Health and Technology
https://doi.org/10.1007/s12553-024-00820-3

ORIGINAL PAPER

Multiple thoracic diseases detection from X-rays using CX-Ultranet

Anwesh Kabiraj[1] · Tanushree Meena[1] · Pailla Balakrishna Reddy[2] · Sudipta Roy[1]

Received: 5 October 2023 / Accepted: 17 January 2024

Abstract

Background and objective Recent developments in deep learning have demonstrated impressive performance in accurately identifying individual diseases from chest X-rays (CXRs). However, multiple diseases, stability of the deep network, and class imbalance problems were not addressed with high accuracy for disease detection and classification. So, the main purpose of this work is to develop a fully automatic computer method to detect thirteen types of thoracic disease from CXRs with high accuracy.

Methods In this research, a CX-Ultranet has been proposed for the classification and detection of 13 different thoracic disorders from plain radiographic images. The baseline model employed is EfficientNet, and a multiclass cross-entropy loss function is utilized within a compound scaling structure. Channel shuffling is implemented at various stages of the network, creating reduction cells and incorporating more skip connections. The loss function algorithm and Adam optimizers work synergistically to stabilize the model and facilitate continuous learning from new data over time.

Results The CX-Ultranet demonstrates an average prediction accuracy of 88% when applied to diverse CXR datasets. In comparison to existing state-of-the-art techniques, the CX-Ultranet exhibits a remarkable improvement ranging from 5% to 15%. Additionally, it shows a reduction in operational time by approximately 30% compared to current cutting-edge models under similar environmental and data conditions.

Conclusion The proposed CX-Ultranet achieves superior overall accuracy and effectively addresses imbalanced classes within the dataset. Furthermore, it significantly reduces the duration of network training in relation to FLOPS, thereby establishing a novel benchmark in the field of CXR-based disease diagnosis.

Keywords Deep learning · CXRs · Multi-thorax diseases · Compound scaling · CX-Ultranet · Medical image · Channel shuffling

1 Introduction

Chest radiography has been a centrepiece of imaging techniques for many decades, and it is still the most performed radiological scan. According to estimates, around 127 million chest plain radiographs were done in the United States alone in 2021 [1]. With the outbreak of COVID-19, there has been a significant surge in chest X-ray data, commonly referred to as plain radiographs, across the globe. The heightened demand for the utilization of CXR images is attributed to their non-invasive nature. Moreover, a single chest radiograph is employed to study a diverse range of diseases and medical disorders, serving purposes such as screening, diagnosing, and managing these conditions. The broad application of chest X-rays extends to the screening, diagnosis, and management of various thoracic diseases and medical conditions. The implementation of deep learning techniques holds considerable promise in achieving high accuracy within limited time frames. This becomes particularly crucial in countries like India, where the doctor-to-patient ratio is notably challenging at 1:100000. The deployment of a comprehensive centralized cloud-based automated analytical tool emerges as a valuable solution, offering substantial benefits. Such a tool can play a pivotal role in assisting individuals residing in remote and rural areas, where access to specialized medical expertise may be limited.

Anwesh Kabiraj and Tanushree Meena contributed equally.

✉ Sudipta Roy
sudipta1.roy@jioinstitute.edu.in

[1] Artificial Intelligence & Data Science, Jio Institute, Navi Mumbai 410206, India

[2] Reliance Jio - Artificial Intelligence Centre of Excellence (AICoE), Hyderabad, India

Published online: 08 February 2024

Clinical diagnosis of chest X-rays (CXR) often encounters challenges and inaccuracies due to limitations in human vision and the potential to overlook errors [2, 3]. The accuracy of radiologists detecting thoracic disease from CXR is 55% - 60% [4]. A detailed study conducted in the greater New York area amongst 636 ambulatory patients and their radiographic findings by a panel of 11 radiologists showed that the accuracy of such detection is less than 38% [5]. Computer-assisted diagnosis (CAD) can significantly increase the accuracy and reduces the false-negative and positive rate [6]. A CAD-based system also processes multiple diseases at once and takes considerably less time than trained medical practitioners with higher accuracy. With the constant development of various deep learning methods and models [7], there are no significant developments that focus on multiple disease detection from the single CXR. Although a substantial amount of research has taken place in single disease detection from CXR like COVID and pneumonia, a lesser number of works have taken place on other disease detection. Therefore, multiple and all primary disease detection using deep learning is essential.

2 Literature review

Transfer learning and multi kernel-based approaches for single disease detections like Pneumonia have been discussed in [8] and [9]. Still, they did not include the negative transfer fact in their deep transfer learning. Also, their multi kernel-based learning was not tested on the multiple diseases where the initial and target were not similar for training. A dual-channel Convolutional Neural Network (CNN) and transfer learning [10] were used to detect eight thoracic diseases from Chest radiograph and achieved an accuracy of approximately 80% for cardiomegaly. Still, they fail to address the class imbalance problem [11] and other thoracic disease classifications. A contour-based morphological feature was used to find the image view classification of Chest X-ray [12] for binary type. It achieved high accuracy, but accuracy decreased with the increase of the classification category. However, the method was implemented only on the 8000 images, and the stability of the process was not considered.

With the recent COVID-19 pandemic, a lot of research was conducted on Pneumonia and COVID-19 (a type of Pneumonia) detection from Chest radiograph at early stages [13–15]. All these studies mainly focus on the single-class classification problem that was related to COVID and pneumonia-based disease. An ensemble COVID-SDNet neural net architecture [15] was proposed to detect and feature matching for disease detection. The model performed well in binary class problems, but the model's effectivity outside the training distribution was not tested properly. Further advancement was made in ensemble methods [16] and another novel architecture, De-Trac, which could detect two diseases from Chest radiograph using ensemble feature learning with background removal to reach very high accuracy. The approach method was highly inefficient and difficult to train an algorithm with medical constraints. The increasing disease types increase the network complexity.

Deep learning-based techniques for detecting COVID were discussed in [17], and most of them do not focus on computational complexity and multiple disease detection. The review also focuses on how the deep learning method fails to achieve higher accuracy, which is needed in multiple disease detection. The authors [18] discussed dual-channel neural networks and ensemble learning to achieve benchmark accuracy without solving any stability issues. The method failed to achieve consistent accuracy as the underlying problem of learning equally from positive and negative classes for each disease remains the same. A three-dimensional U-net in DL was used [19] to detect a different disease from chest X-rays where 3D images were generated from the same stack of 2D images. The method used a minimal training dataset of 4000 images and reached very high accuracy. However, the challenge remained the same as 3D U-net has higher computational requirements than neural networks using 2D images.

Capsule-based networks for the detection of COVID have been discussed in [20], mainly focusing on the fallacy of CNNs in training from small datasets. The method performs exceptionally well in small datasets and achieves very high accuracy scores of 98.3%. However, capsule-based networks are very prone to underfitting and overfitting. Also, such models are usually limited to binary classification, which does not solve our problem domain. Such frameworks on multiclass detection and classification are yet to be tested. Detection of disease with deep learning-based anomaly detection has also been explored in [21], which primarily tries to train their model on anomalies for diseases and finds the disease based on the type of anomalies detected and localized. They achieved an accuracy of 83.61% for detecting COVID. However, training a neural network on anomalies of the disease needs annotated anomalies for a different disease, which is very difficult to find, and annotating data is difficult. Further, developing such a model for detecting multiple diseases is very complex.

UNet and its application have also shown significant improvement in the classification and localization of several anomalies. A novel margin-based min-max surrogate loss function has been discussed in [22] for increasing the true positive rate of their Deep Neural Network. The proposed AUC margin loss performs well, achieving 93% AUC scores. The proposed Deep AUC Maximization (DAM) method outperforms other loss functions before their work. However, this DAM method takes a very long time when training on a large dataset. According to their study, the DAM model

is highly robust to noisy data and easy data, but to achieve this, they train a DenseNet121 model separately. A similar result is achieved with Image Augmentation, and the extra performance received from the DAM model with another DenseNet121 model is not worth the computational complexity one has to encounter.

A substantial amount of research has been done, but the main limitations remain in classifying multi-thorax diseases, the class imbalance problem, and the stability of the network. Most of the time, the researcher's goal is to detect one disease but detecting all 13 diseases from one input image has not been approached till now. Also, there is a chance of higher negative data than positive data containing the disease for many image databases. The process of training a deep neural network using images automatically leads to a considerable increase in the size of the model. Consequently, this results in a high computational strain due to the extensive amount of information that is processed. As a result, the incorporation of negative data for training becomes impractical therefore streamlining the process of feature extraction and model training and enabling the neural network to be self-contained.

In recent studies, the concept of channel shuffling and compound scaling for classifying images on the CIFAR-10 dataset was reported in ShuffleNASNets [23] and achieved a 10% increase in accuracy. In order to find the best trade-off condition between FLOPS, efficiency, and accuracy achieved, the Grid Search Algorithm [24] was deployed to optimize the best hyperparameters. For achieving a high level of precision in terms of true-positive and false-positive rates, we also developed a more robust multi-class cross-entropy loss function which is implemented with the underlying CX-Ultranet. A loss function becomes very vital when a deep model is deployed on huge datasets. To achieve accurate findings, both compound scaling and the loss function constantly evaluate the deviation in the loss. Thus, it constantly improves the learning of the CX-Ultranet over time.

Key findings of the paper:

- Solved the class imbalance problem using the average weight factor.
- Novel multiclass cross-entropy loss function to stabilize the neural net in the long run.
- Compound scaling and channel shuffling are used to decrease the FLOPS of the network.
- Supports multiple disease classification and detection amongst 13 thoracic diseases from a single chest X-ray.
- Outperforms existing standard models in terms of accuracy, time taken, and the number of diseases detected in the same environments.

3 Methodology

3.1 Dataset description

In this study, the CXR dataset prepared by the National Institute of Health (NIH) [25] and Mendeley Data [26] has been utilized. The dataset comprises 112,120 images, each with a resolution of 1024 × 1024 pixels, and includes annotations for 13 different pathologies. These images were sourced from 30,805 unique patients, and annotations, carried out by expert practitioners, were used for anomaly localization or other purposes manually. To validate our model, the Mendeley dataset [26] was employed, encompassing CXRs and Optical Coherence Tomography (OCT) images. This dataset comprises thousands of images, including both healthy and pneumonia X-rays. For evaluation purposes, approximately 390 Pneumonia and 234 healthy CXRs from the test folder were used. The input to the model consists of 2D CXRs, which undergo preprocessing using image augmentation techniques. Subsequently, the processed images are fed into the network that has been trained on multiple diseases. The model is evaluated to detect diseases in CXRs, and the results are reported. The evaluation is performed on the selected subset of the Mendeley dataset, focusing on 390 Pneumonia and 234 healthy CXRs from the test folder. The proposed workflow is illustrated in Fig. 1, providing a comprehensive overview of the steps involved, from dataset selection and preprocessing to model evaluation for disease detection in CXRs.

3.2 Data processing

To reduce data leakage, where a single patient may have undergone testing multiple times at the same hospital,

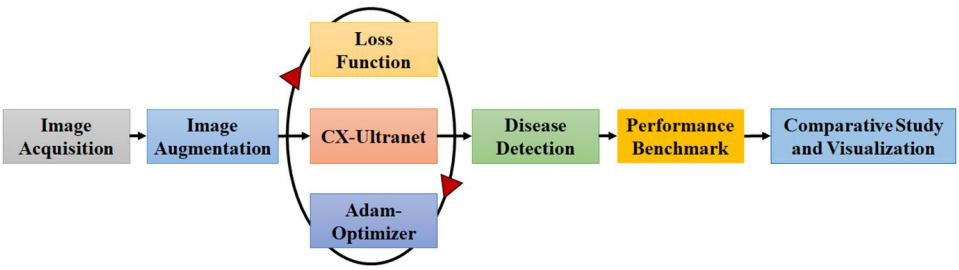

Fig. 1 The overall overview of multiple disease detection and classification from a 2D chest X-ray

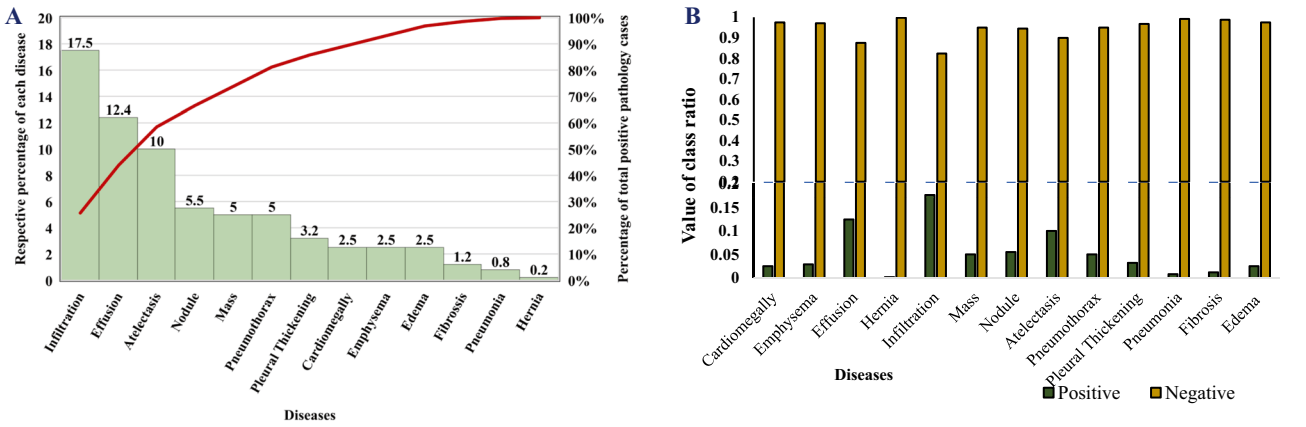

Fig. 2 The Class Imbalance for the 13-disease class corresponding to the Negative Class. **A**) percentage presence of every disease. **B**) Visualizing the corresponding negative class for every disease (Class Imbalance)

duplicate data were systematically removed. Data augmentation was employed using the ImageDataGenerator class from the Keras Framework [27]. The augmentation process included sample-wise centering, standard normalization, and specific transformations such as shear range = 0.1, zoom range = 0.15, rotation range = 5, width shift range = 0.1, and height shift range = 0.05. Horizontal flipping was utilized to facilitate centerline extraction and capture disease-related symmetrical positions. Additionally, the ImageDataGenerator class converted single-channel X-ray images (Grayscale) to a three-channel format by replicating the pixel values across all channels. To manage computational complexity, the image size was set at 320x320. The dataset's mean and standard deviation were normalized, and the input images were shuffled after each epoch. The entire dataset was evenly split with a 1:1 ratio for training and testing. Approximately ≈ 50,000 images were allocated for training, and an equivalent number was used for testing. Given the extensive size of the dataset, aiming to minimize prediction fallacies and stabilize the model, a multiclass cross-entropy loss function was employed.

3.3 Cross entropy loss function

Normal Cross-Entropy for each pathology would give C (X_i) for the i^{th} test case as

$$C_{cross-entropy}(x_i) = -(y_i \log(f(x_i)) + (1 - y_i)\log(1 - f(x_i)) \quad (1)$$

where $f(x_i)$ is the output of the model and x_i are the input features vector, and y_i is the probability of containing the disease or not.

So, for any unique training case either $y_i = 0$ or $(1 - y_i) = 0$. Therefore, one term contributes to the loss function while the other term gets multiplied by 0 and gets nullified. This dataset suffers from a huge class (see Fig. 2). So, we propose our weighted loss function after we solve the class imbalance.

3.4 The problem of class imbalance

Upon conducting exploratory data analysis, which involved cleaning out 0, NaN values, and incomplete or faulty patient details, it was observed that the majority of CXR belong to the negative class, as illustrated in Fig. 2. Notably, the dataset exhibits significant class imbalance, with the highest representation found for the disease "infiltration" (17%) and the lowest for "hernia" (0.2%), as depicted in Fig. 2A. The Pareto line in the graph indicates the cumulative percentage that each subsequent disease contributes to the total dataset. Figure 2B further highlights the substantial class imbalance present in the dataset, emphasizing the need for careful consideration and potential mitigation strategies to address the challenges posed by uneven class distribution. Class imbalance can impact the performance of machine learning models, especially when dealing with diseases that are underrepresented in the dataset.

Infiltration has the highest contribution (17%) out of other diseases. Hence the model starts to learn and incentivize the negative class. We multiplied both the class with class-specific weight factors $weight_{positive}$ and $weight_{negative}$ so that the overall class contribution for both classes is the same. So, with very few positive training cases the loss will be dominated by the negative class. We see that the contribution of either the positive or the negative class could be written as

$$f_{req_p} = \frac{Number\ of\ positive\ examples}{Total\ Cases\ (N)} \quad (2)$$

$$f_{req_n} = \frac{Number\ of\ negative\ examples}{N} \tag{3}$$

Ideally, we need to train our model where contributions from both the classes are equal. Therefore, we multiply both classes with class-specific weight factor w_{pos} and w_{neg} so that the overall class contribution for both the classes is same. Mathematically we want.

$$w_{pos} xfreq_p = w_{neg} xfreq_n \tag{4}$$

To have the above condition satisfied we take

$$w_{pos} = freq_n\ and\ w_{neg} = freq_p \tag{5}$$

I.e., we take samples from the negative class only as much needed to equate the specific weight factor of the positive class which helps us in satisfying (4).

Weighted loss function for multi class cross entropy The loss function is modified by introducing a weight factor, labeled as $weight_{positive}$, for the positive class instead of using y_i. Similarly, the term '1 - y_i' is updated to integrate the weighted average of the negative class, referred to as $weight_{negative}$. As a result, the final formulation of the weighted multi-class Cross-Entropy can be expressed as follows:

$$\mathbb{C}^{w}_{cross-entropy}(x) = -\left(w_p y \log(f(x)) + w_n(1-y)\log(1-f(x))\right) \tag{6}$$

3.5 Compound scaling and reduction cells in CX-Ultranet

When employing high-resolution images with a larger model, an increase in network depth becomes essential [28]. Utilizing high-resolution images is advantageous, and scaling up any dimension of the neural network – be it depth, width, or resolution – leads to improved model accuracy. However, it has been observed that in larger networks, the rate of accuracy improvement tends to slow down after an initial burst of progress. There is a limit to the extent of improvement achievable by merely increasing the network size. Additionally, enlarging the network introduces higher computational costs and the risk of overfitting. To address these challenges, we present the novel CX-Ultranet, designed to maximize performance while efficiently utilizing computational resources.

Floating point operations per second (FLOPS) of a convolutional neural network Floating Point Operations Per Second (FLOPS) of a Convolutional Neural Network: FLOPS serves as a metric for measuring the computational performance of a Convolutional Neural Network (CNN). It quantifies the number of floating-point operations, such as additions and multiplications, that a network can execute in one second. FLOPS is crucial for assessing the computational complexity of a neural network and is commonly used to evaluate the efficiency of hardware accelerators like GPUs or TPUs for deep learning tasks. This metric provides valuable insights into the computational demands of a CNN, allowing for the comparison of different network architectures or configurations to determine which is more efficient in terms of computational resources. Lower FLOPS values indicate more efficient networks requiring less computational power, while higher FLOPS values suggest networks with greater computational demands.

$$FLOPS \propto d, w^2, and\ r^2 (depth\ of\ the\ network) \tag{7}$$

where, d, w, and r are the depth, and width resolution of the model respectively.

Compound scaling method The Compound Scaling Method is a strategic approach employed within the domain of deep learning and neural network building. Its purpose is to attain an optimal balance between accuracy and computational efficiency, with a primary emphasis on the trade-off between time taken and results obtained. The key objective of this approach is to achieve the best possible equilibrium between precision and computing efficiency, placing a specific focus on balancing the time taken against the accuracy achieved. Consequently, we applied this method to identify the most optimized solution, striving for the most ideal trade-off between the time taken and the accuracy achieved.

$$depth\ (d) = \alpha^{\sigma}\ width(w) = \beta^{\sigma}\ resolution\ (r) = \gamma^{\sigma} \tag{8}$$

where σ is the user-specified coefficient. It serves as a control parameter for determining the extent to which resources are allocated for the scaling of a model. The value of "σ" is set by the user and influences the scaling of various model parameters, such as depth, width, and input resolution, as part of the optimization process. It will essentially allow the user to specify the level of resource allocation and, consequently, the trade-off between computational resources and model performance.

$$FLOPS \propto (\alpha.\beta^2.\gamma^2)^{\phi} \tag{9}$$

here, α, β, and γ are parameters that decides the allocation of additional resources to network width, depth, and resolution, and their values are determined through a small grid search. For this work we fixed the condition that the product of $(\alpha.\beta^2.\gamma^2)$ is approximately equal to 2 (≈ 2). Therefore,

$$FLOPS \propto (2)^{\phi} \tag{10}$$

Compound scaling We adopted a similar approach as outlined in [29] for model scaling. Fixing the value of φ at 1, we

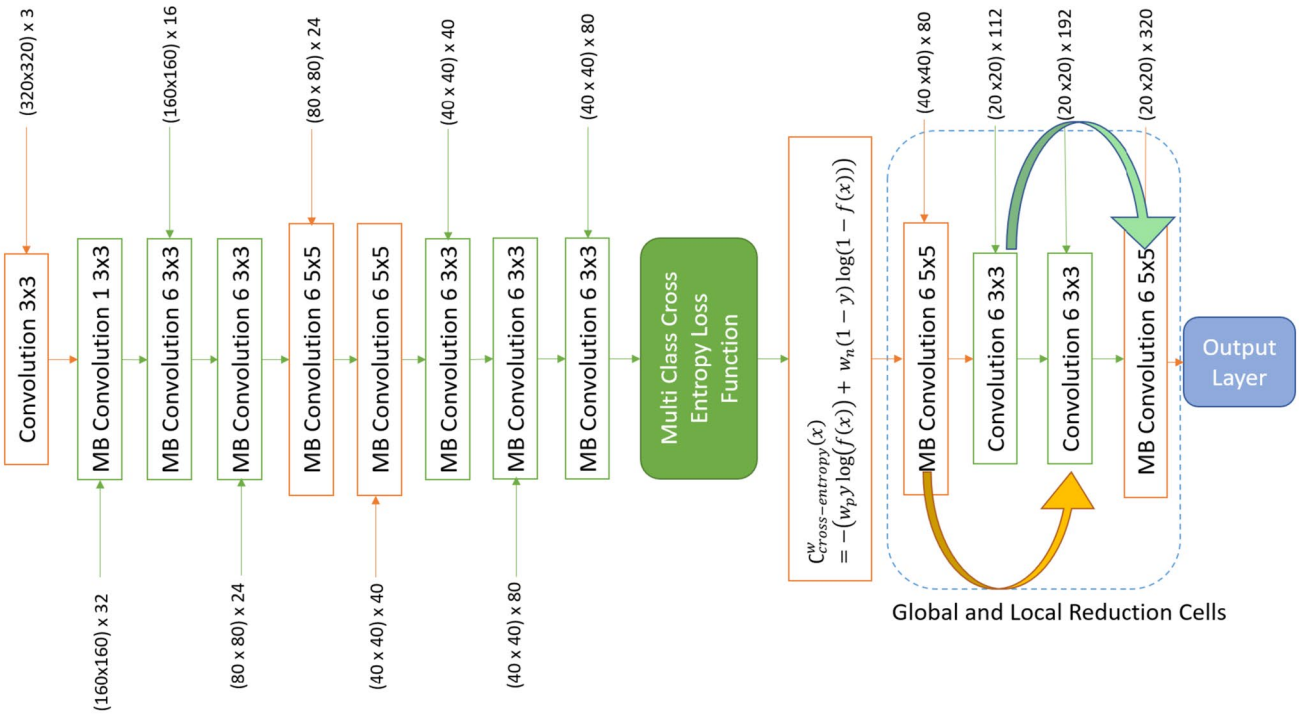

Fig. 3 The schematic flow including the Baseline Model, the Loss Function, and Reduction Cells mechanism

conducted a simple grid search [24] to determine the values of α, β, and γ. Through this search, we identified that α = 1.2, β = 1.1, and γ = 1.14 satisfied the condition (α · β^2 · γ^2) ≈ 2. With these fixed values, we then proceeded to scale up the value of φ (Compound Coefficient), resulting in an improvement in model accuracy. Our optimization strategy aimed at achieving optimal results while minimizing both φ and the Floating Point Operations Per Second (FLOPS) of the model. We achieved this by optimizing the product of model accuracy and the model's execution time, striking a balance between precision and computational efficiency.

$$Optimization\ Goal = Model_{Accuracy} \times Model_{FLOPS} \tag{11}$$

By optimizing the trade-offs between accuracy and FLOPS using the hyper parameter, we aim to maximize our optimization goal with the least amount of scaling.

3.6 Implementation of cell reduction

Our main objective is to improve the efficiency of our model in terms of both time and computational complexity. To do this, we adopt an approach that is similar to the one described in ShuffleNet [23] and FractalNets [30]. This strategy involves designing the layers of the architecture to selectively receive input from only half of the available channels, while disregarding the remaining half. As a result, additional skip connections are included within the model. This design choice avoids the training of identity functions while maintaining equal channel widths, thereby reducing memory access costs. The inclusion of skip connections plays a vital role in mitigating the challenges posed by large group convolutions, which are frequently encountered in larger models with deeper network architectures. Figure 3 provides a visual representation of CX-Ultranet, illustrating the convolutional layers with varying channel widths, alongside the cross-entropy loss function.

A more detailed structure of the CX-Ultranet model is shown in Fig. 4. The different layers in order that are used in the construction of the overall model are shown in Fig. 4A. The layers are then connected in specific ways shown in Fig. 4B to make basic blocks of the compound scaling framework. Sub-block 1 is used as the first sub-block in the first block. Sub-block 2 is used as the first sub-block in all the other blocks.

Sub-block 3 is used for any sub-block except the first one in all the blocks. The complete model denoting the compound scaling blocks is shown in Fig. 5. The skip connections that are established with the pooling layer as a result of channel shuffling are illustrated as in Fig. 5.

We implemented a learning rate scheduler sourced from the Keras Library [27]. The 'Start' and 'Max' parameters were set to 0.000002 and 0.0001, respectively. 0.0001 is the learning rate. For optimization, we used the Adam optimizer [31].

The thoracic diseases in the CXRs are localized using Growing seed algorithm [32]. We have achieved similar

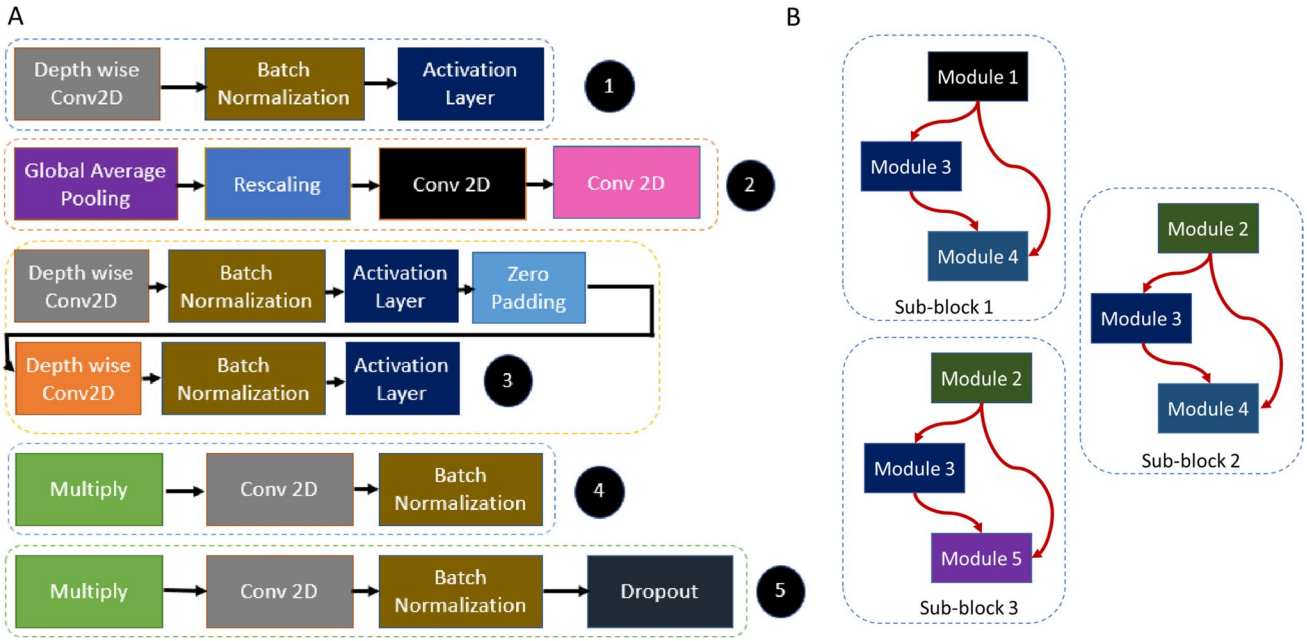

Fig. 4 The inside neural structure of the CX-Ultranet. **A** Module 1-5 that are used in the total architecture, and **B** the different sub-blocks that are used in the overall architecture

training accuracy ≈ 95%. We also used the learning weights used by our model to generate the visualizations based on attention-maps. Since this is a multiclass detection and classification model, we have used it to find out other diseases that are associated with the X-ray. It is very common for medical practitioners that they only focus on the highlighted disease rather than focusing on what other medical conditions they are overlooking at.

4 Results

The rectified class imbalanced model shows (see Fig. 6) no biasness on particular class data and learns equally from both classes. The split of multiclass into simple binary class from the root containing only one disease chest X-ray and a corresponding healthy chest X-ray is shown in Fig. 6.

Now we have plotted the learning curve of the model to check the stability. We have done so by training the model by taking a part of the data and testing the model on cases that were not included in the overall training dataset. The amount of data taken each time is increased by 5% in the subsequent rounds and every time it shows an improvement in accuracy (see Fig. 7). The trained weights and loss function values are not reset for every subsequent round as some part of the data becomes common for every round. We find a very high learning rate at the beginning of the model which is also coupled with the fact that no class imbalance exists, therefore the model learns rapidly (refer Fig. 7.). The model reaches saturation point after it achieves around 70% accuracy which

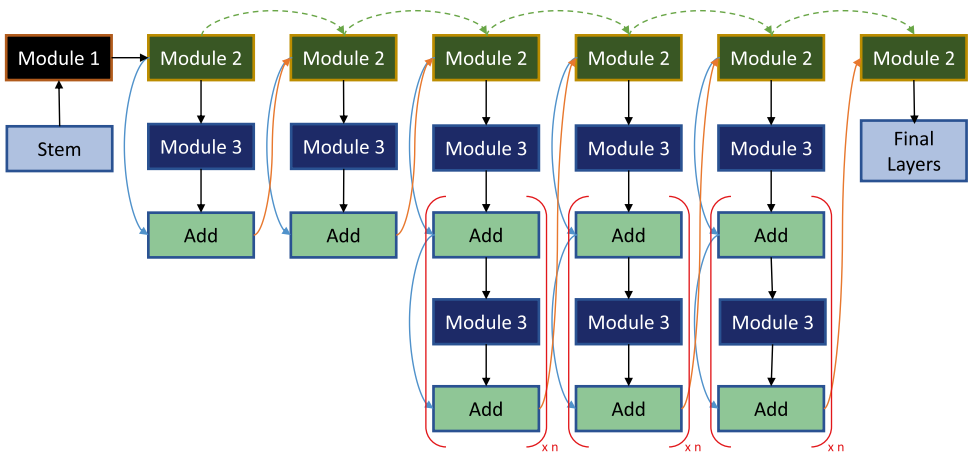

Fig. 5 The architecture of compound scaling method reduction cell mechanism. Architectures in (◌)×n represent they are repeated n times which will subsequently increase the FLOPS and accuracy of the model if a higher accuracy needs to be achieved for a cost of higher computation power. Shows the connected layers which only receive half of the channel

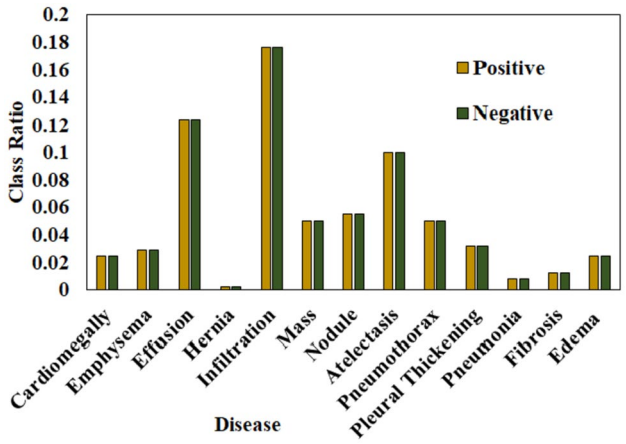

Fig. 6 The contribution of both the class after the class imbalance problem is solved

is close to the highest accuracy achieved by existing state of the art model. The model pushes the accuracy to 85%-90% region with the combined effect of the multiclass cross entropy loss function, learning rate scheduler and Adam optimizer. This plays a very important role as the value of the loss function is not reset every time and it learns by how much the network fails for cumulative addition of new data.

The visual prediction of thorax disease and percentage probability of all thorax diseases in one CXR image is shown in Fig. 8. The major probability of a disease is highest percentage value (here Pleural Thickening) in a single image. The algorithm also predicts various additional diseases like nodules, fibrosis, infiltration, atelectasis and pneumothorax. The heatmap and broad visualization are also shown in Fig. 8. for better understanding. We performed the same for other thorax diseases.

We have used ≈ 60,000 CXR images from outside the training dataset to test our model. The model achieves varied accuracies for different diseases and the receiver operating characteristics curve (ROC) is shown in Fig. 9.

The model attains an average accuracy of 88%, with the highest accuracy of 92% observed for the Atelectasis disease, and the lowest accuracy of 84% for Infiltration. Table 1 displays the mean ± standard deviation (SD), along with the maximum and minimum test accuracies for all 13 diseases. A low SD in these results indicates a high level of consistency and accuracy in the performance of CX-Ultranet.

We find that our CX-Ultranet model performs exceptionally well in terms of sensitivity for most diseases. True positive rate is bad for infections like nodules, mass, and pneumonia as the symptoms related to that disease are common. Enough data to differentiate these three diseases is a challenging task for Radiologists. To ensure that our model is not overfitting or underfitting, we plot the AUC values of the model before and after training (see Fig. 10). A small gap between the two values for every corresponding disease denotes that the model is good and does not underfit or overfit.

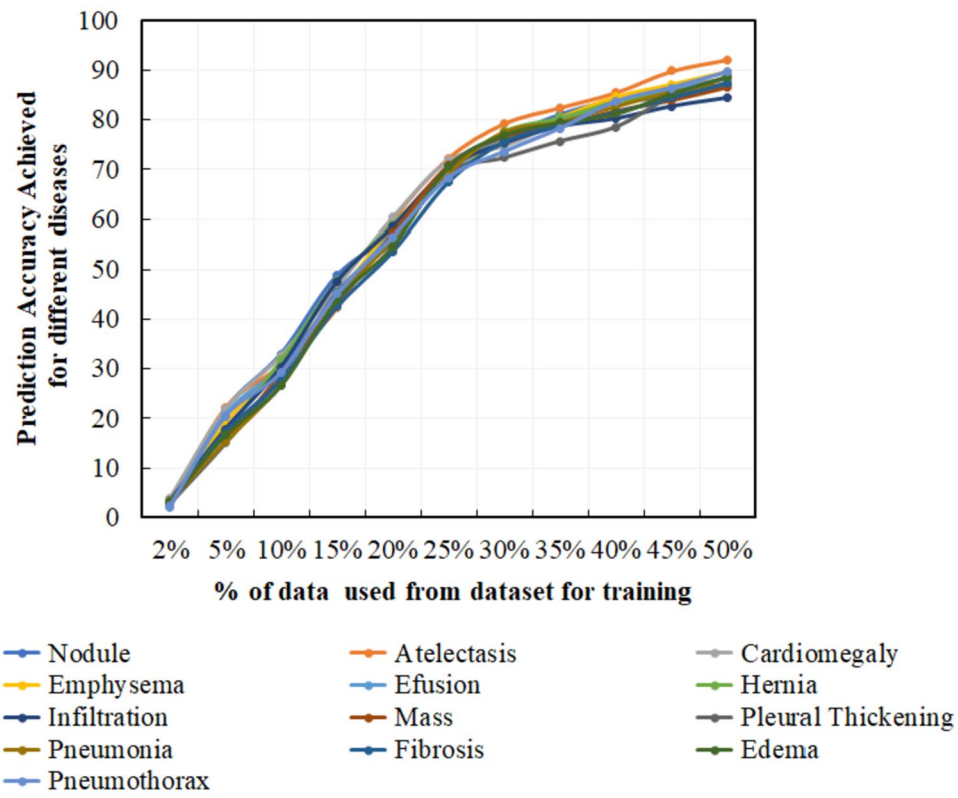

Fig. 7 The learning rate of the model is based on % of data used to train the model

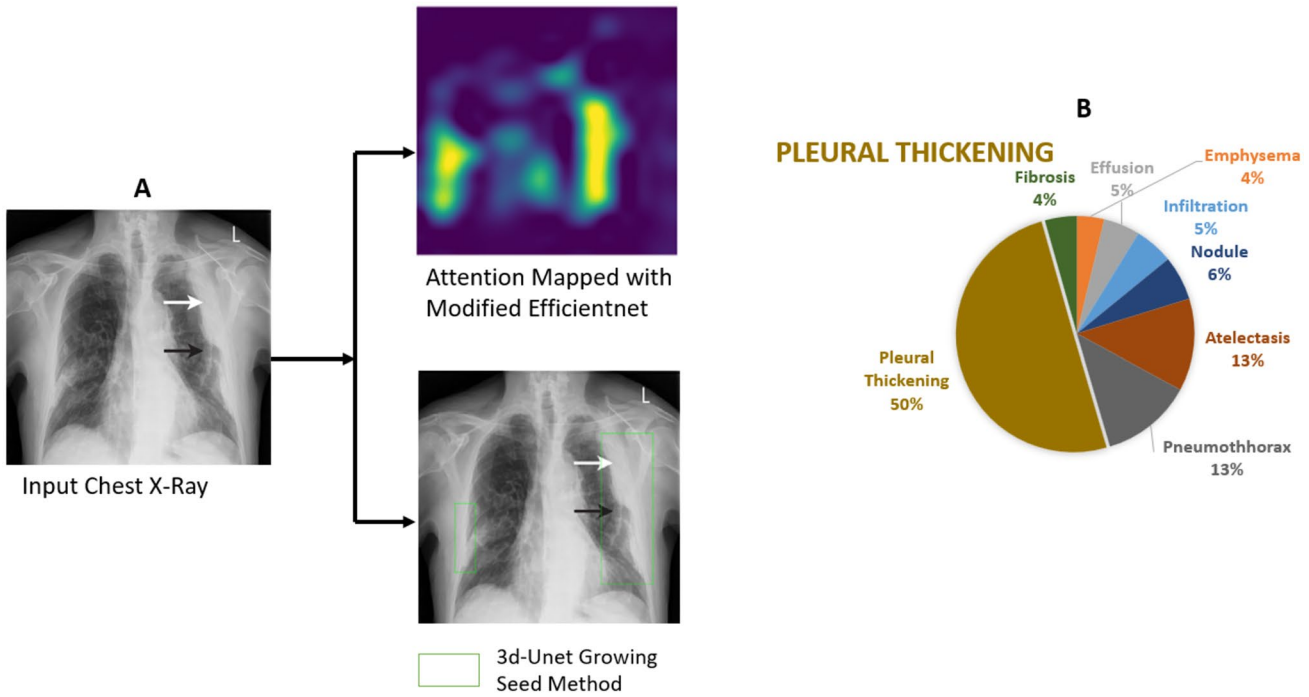

Fig. 8 The prediction of pleural thickening using the CX-Ultranet. **A** The input CXR and corresponding heat map and disease localization, and **B** the % of all thoracic diseases.

5 Discussions

The proposed CX-Ultranet architecture with modified loss function achieves higher accuracy than any other pre-existing recent models due to solved class imbalance problem. The proposed CX-Ultranet model consistently gives an accuracy above 84% and reaches the highest accuracy of 92% for Atelectasis disease. We observe that the initial rate of learning is remarkably high because it learns equally from positive and negative test cases. The learning rate flattens out in the 25% to 35% region, as seen in Fig. 7. The increase in prediction accuracy is entirely due to the loss function and compound scaling method kicking in CX-Ultranet.

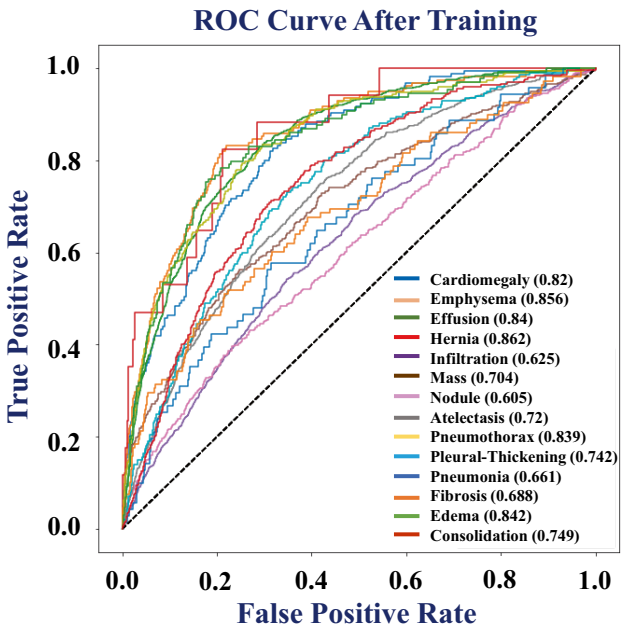

Fig. 9 ROC curve for 13 diseases for the Proposed Model

Table 1 The accuracy was achieved by CX-Ultranet for 13 different diseases

Thoracic Diseases	Mean ± Standard Deviation	Max-Min
Atelectasis	88.55 ± 06.0	92.7 - 84.4
Cardiomegaly	80.2 ± 12.9	92.1 - 68.4
Edema	83.2 ± 09.5	92.0 - 74.4
Effusion	82.85 ± 11.9	93.9 - 71.8
Emphysema	82.95 ± 10.6	92.7 - 73.2
Fibrosis	85.3 ± 07.2	92.3 - 78.3
Hernia	81.8 ± 09.1	89.9 - 73.7
Infiltration	76.05 ± 18.5	93.7 - 58.4
Mass	79.65 ± 12.1	90.7 - 68.6
Nodule	87.3 ± 08.0	94.9 - 79.7
Pneumothorax	84.45 ± 11.5	95.1 - 73.8
Pleural-Thickening	82.65 ± 10.9	92.7 - 72.6
Pneumonia	84.05± 08.0	91.6 - 76.5

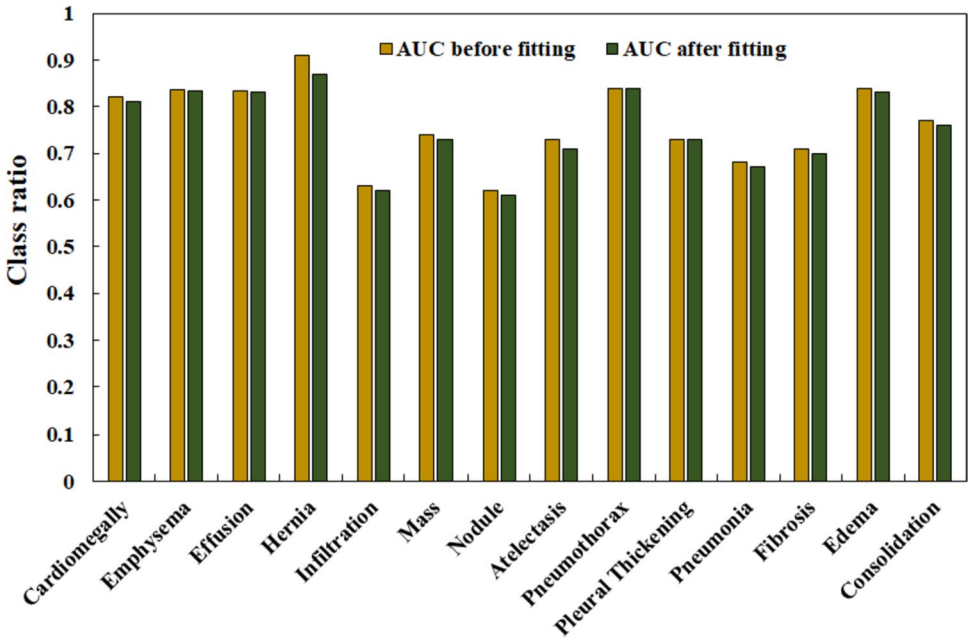

Fig. 10 Comparison of Area under ROC (AUC) before and after fitting the model

CX-Ultranet outperforms all other recent standard models in image classification in terms of accuracy. A similar comparison was performed to evaluate our model against the current state-of-the-art models, as illustrated in Fig. 11. The models included in the comparison were AlexNet [33], U-Net (Encoder), Res-Net (Decoder) [34], and DenseNet121 [35]. DenseNet121 has gained attention for its application in the classification of medical imaging. The U-Net (Encoder) Res-Net (Decoder) model presents a challenge to our model in terms of prediction accuracy. However, due to its dual-channel CNN architecture, it suffers from significantly higher computational complexity. As illustrated in Fig. 12, it also takes considerably more time to execute than our CX-Ultranet.

The model outperforms other existing models when trained on the same dataset under the same conditions in terms of FLOPS and hence the total time taken in seconds. We have used Intel(R) Core (TM) i9-10900K CPU @ 3.70GHz 3.70 GHz with NVIDIA RTX A4000 80 GB high-performance computers and Python for this study. The reduction in time taken is majorly due to two factors: i) A considerable part of the dataset comprising of the Negative Test Cases is eliminated and hence not used in training of the model, and ii) The reduction cell

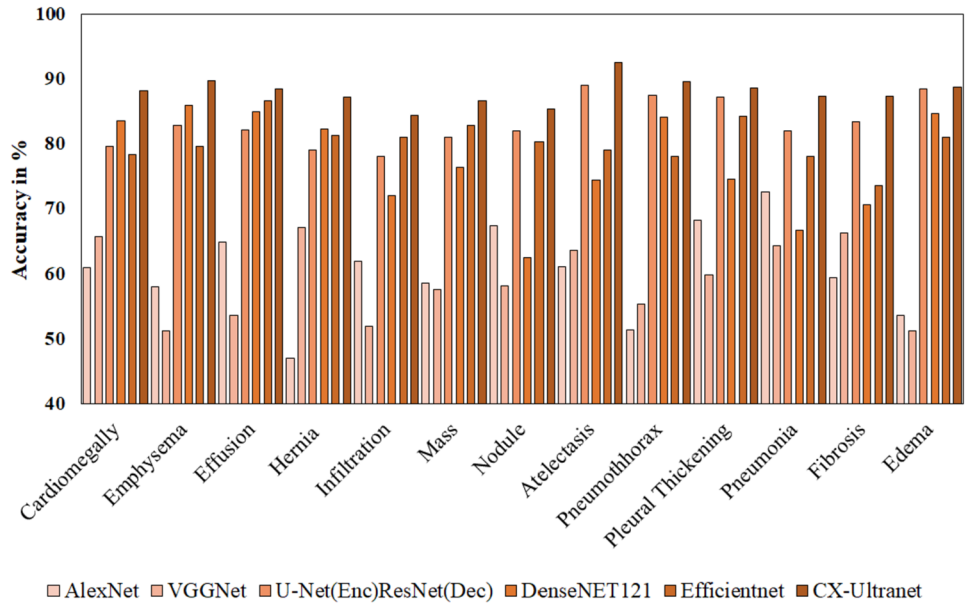

Fig. 11 Comparative analysis of state-of-the-art models with CX-Ultranet in terms of Accuracy

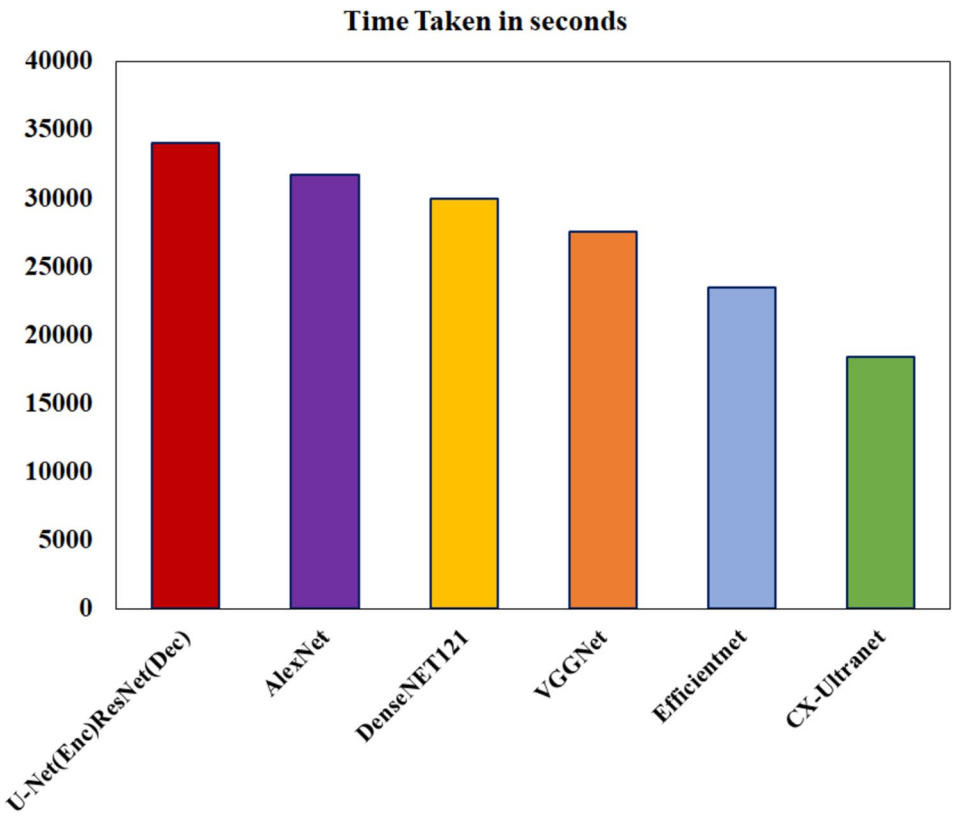

Fig. 12 Time taken in seconds for AlexNet, VGGNet, UNet, DenseNet121, VGGNet, and the proposed CX-Ultranet on NIH dataset

mechanism only takes half channel as input for specific layers which reduces the time taken and makes the model faster. The time gap between EfficientNet and CX-Ultranet is substantial, showing the impact of solving class imbalance and channel shuffling.

To evaluate the effectiveness and practical application of CX-Ultranet, we tested it on the Mendeley dataset that was not used to train our model. Mendeley dataset consists of only a single disease Pneumonia. We evaluated our model on unseen data to ensure reproducibility and observed that our model gives an F1 score of 0.93 for pneumonia disease. We also calculated recall and precision along with the F1 score. The complete analysis is given in Fig. 13.

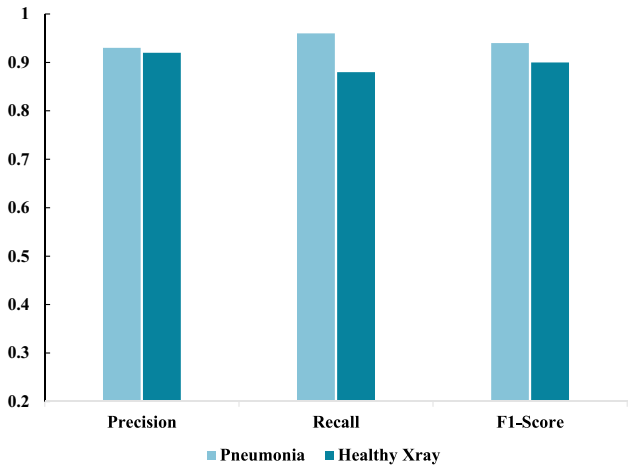

Fig. 13 Precision recall and F1 score for Mendeley dataset

6 Conclusion

Multiple disease detection has not been approached and addressed prior to this work. We built a generalized model for the detection of 13 diseases in real-time from a single chest X-ray input. The class imbalance problem has been addressed and solved. A new multiclass cross-entropy loss function with optimizer functions maintains the learning rate and stability of the model over huge datasets. We have also conducted extensive performance benchmarks against other existing methods. Building high accuracy AI models for medical imaging is challenging and has a major role in this field. The benefits of the same are great and wide for humanity. Reaching people who cannot get benefits from the latest medical facilities in city outskirts and unreachable places becomes possible with advancements in computer technology.

Acknowledgement This research work was supported by the Jio Institute "CVMI-Computer Vision in Medical Imaging" research project fund under the "AI for ALL" research centre. We would like to thank Dr Kalyan Tadepalli, Sir H N Reliance Foundation Hospital for his advice in this CVMI project.

Authors contribution Anwesh Kabiraj: Design, implementation and wrote manuscript; Tanushree Meena: Analysis, part of implementation and wrote manuscript; Pailla Balakrishna Reddy: Conceptualize and revise the manuscript; Sudipta Roy: Supervise, conceptualize, analysis and revise the manuscript.

Funding This research work was supported by RFIER-Jio Institute 'CVMI-Computer Vision in Medical Imaging' research fund under "AI for ALL" research center.

Data Availability The data availability is mentioned in the Data Description subsection under Methodology section. The open-source National Institute of Health (NIH) has been used in this study. Available at: https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345.

Github source code link https://github.com/labsroy007/CX-Ultranet.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Statement of significance A considerable amount of work has been carried out in disease detection from a Chest X-ray, but the main limitations remain in multiple disease classification, stability of the network and solving the class imbalance problem were not focused much. Many a times, the researcher's goal is to detect one disease but detecting all 13 diseases from one input image has not been approached before this work. Recent advances in deep learning have shown many good performances in disease identification from chest X-rays. The implementation of deep learning in medical imaging for disease detection from chest x-rays can help the doctors in early diagnosis and prognosis. The implementation of high accuracy AI models for medical imaging is challenging and has a major role in this field. Understanding the need of today's world, we proposed a novel CX-Ultra net (Chest X-ray Ultranet) to classify and identify thirteen thoracic lung diseases from chest X-rays. Our contribution can have significant benefits for humanity. Reaching people who cannot get benefits of the latest medical facilities in city outskirts and unreachable places becomes possible with advancements in computer technology. We have used a multiclass cross-entropy loss function on a compound scaling framework using EfficientNet as a baseline. We performed channel shuffling in various stages of the network creating reduction cells and more skip connections which resulted in a high performance of the model. Also, there is a chance of higher negative data than the positive data containing the disease for many images database. A deep neural net is already hefty when trained with images that add to the computational complexity coupled with massive data. This eliminates considering the negative data for training purposes and hence the whole procedure of feature extraction and training them to the model, making the neural net heavier.

References

1. IMV Medical Information Division. 2021 X-ray DR CR Market Outlook Report. https://imvinfo.com/product/2021-x-ray-dr-cr-market-outlook-report/ . Visited on 4th May, 2022.
2. Meena T, Roy S. Bone Fracture detection using deep supervised learning from radiological images: A paradigm shift. Diagnostics. 2022;12:2420. https://doi.org/10.3390/diagnostics12102420.
3. Roy S, Meena T, Lim S-J. Demystifying supervised learning in healthcare 4.0: a new reality of transforming diagnostic medicine. Diagnostics. 2022;12(10):2549. https://doi.org/10.3390/diagnostics12102549.
4. Satia I, Bashagha S, Bibi A, Ahmed R, Mellor S, Zaman F. Assessing the accuracy and certainty in interpreting chest X-rays in the medical division. Clinical medicine. 2013;13(4):349–52.
5. Weinstock Michael B, Md Ana Echenique, Md Dabr, Russell Joshua W, Md Msc, Facep Ari Leib, Md Jordan A, Miller Do, Cohen David J, Md Stephen Waite, Md Allen Frye, Np Andfrank A, Illuzzi,. Chest x-ray findings in 636 ambulatory patients with COVID-19 presenting to an urgent care center: a normal chest x-ray is no guarantee. J Urgent Care Med. 2020;14(7):13–8.
6. Chakraborty S, Kumar K, Reddy BP, Meena T, Roy S. An explainable AI based clinical assistance model for identifying patients with the onset of sepsis. Bellevue, WA, USA: 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI); 2023. p. 297–302. https://doi.org/10.1109/IRI58017.2023.00059.
7. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Medical Image Analysis. 2021;72: 102125.
8. Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW. Efficient pneumonia detection in chest xray images using deep transfer learning. Diagnostics. 2020;10(6):417.
9. Hu M, Lin H, Fan Z, Gao W, Yang L, Liu C, Song Q. Learning to recognize chest-Xray images faster and more efficiently based on multi-kernel depthwise convolution. IEEE Access. 2020;8:37265–74.
10. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. IEEE CVPR; 2017. p. 2097–106.
11. Dufour JM, Neves J. Finite-sample inference and nonstandard asymptotics with Monte Carlo tests and R. In: Handbook of statistics, vol. 41. Elsevier; 2019. p. 3–31.
12. Xue Z, You D, Candemir S, Jaeger S, Antani S, Long LR, Thoma GR. Chest X-ray image view classification. Sao Carlos, Brazil: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems; 2015. p. 66–71. https://doi.org/10.1109/CBMS.2015.49.
13. Kabiraj A, Meena T, Reddy PB, Roy S. Detection and classification of lung disease using deep learning architecture from x-ray images. In: International Symposium on Visual Computing, vol. 13598. Cham: Springer International Publishing; 2022. p. 444–55. https://doi.org/10.1007/978-3-031-20713-6_34.
14. Meena T, Kabiraj A, Reddy PB, Roy S. Weakly supervised confidence aware probabilistic CAM multi-thorax anomaly localization network. In: 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI). IEEE; 2023. p. 309–14. https://doi.org/10.1109/IRI58017.2023.00061.
15. Tabik S, Gomez-Rios A, Martin-Rodriguez JL, Sevillano-Garcia I, Rey-Area M, Charte D, Guirado E, Suarez JL, Luengo J, Valero-Gonzalez MA, Garcia-Villanova P, Olmedo-Sanchez E, Herrera F. COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images. IEEE J Biomed Health Inform. 2020;24(12):3595–605. https://doi.org/10.1109/JBHI.2020.3037127. Epub 2020 Dec 4 PMID: 33170789.
16. Abbas A, Abdelsamea MM, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Applied Intelligence. 2021;51(2):854–64.
17. Alghamdi HS, Amoudi G, Elhag S, Saeedi K, Nasser J. Deep Learning Approaches for Detecting COVID-19 From Chest X-Ray Images: A Survey. IEEE Access. 2021;25(9):20235–54. https://doi.org/10.1109/ACCESS.2021.3054484. PMID:34786304;PMC ID:PMC8545235.

18. Minaee S, Kafieh R, Sonka M, Yazdani S, Soufi GJ. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. Med Image Anal. 2020;65:101794.
19. Wu W, Gao L, Duan H, Huang G, Ye X, Nie S. Segmentation of pulmonary nodules in CT images based on 3D-UNET combined with three-dimensional conditional random field optimization. Medical Physics. 2020;47(9):4054–63.
20. Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. Pattern Recognition Letters. 2020;138:638–43.
21. Zhang J, Xie Y, Li Y, Shen C, Xia Y. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint; 2020. arXiv:2003.12338.
22. Yuan Z, Yan Y, Sonka M, Yang T. Large-scale robust deep auc maximization: a new surrogate loss and empirical studies on medical image classification. Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 3040–9. arXiv:2012.03173.
23. Laube KA, Zell A. ShuffleNASNets: efficient CNN models through modified efficient neural architecture search. Budapest, Hungary: 2019 International Joint Conference on Neural Networks (IJCNN); 2019. p. 1–6. https://doi.org/10.1109/IJCNN.2019.8852294.
24. Hesterman JY, Caucci L, Kupinski MA, Barrett HH, Furenlid LR. Maximum-likelihood estimation with a contracting-grid search algorithm. IEEE transactions on nuclear science. 2010;57(3):1077–84.
25. National Institutes of Health - Clinical Center, ChestXray-NIHCC, Available at: https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345, (Accessed on 10 January 2022)
26. Kermany D, Zhang K, Goldbaum M. Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley Data. 2018;2(2):651. https://doi.org/10.17632/rscbjbr9sj.3.
27. Gulli A, Pal S. Deep learning with Keras. Packt Publishing Ltd; 2017.
28. Sabottke CF, Spieler BM. The effect of image resolution on deep learning in radiography. Radiol Artif Intell. 2020;2(1):e190015.
29. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019. p. 6105–14.
30. Larsson G, Maire M, Shakhnarovich G. Fractalnet: ultra-deep neural networks without residuals. arXiv preprint; 2016. arXiv:1605.07648.
31. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint; 2014. arXiv:1412.6980.
32. Cui J, Guo H, Wang H, Chen F, Shu L, Li LC. Fully automatic segmentation of coronary artery using growing algorithm. Journal of X-ray Science and Technology. 2020;28(6):1171–86.
33. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. arXiv preprint; 2016. arXiv:1602.07360.
34. Chu Z, Tian T, Feng R, Wang L. Sea-land segmentation with Res-UNet And fully connected CRF. Yokohama, Japan: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium; 2019. p. 3840–3. https://doi.org/10.1109/IGARSS.2019.8900625.
35. Sarker L, Islam MM, Hannan T, Ahmed Z. COVID-DenseNet: a deep learning architecture to detect COVID-19 from chest radiology images. preprint. 2020;2020050151.