# Abstract

Due to their exceptional performance in classification challenges, such as the automatic analysis of X-rays, and their capacity to extract features, convolutional neural networks (CNNs) have recently come to dominate the area of computer vision. It is urgently necessary to use cutting-edge technology, such as deep learning, to detect anomalies in chest X-ray images in order to increase productivity and diagnosis precision. The over-fitting and low transfer rate of existing neural networks makes them difficult to fine-tune for medical image processing. To overcome these challenges, the proposed study designed an attention-based CNN model ‘Feature Extraction Classifier Network’ (FEC-NET) to automatically extract the features from customized inception blocks for efficient classification of chest X-ray images. The proposed model is small and lightweight compared to the existing pre-trained model. Moreover, we also turned the proposed model to achieve better performance of the model. According to the experimental findings, our light model FEC-NET had improved accuracy, recall, F1 score, and AUC values. It demonstrates that developing a suitable CNN is preferable to optimizing deep networks and that increasing the training data is essential to improving CNN performance.

# Introduction

1In recent years, the medical field has been troubled by two significant issues that impede patient care: medical staff burden and subjectivity in interpreting the findings of medical tests [1, 2]. There is no simple solution to these problems, which is particularly hazardous in medicine, where procedural mistakes can result in severe health complications. First, overwork in medicine, exacerbated in recent years by the global COVID-19 pandemic, can result in diagnostic errors and treatment delays. As mentioned previously, the interpretation of certain medical tests is also subjective. The expert interpreting these tests, such as X-rays, may make an incorrect diagnosis due to, for instance, the presence of signs of multiple disorders to varying degrees [3]. Due to its low cost, rapidity of acquisition, and lack of preparation requirements, this form of imaging test is one of the most frequently used for a variety of diagnoses [4]. thoracic X-rays are beneficial for detecting a variety of thoracic diseases associated with organs such as the heart, lungs, and bones. X-rays are appropriate for analysis with convolutional neural networks (CNN) [5] due to their characteristics. The combination of AI algorithms and medical knowledge can enhance the performance of medical personnel [6] and reduce patient wait times by accelerating the diagnostic process and reducing physicians' burden.

2The chest X-ray is one of the most accessible examinations for lung disease screening and diagnosis, such as Infiltration, Atelectasis, Mass, Pneumothorax, Consolidation, Cardiomegaly, Emphysema, Pneumonia, and Hernia [1]. In the United States alone, over 35 million chest X-ray images are obtained annually, and radiologists must interpret over 100 X-ray studies daily [2]. Due to the vast population and increased awareness of health in China, there are significantly more patients and chest X-rays. Advanced technologies and automated methods can aid radiologists in accurately and efficiently diagnosing disease. Normal and abnormal chest X-ray images can be automatically classified into two categories by trained programs, allowing radiologists to focus more on abnormal X-ray images. In addition, trained algorithms may group chest X-ray images into additional categories based on various diseases. In addition, trained instruments can aid in disease localization and visualization.

Deep learning, particularly the classification approach based on the architecture of convolutional neural networks, has acquired popularity in recent years due to its capacity to learn appropriate image features via automatic backpropagation [18]1. CNNs have revolutionized computer vision due to their ability to derive image features. These architectures consist of multiple layers. The first model's convolutional layers are derived from the concept of cells in visual neuroscience. The designs were inspired by the visual cortex of animals. These architectures have distinguished themselves due to their exceptional ability to extract patterns from data, thereby enhancing the performance of preceding systems based on Machine Learning models. Due to their high performance in a variety of tasks, including speech recognition, computer vision, and text analysis, this advantage has made them a benchmark in Deep Learning [7]2.

Chest X-rays are susceptible to analysis by this form of the algorithm due to their characteristics. It is not essential to manually extract image features or conduct segmentation with CNNs, and their ability to learn from large amounts of data enables them to recognize patterns that are challenging for human beings to detect. Although this article focuses on classification problems, other problems, such as X-ray segmentation [8]1, localization, regression (such as drug dosage prediction), and others, can also be solved. CNNs are a possible analysis instrument for chest radiographs based on the background of big data and automated feature extraction. But large training sets are typically unavailable in the medical field due to patient privacy concerns. The OpenI was the largest dataset that have 3,955 radiology reports and 7,470 chest x-rays. In recent years, ChestX-ray14, an archive containing more than 30,000 patients and 112,120 classified chest x-ray images, was released [1]2. Despite being unbalanced and defective, ChestX-ray14 is sufficiently large for training, so we select it as our dataset.

2For processing medical images, including chest x-ray images, transfer learning is considered as the preffrered method in numerous publications for the classification task [21]. Researchers train ChestX-ray14 by improving existing ImageNet-trained deep neural networks. Four standard CNN architectures (AlexNet, VGGNet, GoogleNet, and ResNet) were fine-tuned in the only peer-reviewed publication, and ResNet obtained the best results [1], [22]– [24]. However, fine-tuning existing deep neural networks is not always the optimal solution for low transmission performance caused by the variation between medical and natural images. Moreover, excessive parameters may lead to overfitting and unnecessary space consumption. ChestX-ray14 is sufficiently large for the research community and necessitates the training of a customized lightweight model without excessive time or memory consumption. In this work, we propose a novel CNN architecture, FEC-NET, and train it on ChestX-ray14 from its founding.

# Related Work

1The usage of these algorithms has increased significantly from the first use of AI approaches in medicine in the 1980s, notably in recent years. The techniques of deep learning have been applied to all kinds of clinical data [17]: biosignals, involving electrical [18], [19], mechanical [20], [21] and heat signals [22], [23]; biological medicine, examining the molecules of biological function [24], [25], [26]; electronic health records (EHR), aimed at improving diagnosis [27], [28], [29], [30]; and clinical scans, widely used in the identification of multiple disease [31], [32], [33], [34], as is the case with our problem. Evidence-based medicine has replaced observation-based medicine in the delivery of healthcare. Due to their ability to recognize some radiological signals that medical personnel are unable to notice, deep learning and big data algorithms are particularly helpful in this sector [35]. Although the focus of this work is related to classification, there are papers that use these algorithms for regression issues, including estimating drug dosage [36], producing medical reports from clinical trials [37], assisting healthcare management [38], or image processing, including segmentation [39] and reconstruction of images [40].

1The study into the use of deep learning and machine learning in medical image processing has been significantly influenced by the COVID-19 epidemic. As might be predicted, a large portion of the analyzed categorization schemes have concentrated on identifying symptoms of bilateral COVID-19-associated pneumonia. In Ahmed et al. [4], they categorize chest X-rays using VGG16 and ResNet, two distinct pre-trained structures, and optimize the hyperparameters. In Pham [41], they train three alternative pre-trained architectures—AlexNet, GoogleNet, and SqueezNet—with six datasets separately, evaluating various train set sample sizes (50 and 80%), and SqueezeNet achieves an accuracy of 99.85%. The ensemble system developed by Ahmad et al. [42] obtains 96.49% accuracy and is based on MobileNet and InceptionV3. The ability to distinguish whether pneumonia is brought on by COVID-19 or another virus/bacteria or if the patient is healthy was soon made feasible by the extension of binary classification to multiclass situations. Similar to binary classification issues, several studies, including that by Avola et al. [43], search through the most effective topologies, including AlexNet, GoogleNet, ResNet, and ShuffleNet, among others. On a dataset of 6330 samples, MobilNet\_v3 produces the hoghest result with an accuracy of 84.92%. Zebin and Rezvy [44] train a pre-trained model in addition to generating a heatmap-based visualisation that displays two pictures for each sample. First is for the orignal X-ray image and second for the class activation areas which is more important for classification model. However, because the pictures do not overlap, interpretation is challenging. Other studies, like Teixeira et al. [45], use segmentation approach to eliminate any system components that are not necessary, which should to enhance performance and visualization. Three separate datasets make up their dataset labelled as COVID-19, normal, and lung opacity.

1Although samples of a specific ailment and healthy samples make up the majority of medical datasets, chest X-rays frequently show symptoms of many pathologies. Due to this, several scientists have released multilabel radiological datasets during the past five years. With the extra difficulty of class imbalance, these datasets are more realistic than binary ones. The prevalence of disease in society should determine the proportion of each class in a real dataset, meaning that certain classes should be greater in number than others. These datasets' intriguing qualities must be carefully examined in order to properly solve the issue. Since its release in 2019 [54], the ChestX-ray 14 dataset has become one of the most frequently used datasets in the area of chest X-ray classification. By using DenseNet-121 to optimize its hyperparameters, for instance, Wang et al. [56] were able to achieve an average AUC of 0.82. The pneumonia class's AUC was 0.662, the lowest possible, while the hernia class' AUC was 0.923, the best possible. Other researchers have achieved an AUC for pneumonia of 0.73 using other designs like InceptionResNet\_v2 and ResNet152\_v2 [57]. A lot of the work on this dataset involves retraining cutting-edge architectures, however there are various methods for enhancing classification performance. For instance, Almezhghwi et al. [58] switched the classification algorithm from AlexNet and VGG16 to SVM in an effort to enhance the findings of earlier manuscripts, and both architectures achieved an AUC for pneumonia of 0.98. The performance of the classification model can be increased with different types of X-ray images or with the customized deep learning models.

2Because of the property that demands a large amount of labelled training data and a considerable degree of experience to achieve convergence, training a deep convolutional neural network (CNN) from inception is difficult [43]. Fine-tunned networks that have already been trained on huge datasets are a viable option. For semantic segmentation, the study [44] fine-tuned all layers using backpropagation over the whole FCN-AlexNet, FCN-VGG16, and FCN-GoogLeNet. The study [45] used a self-paced fine-tuning network (SPFTN)-based architecture to separate objects into categories in movies with sparse labelling. A fine-tuning approach to update the network that was previously trained on photos of urban environments was provided in a study [46], and it successfully transfers semantic properties to a different context. Regarding medical pictures that diverge greatly from natural ones, several studies have varied opinions. The study [47] showed that in the multi-class classification evaluation of knee osteoarthritis, fine-tuning considerably outperformed feature extraction from scratch. However, [48] showed that in the categorization of cytopathology images, a novel CNN that extracts features surpassed fine-tuning.

# Proposed Methodology

The conventional method employed to enhance the performance of image classification entails the stacking of convolutional layers within convolutional neural networks (CNNs). Although this technique has proven effective to some extent, it brings about certain drawbacks that limit its overall efficacy. Layer-stacking in traditional image classification methods significantly increases parameters and computational complexity, posing challenges for scalability, real-time processing, and resource utilization. Alternative approaches are needed to mitigate these drawbacks while maintaining or improving performance. The elevated parameter count and computational demand can hinder real-time processing capabilities and impede the scalability of the network architecture. Moreover, the traditional CNNs such as VGG [], ResNet [], and AlexNet[], use the single kernel size which results in the loss of fine-grained features important for analyzing abnormal spots found in chest X-ray images. Consequently, the compromised retention of significant feature details adversely affects the accuracy and robustness of disease recognition algorithms utilized within the context of chest X-ray classification.

To enhance chest X-ray image classification, we introduce FE-unit, a feature extraction module comprising three convolutional units designed to address the identified challenges. RIBFE-unit emphasizes network width to overcome limitations in existing methods. The first convolutional block employs two 1D 3 Conv receptive fields, achieving an equivalent 5 × 5 receptive field. This configuration enables the detection of larger disease spots, capturing global feature information vital for disease recognition. The second unit utilizes 1D 3 Conv convolutions to detect smaller disease spots and gather local feature information. These units together reduce information loss and enhance recognition accuracy. To address issues of gradient vanishing and accelerate optimization, the third unit adopts a residual structure [1]. It overcomes depth-related degradation by using a mapping relationship expressed as H(x) = F(x) + x, allowing small weight adjustments for optimal network performance. 1D 1 Conv convolutions are employed in all units to enhance compatibility with input data, facilitating downsampling, upsampling, and flexible adjustment of output size. Figure 1 shows the novel Inception unit and modified inception unit called FE-unit.

Novel Inception structure tries to reduce the model parameters but still suffers from overfitting in complex image scenarios like medical image classification where data is always limited. To overcome these challenges associated with the Inception network, we have devised an alternative Chest diseases classification network called FENet. FENet consists of three simplified FE-unit1, each designed to address specific shortcomings of the Inception network. Specifically, the FE-unit1 produces 512 feature maps, the FE-unit2 generates 1024 feature maps, and the FE-unit3 yields 2048 feature maps. This hierarchical design enables the feature extraction module, consisting of the three FE units, to perform a localized perception of Chest disease present in the input images. By leveraging the varying channel outputs and increasing feature maps, the network becomes capable of capturing a diverse range of local feature details pertaining to Chest diseases. Following the extraction of local features, the FENet incorporates higher-level synthesis operations to integrate the acquired local information and derive comprehensive feature representations. This synthesis process aggregates and synthesizes the local feature information from multiple FE-Blocks, allowing the network to capture more holistic and discriminative representations of Chest diseases. By combining localized perception with higher-level synthesis, the FENet effectively captures both detailed local features and their contextual relationships, enhancing the overall capability of the network for accurate Chest disease classification. By incorporating the FE-Block module, the resulting FE-Net network possesses a broader network structure, enabling the selection of fine-grained information while reducing the model complexity. Figure 2 illustrates the structure of FE-Net.

The FE-Net (Feature Extraction Network) encounters challenges in accurately detecting Chest disease due to the distinctive characteristics of Chest disease spots. These characteristics manifest as greater inter-disease and minimal intra-disease differences. The presence of such variations in disease patterns complicates the accurate identification and classification of Chest diseases within the FE-Net network. Overcoming these challenges requires sophisticated feature extraction mechanisms capable of capturing the subtle differences between different Chest disease types. We propose integrating a modified Convolution Block Attention Module (CBAM) [2] into the FE-Net network to reduce the redundancy in the information. The creation of a channel attention map is facilitated through the utilization of the inter-channel associations within features. Given that each channel within a feature map is construed as a feature detector, the focus of channel attention is on discerning the significant elements of the input image. The reduction of the spatial dimension of the input feature map is instrumental in computing channel attention efficiently. It has been generally recognized that average-pooling is an effective method for the consolidation of spatial information." This module enhances disease spot localization, leading to improved recognition accuracy. The modified CBAM module seamlessly integrates with the FE-Net network as a channel and spatial attention mechanism. The channel attention module operates on an input feature map encompassing height (H), width (W), and channel (C) dimensions. Through average pooling and maximum pooling, the feature map (F) is transformed into distinct feature representations, namely and . To preserve disease region features, the original Dense layers are replaced with two 1D Conv layers, which serve as feature information detectors rather than dimensionality reducers. Qualitative analysis of kernel size ensures effective detection of disease spot information across various sizes in and . The feature representations are undergoing a join operation and generates the channel attention map . This map is fed to spatial attention to produce respective output map Experimental observations indicate that reflects scores associated with the diseased region and non-diseased region in the original representation Z, reflecting the proportion of such region within the image. To preserve unaffected convolution calculations in subsequent modules, a weighting operation is introduced involving a channel-wise weighting of output by multiplying it by the original image, to focus on the crucial of lesions. This magnifies disease region information while suppressing non-disease features and eventually enhancing robustness. The modified attention module is depicted in Figure 3, while Figure 4 illustrates the modified convolutional attention module.

To further enhance the performance of our FE-Net for Chest regions with varying shapes and sizes and considering the limited adaptability of the CBAM [2] to FE-Net, we propose the FEC-Net. FE-Net given the limitation of original CBAM in generalization. FEC-Net is built upon the foundation of FEC-Net, incorporating modified CBAM. The middle layer of FEC-Net employs four FE-Blocks, and the output features from each FE-Block undergo further feature extraction using modified CBAM, ensuring a more fine-grained representation of the input X-ray image. The SoftMax activation is employed in the final stage to compute the recognition probabilities for each disease category. FEC-Net operates in both the channel-wise and spatial domains to get finer identification of Chest diseases and effectively filters non-disease feature information, thereby enhancing the accuracy of disease identification. Figure 6 provides an illustration of the FEC-Net network architecture.

# Results and Discussion

## Dataset and Preprocessing

ChestX-ray14 is the most useful database that is currently available for training a classification algorithm for chest X-ray images. The collection contains information on almost 30,000 individuals, 297,541 labeled chest X-rays, and 14 distinct categories of abnormal images. The statistical information on the dataset can be found in Table #. On the other hand, as can be shown in Table #, there is a significant disparity in the number of chest X-ray images that belong to each of the categories in Chest X-ray 14. Due to the fact that some of the images in the dataset have multi-label data, there are more than 15 categories present in those images. The work that was proposed chose images that had any anomaly, and the number of samples that had a anomaly label and belonged to each category is shown in Table #. In addition to this, each and every image was resampled to the dimensions of 224 by 224.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Category | Infiltration | Effusion | Atelectasis | Nodule | Mass |
| Number of Samples | 19894 | 13317 | 11559 | 6331 | 5782 |
| Category | Pneumothorax | Consolidation | Pleural Thickening | Cardiomegaly | Emphysema |
| Number of Samples | 5296 | 4643 | 3385 | 2778 | 2516 |
| Category | Edema | Fibrosis | Pneumonia | Hernia |  |
| Number of Samples | 2305 | 1686 | 1431 | 227 |  |

## Training and Testing

We split the Chest ChestX-ray14 dataset into the training, testing and validation set with the ratio of 70%, 20% and 10% respectively. The 208278 samples of the training set and 29745 samples of the validation set were used for the training of the model. Two different architectures of the proposed model (proposed model without attention and proposed model with attention blocks) were used to classify the chest X-ray images. Each architecture was trained multiple times on different hyperparameters. Each time the X-ray images of shape 224\*224 were used as input. The model showed the best results with Adam optimizer with 0.003 learning rate, 64 batch size and categorical cross entropy as loss function. All the experiments were performed in the virtual environment of Python with the TensorFlow framework. Moreover, the GeForce GTX 1080 of 11GB on the Ubuntu operating system was used to perform the experiments. By following every epoch, validation samples were used to evaluate the performance of the model. The decay was also set by a factor of 2, 5 and 10 after every 25 epochs. In each experiment, the model was trained on 100 epochs with a training and validation set.

## Classification Results

The proposed model was trained without an attention mechanism and with an attention mechanism on multiple hyperparameter values. Both models showed 0.78% and 0.8% training accuracy during the training of the proposed model. The accuracy and loss curve on the training and validation set during the training of the model is shown in Figure #. By following the training of the model, 59,508 test samples of the chestXray14 dataset were used to evaluate the performance of the model. The evaluation process showed that the proposed model with attention mechanism outperforms by the other variants of the proposed model or by the available pretrained model. The trained model showed the 0.98% highest accuracy score for hernia class and 0.68% lowest accuracy score for pneumonia class (Table #). The overall accuracy of the proposed model for ChestX-ray14 dataset was significantly considerable compared to pretrained model on ImageNet dataset. The complete comparison of the achieved results is shown in Table #.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

The overall accuracy of the proposed model for ChestX-ray14 dataset was significantly considerable compared to pretrained model on ImageNet dataset. The complete comparison of the achieved results is shown in Table #.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

The results of the Table # showed the performance of the proposed model over the pretrained model with parameter tunning options. The results reveal that the utilization of already trained deeper model with parameter tunning is less significant over the newly designed CNN model. The chest X-racy images for anomalies classification are requiring the light weight CNN model with timely attention of integrating the convolutional layers in the model. The proposed model also contributed in this regard to achieve the good accuracy score compared to the available models. We also achieve the best accuracy score (Table #) with our proposed attention based FEC-NET. There is need to investigate the weights of the model to tunned the model in future for better experimental results.

# Conclusion

In this study, we used deep learning approaches, concentrating on convolutional neural networks (CNNs), to meet the pressing demand for improved technology to detect abnormalities in chest X-ray pictures. CNNs have gained popularity in the field of computer vision because to their outstanding performance in a variety of classification tasks, including the automated analysis of X-rays. When used for medical image processing, the current neural networks suffer from overfitting and poor transferability. To effectively classify chest X-ray pictures, we presented an attention-based CNN model named "Feature Extraction Classifier Network" (FEC-NET). Customized inception blocks were inserted into our model to extract pertinent characteristics, and attention mechanisms were added to improve the model's capacity to concentrate on prominent areas of the images. Notably, compared to current pre-trained models, FEC-NET showed greater performance despite being smaller and lighter. We conducted considerable experimentation, and the findings were compelling, validating the efficacy of our suggested FEC-NET paradigm. When compared to earlier methods, the model showed considerable gains in accuracy, recall, F1 score, and AUC values. Our model has the capacity to accurately detect anomalies in chest X-ray pictures as seen by the accuracy score of 0.81% that was acquired. This result represents a significant improvement in classification performance. We stress that creating an appropriate CNN architecture, such as FEC-NET, is superior to focusing only on deep network optimization. Personalized feature extraction blocks and attention methods may help the model successfully extract pertinent data from chest X-ray pictures, according to our findings. Additionally, our work emphasizes the need of expanding the training data set to improve CNN performance in medical image classification tasks. To capture the intrinsic complexity and diversity contained in medical pictures and enable the model to build robust representations and enhance its accuracy, sufficient training data is crucial.