International Journal of Data Science and Analysis
2024; Vol. 10, No. 1, pp. 11-19
https://doi.org/10.11648/j.ijdsa.20241001.12

Research Article

Deep Convolutional Neural Networks with Augmentation for Chest X-Ray Classification

Hannah Kariuki*, Samuel Mwalili, Anthony Waititu

Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Abstract

The recent release of large amounts of Chest radiographs (CXR) has prompted the research of automated analysis of Chest X-rays to improve health care services. DCNNs are well suited for image classification because they can learn to extract features from images that are relevant to the task at hand. However, class imbalance is a common problem in chest X-ray imaging, where the number of samples for some disease category is much lower than the number of samples in other categories. This can occur as a result of rarity of some diseases being studied or the fact that only a subset of patients with a particular disease may undergo imaging. Class imbalance can make it difficult for Deep Convolutional Neural networks (DCNNs) to learn and make accurate predictions on the minority classes. Obtaining more data for minority groups is not feasible in medical research. Therefore, there is a need for a suitable method that can address class imbalance. To address class imbalance in DCNNs, this study proposes, Deep Convolutional Neural Networks with Augmentation. The results show that data augmentation can be applied to imbalanced dataset to increase the representation of the minority class by generating new images that are a slight variation of the original CXR images. This study further evaluates identifiability and consistency of the proposed model.

Keywords

Deep Convolution Neural Networks, Data Imbalance, Data Augmentation, Chest X-Ray Imaging

1. Introduction

Chest X-ray (CXR) is the initial diagnostic imaging tool for thoracic abnormalities such as pneumonia, tuberculosis, and lung cancer in the clinical setting and remains central to screening and diagnosis. The demand for CXR images is due to their reasonable sensitivity to various pathologies, low radiation dose, and cost-effectiveness [1]. However, the volume of CXR images acquired is higher compared to the number of qualified radiologists.

Shortage of qualified radiologists, complexity of interpreting the CXR images, and, their value in clinical practice has motivated researchers to build automated algorithms for Chest X-Ray analysis. Automated CXR analysis has been shown to help control variability among radiologists, increase sensitivity for subtle findings, provision of analysis where radiologists are not available, and, automation of tedious daily tasks [1].

Deep Learning (DL) methods have widely been adopted for the analysis of images and Computer Vision tasks such as object segmentation, detection, and classification. Compared to other conventional machine learning methods, deep learning can perform feature extraction eliminating the need for human-aided feature extraction which makes DL methods

*Corresponding author: hkariuki059@gmail.com (Hannah Kariuki)

Received: 23 October 2023; **Accepted:** 6 March 2024; **Published:** 19 March 2024

well-suited for medical image analysis. Deep Convolutional Neural Networks (DCNNs), deep learning technique have the ability to automatically learn features through several network layers from large labelled datasets. In medical image analysis, DCNNs have been successfully utilized for various tasks such as skin lesion classification [2], suspicious region detection [3] classification of diabetic retinopathy on fundoscopic images [4], and automatic classification of breast cancer histopathological images [5].

Deep Learning-based solutions serve as a second opinion tool in the medical field in diagnosis and proper treatment planning. However, medical image datasets suffer class imbalance. This makes it difficult for researchers to develop models that are unbiased and accurate [6]. Training a DCNN model on imbalanced data regularly results in overfitting of the majority class [7]. This is because they aim to optimize the overall accuracy without considering the relative distribution of each class. To mitigate this problem, this work considers data augmentation where the goal is to increase the representation of the minority classes in training data to completely eliminate the imbalance.

2. Literature Review

2.1. Deep Convolution Neural Networks DCNN

LeCun et al. introduced Convolutional Neural Networks (CNNs), a typical deep neural network and applied it for handwritten zip code digit recognition and has since been successfully used for many classification, recognition and segmentation tasks [8]. The powerful learning ability of Deep Convolutional Neural Networks (DCNNs) is primarily attributed to the use of multiple feature extraction stages that can automatically learn representations from the data [9].

According to Zeiler and Fergus, the availability of large amounts of labeled data, and implementation of GPUs has made the training of DCNNs possible, and recently more DCNNs architectures have been explored [10].

With CNN becoming the state-of-the-art method in the computer vision field, advancements in CNNs have been explored to improve on the CNN architecture originally introduced by [11] to achieve better accuracy.

According to Khan et al., the significant improvement in the representational capacity of the DCNNs has been achieved through architectural innovations [9]. Sermanet et al. utilized smaller strides, (2 instead of 4) in the first convolutional layer for the best-performing submissions to the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC2013) [12]. Simonyan and Zisserman used very small receptive fields (3 × 3) with a stride of 1 throughout the whole net compared to the previous studies where (11 × 11) with 4 strides and (7 ×7) with 2 strides were used [13].

Szegedy et al. proposed a DCNN architecture (Inception) in which the depth and width of the network was increased at the same time keeping the computational cost constant and lead to a 22-layer deep model in the case of the GoogLeNet model [14]. Rajpurkar et al. developed CheXNeXt, a convolutional neural network with 121-layer DenseNet architecture to detect the presence of 14 different pathologies in frontal-view chest radiographs [15]. The algorithm was shown to perform as well as the radiologists for 10 pathologies and performed better than the radiologists on 1 pathology. However, it performed poorest in the detection of emphysema and hiatal hernia.

In medical image analysis, one of the biggest challenge faced by researchers is shortage of labeled image datasets. To counter the data shortage, researchers are prompted to use transfer learning. Transfer learning is based on the concept of reusability. The weight and bias of a pretrained model are transferred to a new model for training or testing. Transfer learning is often implemented with CNN in the way that all or some layers of the pretrained model are kept except the last layer, which is trained for the specific problem.

Han proposed a DCNN model with 27 convolutional layers alternated with 7 pooling and un-pooling layers. The model was based on transfer learning by initializing the model weights from a pre-trained model for learning mappings from magnetic resonance imaging (MRI) to their corresponding computerized tomography scan (CTS) for limited data [16].

Asif et al. proposed DCNN based model Inception V3 (Google's CNN architecture) with transfer learning for automatic detection of COVID-19 using chest X-ray radiographs [17]. The transfer learning proved to be effective for the classification of COVID-19 Chest X-ray images to normal, viral pneumonia and COVID-19 classes and achieved a classification accuracy of more than 98%. Chakraborty et al. applied transfer learning approach with pre-trained VGG-19 architecture for the classification of COVID-19, Pneumonia, and Healthy cases from the chest X-ray images and obtained an accuracy of 97% [18].

2.2. Data Imbalance

The problem of data imbalance is prevalent in medical image analysis. Class imbalance occurs when classes in a dataset are not equally distributed. According to Banik and Bhattacharjee, class imbalance occurs when a specific type of disease has a significantly lower number of samples in the training set compared to other disease categories [19]. For instance, in mammography screening, majority of women who undergo the screening do not have breast cancer, and therefore the malignant samples will be less than benign samples [20]. Class imbalance has been shown to affect the disease classification task as it affects both convergence during the training phase as well as generalization of the model on the test set. It also results in a DCNN model with poor predictive performance, specifically for the minority class. This is because most machine learning algorithms assume that the number of examples in considered classes to be roughly similar. However, in many real-life situations the

distribution of examples is skewed since representatives of some classes appear much more frequently than others. This poses a difficulty for learning algorithms, as they will be biased towards the majority group.

Banik and Bhattacharjee observed that a standard classification algorithm can be biased towards the majority class and ignore the importance of the minority class, which can lead to wrong diagnosis [19]. According to Zhang et al., in highly class imbalanced learning, majority classes will have dominant effect during the learning process [21]. This implies that, the classification costs of the majority classes and minority classes are not equal, which leads to the classifier being biased towards the majority class. Buda et al. showed that a CNN for brain tumor segmentation overfits more on the majority class with a higher level of imbalance [22].

Li et al. observed that when training with limited and imbalanced data, the distribution of logit activations may shift across the decision boundary at test time, while samples of the well-represented class seem unaffected [6].

Obtaining more data for the minority groups in medical research is not feasible due to the rarity of diseases, patient privacy, labeling of images by experts, and cost of conducting medical imaging processes. Thus, methods like class weighting, and sampling have been presented in literature to address the problem of class imbalance. However, their effectiveness might vary based on the complexity of the task at hand and the distribution of the dataset. In the class weighting approach, a higher weight is applied to the minority class during training when calculating the loss.

Sampling is one of the main methods used in medical imaging to tackle the problem of class imbalance. Either the classes with minority samples are over-sampled or the classes with majority samples are under-sampled before training to create a balanced dataset. The downside to under-sampling is that some important and informative examples might be removed from the dataset. The oversampling techniques are widely employed in deep learning and have been shown to be reliable.

In a study to address class imbalance in CNNs for small lesion detection, Bria et al., compared various methods for dealing with class imbalance and found that oversampling the minority class (malignant class) to be the most effective approach [23]. Walsh and Tardy compared various techniques for dealing with class imbalance in deep learning classification of breast cancer namely, class weighting, oversampling, under-sampling, and synthetic lesion generation approach [20]. In their study, they noted that oversampling appeared dominant and should be applied to a level that entirely eliminates the class imbalance.

This study considers data augmentation as the over sampling technique applied to minority classes that involves creating new examples that are variations of existing ones in order to eliminate class imbalance.

3. Methodology

Given $X \in \mathbb{R}^{m \times n \times n_c}$ is an imbalanced image dataset with the labels $y \in \mathbb{R}^p$ where m and n represents the image width and height respectively, n_c is the number of channels, p is the number of classes.

3.1. Data Augmentation

Let T(.) be a data augmentation function that takes in an input sample $x_i, i = 1, \ldots, n$ from the minority class and applies a series of transformations. Denoting the augmented dataset as

$$X' = T(x_1), T(x_2), \ldots, T(x_n) \tag{1}$$

The augmentation process is formulated in the following equations:

$$x_n = T_i(x) \tag{2}$$

$$\hat{X} = x_1, \ldots, x_n = \cup_{n=1}^{n} \cup_{j=1}^{i} T_i(x_n) \tag{3}$$

$\hat{X}$ is the expanded training set, which is then employed to train a deep CNN.

3.2. Convolutional Neural Network

A kernel (filter) $k_{u,v}^{p,q}$ is slid over the input image $x_{m,n}$ with a stride of 1 and zero padding projecting the element-wise dot product as the feature maps. Where m, n represent the m^{th}, n^{th} pixels. u, v Pixels of kernel, p is the number of convolution kernel, q is the number of convolution layers, b is the bias.

The feature maps of the convolution layer $C_{m,n}^{p,q}$ are computed by,

$$C_{m,n}^{p,q} = \sum_{m=1}^{m} \sum_{n=1}^{n} x_{i(m-u,n-v)} * K_{u,v}^{p,q} + b^{p,q} \tag{4}$$

The feature maps are passed through a non-linear activation function ϑ for non-linear transformation, where ϑ is the Rectified Linear Unit (ReLU) activation function. The function can be expressed as $f(x) = max(0, x)$

$$C_{m,n}^{p,q} = \vartheta\left(\sum_{m=1}^{m} \sum_{n=1}^{n} x_{i(m-u,n-v)} . K_{u,v}^{p,q} + b^{p,q}\right) \tag{5}$$

Max pooling operation is applied to condense the image. The pooling layer $P_{m,n}^{p,q}$ is developed by taking out the maximum valued pixels m, n in the convolution layers. The pooling layer is calculated by,

$$P_{m,n}^{p,q} = \max\left(C_{m,n}^{p,q}\right) \tag{6}$$

The pooling layer $P^{p,q}$ is concatenated to form a long

vector with the length of $p * q$ and is fed into the fully connected dense layers for the classification, then the vectorized data points a x_i^{l-1} layer is given by,

$$x_i^{l-1} = f(P^{p,q}) \tag{7}$$

This long vector is fed into fully connected dense layers from *l* layer to *L+1*. Where *l* is the first layer, *L* is the last layer and (*L+1*) is the classification layer.

the forward run between the layers are given by,

$$\begin{bmatrix} z_1^l \\ z_2^l \\ \vdots \\ z_i^l \end{bmatrix} = \begin{bmatrix} w_{11}^l & w_{12}^l \dots & w_{1n}^l \\ w_{21}^l & w_{22}^l \dots & w_{2n}^l \\ \vdots & \vdots & \vdots \\ w_{i1}^l & w_{i2}^l \dots & w_{in}^l \end{bmatrix} \begin{bmatrix} x_1^{l-1} \\ x_2^{l-1} \\ \vdots \\ x_i^{l-1} \end{bmatrix} + \begin{bmatrix} b_1^l \\ b_2^l \\ \vdots \\ b_i^l \end{bmatrix} \tag{8}$$

The input values x_i^{l-1} are multiplied by weights w_i and bias values b_i^l are added. The output value of last layer L is given by,

$$x^L = \vartheta((W^L)^T x^L - 1 + b^L) \tag{9}$$

where, $x^L = \vartheta(z^L)$

The output of l^{th} layer is, $x^l = \vartheta((W^l)^T x^{l-1} + b^l) x^l = \vartheta(z^l)$ where ϑ is the logistic activation function. The final output predicted value $\hat{Y}_i^{L+1}$ at $L+1$ layer can be expressed as,

$$\hat{Y}(X;\theta) = \vartheta(W^L, \dots, \vartheta(W^2(\vartheta(W^1 x^1 + b^1) + b^2 \dots + b^L))) \tag{10}$$

The output is the probability $p \in [0,1]$

$$p_i = \frac{e^{x_i}}{\sum_{i=j}^{n} e^{x_j}}$$

The statistical model relating Y to $\hat{Y}(X;\theta)$ is:

$$Y = Y(X;\theta) + \epsilon \tag{11}$$

The performance of the model is computed by the following loss function equation as shown, where $\hat{Y}_i^{L+1}$ is the predicted value and Y_i actual value.

The cross-entropy loss function is given as,

$$J(Y,X;\theta) = -\frac{1}{n}\sum_i^n (Y_i) ln(Y(X_i;\theta)) + (1 - Y_i) ln(1 - Y(X_i;\theta)) \tag{12}$$

Where n is the total number of training samples and θ represents all the parameters. In order to minimize the loss function, L-BFGS is adopted for this study where the learning parameter is updated at every iteration process a until termination criterion is met.

Stopping Rule

1. $\|\theta^{r+1} - \theta^r\| < \epsilon$, for $\epsilon > 0$
2. r > MAX where MAX is pre-specified number of iterations
3. $|J(Y,X;\hat{\theta}^{(r+1)}) - J(Y,X;\hat{\theta}^{(r)})| < \epsilon$ where ε > 0 but small
4. $J(Y,X;\hat{\theta}^{(r)}) < E_{min}$ which is a pre-specified lower bound for the training error

3.3. Statistical Properties of DCNN with Augmentation

3.3.1. Model Identifiability

According to Hwang and Ding, identifiability of parameters is a fundamental problem in neural networks [24]. Consider a DCNN with an activation function ϑ and H hidden nodes. This paper considers different sets of parameters with their corresponding distributions being identical. Therefore, the parameters are not unique. The model weights are represented as follows:

$$\alpha_0 \text{ and } \Phi_i = (\alpha_i, W_i) \text{ for } i = 1, \dots, H \text{ where } W_i = (W_{i0}, \dots, W_{in})$$

To assess model identifiability, consider the following two transformations that leave the input-output map of DCNN with augmentation model unchanged.

(i) The Permutations of $\Phi_i's$:

Consider two hidden nodes, if they are interchanged, say h_f and h_g where f and g denote the node's position, and relabel them as h_g and h_f and also relabel the corresponding weights as α_g and α_f, and W_g and W_f, the output $Y(X;\theta)$ remains unchanged.

(ii) Symmetry of $\vartheta(x)$:

That is:

$$\vartheta(x) = -\vartheta(x) \tag{13}$$

Select a hidden node h_g and negate the weights W_g as well as α_g, then the input-output map does not change. This implies that, $(\alpha_0, \Phi_1, \dots, \Phi_i, \dots, \Phi_H)$ and $(\Phi_0, \Phi_1, \dots, -\Phi_i, \dots, \Phi_H)$ have the same input-output map. These transformations yield 2^H different models with similar input-output map [24].

Taking τ to denote the transformations, similar to Hwang and Ding, each transformation is a composite function of $\tau_1, \dots \tau_H$ [24].

Where

$\tau_1(\alpha_0, \Phi_1, \dots, \Phi_i, \dots, \Phi_H) = (\Phi_0, -\Phi_1, \dots, \Phi_i, \dots, \Phi_H)$

and

$$\tau_i(\Phi_0, \Phi_1, \dots, \Phi_i, \dots, \Phi_L) = (\Phi_0, \Phi_i, \Phi_2, \Phi_{i-1}, \Phi_1, \Phi_{i+1}, \dots, \Phi_L) \tag{14}$$

for $i = 2, \dots, H$

Theorem 1: Assume that the model's output and the activation function ϑ, and further that ϑ is a continuous function satisfying condition A of [24]. Suppose that the parameter θ is irreducible. Then, θ is identifiable up to the family of transformations generated by equation (14).

3.3.2. Model Consistency

Consider the loss function (12), let $J_0(\theta) = E[J(\theta)]$ denote the expectation of $J(\theta)$. $J(\theta)$ is minimized at true parameter value θ_0.

The consistency of the estimator $\hat{\theta}$ is obtained by minimizing the objective function equation (12). This then means that $\hat{\theta} \to \theta_0$ for $n \to \infty$ in probability.

Rewrite the model as

$$Y_i = p(x_i) + \epsilon_i, i = 1, \ldots, n \tag{15}$$

Where ϵ_i is defined as

$$\epsilon_i = Y_i - p(X_i) \tag{16}$$

And $E[\epsilon_i] = 0$

$$Var[\epsilon_i] = \sigma_\epsilon < \infty \tag{17}$$

Consider the uniform law of large numbers (ULLN).

Theorem 1: Let $S_1, S_2, \ldots$ be independent and identically distributed random vectors in $\mathbb{R}^d$, $\phi \subseteq < M$ compact, $\delta : < \mathbb{R}^d \times \phi \to \mathbb{R}$ measurable such that;

$$E|\delta(S_1; \theta)| < \infty \text{ for all } \theta \in \Phi$$

$\delta(u; \theta)$ is Lipschitz continuous in θ

$$E[L(S_1)] < \infty$$

Then, $\sup_{\theta \in \phi} \left| \frac{1}{n} \sum_{i=1}^{n} \delta(S_i : \theta) - E\delta(S_i; \theta) \right| \to 0$ in probability

Assumptions:

(A_1) The activation function ϑ is bounded and twice continuously differentiable with bounded derivatives [25].

(A_2) $J_0(\theta)$ has a unique global minimum at θ_0 lying in the interior of θ, and with the Hessian

$A(\theta_0) = (\frac{\partial^2}{\partial \theta_k \partial \theta_l} J_0 \theta) = \nabla^2 J_0 \theta$ where $A(\theta_0)$ is positive definite.

(A_3) Let ϕ be such that for some $\Delta > 0$, then, $\Delta \leq Y(X; \theta) \leq 1 - \Delta \; \forall X \in \mathbb{R}^d, \theta \in \phi$

(A_4) $(Y_i, Xi), i = 1, \ldots, n$ be i.i.d with density ζ (x) and $E\|X_1\|^2 < \infty$

(A_5) p(x) is continuous in x and $0 < \eta \leq p(x) \leq 1 - \eta <$ 1 for some η > 0

Theorem 2: Let $(Y_i, X_i), i = 1, \ldots, n$ be i.i.d with $L(Yi|Xi) = B(1, p(Xi))$. Suppose that assumptions A_1 to A_5 are satisfied, then, for $n \to \infty$, with $\hat{\theta}, \theta_0$, as previously shown,

$$\sqrt{n(\hat{\theta} - \theta_0)} \xrightarrow{L} N(0, \Sigma_1 + \Sigma_2)$$

where $\Sigma_1 = A^{-1}(\theta_0) B_1(\theta_0) A^{-1}(\theta_0)$
and $\Sigma_2 = A^{-1}(\theta_0) B_2(\theta_0) A^{-1}(\theta_0)$

$$B_1(\theta_0) = E \frac{(p(X_1) - Y(X_1; \theta_0))^2}{Y^2(X_1; \theta_0)(1 - Y(X_1; \theta_0))^2} \nabla$$

$$Y(X_1; \theta_0) \nabla^T Y(X_1; \theta_0)$$

$$B_2(\theta_0) = E \frac{p(X_1)(1 - p(X_1))}{Y^2(X_1; \theta_0)(1 - Y(X_1; \theta_0))^2} \nabla$$

$$(X_1; \theta_0) \nabla^T Y(X_1; \theta_0)$$

And $A(\theta_0)$ is as defined. The proof of this theorem can be found in [25].

4. Results

4.1. Data

The Chest X-ray Dataset was obtained from [26]. The dataset contains CXR images (anterior-posterior) in JPEG format selected from historic cohorts of pediatric patients between the age of one to five from Guangzhou Women and Children's Medical Center, Guangzhou. The training set has a total of 5,232 chest X-ray images and the test set a total of 624 CXR images characterized as depicting pneumonia (bacterial and viral) and normal (no Pneumonia). The training data was split into 80% training and 20% validation based on Patient's unique ids to prevent data leakage. The distribution of the data in the training, validation and the testing sets is as follows.

Table 1. *Distribution of the dataset.*

Counts			
Class	**Train**	**Validation**	**Test**
Bacterial Pneumonia	2053	486	242
Viral Pneumonia	1083	276	148
No Pneumonia	1071	268	234

From Table 1 the data is imbalanced as majority of the samples have Bacterial pneumonia. Training a DCNN model on the data will result in mis-classification of samples, as most of the samples will be classified as "Bacterial Pneumonia".

Figure 1 shows a sample of the Chest X-ray images in the

dataset.

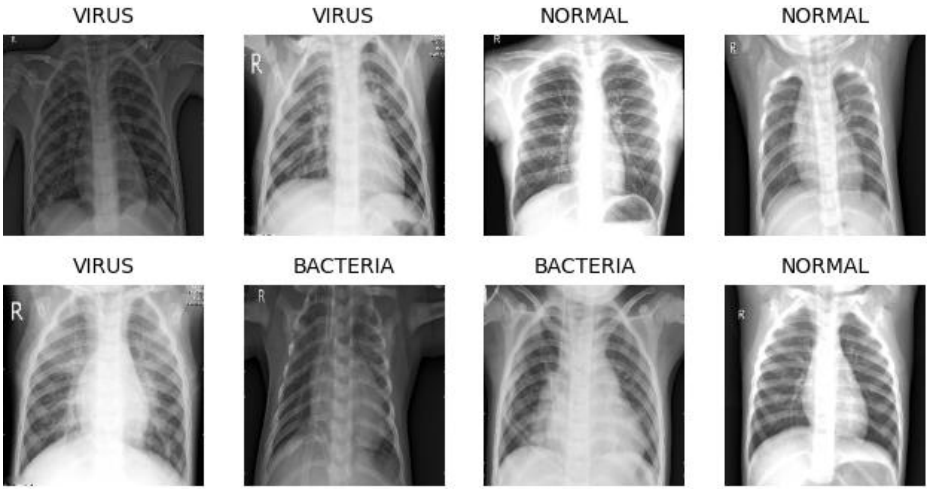

Figure 1. *Sample images.*

4.2. Data Augmentation

Table 2. *Data Augmentation Techniques.*

Technique	Parameter
Horizontal flip	True
Rotation	[-15 °, +15 °]
Height shift	0.1
Zoom	0.1
CLAHE	Clip limit 0.02

Data augmentation was applied to the training set images to oversample the minority classes (No Pneumonia and Viral Pneumonia) to eliminate class imbalance. The testing and the validation sets are not balanced as they represent real world data distribution. For this paper, the following augmentation techniques were applied to the minority classes shown in Table 2.

The minority classes were oversampled to be equal to the majority class. A total of 1952 new CXR images were created for the "No Pneumonia" and "Viral Pneumonia" groups. Figure 2 shows a sample of the augmented images.

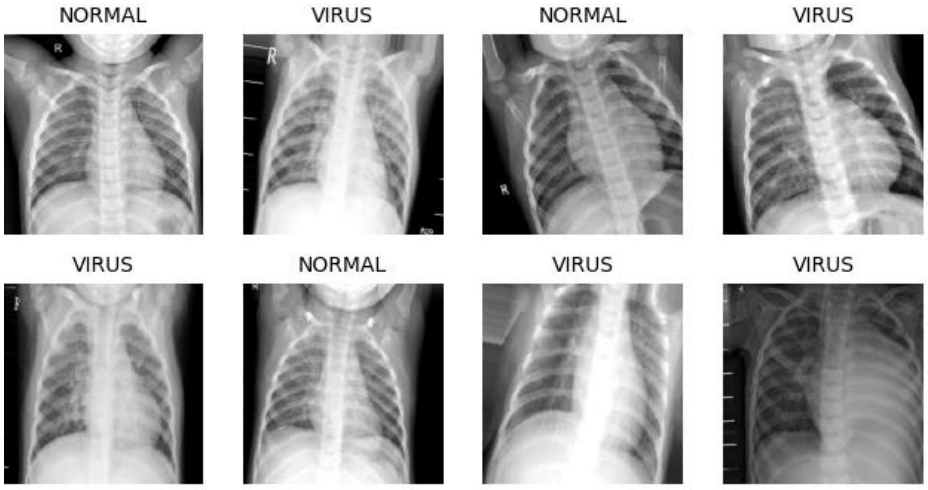

Figure 2. *Sample of augmented images.*

4.3. Model Analysis and Evaluation

The CXR images were of different sizes, and therefore, they were resized from their original resolution to 224×224 pixels to make the classification problem easier. A batch size of 128 was used for both training and testing and a learning rate 0f 0.01 was selected.

A pretrained ResNet18 via fine-tuning was first fit using the imbalanced training data and evaluated using the validation data while tuning the hyperparameters. The fully connected (FC) layer head was removed from the pre-trained network, and a new FC layer head with 2 dense layers constructed and placed on top of the original body of the network. In the first training, the weights in the body of the DCNN were frozen, and the new layer head was trained. In the second training the weights were unfrozen and trained the entire network to allow the model to learn and optimize new parameters. The model was then fit using the balanced dataset and evaluated using the validation set.

The DCNN architecture is represented in Figure 3.

The model was trained for 20 epochs and then evaluated on the validation set. The training and validation accuracy and error curves are presented in Figure 4.

The out-of-sample predictions were done on the test set where an overall accuracy of 88% was obtained by the model trained on the imbalanced dataset (without augmentation) and, an overall accuracy of 91% from the balanced dataset (with augmentation) which was an improvement of +3 compared to the imbalanced data.

The proposed model was evaluated on various performance metrics which include sensitivity, specificity, Precision, F1-score, Negative predictive value on the test set. The proposed model correctly identified and classified the CXR images into 3 distinct classes and achieved an overall accuracy of 91%.

Layer Name	Output size	Layers
Conv1	$112 \times 112 \times 64$	$7 \times 7, 64$
Conv2_x	$56 \times 56 \times 64$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
Conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
Conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
Conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
Average pooling	$1 \times 1 \times 512$	7×7
Fully connected	256	512
Fully connected	128	256
Fully connected	3	128
Softmax	3	

Figure 3. *DCNN architecture.*

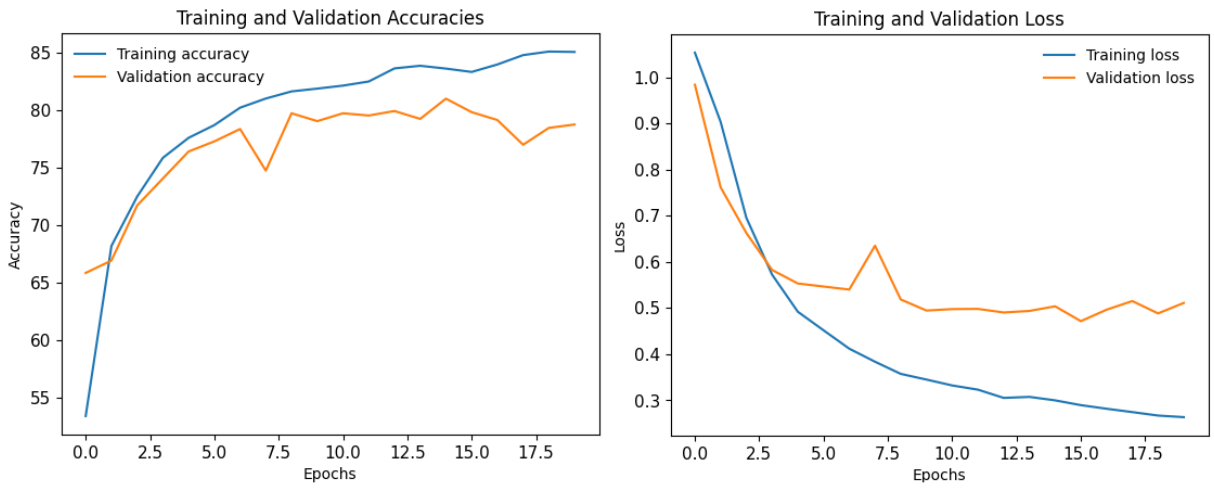

Figure 4. *Training and validation accuracy and error curves.*

The confusion matrix is as shown in Figure 5.

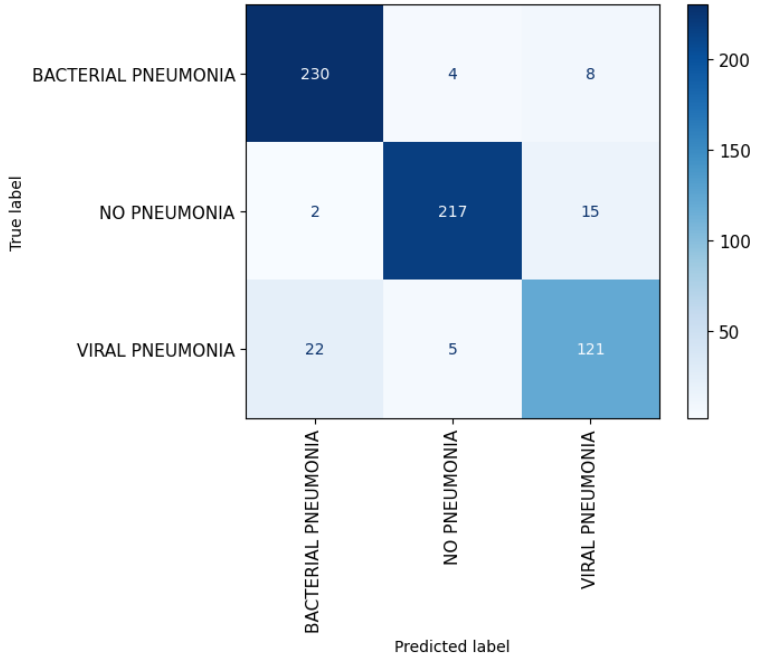

Figure 5. *Confusion matrix.*

Table 3. *Evaluation metrics.*

Class	Sensitivity	Specificity	Precision	Negative Predictive Value	F1-Score
Bacterial Pneumonia	0.9504	0.9372	0.9055	0.9676	0.9274
No Pneumonia	0.9274	0.9769	0.9602	0.9573	0.9435
Viral Pneumonia	0.8176	0.9516	0.8403	0.9438	0.8288

From Table 3, DCNN with augmentation was able to correctly identify and classify the CXR images achieving significant results for all classes. The high sensitivity, specificity, precision, F1-score and negative predictive value, demonstrates the effectiveness of DCNN with augmentation.

5. Conclusion

In this study, data augmentation is presented as a preprocessing technique to address the problem of class imbalance in the dataset. The results from this study show that data augmentation can help solve the problem of class imbalance by generating new images from minority groups that are a slight variation of the original images. Deep Convolutional Neural Network was thereafter used to identify and classify CXR images into 3 distinct categories namely; Bacterial Pneumonia, No Pneumonia (Normal) and viral Pneumonia. Eliminating class imbalance completely from the training data helped the proposed model not to biased towards the majority group. The DCNN with augmentation can therefore be used to predict pathologies in CXR images and this can assist radiologists to automate repetitive tasks and overcome reader variability.

In this paper, a medical image classification algorithm based on neural networks was presented. This project shows the true power of machine learning in real-life use. Since the model clearly has satisfying metrics such as accuracy, precision, F1 score etc., it could, with further parameter adjustment which would resolve the issue of overfitting, be used in a real-life environment. The findings of published papers that neural networks can be of exceptional help in healthcare, have been repeated and confirmed.

Future studies should explore other data augmentation techniques for imbalanced data to improve the performance of the proposed model. Further extension of this research work should focus on using much larger CXR datasets with both pediatric and adult CXR images for various chest pathologies and higher class imbalance. In addition, the success of the proposed model can be extended to real clinical trials.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen and K. Murphy, "Deep learning for chest X-ray analysis: A survey," Medical Image Analysis, vol. 72, p. 102125, 2021.

[2] T.-C. Pham, C.-M. Luong, M. Visani and V.-D. Hoang, "Deep CNN and data augmentation for skin lesion classification," in Intelligent Information and Database Systems: 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part II 10, 2018.

[3] Y. Zheng, C. Yang and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," in Computational Imaging III, 2018.

[4] S. H. Khan, Z. Abbas, S. D. Rizvi and others, "Classification of diabetic retinopathy images based on customised CNN architecture," in 2019 Amity International conference on artificial intelligence (AICAI), 2019.

[5] Y. Hou, "Breast cancer pathological image classification based on deep learning," Journal of X-ray Science and Technology, vol. 28, p. 727–738, 2020.

[6] Z. Li, K. Kamnitsas and B. Glocker, "Analyzing overfitting under class imbalance in neural networks for image segmentation," IEEE transactions on medical imaging, vol. 40, p. 1065–1077, 2020.

[7] W. Al-Dhabyani, M. Gomaa, H. Khaled and F. Aly, "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images," Int. J. Adv. Comput. Sci. Appl, vol. 10, p. 1–11, 2019.

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, p. 541–551, 1989.

[9] A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artificial intelligence review, vol. 53, p. 5455–5516, 2020.

[10] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13, 2014.

[11] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing 25, MIT Press, Cambridge, MA, 2012.

[12] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," arXiv preprint arXiv: 1312.6229, 2013.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409.1556, 2014.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[15] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz and others, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," PLoS medicine, vol. 15, p. e1002686, 2018.

[16] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," Medical physics, vol. 44, p. 1408–1419, 2017.

[17] S. Asif, Y. Wenhui, H. Jin and S. Jinhai, "Classification of COVID-19 from chest X-ray images using deep convolutional neural network," in 2020 IEEE 6th international conference on computer and communications (ICCC), 2020.

[18] S. Chakraborty, S. Paul and K. A. Hasan, "A transfer learning-based approach with deep cnn for covid-19-and pneumonia-affected chest x-ray image classification," SN Computer Science, vol. 3, p. 1–10, 2022.

[19] D. Banik and D. Bhattacharjee, "Mitigating data imbalance issues in medical image analysis," in Data preprocessing, active learning, and cost perceptive approaches for resolving data imbalance, IGI Global, 2021, p. 66–89.

[20] R. Walsh and M. Tardy, "A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer," Diagnostics, vol. 13, p. 67, 2022.

[21] L. Zhang, C. Zhang, S. Quan, H. Xiao, G. Kuang and L. Liu, "A class imbalance loss for imbalanced object recognition," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, p. 2778–2792, 2020.

[22] M. Buda, A. Maki and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural networks, vol. 106, p. 249–259, 2018.

[23] A. Bria, C. Marrocco and F. Tortorella, "Addressing class imbalance in deep learning for small lesion detection on medical images," Computers in biology and medicine, vol. 120, p. 103735, 2020.

[24] J. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," Journal of the American Statistical Association, vol. 92, p. 748–757, 1997.

[25] A. W. Gichuhi, "Nonparametric changepoint analysis for bernoulli random variables based on neural networks," 2008.

[26] D. Kermany, Labeled Optical Coherence Tomography (OCT) for Classification, Mendeley, 2017.