Academic Journal of Research and Scientific Publishing | Vol 5 | Issue 54
Publication Date: 05-10-2023 ISSN: 2706-6495

Detection of Pneumonia from Chest X-ray Images Using Transfer Learning and Ensemble Learning

Bahatheq Tariq Ahmed S[1], Jayanth Balasubramanian[2], Koh Shang Hui[3]

Undergraduate, Computer Science, National University of Singapore, Singapore [1,2,3]

Email: Tariqbahatheq@outlook.com

https://colab.research.google.com/drive/1YGE9p2B4Fgx5dN2omxA0wa4djPkZigpI?usp=sharing

Abstract

Pneumonia, characterized by lung inflammation and caused by various strains of bacteria and viruses, consistently ranks as one of the top three causes of death in Singapore. Despite chest X-ray being the standard imaging test for pneumonia, its diagnostic accuracy has limitations due to factors like low specificity and subjective variance among radiographers. This research aims to enhance the diagnosis by predicting pneumonia based on chest X-ray images sourced from a Kaggle dataset with 5856 images, primarily from pediatric patients at the Guangzhou Women and Children's Medical Center. A novel approach modifies the Kaggle data split to introduce more validation data and employs data augmentation techniques such as random flips and rotations to stabilize the model. Convolutional Neural Networks (CNNs) form the core of the prediction methodology. Simple CNNs are evaluated, followed by transfer learning models using ImageNet architectures. The models are assessed based on their weighted F1 scores, with a special emphasis on recall to ensure critical cases aren't missed. Ensemble techniques, such as voting, are explored to further enhance prediction capabilities and robustness. The paper culminates by comparing these models against existing frameworks and provides insights into their potential application in real-world medical diagnostics to further enhance the medical field.

Keywords: AI, Machine learning, Deep Learning, Neural Networks, Convolutional Neural Networks, Transfer learning, Ensemble models, Explainable AI, Classification, Pneumonia, AI & Medicine

1. Introduction:

Pneumonia is an inflammatory condition affecting the lungs, predominantly the tiny air sacs known as alveoli. Typically, these sacs fill with pus or fluid, making breathing painful and limiting oxygen intake. Caused by a plethora of bacteria, viruses, and even fungi, pneumonia presents a significant healthcare challenge globally.

In Singapore, the gravity of this disease is further underscored, ranking consistently among the nation's top causes of death. Traditional diagnostic methods, such as chest radiography, have been employed as the frontline defense to detect this ailment. However, the accuracy of these methods is often marred by various challenges, ranging from the inherent limitations of the imaging technique to the subjective interpretations of the radiographers.

Amidst these challenges, the advent of artificial intelligence and machine learning offers a beacon of hope. By leveraging vast datasets and intricate algorithms, researchers are continuously striving to improve the accuracy and speed of pneumonia diagnosis. This paper delves into one such endeavor, exploring the potential of convolutional neural networks and ensemble learning to detect pneumonia from chest X-ray images. Through a nuanced exploration of methodologies, datasets, and evaluation criteria, this research aims to contribute to the ongoing efforts in combating this formidable disease.

Motivation

Pneumonia consistently ranks among the top 3 principal causes of death in Singapore, as reported by the Ministry of Health. It is characterized by lung inflammation, particularly in the air sacs, and is caused by various strains of bacteria and viruses.

A chest radiography, or chest X-ray, is the standard imaging test for pneumonia. However, it has been noted that the diagnostic accuracy of chest radiography is limited (Self et al. 2013, van den Berk et al. 2022). One reason is its low specificity (Esayag et al. 2010). Additionally, subjective variance among radiographers induces significant uncertainty in chest radiography reports, further complicating clinical diagnoses.

Comparatively, a computed tomography (CT) scan displays slightly better sensitivity and significantly higher specificity. A CT scan is often employed to rule out harder-to-detect cases of pneumonia. However, it remains prohibitively expensive to be used as a first-line test. Thus, an

accurate automated computer-aided diagnosis system using chest X-ray images has both medical and economic benefits.

Dataset

The chest X-ray images used are from a Kaggle challenge, and they are sourced from research done by Kermany in 2018. 5856 images consisting of 4273 positive 'pneumonia' and 1583 negative 'normal' samples were selected from one to five years old pediatric patients at the Guangzhou Women and Children's Medical Center, Guangzhou.

The Kaggle data split was modified to introduce more validation data. 20% of the testing dataset was randomly split and used to supplement the small validation dataset. A fixed seed was used for this split to ensure consistency across all trained models and to facilitate comparison between them. The decision to split the validation set from the testing set was motivated by an observation that the training and testing sets were dissimilar, complicating early attempts at model selection. After the split, more consistency between validation and testing was achieved.

Split\ Label	Normal	Pneumonia	Total
Train	1341	3875	5216
Validation	8	8	16
Test	234	390	624
Total	1583	4273	5856

Figure 1: Structure of the Kaggle dataset

Split\ Label	Normal	Pneumonia	Total
Train	1341	3875	5216
Validation	60	81	141
Test	182	317	499
Total	1583	4273	5856

Figure 2: Structure of the modified Kaggle dataset

Weighted random sampling was applied for the training dataset to account for label imbalances. As before, a generator with a fixed seed was used to ensure uniform training conditions for all models.

The following transformations were applied to the images: resizing, normalization with respect to the mean and standard deviation of the unaugmented training dataset, and a 80% center crop. The center crop removes large empty spaces and focuses on the lungs.

Figure 3: Sample resized chest X-ray image with (left) and without (right) center crop.

Augmentation

To make the model more stable and less prone to noise, new data was generated by augmenting the current data. This was also done to make up for the amount of data we have. The following transformations were applied to the images: random horizontal flip, random rotation, and random brightness change.

Strategy

Batch gradient descent was employed to train all the networks, with a typical batch size of 48 to 64. To speed up convergence, AdamW was used to adapt the per-parameter learning rate. The models' accuracies on the validation dataset were calculated after each epoch. Early stopping was implemented to prevent model overfitting by stopping training if validation accuracy dropped by more than a threshold of 0.1 (which worked best in our testing) over 3 consecutive epochs.

Various learning rate schedulers were employed to adjust the learning rate based on validation accuracy. For ResNet based models, ReduceLROnPlateau was found to work best. This required some fine-tuning: for example, one combination of schedulers given as an example on the PyTorch docs, ExponentialLR followed by MultiStepLR, was found to make the accuracy of the trained Resnet based model worse. Other models were trained using StepLR to reduce the learning rate after a fixed number of iterations.

Evaluating the Models

Many criteria were considered to be used as the basis for choosing the best models. The first option was to choose recall. If the model is used as a diagnostic aid for medical workers, missing a pneumonia case can be critical. The next option was to use accuracy, the most straightforward option given that the dataset was balanced using weighted random sampling. Moreover, it is transparent for workers in the medical field and is widely used in that field. Another option was

precision, or the positive predictive value. For such a model, a positive prediction is most likely to be an actual positive. A potential drawback might be a larger proportion of false negatives.

We reason that solely relying on recall or precision alone is inadequate for this computer-aided diagnosis as the other scores can be very low. A weighted F1 score focused more on recall was chosen. We prefer recall because, for this critical disease, we prefer that the model doesn't miss any cases and classifies them as negative. The beta value chosen in this case is 2.

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$

Baseline CNN

To provide a baseline accuracy with which to compare later models with, a basic deep CNN was trained using the following architecture:

Convolution layers:

Conv (3, 32)→ ReLU→ BN(32)→ maxpool(2, 2)→ Conv(32, 64)→ ReLU→ Dropout(0.1)→ BN (64)→ maxpool(2, 2) → Conv(64, 64)→ ReLU→ BN(64)→ maxpool(2, 2)→ Conv (64, 128)→ ReLU→ Dropout(0.2)→ BN(128)→ maxpool(2, 2)→ Conv (128, 256)→ ReLU→ Dropout(0.2)→ BN(256)→ maxpool(2, 2)

Classifier:

Linear (9216, 128)→ ReLU→ Dropout(0.2)→ Linear(128, 2)

Training Parameters:

Kernel size : 3
Optimizer : AdamW
Learning rate : 0.00001
Loss Function: CrossEntropyLoss
Scheduler : StepLR - size 7, gamma 0.1
Epochs : 20

Figure 4: Architecture for the baseline CNN

Transfer Learning

From 2012 to 2017, increasingly deeper CNN architectures showed drastic improvements in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

The availability of a dataset of over 1.2 million training images, orders of magnitude above past datasets, was a key factor in enabling the training of larger yet more accurate models.

A direct application of top-performing ILSVRC models to the Kaggle dataset is likely to result in model overfitting as the dataset provides only 5216 training images. Instead, an inductive transfer learning was used to leverage existing ConvNet architectures. By replacing only the classifier layers, the hypothesis space of possible models is restricted beneficially.

In total, 7 classifiers were adapted for transfer learning: AlexNet, InceptionNetV3, EfficientNetV2, DenseNet121, GoogLeNet, ResNet50, and VGG19BN. For all models, the final classifier layer was replaced with a three-layer classifier as described in Figure 5. The parameter *n* was varied to match the output size of the previous layers. *k* =128 or 256 and *d* was varied in pursuit of higher accuracy. The model parameters in all other layers were frozen.

Classifier:

Linear(n, 512)→ ReLU→ Dropout(d) → Linear(512,k)→ ReLU→ Linear(k,2)

Training Parameters:

Optimizer : Adam/AdamW

Learning rate : 0.00001 for AdamW, 0.001 for Adam

Loss Function: CrossEntropyLoss

Scheduler : StepLR - size 7, gamma 0.1/ ReduceLROnPlateau

Epochs : 20

EarlyStopper : patience 3, delta 0.1

Figure 5: Classifier architecture and model parameters used for transfer learning

Frequency Domain Learning

Xu et al. noted that existing structures of notable neural networks can be leveraged to perform deep learning in the frequency domain. An exploration of these ideas was performed by analyzing an instance of transfer learning on the frequency space. After a 2D FFT, the modulus of the real and complex parts was taken. The frequency image was then resized to 512x512 and normalized to a mean of 0 and standard deviation of 1. Transfer learning was then conducted using VGG19BN with the classifier and parameters described in Figure 5.

2. Discussion and Analysis

The accuracy, precision and recall of the models tested are shown in Figure 6.

Architecture	Overall Accuracy	Precision	Recall	Weighted F1
Baseline CNN	76.95	73.60	99.37	92.87
Fourier VGG	89.38	89.31	96.53	94.99
AlexNet	91.78	90.12	97.79	96.15
InceptionNetV3	80.76	77.56	98.11	93.17
EfficientNetV2	91.58	91.79	95.27	94.55
DenseNet121	89.78	88.00	97.16	95.18
GoogleNet	89.38	88.82	95.27	93.91
ResNet50	91.58	90.56	96.85	95.52
VGG19BN	90.18	90.12	94.95	93.94

Figure 6: Accuracy, precision, recall, and weighted F1 score (β=2) of different ML models in %

All models tested showed excellent recall with very few false negative identifications. Precision was less noteworthy, with a significant false positive rate. In contrast, there were near-zero false negative identifications.

With the exception of InceptionNetV3, the transfer learning models substantially outperformed the baseline CNN in terms of accuracy. The Fourier VGG model trained on the frequency space was competitive with the traditional transfer learning models.

Evaluating our model against other state-of-the-art models can provide insights into its competence and resource utilization. On Kaggle, the source of our dataset, most models achieve a testing accuracy between 85 and 90%. However, two models outperform the rest. The first is a CNN trained from scratch, achieving 92.63% accuracy (Mathur, 2020). The second employs a fine-tuned ResNet152 model, attaining 91.99% accuracy (Barbosa, 2022).

Our best-performing model achieved 91.78% accuracy, placing it among the top-performing models for this specific dataset. On the other hand, in 2022, Li and Li demonstrated a 99.62% accuracy using ensemble methodologies for COVID-19 pneumonia detection on a separate dataset. As a result, our model may not be suitable for deployment in the medical field as a standalone diagnostic tool. Instead, it could be used to assist radiologists in their decision-making process, rather than serving as the sole determinant.

Explaining Model Outputs

With deep learning models becoming increasingly deep (even the baseline CNN has 26 layer) and complex, such as with skip connections between layers, they become even more opaque in their decision making. The tens or even hundreds of thousands of parameters seemingly renders these networks to be all but a black box classifier. Two problems naturally arise: (1) difficulty in trusting the model's outputs (2) difficulty in monitoring and debugging the model.

In this section, feature (pixel) attribution is examined. Each pixel's contribution to the model's classification, a weaker but useful notion of explainability, is calculated. Formally, given a complex model *f* and an input instance (image) *x*, one aims to

- Map *x* to the simplified input *x'* through the mapping function *h: h(x') = x* (In our case, since the input instance is an image, a sample mapping function h could map a vector of 0s and 1s indicating the presence or absence of a group of pixels to an image (where 0 means that "superpixel" is replaced with the average RGB value of its neighbors).
- Create a simplified model g such that $g(z') \approx f(h(z'))\ \forall\ z' \approx x'$. (Lundberg, Lee 2017) LIME, an additive explanation model, is utilized. It assigns a contribution ϕ_i to each simplified feature $z_i' \in z'$ (where z has M dimensions) such that $g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \quad \forall\ z' \approx x'$

LIME uses penalized linear regression to minimize Least Squared Loss between $f(h(z'))$ and $g(z')$, ensuring that *g* is faithful to the original model *f* around the given input instance (Lundberg, Lee 2017). Figures 7 and 8 show the outputs from running LimeImageExplainer on two of our best models, and ResNet50 and EfficientNet.

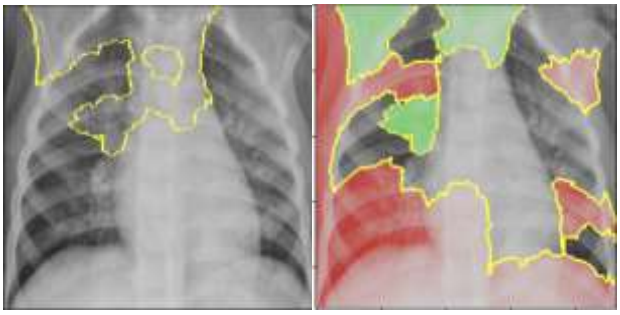

Figure 7: Chest X-Ray of a person with pneumonia. Left: yellow boundary indicates the top 5 features that cause the trained ResNet50 model to classify the image as "Pneumonia". Right: green and red regions indicate features that contributed positively (and negatively, respectively) towards "pneumonia" classification

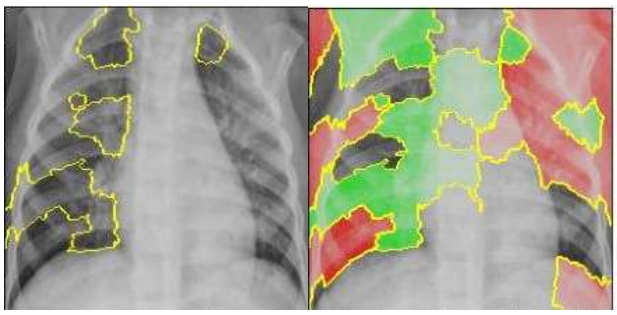

Figure 8: Chest X-Ray of a person with pneumonia. Left: yellow boundary indicates the top 5 features that cause the trained EfficientNet model to classify the image as "Pneumonia". Right: green and red regions indicate features that contributed positively (and negatively, respectively) towards "pneumonia" classification

Despite the identical positive classification, the two models pay attention to different regions to make the classification. Comparing the yellow boundaries in Figure 7 and 8, we can see that the top and middle of the left lung contributed positively to classification as Pneumonia for both models. However, if we compare the red and green regions for the two models, we notice that the models disagree on the contribution of the bottom right of the lung (contributes positively to EfficientNet classification, negatively to ResNet classification). Unfortunately, our team does not have the medical expertise to determine which model is correct regarding this.

From the above discussion, it is clear that pixel attribution using tools like LIME can give more information to a doctor than a binary classification to help a doctor reach an informed conclusion. However, disagreement between the different models suggests that the information could potentially be misleading,

and a similar approach with LIME plots of multiple models could be used by the doctor to find common regions that both models found significant in making a prediction.

Voting Ensembles

The training of multiple independent deep learning models motivates the use of ensemble learning methodologies. A combination of multiple hypotheses may produce higher accuracy than any individual constituent hypothesis.

Bayes classifiers in the form of unweighted voting ensembles were tested first using all 7 trained transfer learning models and next using the top 5 models by weighted F1 score. In light of the significant false positive rate of the transfer learning models, three levels of bias were evaluated. The first, an unbiased voting, was a simple majority algorithm requiring 4 of 7 or 3 of 5 votes. The second required one additional positive vote while the third required two more (effectively a unanimous decision for the ensemble of 5) for an overall positive classification respectively. The stricter requirements for an overall positive result represent an increasing negative bias, or bias towards the negative or 'normal' classification. The results are shown in Figures 10 and 11.

We observe a positive trend in the precision and negative trend in recall as the negative bias increases. This is a result of a decreasing false positive but increasing false negative rate. The unbiased, simple majority ensemble was the best performing ensemble. However, it did not improve upon AlexNet, the best performing transfer learning model. This indicated a high degree of similarity between the transfer learning models.

Ensemble	Overall Accuracy	Precision	Recall	Weighted F1
Simple Majority	91.78	90.35	97.48	95.97
5 or more	92.79	93.00	95.90	95.31
6 or more	92.38	94.86	93.95	94.13

Figure 9: Accuracy, precision, recall and weighted F1 score (β=2) for the unweighted voting ensembles using 7 models in %

Ensemble	Overall Accuracy	Precision	Recall	Weighted F1
Simple Majority	91.98	91.10	96.85	95.64
4 or more	92.18	93.71	93.86	93.83
Unanimous	87.37	95.68	89.41	90.60

Figure 10: Accuracy, precision, recall and weighted F1 score (β=2) for the unweighted voting ensembles using 5 best-performing models (by weighted F1 score) in %

3. Recommendations:

1. **Further Exploration of Transfer Learning**: Given the promising results observed from transfer learning, it is recommended to explore more recent architectures and fine-tuning techniques. This could enhance model performance and adaptability to varied datasets.
2. **Data Augmentation and Expansion**: To mitigate the risk of overfitting, especially when adapting large models to relatively smaller datasets, it would be beneficial to consider more advanced data augmentation techniques. Additionally, collaborations with medical institutions could provide a larger and more diverse dataset, further enhancing model robustness.
3. **Ensemble Learning**: The research demonstrated the potential of ensemble learning in boosting accuracy. Exploring different ensemble techniques, such as stacking or bagging, could potentially lead to even better results.
4. **Explainability and Trust**: As machine learning models become integral to medical diagnostics, ensuring their transparency and trustworthiness is crucial. Further research into model explainability, beyond pixel attribution, is recommended. This could enhance trust among medical professionals and facilitate the adoption of these tools in clinical settings.
5. **Collaboration with Radiologists**: The model, while promising, should not be viewed as a standalone diagnostic tool, given the current accuracy levels. Collaborations with radiologists could provide insights into practical challenges and needs, leading to models better tailored for real-world application.
6. **Frequency Domain Analysis**: The exploration into frequency domain learning showed potential. It's recommended to delve deeper into this area, perhaps looking at different transformation techniques or leveraging domain-specific knowledge

4. Conclusion and Future Work

Based on the weighted F1 scores from all the models and ensembles, AlexNet achieved the highest score.

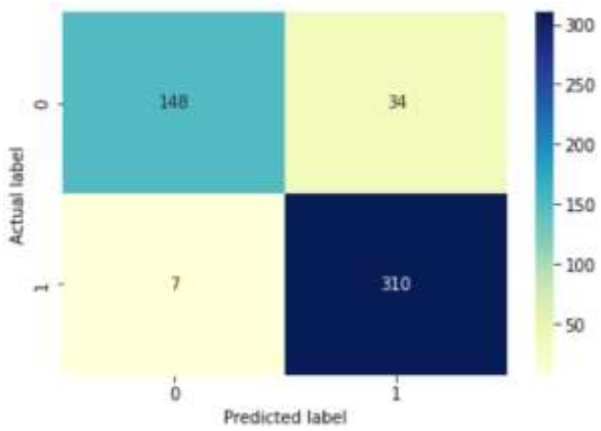

Figure 11: Confusion matrix for AlexNet

Further classification of positive pneumonia cases by cause (bacterial or viral) can be explored. While the dataset explanation on Kaggle notes that a typical bacterial pneumonia exhibits a local anomaly in the chest X-ray and a viral pneumonia is characterized by diffuse patterns, it is difficult for an untrained eye to differentiate the two in practice.

Many new large models being deployed in 2022 and 2023 claim to be extremely efficient on data. These include the recently released and commercially available Amazon Titan foundation models. Training using such models can be explored to further enhance the results of the project at the expense of the project budget.

There is ample room for fitting other architectures from ILSVRC or elsewhere into the frequency space. In 2021, Han and Hong introduced specialized Fourier CNNs and experimented with shallow networks. Adapting their ideas to deeper networks may yield better results.

5. Ethical Consideration:

This research prioritizes the ethical standards integral to scientific endeavors, particularly in the realm of medical research. Herein, we outline the key ethical considerations that were adhered to:

1. **No Human Experiments**: At no point did this study involve direct experiments on humans by the authors. The research was fundamentally computational, focusing on the analysis of pre-existing datasets.
2. **Data Sources**: The data used in this research is sourced from a Kaggle challenge, which in turn was based on research done by Kermany in 2018. It's crucial to understand that all data

used was anonymized and void of any personal identifiers, ensuring the privacy and confidentiality of the patients from whom the X-ray images were derived.

3. **Patient Consent & Approvals**: While the authors did not directly conduct experiments or collect data, it's implicit that the original data collectors sought necessary consent from patients or their guardians, especially given the sensitive nature of medical imaging. Moreover, the necessary approvals from relevant authorities would have been secured by the primary data collectors.
4. **Protocols Followed**: The research strictly followed data handling and analysis protocols to ensure the integrity of the results. Furthermore, while the models and findings show promise, it's crucial to emphasize their supplementary role in medical diagnosis. Decisions based on these findings should be made with caution, ideally in tandem with expert human judgment.
5. **Transparency & Openness**: The research aims to contribute to the broader scientific community. As such, efforts have been made to ensure transparency in methodology, findings, and potential limitations. This open approach facilitates peer review and collective advancements in the field.

By adhering to these principles, this research aims to be both scientifically rigorous and ethically responsible, ensuring that advancements made contribute positively to patient care and the broader medical community.

6. References

Bai, T.; Luo, J.; Zhao, J.; Wen, B.; Wang, Q. (2021). Recent Advances in Adversarial Training for Adversarial Robustness. arXiv:2102.01356.

van den Berk, I. A. H.; Kanglie, M. M. N. P.; van Engelen, T. S. R. et al. (2022). Ultra-low-dose CT versus chest X-ray for patients suspected of pulmonary disease at the emergency department: a multicentre randomised clinical trial. *Thorax*. doi.org/10.1136/thoraxjnl-2021-218337

Esayag, Y.; Nikitin, I.; Bar-Ziv, J.; Cytter, R.; Hadas-Halpern, I.; Zalut, T.; Yinnon, A. M. (2010). Diagnostic Value of Chest Radiographs in Bedridden Patients Suspected of Having Pneumonia. *The American Journal of Medicine*, *123*(1), 88. doi.org/10.1016/j.amjmed.2009.09.012

Han, Y., Hong, B. (2021). Deep Learning Based on Fourier Convolutional Neural Networks Incorporating Random Kernels. *Electronics* 2021, *10*(16), 2004. doi.org/10.3390/electronics10162004.

He, K., Zhang, X.; Ren, S.; Sun, J. (2015). Deep Residual Learning for Image Recognition. arXiv:1512.03385.

Hosna, A.; Merry, E.; Gyalmo, J. et al (2022). Transfer learning: a friendly introduction. *J Big Data*, *9*(102). doi.org/10.1186/s40537-022-00652-w.

Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K. Q. (2016). Densely Connected Convolutional Networks. arXiv:1608.06993.

IBM. (2020). Adversarial Robustness Toolbox: One Year Later with v1.4. https://research.ibm.com/blog/adversarial-robustness-toolbox-one-year-later-with-v1-4. Accessed: 2023-04-18

Kaggle. (2018). Chest X-Ray Images (Pneumonia). https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia. Accessed: 2023 03-11.

Kaggle. (2020). Pneumonia Detection using CNN(92.% Accuracy). https://www.kaggle.com/code/madz2000/pneumonia-detection-using-cnn-92-6-accuracy. Accessed: 2023-04-11.

Kaggle. (2022). Chest X-Ray (Pneumonia) - CNN & Transfer Learning. https://www.kaggle.com/code/jonaspalucibarbosa/chest-x-ray-pneumonia-cnn-transfer-learning. Accessed: 2023-04-11.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. doi.org/10.1145/3065386.

Li, D., Li, S. (2022). An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images. *iScience*, *25*(4). doi.org/10.1016/j.isci.2022.104031. arXiv:1705.07874.

Lundberg, S., Lee, S. (2017). A Unified Approach to Interpreting Model Predictions.

Ministry of Health Singapore. 2022. Principal Causes of Death. https://www.moh.gov.sg/resources-statistics/singapore-health-facts/principal-causes-of-death. Accessed: 2023-04-04.

Self, W. H.; Courtney, D. M.; McNaughton, C. D.; Wunderink, R. G.; Kline, J. A. (2013). High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia. *The American journal of emergency medicine*, *31*(2), 401–405. doi.org/10.1016/j.ajem.2012.08.041

Simonyan, K., Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. (2014). Going Deeper with Convolutions. arXiv:1409.4842.

Tan, M., Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946.

Xu, K; Qin, M.; Sun, F.; Wang, Y.; Chen, Y.; Ren, F. (2020). Learning in the Frequency Domain. arXiv:2002.12416

Team Member Roles

Tariq: Augmentation, GoogleNet, EfficientNet, AlexNet, Inceptionv3, SimpleCNN, DenseNet121, Evaluation, Meeting/Team coordination. Report: Augmentation, Baseline CNN, Evaluating the models, Discussion (part), Conclusion (part).

Jayanth: Wrote main training, validation and testing loops, SimpleCNN, EarlyStopper, ResNet-50, ResNet-34, LIME explainability, adversarial-robustness-toolbox for evasion attacks. Report: Strategy, Explaining model outputs, Improving model robustness

Shang Hui: VGG19BN, Fourier VGG, Unbiased and biased voting ensembles. Report: Abstract, Motivation, Dataset, Voting Ensembles, Conclusion (part), Frequency Domain Learning. Referencing and formatting.

Doi: https://doi.org/10.52132/Ajrsp.e.2023.54.3