepted: date

Abstract Pneumonia is the primary cause of death in children under the age of 5 years. Faster and more accurate laboratory testing aids in the prescription of appropriate treatment for children suspected of having pneumonia, lowering mortality. In this work, we implement a deep neural network model to efficiently evaluate pediatric pneumonia from chest radio graph images. Our network uses a combination of convolutional and capsule layers to capture abstract details as well as low level hidden features from the the radio graphic images, allowing the model to generate more generic predictions. Furthermore, we combine several capsule networks by stacking them together and connected them with dense layers. The joint model is trained as a single model using joint loss and the weights of the capsule layers are updated using the dynamic routing algorithm. The proposed model is evaluated using benchmark pneumonia dataset[14], and the outcomes of our experimental studies indicate that the capsules employed in the network enhance the learning of disease level features that are essential in diagnosing pneumonia. According to our comparison studies, the proposed model with Convolution base from InceptionV3 attached with Capsule

layers at the end surpasses several existing models by achieving an accuracy of 94.84%. The proposed model is superior in terms of various performance measures such as accuracy and recall, and is well suited to real-time pediatric pneumonia diagnosis, substituting manual chest radiography examination.

Keywords Capsule layers · Capsule networks · Dynamic routing · Transfer Learning · Pneumonia Prediction

1 Introduction

Pneumonia is a type of respiratory infection caused by bacterial infection [16]. With the world under the grip of COVID-19, pneumonia has once again taken centre stage due to symptoms similar to COVID-19 [10]. According to the World health Organization (WHO) statistics, pneumonia accounts for 15% of all deaths in children under the age of five. Majority of these deaths are avoidable if the disease is appropriately diagnosed [22]. The most widely used method for diagnosing pneumonia is manual chest radio graph analysis, but it is impractical and less reliable on a large scale [9]. Hence, there is a strong need for Computer Aided Diagnosis (CAD) tools that aid in the diagnosis of pneumonia.

In recent years, research community shifted their focus on building fully automated tools for “reading chest radio graphs” and this has been an active research area. Following the advent of deep learning, numerous Convolutional Neural Network (CNN) models have been developed and successfully used for diagnosing pneumonia from radio graph images [13]. According to the empirical studies in the literature, adding more convolutional layers to the model improves the precision of

Jyostna Devi Bodapati
Vignan’s Foundation for Science Technology and Research,
Guntur 522213, India
E-mail: jyostna.bodapati82@gmail.com

VN Rohith
Vignan’s Foundation for Science Technology and Research,
Guntur, India
E-mail: rohithkarthikeya48851@gmail.com

Venkatesulu Dondeti
Vignan’s Foundation for Science Technology and Research,
Guntur 522213, India
E-mail: hodcse@vignan.ac.in

Table 1 Recent Neural network Models on chest Radio graphs

| Ref | Description | Scores |
|--------------------------------|--|-------------------------------------|
| Ozturk[18] | Customised Darknet Model | Avg Recall:85.35% |
| Ayan [4] Muhammad Irfan[12] | Uses Pretrained VGG16 model with ImageNet weights DNN with CNN and LSTMS layers | Accuracy: 87.00%. Recall: 88.10% |
| Sharma [23] | Uses two convolutional Neural Networks; One for feature extraction; Other for classification of pneumonia | Accuracy: 90.68% |
| Narin[17] | Transfer learning on pretrained models | Recall: 92.10% |
| Kermani [14] | Follows transfer learning approach; Fine-tuning of InceptionV3; Critical pre-processing approaches | Accuracy: 93.20 |
| Stephen [24] | Customised CNN from scratch | Accuracy: 93.73% |
| Samir[25] | Fine-tuned VGG16 with dropouts | Accuracy: 93.80% |
| Savaria [22] | Convolutional neural network model developed from scratch; Trained using gray-scale converted, normalized and re-scaled chest images | Accuracy: 95.07% |
| Ibrahim [11] RajPurkar[19] | Uses AlexNet as a reference model 121 layered custom CNN Model | Accuracy: 94.15% F1-Score:95.12% |
| Altan[3] | 2d curvlet images On EfficientNet-B0 | Recall:95.68 |
| Chowdhary [7] | ChexNet | Recall:96.61% |
| Almalki[2] | Transfer learning with IRV2 | Recall : 97.02% |

these CNN models [22]. Deep pre-trained CNN models, such as VGG16, VGG19, ResNet50, and Inception-v3, have been shown to be effective in solving variety of real-world applications. These pre-trained models are extensively used and have shown considerable gains in pneumonia recognition from chest radio graph images [13]. According to empirical studies in the literature, fine-tuning the pre-trained models leads to faster convergence and higher scores than constructing models from scratch [8]. Further research shows that the performance of deep learning models such as Mobile-Net and VGG-16 outperform conventional machine learning models such as Logistic regression, Support vector machines (SVMs), K-Nearest Neighborhood (KNN), Random Forest, XGBoost and Cat-Boost [5]. A deep Convolutional Neural Network (CNN) architecture with 10 layers [22] was developed to detect pneumonia in children under the age of five, which was trained using gray-scale converted, normalized and re-scaled images. Majority of the existing models rely on Convolutional Neural Networks, which set benchmark performance for the diagnosis of diseases such as pneumonia from images [1]. Convolution operations used in CNNs extract higher level features from input images, followed by

sub-sampling operations which aids in increasing network depth. Table 7 outlines few more remarkable neural network models for pneumonia prediction from the literature.

CNNs are strong in identifying a hierarchy of spatial features from images they are not very effective at investigating the spatial relationships between those features and the same has been pointed out by Geoffrey Hinton [20]. On the other side, for medical imaging applications, such low-level details can not be disregarded since they are likely to play a vital role in disease prediction. Addressing these issues, we present deep capsule network models for predicting pneumonia from radio graph images of the newborns. The proposed deep learning architecture combines the features of CNNs and deep capsule networks, captures disease details from the chest images including low level features. We further stacked up multiple such capsule networks and an ensemble model is implemented to create a much robust model. Empirical studies on bench mark datasets [14], demonstrate that the proposed model is superior than existing models, improving pneumonia prediction accuracy from 92.80% to 95.46%. The model is also evaluated in terms of various performance measures such as pre-

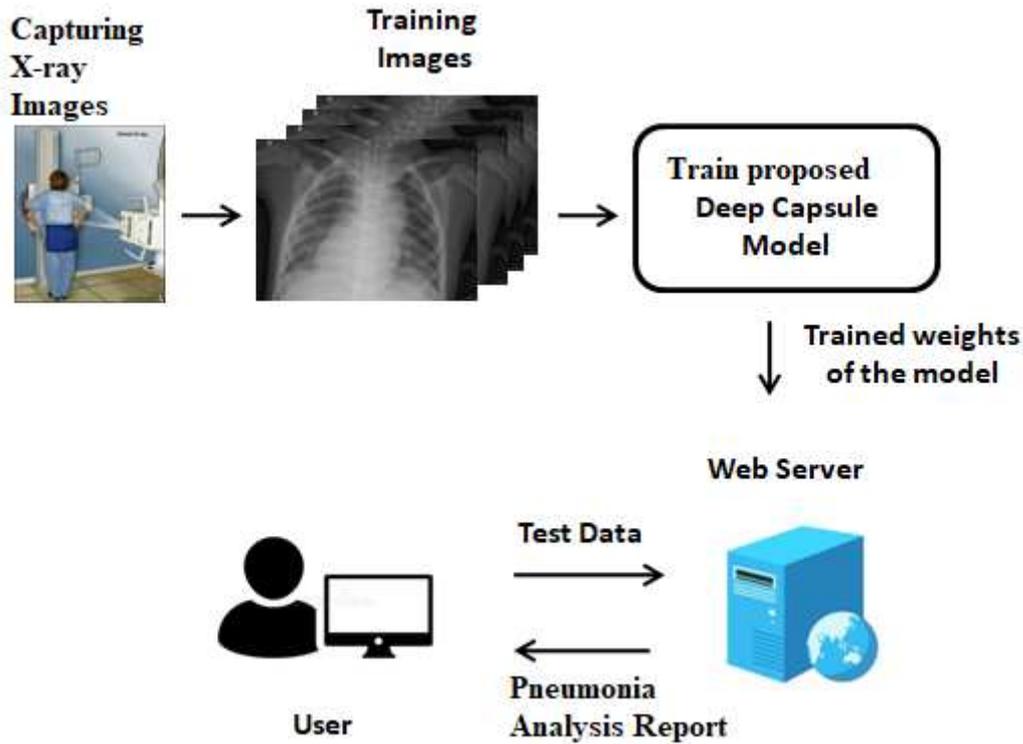


Fig. 1 Demonstration of the Pneumonia prediction system in Healthcare industry

recision, recall, F1-score, and AUC-ROC, with improvements noted. Figure 1 depicts the abstract view of the pneumonia prediction system in the medical field and figure 4 depicts the detailed architecture of the proposed system.

2 Methodology

This section gives a summary of the dataset used along with comprehensive details of the proposed model.

2.1 Dataset

For the experimental studies, we use a pediatric pneumonia benchmark dataset developed by Kermany [14]. The dataset consists of 5863 Chest radio graph images in JPEG format from the Normal and Pneumonia classes. All of the images in the dataset were obtained from the Guangzhou Medical centre, where the radiographic image impressions were captured from children aged 1 to 5 years old suffering from pneumonia. Figure 2 shows sample radio graph images from either classes. The dataset comprises a training set with 5216 images, a validation set with 16 images, and a test set with 640 images.

| Class | Train set | Validation set | Test Set |
|-----------|-----------|----------------|----------|
| Normal | 1073 | 268 | 242 |
| Pneumonia | 3100 | 775 | 398 |
| Total | 4173 | 1043 | 640 |

Table 2 Summary of the Dataset

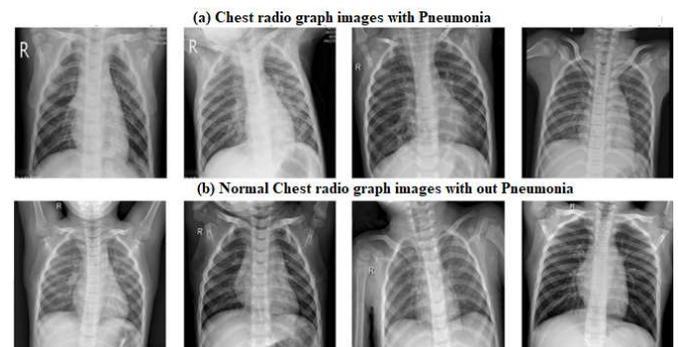


Fig. 2 Sample chest radio graph images from the Kermani dataset (a) Pneumonia affected (b) Healthy (Normal)

The validation set of the set comprises only 16 images, which is insufficient for model hyper parameter tuning, we consider 20% of the images from the actual training dataset to validate the model. Table 2 showcases summary of the dataset. Class imbalance present in the dataset is visualized using a doughnut plot in figure 3,

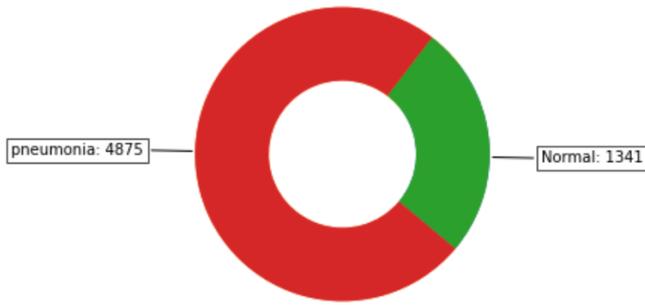


Fig. 3 doughnut chart depicting class wise samples of the Kermani dataset

depicts the number of normal chest radio graph images and Pneumonia infected images.

2.2 Model

The primary objective of this study is to develop a high-precision model for automated diagnosis of diseases from medical images, with an emphasis on detecting infant pneumonia from chest radio graph images. Realizing our set objective, we introduce ChxEnsCapsNet, a deep learning framework based on capsule networks for the effective identification of pediatric pneumonia from chest radio graph images. The complete block diagram of the proposed work is depicted in figure 4. The proposed ChxEnsCapsNet consists of five different blocks, which are extensively discussed in the following sections: data augmentation and preprocessing block, Spatial feature extraction, Convolution base, Capsule base and disease Prediction block.

2.3 Data augmentation and preprocessing

Medical image datasets are notoriously unbalanced, with extremely few examples in certain classes compared to the rest of the classes. It is critical to address the class imbalance issues and if not addressed may lead to a bias in the model favoring the majority class. We handle this problem by applying variety of augmentation approaches such as zoom, shear, horizontal flipping to generate synthetic images and the model is trained using these extended dataset which ultimately helps to reduce the problem of over-fitting. In addition, all the images are resized to bring them to uniform shape before they are passed to the models. Each image has been re-scaled such that every pixel value of the image lies between 0 and 1, this a common practice followed for neural network training and this ultimately helps the model to converge faster.

2.4 Spatial Feature Representation Learning from Chest radio graph images

Feature representation of the images is a key to the performance of a machine learning model. Deep neural networks are successful as they capture hierarchical representation of such features from the raw images provided at the input layer. Convolutional Neural Networks (CNN) are specifically capable of summarizing the presence of patterns and aggregates the presence of discovered patterns by applying a sequence of Convolution operations followed by Pooling operations on the input image [6]. The filters used in the convolution operations of CNNs are called as kernels, and are automatically learned during the training process. In this work, we propose to leverage the power of deep pre-trained models such as VGG16, Xception, MobileNet, Inception ResNetV2, Inception3 for feature extraction.

The advantage of using pre-trained models is two fold: first it avoids the training of deep networks from scratch and the second benefit is this brings the power of the deep networks for the extraction of feature representations with spatial information from radio graph images. Uniqueness of this module is unlike the existing models the features are extracted from the convolution layers of the deep pre-trained models. Feature descriptors extracted from the fully connected layers lose the local feature information while the convolution base preserves the local feature information from radio graph images which ultimately helps in identifying pediatric pneumonia recognition.

2.5 Capsule Networks

The spatial feature representation module produces a set of feature maps, where each feature map represents certain spatial information about the input radio graph images. These spatial representations obtained from the convolutional layers of deep pre-trained models are passed to the subsequent Capsule Layers to extract much low-level details pertaining to the pneumonia disease.

Neurons are the basic elements of conventional neural network models, where the input to a neuron is a scalar and the output of a neuron is another scalar. This restricts the ability of CNN to learn the direction of the captured objects and therefore CNNs are not robust to transformation and rotation. On the other hand Capsule Networks (CapsNETs) consists of a series of capsule layers and are invariant to transformation and rotation by preserving the spatial information throughout the network. Each capsule layer consists of multiple

capsules. The core unit of these CapsNets is a capsule, which is characterized as a collection of neurons that determine the likelihood of whether or not particular objects (entities) are present in a picture, as well as their orientation. Unlike traditional neurons, the input to a capsule is a vector, u_i , and by using a non-linear squash function it results in a vector v_j as output. The norm of the pose vector encodes the probability that an object of interest is present, while its direction represents the object's pose information, such as location, size, and orientation. Each capsule performs several complex operations such as affine transformation, weighted sum, squashing function on the input and encapsulates the result into a small vector of highly informative output. Let u_i^l be the input vector from a low level capsule, i at layer l and, v_j^{l+1} be the output vector at higher level capsule, j at layer $l + 1$. The following affine transformation takes place to encode the lower level features into a higher level abstract representation:

$$\hat{u}_{j|i}^{l+1} = W_{ij}^l u_i^l \quad (1)$$

Where W_{ij}^l is the weight matrix and learns the spatial relationships between the input features during the model training. A weighted sum operation is applied on the resultant vector, $\hat{u}_{j|i}$, where the lower level capsule will send its input to the higher level capsule that "agrees" with its input. The output of Capsule j , at layer $l + 1$ is denoted by s_j^{l+1} , is a weighted sum of all the incoming predictions $\hat{u}_{j|i}$:

$$s_j^{l+1} = \sum_i c_{ij} \hat{u}_{j|i}^{l+1} \quad (2)$$

where c_{ij} are the coupling coefficients such that $\sum_i c_{ij} = 1$ and $c_{ij} \geq 0 \forall j$. These coefficients are determined by the routing by agreement procedure or the dynamic routing algorithm.

At the end, a squash function, is applied on the output vector s_j^{l+1} to get non-linearity and to squash the values between 0 and 1.

$$v_j^{l+1} = \frac{\|s_j^{l+1}\|^2}{1 + \|s_j^{l+1}\|^2} \frac{s_j^{l+1}}{\|s_j^{l+1}\|^2} \quad (3)$$

In equation 3, the first term on the right hand side of the equation does the scaling while the second term normalizes the data.

For every capsule i , at layer l :

$$c_{ij} = \frac{\exp(b^{ij})}{1 + \exp(b^{ij})} \quad (4)$$

where b_{ij} s are the logit values that indicate whether capsules i at layer l and the layer j at layer $l + 1$ have

strong coupling. In other words, it is a measure of how much the presence of the capsule j is explained by the capsule i . Initially, all b_{ij} should be equal. Over the iterations, update b such that their sum of weighted inputs will be squashed.

2.6 Deep Capsule Networks for the Detection of Pneumonia in Chest radio graphs

Observing the representation power CNNs, the constraints of sub-sampling layers of CNNs and capabilities of capsule networks, we propose a deep neural network framework that leverages the benefits from both these models.

The proposed model comprises of convolution base along with capsule layers at the end of the network. The convolution base of the proposed framework is taken from the deep pre-trained models like VGG16, Inception ResNetV2, InceptionV3, MobileNet and Xception. The layers close to the output layers of these models are removed, the remaining layers are frozen, and the weights from the ImageNet-trained models are utilized for these layers. These frozen layers serve as the convolution base of the model to get the spatial representations of the chest radio graph images. Then capsule layers are added on top of the convolution base to get the low-level pneumonia specific details from the spatial representations of radio graph images. Finally, at the end of the network, fully connected layers are attached to identify the presence of pneumonia when an input chest radio graph images is fed to the network. These models are then trained in an end-to-end manner on the chest radio graph images for the identification of Pneumonia.

Each of the pre-trained model produces spatial representation of the radio graph images in different sizes. The number of capsules and dimensions of the capsules used in the model are decided according to the dimensions of the spatial representations produced by the deep pre-trained models.

Each radio graph image passes through a sequence of non-linear layers where the first set of non-linearity comes by using convolution base of the deep pre-trained models and helps to grab the spatial representations from the radio graph images while the second set of non-linearity comes by using the capsule layers which in turn helps to capture the pneumonia disease specific details. The same can be represented in mathematical terms using the below equations:

$$\hat{x} = g(h(x)) \quad (5)$$

where x is the input, $h(\cdot)$ and $g(\cdot)$ refers to the non-linear mapping of the input values while \hat{x} is the non-linear representation of the input x .

The conventional deep neural network models that uses convolution base alone are most likely to loose the low-level details from the input images that are most critical for the disease detection. The significance of the proposed model is that it attempts to derive a model that uses capsule layers which is capable of obtaining low level details from the chest images. Instead of just relying on convolutional layers, the proposed model is designed with a right mix of convolutional and capsule layers and is trained in end-to-end manner targeting to achieve superior performance over existing models.

2.7 Stochastic Ensemble of Deep Capsule Networks

We use ensemble of capsule networks to further improve the performance of the models that we have developed. It is a known fact that ensemble of multiple machine learning models results in better performance compared to the performance of any of the single model involved. This is because of the reason that the ensemble of multiple models gets the benefit from all the individual capsule networks and helps in boosting the model performance. Commonly used approaches to neural networks ensemble include concatenation, average, weighted average. In case of concatenation ensemble, the outputs from the models involved in ensemble process are simply concatenated to form a single tensor while in average ensemble approach the outputs are averaged to produce a single tensor. Weighted ensemble approach is the most successful approach where each model outputs are multiplied with weights and then the output is formed as a linear combination of the weighted values. Each capsule network is attached with a softmax layer to produce the probability of the type of disease. Let us consider $y_i = \{y_{i1}, y_{i2}, \dots, y_{iC}\}$ are the scores produced by the softmax layer of the i^{th} capsule network where the model is designed to predict C different diseases. Each of the ensemble approaches can be mathematically expressed as:

$$y = \sum_{i=1}^k y_i = \sum_{i=1}^k y_{i1}, y_{i2}, \dots, y_{iC} \quad (6)$$

$$y = \prod_{i=1}^k y_i = \prod_{i=1}^k y_{i1}, y_{i2}, \dots, y_{iC} \quad (7)$$

$$y = \sum_{i=1}^k \lambda_i y_i \quad \text{where} \quad \sum_{i=1}^k \lambda_i = 1 \quad (8)$$

In equations 6, 7 and 9, y refers to a tensor that is a fused representations of the outputs from multiple capsule networks.

Limitation of these ensemble approaches is, which type of approach is suitable for the given application has to be decided empirically. Addressing this limitation we propose to use a stochastic ensemble approach which enables the model to decide on the type of fusion approach to be used. This model is termed as a stochastic model as it is capable of learning the appropriate weights suitable for fusing the output scores of the capsule networks. We design this stochastic ensemble approach using a fully connected neural network model and the computation can be represented as:

$$y = \sum_{i=1}^k \lambda_i y_i \quad \text{where} \quad \sum_{i=1}^k \lambda_i = \{-\infty, +\infty\} \quad (9)$$

In equations 6, 7 and 9, y refers to a tensor that is a fused representations of the outputs from multiple capsule networks.

Another benefit of using neural network layers for stochastic ensemble is these layers can be attached to the outputs layers of the capsule networks.

Thus the stochastic ensemble receives disease prediction scores and produces the type of disease with high precision. We call the proposed ensemble model as ChxEnsCapsNet, an ensemble of stacked capsule networks for detection of pneumonia from chest X-ray images and the complete architecture of the proposed ChxEnsCapsNet is shown in Figure 4.

3 Model training

All the experiments are conducted in the Google Colaboratory with the Keras and TensorFlow libraries. All of the baseline models, as well as the proposed model, are trained using data from the train split of the dataset, and the results are reported for the test split.

We empirically determined the values of hyper parameters by observing the model performance while altering the parameters within defined ranges. The learning rate is set to 0.0001 for all models, while the weight and bias regularization rates are set to 0.001 to 0.1. Stabilizing the performance of the models in all in all the experiments, the models are allowed to run for 200 epochs. The model loss is computed by taking into account the cross-entropy loss between the predictions and the

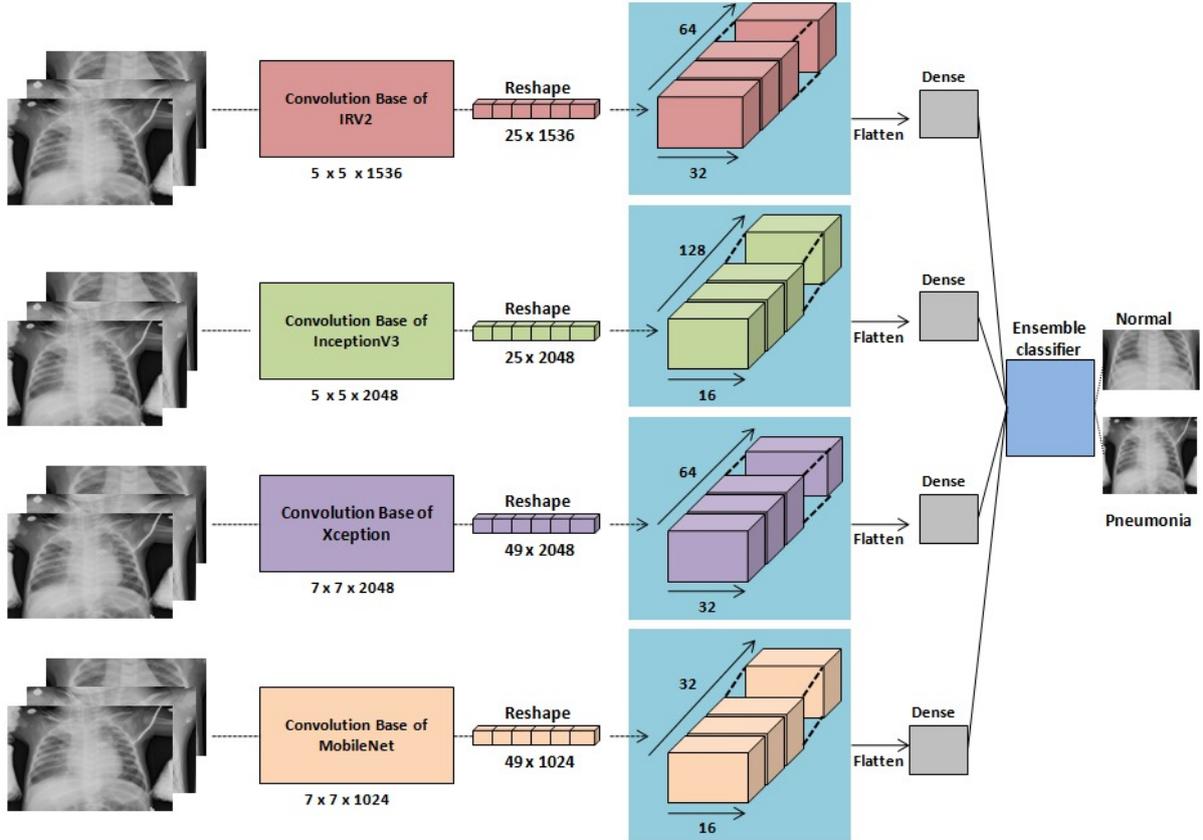


Fig. 4 Architecture of Proposed ChxEnsCapsNet: Transfer learning based Deep Capsule Networks for detection of Pneumonia in Chest radio graph images

ground truths, and the model parameters are optimized using ADAM optimizer. The model loss is evaluated during training, and if there is no change in loss after 5 epochs, the training is terminated since the patience value is set to 5. We use the checkpoint method to select the model with the best validation split performance.

3.1 Evaluation and statistical Analysis

Different classification metrics such as Accuracy, Precision, Recall and F₁ Score are used to determine the efficiency of the proposed model for the detection of pneumonia from lung images. Accuracy is defined as the proportion of properly recognized images to the total number of images analyzed. Precision shows how many of the total images projected as pneumonia are really influenced by pneumonia. Recall indicates how many of the actual pneumonia images are predicted as pneumonia. The harmonic mean of Precision and recall is used to get the F₁ Score and following mathematical

expression is used to compute:

$$F1_{Score} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (10)$$

Furthermore, we report measures such as AUC (Area under the curve) of the ROC (Receiver operating characteristic) popularly known as AUC-ROC, which is especially useful when working with datasets with class imbalance. ROC is a probability curve and is plotted with True Positive Rate (TPR) against the False Positive Rate (FPR). AUC of the ROC aids in quantifying how well the proposed model performs at various threshold levels. The greater the AUC, the better the model distinguishes radio graph images between the pneumonia and normal [15]. Following are the mathematical expressions for TPR and FPR measures:

$$TPR = \frac{TP}{TP + FN} \quad (11)$$

$$FPR = 1 - Sensitivity = \frac{FP}{TN + FP} \quad (12)$$

In the above expressions TP and TN refers to True Positive and True Negative cases similarly FP and FN

Table 3 Performance comparison of deep pre-trained CNN models for pediatric Pneumonia Prediction

| Model | Accuracy | Precision | Recall | F1.Score | AUC |
|--------------------|----------|-----------|--------|----------|-------|
| Inception ResNetV2 | 86.87 | 86.87 | 95.22 | 90.02 | 84.18 |
| InceptionV3 | 86.71 | 86.71 | 79.64 | 88.17 | 88.99 |
| VGG16 | 90.62 | 90.62 | 97.48 | 92.82 | 88.41 |
| MobileNet | 90.93 | 90.93 | 99.49 | 93.17 | 88.17 |
| Xception | 91.25 | 91.25 | 98.49 | 93.33 | 88.91 |

refers to the total number of False Positive and False Negative cases.

4 Results

This section provides the experimental outcomes of the baselines as well as the proposed model.

4.1 Transfer Learning approaches for pediatric Pneumonia Prediction

The chest radio graph images of the dataset are used to fine-tune deep pre-trained models such as VGG16, Xception, MobileNet, Inception ResNetV2, Inceptionv3, and these models serve as the baselines for the proposed model. The convolutional base of these deep pre-trained models is loaded after removing the final softmax layer. This convolution base of the deep networks produces spatial representations corresponding to the the input radio graph images. A neural classification model is attached to the convolution base in order to confirm the presence or absence of pneumonia in the radio images. The neural classification model is designed with an optional Global Average Pooling (GAP) layer followed by a sigmoid activated dense layer. The weights of the convolutional layers are frozen and are loaded with ImageNet weights and the dense layers of the model are fine-tuned using the chest radio graph images from the dataset. This neural classification head predicts whether pneumonia is present or not in the given radio graph image. Prior to transferring the chest radio graph images to deep pre-trained models, they are passed through a pre-processed module where feature-wise normalization takes place and are further reshaped so that they are accepted by the pre-trained models.

Table 3 compares the effectiveness of various baseline models for pediatric pneumonia prediction using chest radio graph images. According to the table, fine-tuning Xception network results in an accuracy of 91.25%, placing it on top among other models in terms of performance. Similar trends may be noticed in the other scores as well. This might be due to the depth-wise convolutions used in the Xception network, which allow the

model to learn hidden spatial details from the input radio graph images. We can even observe that VGG16 and MobileNet outperform InceptionV3 and IRV2 in terms of performance.

4.2 Capsule Networks for Pneumonia Prediction

In this experiment, we evaluate the efficiency of capsule networks in detecting pneumonia from chest radio graph images. We hypothesize that the capsule layers in the capsule networks have the potential to extract minute details from the input radio graph images. Initially, spatial representations from the radio graph images are extracted by passing them to the deep pre-trained models after removing the final dense layers. Then we attach capsule layers on to the convolution base and then we attach dense layers for classification of the disease. Attaching the capsule layer directly on the convolution base is not feasible and hence we reshape the spatial tensor such that we can add capsule layers on top of that. The weights of the convolutional layers are frozen and are loaded with ImageNet weights and the capsule layers along with dense layers of the model are fine-tuned using the chest radio graph images from the dataset.

For instance, if the convolution base of VGG16 is used, then from its final convolution layer we receive a tensor of size $7 \times 7 \times 512$, which is the spatial representation of the input image. This 3-D tensor is then reshaped into 49×512 representation and then a capsule layer is attached to get the tiny details about the disease. The primary capsule layer contains single capsule with each of dimension 512. This primary layer produces a 49×512 output. This is flattened and passed through a 512 neuron dense layer with ReLU activation for non-linearity and finally a sigmoid neuron is attached at the end for pneumonia classification. Similarly Capsule Networks are developed for each of the pre-trained models. The hyper-parameters used for these models are same as those used with baseline models.

Table 4 compares the effectiveness of capsule networks for pneumonia detection. From table 4, we can see that

Table 4 Performance of deep Capsule Networks for pediatric Pneumonia Prediction

| Model | Accuracy | Precision | Recall | F1-Score | AUC |
|------------------------|----------|-----------|--------|----------|-------|
| VGG16Caps | 91.06 | 91.09 | 90.70 | 92.68 | 91.21 |
| XceptionCaps | 92.03 | 92.03 | 95.54 | 93.30 | 90.90 |
| MobileNetCaps | 93.75 | 93.75 | 97.23 | 95.08 | 92.62 |
| Inception ResNetV2Caps | 94.06 | 94.06 | 93.21 | 95.12 | 94.33 |
| InceptionV3Caps | 94.84 | 94.84 | 97.73 | 95.93 | 93.91 |

Table 5 Performance analysis of Ensemble stacked baseline neural network Models

| Model | Accuracy | Precision | recall | F1 Score | AUC |
|----------------------------------|----------|-----------|--------|----------|-------|
| Baseline Classifier ^a | 91.87 | 90.04 | 97.73 | 93.73 | 89.9 |
| Baseline Classifier ^b | 92.81 | 92.71 | 95.97 | 94.32 | 91.79 |
| Baseline Classifier ^c | 92.96 | 91.14 | 98.24 | 94.55 | 91.26 |
| Baseline Classifier ^d | 93.28 | 93.39 | 95.97 | 94.67 | 92.4 |
| Baseline Classifier ^e | 93.59 | 92.80 | 97.23 | 94.96 | 92.41 |

the network with convolution base of InceptionV3 (InceptionV3Caps) outperforms other models when compared to the rest of the models. When compared to the baseline models, shown in table 3, networks with capsule layers offer superior performance. The convolution base of the models learn spatial representations from radio graph images while the pneumonia disease specific features that are concealed in the spatial representations are learned by the capsules utilized in these networks and helps in improving the performance. Among all the models InceptionV3Caps network achieves better accuracy and when compared to VGG16Caps, XceptionCaps, MobileNetCaps, and Inception ResNetV2 Caps.

4.3 Ensemble Stacked Deep Neural Networks for Pneumonia Prediction

From the previous two sets of experiments, we may deduce that capsule layers used in the capsule networks aid in extracting minute details from chest X-rays and helps in exact detection of pediatric pneumonia. Ensemble of multiple models is commonly used to boost up the classification performance of these individual classification models. We hypothesize that ensemble of our individual models also leads to offer superior performance. Unlike the conventional ensemble approaches like max voting, bagging or boosting, our ensemble approach is quiet different. We extract the predicted probabilities from individual models and pass them to a multi layer perceptron (MLP). Instead of training these two models in isolation, we join these two models, the individual classification models and the MLP model, and are trained in an end to end manner. Table.5 represents classification performance of ensemble classifier by varying the baseline models. In the first experiment

we ensemble the baseline models and then later we ensemble the capsule networks.

Table 5, shows the performance obtained by the models that are the ensemble of baseline Classifiers. Instead of blindly stacking all the baseline CNN models, we experimented with different combinations. Table 5 shows five different combinations of baseline models. The model named Baseline Classifier^a refers to the ensemble of VGG16, Xception, MobileNet, Inception ResNetV2, InceptionV3. Baseline Classifier^b is the ensemble of InceptionV3, Xception, MobileNet, Inception ResNetV2. Baseline Classifier^c is ensemble of Xception, Inception V3, MobileNet. Baseline Classifier^d is ensemble of VGG16, Inception ResNetV2, Xception. Baseline classifier ^e is the ensemble of VGG16, Inception V3, Xception, MobileNet. Baseline Classifier^e achieves an accuracy of 93.59% which is superior to other models. When we compare the results in 5 with those of 3, we can understand that the ensemble models significantly improve the accuracy compared to those of the single individual models. Baseline Classifier^a achieves an accuracy of 91.87%, while Baseline Classifier^b, Classifier^c and Classifier^d achieve accuracy of 92.81%, 92.96% and 93.28% respectively. In the next experiment we ensemble the capsule network models developed for pneumonia classification. The approach to ensemble is similar to those of the baseline models. Classifier^a is the ensemble of XceptionCaps, InceptionV3Caps, Inception ResNetV2Caps, and VGG16Caps. Classifier^b is the ensemble of XceptionCaps, MobileNetCaps, VGG16Caps, InceptionV3Caps, and Inception ResNetV2Caps. Classifier^c is the ensemble of MobileNetCaps, Inception ResNetV2Caps, and InceptionV3Caps. Classifier^d is the ensemble of XceptionCaps, MobileNetCaps, and Inception ResNetV2Caps. Classifier^e refers to the ensemble of XceptionCaps, Inception ResNetV2Caps, InceptionV3Caps, and MobileNet-

Table 6 Performance analysis of Ensemble stacked Capsule neural network Models

| Model | Accuracy | Precision | recall | F1 Score | AUC |
|-------------------------------|----------|-----------|--------|----------|-------|
| <i>Classifier^a</i> | 94.58 | 96.67 | 94.72 | 95.55 | 94.50 |
| <i>Classifier^b</i> | 94.84 | 95.28 | 96.48 | 95.88 | 94.33 |
| <i>Classifier^c</i> | 95.00 | 95.20 | 96.48 | 96.0 | 94.52 |
| <i>Classifier^d</i> | 95.15 | 95.53 | 96.73 | 96.12 | 94.84 |
| <i>Classifier^e</i> | 95.46 | 96.47 | 96.23 | 96.25 | 95.22 |

Table 7 Comparison study of the proposed ChxEnsCapsNet with state-of-the-art models for Pneumonia Classification

| Model | Accuracy | Precision | recall | F1 Score | AUC |
|-----------------------------|--------------|--------------|--------------|--------------|-------|
| Kermany[14] | 92.8 | – | 93.2 | – | 96.8 |
| Ayan[4] | 87.0 | – | 82 | – | – |
| Sharma[23] | 90.68 | – | – | – | – |
| Sahlol[21] | 94.18 | – | 87.22 | – | – |
| Ibrahim[11] | 94.15 | – | – | – | – |
| Samir S Yadav[25] | 93.8 | 94.4 | 93.8 | – | – |
| O Stephen[24] | 93.73 | – | – | – | – |
| Saraiva[22] | 95.07 | 95.72 | 94.85 | 95.3 | – |
| ChxEnsCapsNet (Ours) | 95.46 | 96.47 | 96.23 | 96.25 | 95.22 |

Caps. The performance obtained by each of these ensemble approaches is reported in table 6.

From table 6, we can observe the performance of models that follow ensemble of stacked capsule networks. These ensemble models Classifier^a, Classifier^b, Classifier^c, Classifier^d and Classifier^e achieve accuracy of 94.58%, 94.84%, 95.00%, 95.15% and 95.46% respectively. When we compare the results in 6 with those in table3, table5, and table 4, we can infer that capsule networks are superior compared to those of conventional CNN models. We can even observe that ensemble of models significantly helps in improving the performance of the models and are most suitable for pneumonia detection. Especially the recall of these capsule networks is stable compared to those of the CNN models. The Classifier^e, which is an ensemble of stacking XceptionCaps, Inception ResNetV2Caps, Inceptionv3Caps, and MobileNetCaps is superior in terms of accuracy, precision, recall, F_1 score and AUC. We call this model as ChxEnsCapsNet and we claim that this model is more appropriate for pediatric pneumonia detection.

5 Discussion

From the previous set of experiments and their outcomes, we understand that capsule networks are superior to the deep CNN models as they are capable of extracting the minute details from chest X-ray images of kids. In order to validate the model, we compare our proposed ChxEnsCapsNet model with the baselines. We also compare our ensemble model performance with numerous state-of-the-art models available in the liter-

ature on the benchmark kermany dataset [14]. Most of the previous researches rely on the concept of transfer learning for detection of pneumonia from chest x-rays. This is the first time with the concept of transfer learning we fine-tuned Pretrained Neural Networks by adding capsule layers at the end of the deep networks.

In table 7, we reported the performance of various state-of-the-art deep neural network models and compared the performance of our proposed ChxEnsCapsNet. The model by Kermany [14] which follows transfer learning approach that uses fine-tuning of InceptionV3 achieves 92.8% accuracy. The pre-processing approached they have used is much critical and helped in achieving that accuracy. Ayan [4] used pretrained VGG16 model with ImageNet weights for classification of pneumonia and achieves an accuracy of 87.00%. Sharma [23] used two convolutional Neural Networks, one for feature extraction and the other for classification of pneumonia in chest x-rays and his classification model produced an accuracy of 90.68%. Ibrahim [11] worked for classification of pneumonia by using AlexNet as a reference model and achieved 94.15%. Stephen [24] developed a customised CNN from scratch which was able to achieve an accuracy of 93.73%. Samir[25] fine-tuned VGG16 with dropouts which was able to achieve an accuracy of 93.80%, which was significantly better compared to that obtained by the standard VGG16 model. A Convolutional neural network model developed from scratch [22] and trained using gray-scale converted, normalized and re-scaled chest images achieved an accuracy of 95.07%. Our proposed Ensemble Capsule network is superior in

terms of accuracy and various other metrics when compared to other models. Though the accuracy achieved by [22] is almost close to ours, their recall is inferior to the proposed ChxEnsCapsNet model. With this we claim that the proposed ChxEnsCapsNet model is more robust compared to other models and much appropriate for pediatric pneumonia detection.

6 Conclusion

In this work, we introduced deep capsule neural network model for effective prediction of pediatric pneumonia from chest radio graph images. The spatial representations from the pre-trained CNN models capture multiple levels of abstract details from chest radio graph images while they fail to capture the low level hidden details. We develop networks with a combination of convolutional and capsule layers that are suitable for capturing high level as well as disease level details from the spatial representations that are critical for the prediction of pneumonia. These capsule layers used in the model allows it to learn low-level features that were ignored by convolutional layers and allows the model to produce generic predictions. From the outcomes of our experimental studies on benchmark Kermany [14] dataset, we claim that InceptionV3 as a convolutional base along with capsules our model, designated as ChxEnsCapsNet, is superior to the vanilla convolutional models. Our experimental results reveal that the proposed model is capable of predicting pneumonia in children with improved accuracy, recall and AUC-RoC, and is well suited to real-time pediatric pneumonia diagnosis, substituting manual chest radiography examination.

Limitation of the proposed model is, if the model is trained from scratch with out applying transfer learning approach, then the model does not behave as intended due to lack of sufficient data. In future, this model can be extended to include attention mechanism along with capsule layers which is believed to give distinguished focus to the regions critical to the discrimination of disease. Further we are interested in designing a composite model with an encoder comprising convolutional layers and the other encoder consisting of Capsule layers and train the joint model with a single loss function for pneumonia prediction. We are interested to check the performance of these models in a normal patient population where other respiratory conditions would be present.

References

1. Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N Plataniotis, and Arash Mohammadi. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *Pattern Recognition Letters*, 138:638–643, 2020.
2. Yassir Edrees Almalki, Abdul Qayyum, Muhammad Irfan, Noman Haider, Adam Glowacz, Fahad Mohammed Alshehri, Sharifa K Alduraibi, Khalaf Alshamrani, Mohammad Abd Alkhalik Basha, Alaa Alduraibi, et al. A novel method for covid-19 diagnosis using artificial intelligence in chest x-ray images. In *Healthcare*, volume 9, page 522. Multidisciplinary Digital Publishing Institute, 2021.
3. Aytac Altan and Seçkin Karasu. Recognition of covid-19 disease from x-ray images by hybrid model consisting of 2d curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, Solitons & Fractals*, 140:110071, 2020.
4. Enes Ayan and Halil Murat Ünver. Diagnosis of pneumonia from chest x-ray images using deep learning. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, pages 1–5. Ieee, 2019.
5. Tathagat Banerjee, Dhruv Batta, and Aditya Jain. Deep belief convolutional neural network with artificial image creation by gans based diagnosis of pneumonia in radiological samples of the pectoralis major. Technical report, EasyChair, 2020.
6. Jyostna Devi Bodapati, Nagur Shareef Shaik, and Veeranjaneyulu Naralasetti. Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–15, 2021.
7. Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al Emadi, et al. Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8:132665–132676, 2020.
8. Venkatesulu Dondeti, Jyostna Devi Bodapati, Shaik Nagur Shareef, and Naralasetti Veeranjaneyulu. Deep convolution features in non-linear embedding space for fundus image classification. *Rev. d’Intelligence Artif.*, 34(3):307–313, 2020.
9. Zhichao Feng, Qizhi Yu, Shanhu Yao, Lei Luo, Wenming Zhou, Xiaowen Mao, Jennifer Li, Junhong Duan, Zhimin Yan, Min Yang, et al. Early prediction of disease progression in covid-19 pneumonia patients with chest ct and clinical characteristics. *Nature communications*, 11(1):1–9, 2020.
10. Peter M George, Shaney L Barratt, Robin Condliffe, Sujal R Desai, Anand Devaraj, Ian Forrest, Michael A Gibbons, Nicholas Hart, R Gisli Jenkins, Danny F McAuley, et al. Respiratory follow-up of patients with covid-19 pneumonia. *Thorax*, 75(11):1009–1016, 2020.
11. Abdullahi Umar Ibrahim, Mehmet Ozsoz, Sertan Serte, Fadi Al-Turjman, and Polycarp Shizawaliyi Yakoi. Pneumonia classification using deep learning from chest x-ray images during covid-19. *Cognitive Computation*, pages 1–13, 2021.
12. Muhammad Irfan, Muhammad Aksam Iftikhar, Sana Yasin, Umar Draz, Tariq Ali, Shafiq Hussain, Sarah Bukhari, Abdullah Saeed Alwadie, Saifur Rahman,

- Adam Glowacz, et al. Role of hybrid deep neural networks (hdnns), computed tomography, and chest x-rays for the detection of covid-19. *International Journal of Environmental Research and Public Health*, 18(6):3056, 2021.
13. Rachna Jain, Preeti Nagrath, Gaurav Kataria, V Sirish Kaushik, and D Jude Hemanth. Pneumonia detection in chest x-ray images using convolutional neural networks and transfer learning. *Measurement*, 165:108046, 2020.
 14. Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131, 2018.
 15. Jayawant N Mandrekar. Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9):1315–1316, 2010.
 16. Veeranjanyulu Naralasetti, Reshmi Khadherbhi Shaik, Gayatri Katepalli, and Jyostna Devi Bodapati. Deep learning models for pneumonia identification and classification based on x-ray images. *Traitement du Signal*, 38(3), 2021.
 17. Ali Narin, Ceren Kaya, and Ziyet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, pages 1–14, 2021.
 18. Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U Rajendra Acharya. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in biology and medicine*, 121:103792, 2020.
 19. Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.
 20. Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. *arXiv preprint arXiv:1710.09829*, 2017.
 21. Ahmed T Sahlol, Dalia Yousri, Ahmed A Ewees, Mohammed AA Al-Qaness, Robertas Damasevicius, and Mohamed Abd Elaziz. Covid-19 image classification using deep features and fractional-order marine predators algorithm. *Scientific Reports*, 10(1):1–15, 2020.
 22. Arata Andrade Saraiva, Nuno M Fonseca Ferreira, Luciano Lopes de Sousa, Nator Junior C Costa, José Vignó M Sousa, DBS Santos, Antonio Valente, and Salviano Soares. Classification of images of childhood pneumonia using convolutional neural networks. In *BIOIMAGING*, pages 112–119, 2019.
 23. Harsh Sharma, Jai Sethia Jain, Priti Bansal, and Sumit Gupta. Feature extraction and classification of chest x-ray images using cnn to detect pneumonia. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 227–231. IEEE, 2020.
 24. Okeke Stephen, Mangal Sain, Uchenna Joseph Maduh, and Do-Un Jeong. An efficient deep learning approach to pneumonia classification in healthcare. *Journal of health-care engineering*, 2019, 2019.
 25. Samir S Yadav and Shivajirao M Jadhav. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6(1):1–18, 2019.