Disease Detection in Chest X-Rays Using CNN with Squeeze-and-Excitation Block

Sebastian Carp[1]

Other group members:
Marco Gizzi[2], **Hanzla Ilyas**[3], **Harmanpreet Singh**[4]

Abstract. This study explores the application of an enhanced Convolutional Neural Network (CNN) model integrated with a Squeeze-and-Excitation (SE) mechanism for disease detection in medical X-ray images. The model leverages a pipeline of data augmentation, grayscale image preprocessing, and k-fold cross-validation to improve classification accuracy and robustness. By integrating SE blocks, the model dynamically recalibrates feature maps to emphasize relevant image features, boosting diagnostic performance. Comprehensive experiments using the NIH chest X-ray dataset demonstrated the model's effectiveness, achieving an average test accuracy of 67%+ across folds. These findings highlight the potential of SE-enhanced CNNs for early and reliable disease detection in healthcare applications.

1 Introduction

Artificial Intelligence (AI) in healthcare is transforming disease detection, diagnosis, and treatment by leveraging complex algorithms and large data sets to identify diseases earlier and more accurately. This can lead to better patient outcomes, lower healthcare costs, and more efficient use of resources.

However, challenges remain, such as the need for high-quality, diverse datasets to avoid biases and ensure fairness in healthcare outcomes. Additionally, the lack of transparency in AI models can hinder clinician trust and adoption, particularly when the decision-making process is unclear [7].

2 Background

2.1 Introduction to Alternative Models

To address disease detection in medical imaging, my teammates explored several advanced deep learning models. ResNet50 improves accuracy by using skip connections to avoid the vanishing gradient problem, maintaining feature map integrity in deeper layers. DenseNet121 ensures efficient parameter use and strong feature propagation by connecting every layer to all subsequent ones, enhancing performance. Additionally, autoencoders support unsupervised feature learning, helping with pretraining and reconstruction of complex medical images for classification tasks.

2.2 Background on Simple CNNs

A simple Convolutional Neural Network (CNN) is foundational in deep learning for image analysis. It consists of convolutional layers to extract spatial hierarchies, pooling layers to reduce dimensionality, and fully connected layers for final classification. CNNs are highly effective in capturing spatial and hierarchical features from medical images but often struggle with prioritizing relevant features, particularly in complex datasets like X-rays [4].

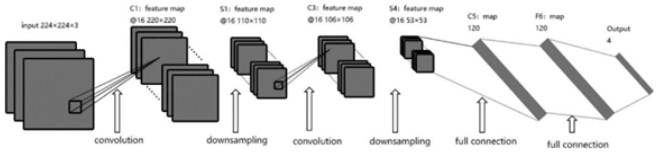

Figure 1. CNN Architecture [1]

2.3 Background on SE Implementation

The Squeeze-and-Excitation (SE) mechanism, introduced in SE-Nets, refines CNN performance by adaptively recalibrating channel-wise feature responses. The squeeze operation performs global average pooling to capture channel-level global context, while the excitation step learns channel interdependencies through fully connected layers. The recalibrated feature maps emphasize informative channels, significantly improving the model's ability to focus on disease-relevant features in medical images [4].

3 Methodology

To better generalize the data fed during training, my teammates have chosen to use pre-trained models; however, a basic CNN can have tremendous potential if it is modified appropriately for the task at

[1] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: sc5376d@greenwich.ac.uk
[2] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: mg1232y@greenwich.ac.uk
[3] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: hi3686c@greenwich.ac.uk
[4] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: hs5970g@greenwich.ac.uk

hand. The general formula for a single convolutional layer is as follows:

$$y_{i,j,k} = \sigma\left(\sum_{m=1}^{M}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1} x_{i+u,j+v,m} \cdot w_{u,v,m,k} + b_k\right) \quad (1)$$

This will appear as follows when used with the SE block, which is intended to enhance CNNs by adaptively recalibrating channel-wise feature responses:

$$\tilde{y}_{i,j,k} = s_k \cdot y_{i,j,k} \quad (2)$$

Where s_k is the channel-wise excitation scalar derived as:

$$s_k = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot z + b_1) + b_2) \quad (3)$$

And z_k is the squeezed representation of the feature map:

$$z_k = \frac{1}{H_{\text{out}} \times W_{\text{out}}} \sum_{i=1}^{H_{\text{out}}}\sum_{j=1}^{W_{\text{out}}} y_{i,j,k} \quad (4)$$

The implementation has been made possible using the PyTorch framework, multiple pieces and concepts from GitHub Repositories, along with various structural ideas given by OpenAI [3]

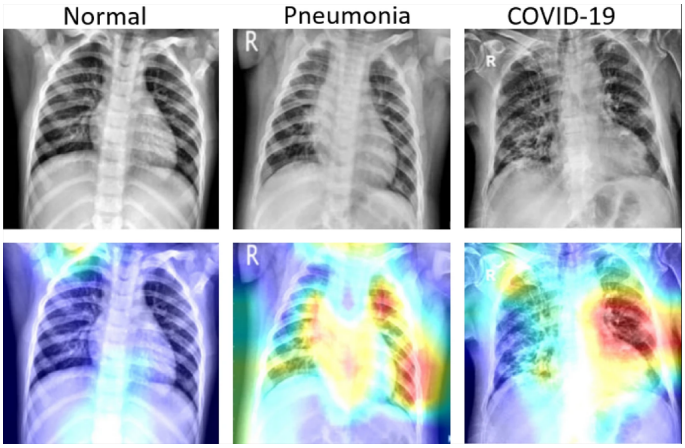

Figure 2. SE Architecture [6]

4 Experiments and Results

4.1 Experimental Settings

The training and testing were conducted on a local system using an Intel Core i7-7700 CPU @ 3.60GHz.

- **Initial Learning Rate**: The learning rate is set to 1×10^{-4} using the Adam optimizer. Additionally, a StepLR scheduler reduces the learning rate by a factor of 0.7 every 2 epochs to ensure gradual and stable training. The learning rate controls how much to update the model in response to estimated error.[5]
- **Batch Size**: The batch size is set to 32. This means 32 training examples are processed in one forward pass through the network. This smaller batch size is used to balance memory usage and model convergence.[5]
- **Epochs**: The training is conducted over 30 epochs, which represents the number of times the model iterates through the entire dataset. This choice reflects the computational constraints of the training environment and is smaller than typical deep learning experiments (e.g., 200 epochs).[5]
- **Weight Decay**: A regularization technique is used to minimize the loss function by penalizing large weights. The Adam optimizer implicitly supports weight decay, contributing to generalization during training.[5]
- **5-Fold Cross-Validation**: The dataset is split into 5 folds, with one fold used as a validation set and the remaining 4 used for training. This process is repeated 5 times, each time with a different validation fold, ensuring a robust evaluation of the model.[5]
- **Image Transformations**: Transformations include resizing to 224×224, grayscale conversion, Gaussian blur for denoising, and normalization with a mean and standard deviation of 0.5.[5]
- **Validation**: Model performance is evaluated on the validation fold after each epoch. Accuracy and loss are calculated to monitor training progress. The best-performing model is saved for each fold based on the highest training accuracy.[5]

4.2 Evaluation Criteria

The NIH Chest X-rays dataset (˜5,600 labelled images) was used for training and evaluation. Evaluation metrics include Validation Accuracy and Loss for cross-validation performance, and Training Accuracy for model performance during training [2]. Testing Accuracy was assessed using 1,000 filtered images compared to ground truth. The goal is to minimize validation loss while achieving high accuracy during training and testing to ensure strong generalization.

4.3 Results

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
Simple CNN	Up to 99%	Up to 90%	50%
Improved CNN	Up to 99%	Up to 63%	57% - 62%
Improved CNN + SE	Up to 99%	Up to 63%	66%

Table 1. Model performance across different stages.

4.4 Discussion

The results from training and testing the Improved CNN with the SE mechanism show promising progress, with enhancements at each stage of development.

Training and Validation Accuracy: The model shows consistently high training accuracy across epochs, indicating effective learning. However, the validation accuracy is 5-10% lower than training accuracy, suggesting that the model is generalizing well but may benefit from regularization techniques like dropout or early stopping to prevent overfitting and improve generalizability.

Improvement with the SE Mechanism: Incorporating the SE block significantly improved testing accuracy. Initially, the simple CNN struggled with complex datasets, but the SE block's adaptive reweighting of channel features enabled the model to focus on more relevant aspects, improving its ability to distinguish between classes. This is crucial in medical image analysis, where subtle feature differences matter.

Model's Behaviour on the Test Set: When evaluated on a separate test set, the model correctly classified approximately three out of four images, showing a positive but improvable result. The slight decline in accuracy suggests challenges with specific images, possibly due to variations in image quality, artefacts, or the difficulty in differentiating between certain chest illnesses.

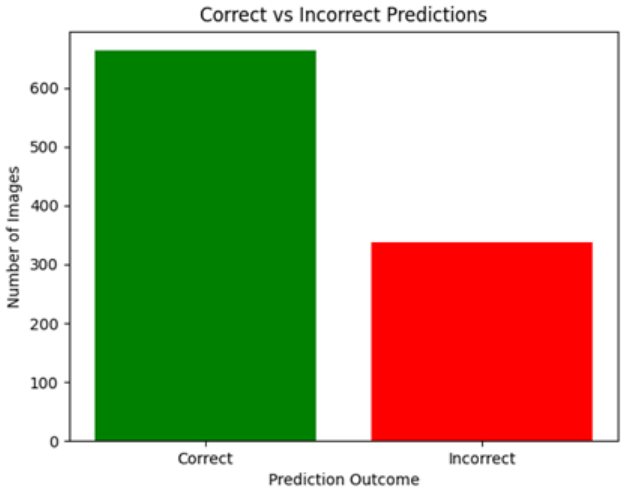

Figure 3. Model Results

5 Conclusion and Future Work

In conclusion, the CNN + SE model showed improvements in handling the complex NIH dataset compared to other models. The final results were satisfactory, but further work can enhance test accuracy through more training data, regularization to prevent overfitting, and better hardware, such as a GPU, to speed up training and gain more room for experimentation.

Additionally, integrating AI in healthcare requires addressing legal, social, ethical, and professional issues (LSEPI), ensuring privacy, fairness, transparency, and accountability in AI-driven decisions.

ACKNOWLEDGEMENTS

I would like to thank my teammates for helping me with ideas from their work and moral support, my professors and lab tutors for always providing insightful feedback on any questions and to the providers of the NIH dataset that helped so many people carry out research and find solutions to this real-world problem.

REFERENCES

[1] Guanxiong Feng, Bo Li, Mao Yang, and Zhongjiang Yan, 'V-cnn: Data visualizing based convolutional neural network', in *2018 IEEE international conference on signal processing, communications and computing (ICSPCC)*, pp. 1–6. IEEE, (2018).
[2] National Institutes of Health (NIH). Random sample of nih chest x-ray dataset, 2017.
[3] OpenAI. Chatgpt, 2024.
[4] Hannes Stärk, Octavian Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola, 'Equibind: Geometric deep learning for drug binding structure prediction', in *International conference on machine learning*, pp. 20503–20521. PMLR, (2022).
[5] Adel Sulaiman, Vatsala Anand, Sheifali Gupta, Yousef Asiri, M. A. Elmagzoub, Mana Saleh Al Reshan, and Asadullah Shaikh, 'A convolutional neural network architecture for segmentation of lung diseases using chest x-ray images', *Diagnostics*, **13**(9), (2023).
[6] Zahid Ullah, Muhammad Usman, Siddique Latif, and Jeonghwan Gwak, 'Densely attention mechanism based network for covid-19 detection in chest x-rays', *Scientific Reports*, **13**(1), 261, (2023).
[7] Qamar Zaman et al., 'The role of artificial intelligence in early disease detection: Transforming diagnostics and treatment', *Multidisciplinary Journal of Healthcare (MJH)*, **1**(2), 43–54, (2024).