PneumoStack: A Novel Approach to Automated Pneumonia and COVID-19 Diagnosis with Chest X-Ray Analysis via Convolutional Neural Networks and Stacked Generalization

Sonya Jin

Motivation

Improving Diagnostic Accuracy and Efficiency with Automated Diagnostic Systems

- Pneumonia accounts for 15% of all deaths of children under 5 years old - single largest infectious cause of death in children worldwide [1]
- Diagnosed with chest X-ray (CXR) analysis (gold standard), along with pulse oximetry, blood tests, physical exam, and medical history evaluation [2]
- Automated diagnosis methods can extract characteristic features of the disease on the CXR scan such as lobar and lobular consolidation, interstitial opacities, and ground glass opacities, minimizing false predictions from human intervention.

The Need to Enhance Current Diagnostic Methods of Pneumonia and COVID-19

- Convolutional Neural Networks (CNNs) have shown unsurpassed success in image classification due to its capabilities of automated unsupervised feature extraction and model preservation with parameter reduction [3]
- For all chest CT scans (n = 424), the accuracy of the two radiologists from China in differentiating COVID-19 from non-COVID-19 viral pneumonia was 80% (338 of 424) and 60% (255 of 424) [4], emphasizing the need for AI applications in medical image analysis
- Recent reports have revealed that RT-PCR has a sensitivity as low as 60%–71% for helping detect COVID-19, while CXRs have a sensitivity of 69% [4], presenting the possibility for CXR analysis rectifying false negative findings in RT-PCR in COVID-19 diagnosis
- Automated diagnosis can reduce child mortality rates in regions where pneumonia is most prevalent - South Asia and sub-Saharan Africa [1] - in the event that trained radiologists are limited
- Automating CXR analysis may expedite diagnosis in improving both accuracy and efficiency, allowing treatment to be prescribed sooner

Relevant Work

- Wang et al. [16] - COVID-Net. Tailored CNN with PEPX design pattern for three-class classification
- Apostolopoulos et al. [17] - VGG-19 as base model for three-class classification
- Umer et al. [18] - COVINet. Sequential CNN approach for three-class classification
- Nishio et al. [19] - VGG-16 for three-class classification with combined data augmentation methods
- Singh et al. [20] - MADE-based CNN for binary classification
- Sahinbas and Catak - ResNet, DenseNet, InceptionV3 transfer learning approach for binary classification

Summary of Current Research:

- Individual transfer learning models and ensembling by methods such as bootstrap aggregation and weighted voting have been explored
- Stacking has also been done, but with the same model, unable to reap the benefits of multiple architectures
  - Meta-learner is also based on logistic regression or other algorithms. Using a CNN may result in higher performance due to its state-of-the-art performance in image classification

Aim

In contrast to other ensembling methods and the use of individual transfer learning models, the aim of this study is to present a stacked CNN meta-learner of transfer learning CNNs with stacked generalization in effort to achieve higher performance than any one of its constituent classifiers and existing individual models in binary and multiclass pneumonia CXR classification.

Introduction (cont.): Transfer Learning

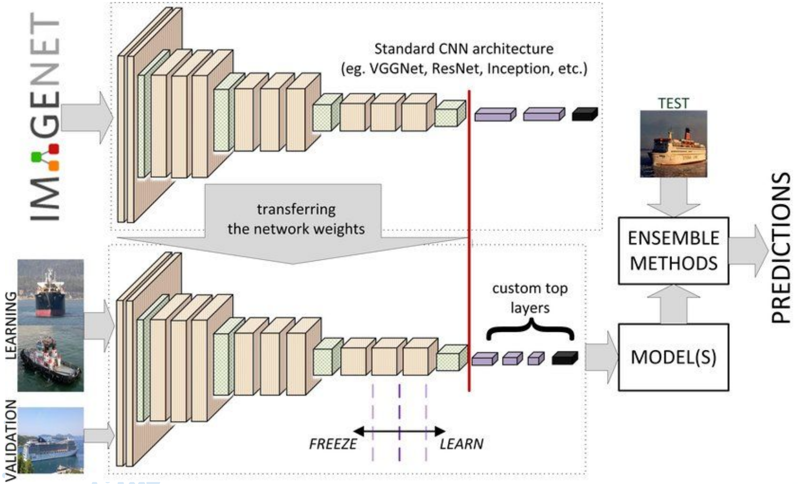


Figure 3. *Ensemble transfer learning using pretrained CNN model initialized with weights trained on ImageNet* [7]

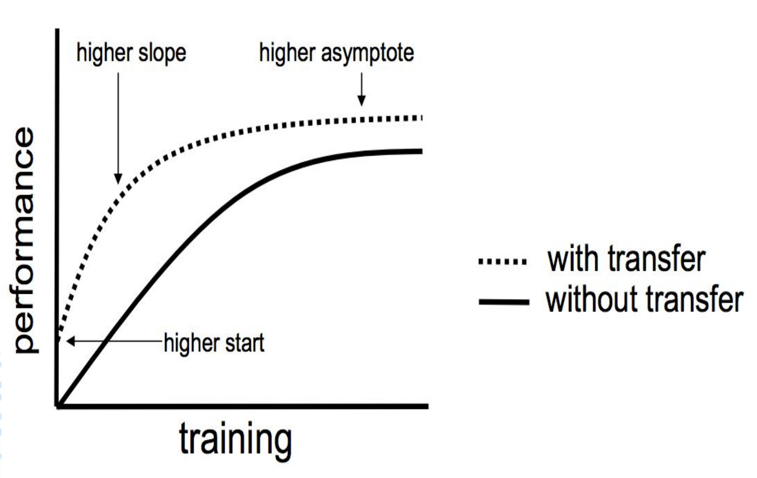


Figure 4. *The benefits of transfer learning* [8]

- ➢ Machine learning method where a model developed for a task is reused as the starting point for a model on a second task [6]
- ➢ Deep learning model pre-trained on the *ImageNet* 1000-class classification competition with 1,000,000 images, optimal for the use with a small dataset to maximize accuracy [6]
- ➢ Early layers - CNN features are generic, later layers - dataset-specific → finetune model [6]
- ➢ Three benefits: higher start, higher slope, higher asymptote [8]

Introduction (cont.) Ensembling via Stacked Generalization

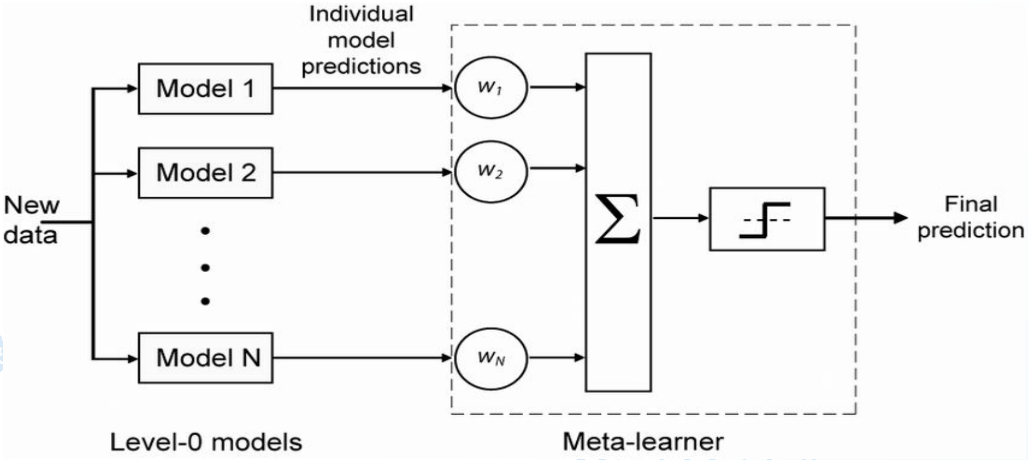


Figure 5. *Schematic diagram of stacked generalization* [10]

- Harness the capabilities of a range of well-performing models on a classification task → make predictions that have better performance than any single model in the ensemble [9]
- **Stacked generalization**: meta-model is trained on the predictions made by base models on out-of-sample data [9]
  - *Level-0 Models* (Base-Models): Models fit on the training data and whose predictions are compiled.
  - *Level-1 Model* (Meta-Model): Model that learns how to best combine the predictions of the base models.
- Three base models chosen: Xception, InceptionResNetV2, ResNet50 after testing multiple models

Constituent Models of Stacked Ensemble

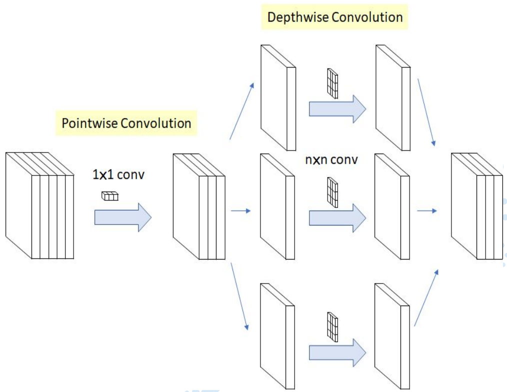


Figure 8. *Depthwise separable convolution* [12]

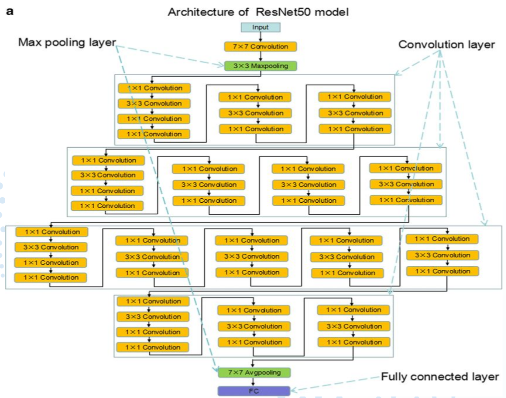


Figure 9. *ResNet50 Architecture* [13]

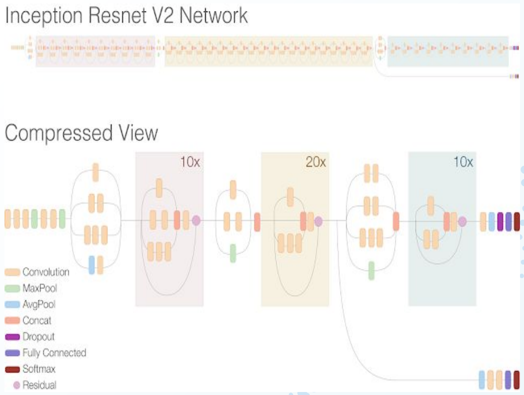


Figure 10. *InceptionResNetV2 Architecture* [14]

Xception

- Modified depthwise separable convolution [10]
  - Pointwise convolution with depthwise convolution
- Outperforms VGGNet, ResNet, and Inception-v3 in ImageNet with 94.5% accuracy [15]

ResNet50

- Skip-wise connections [13]
  - Train extremely deep neural networks with 50+layers successfully
- Winner of ImageNet 2015 with 93% accuracy [15]

InceptionResNetV2

- Residual inception blocks [14]
  - avoids degradation problem and reduces the training time
- 94.6% performance on ImageNet, outperformed Inception V3 and ResNet152 [15]

Proposed Stacked Model

Meta-model

- Trained with stacked dataset -> predictions of constituent models
- Flattened inputs into one vector with Flattening layer
- Added Dense layer with ReLU activation
- **Binary classification**
  - Dense layer with sigmoid activation
  - Compile with binary cross-entropy loss function defined below:

$$Error = \sum_{i=1}^{n} -(p_i \log q_i + (1 - p_i) \log(1 - q_i))$$

- **Multiclass classification**
  - Dense layer with softmax activation
  - Compile with categorical cross-entropy loss function defined below:

$$CCE(p, t) = -\sum_{c=1}^{C} t_{o,c} \log (p_{o,c})$$

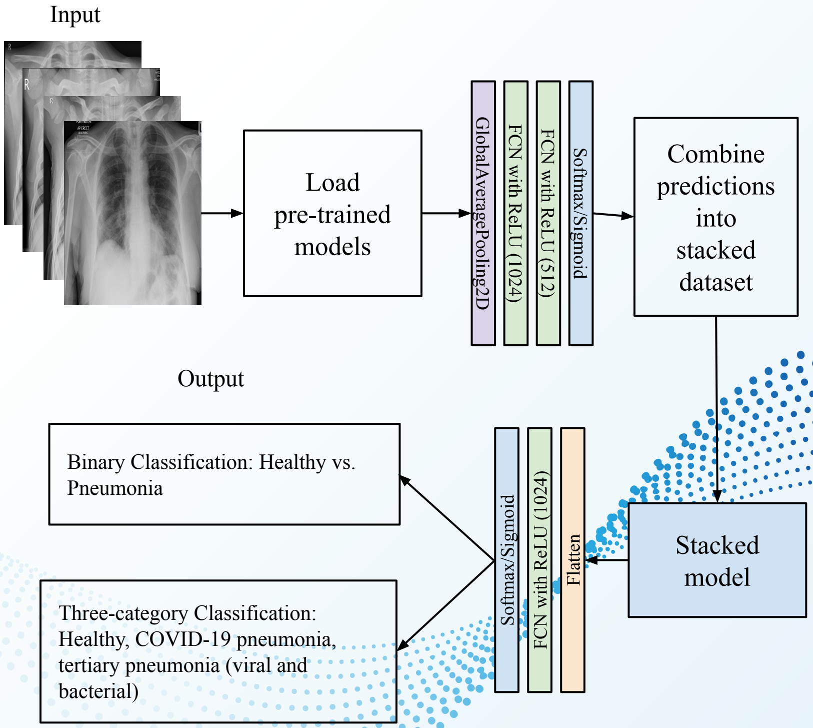


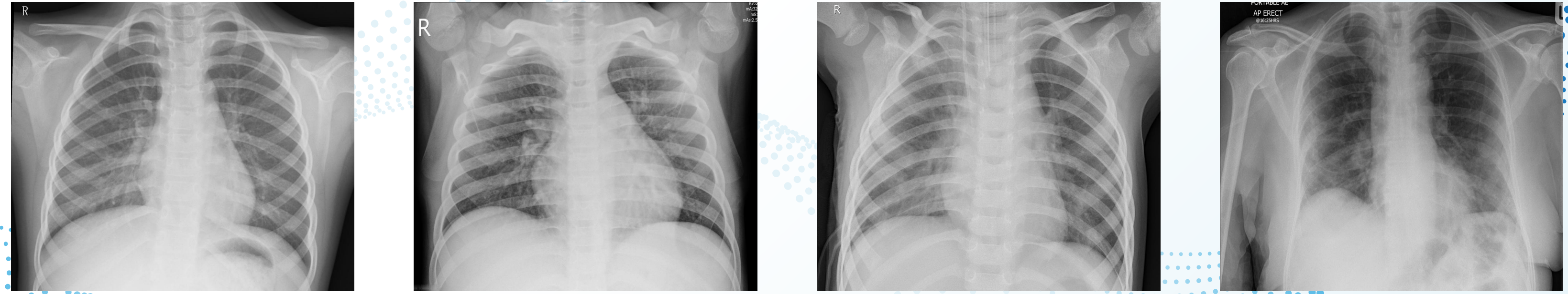
Figure 11. *Schematic Diagram of Proposed Method*

Dataset

- Dataset
  - 5829 X-ray images collected at Guangzhou Women and Children's Medical Center [11]
  - X-ray images of the fungal Pneumocystis pneumonia and lipoid pneumonia were removed for this study

Label	**Number of Images**
Normal	1433
Viral Pneumonia Viruses included: *Influenza, SARS*	1414
Bacterial Pneumonia Bacteria included: *Chlamydia pneumoniae, Streptococcus pneumoniae, E. coli, Klebsiella pneumoniae, Legionella, Mycoplasma pneumoniae*	2521
COVID-19	461

Table 1. *Dataset Distribution*



Normal Viral Bacterial COVID-19

Figure 7. *Data samples*

Methods

1. Dataset was selected and preprocessed with the splitting of training and testing data, class stratification, and augmentation
2. Transfer learning models (Xception, InceptionResNetV2, ResNet50) were constructed, fine-tuned via layer freezing iteration, compiled with binary/categorical cross entropy and ADAM optimization
3. Models were trained on training data
4. Model evaluation
   - 4.1. Binary classification: F1, ROC-AUC, accuracy performance metrics were evaluated
   - 4.2. Multiclass classification: precision, recall, accuracy performance metrics were evaluated
5. CNN meta-learner and stacked dataset consisting of the predictions of constituent models were constructed
6. Meta-learner was trained on stacked dataset
7. Model was evaluated with out-of-sample data. Refer to Step 3.
8. Results were analyzed.

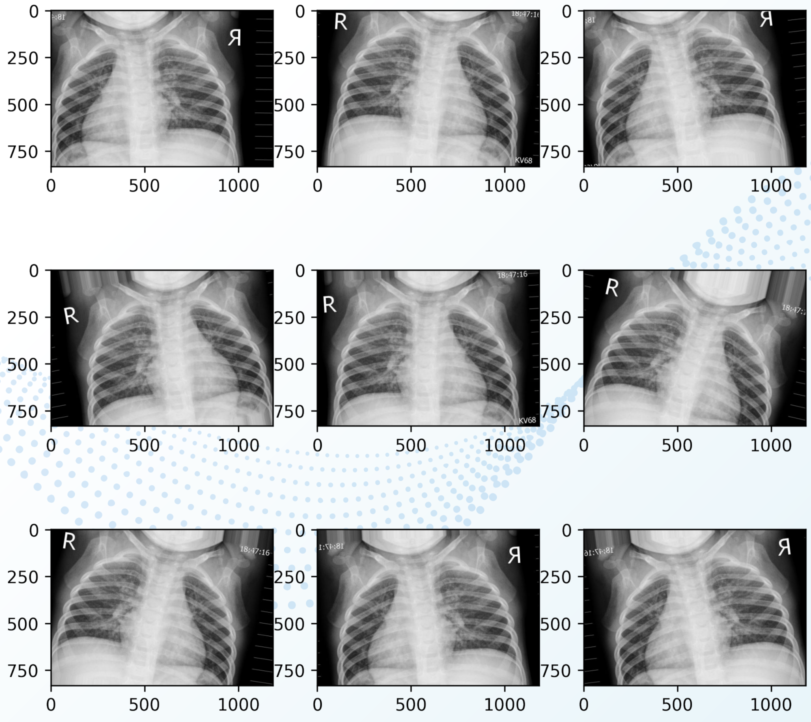


Figure 6. *9 images with applied data augmentation transformations: horizontal flip, clockwise rotation by 15 degrees, vertical shift 10%, horizontal shift 10%*

Results and Discussion: Constituent Model and Stacked Model Evaluation

Model (Binary)	ROC-AUC	F-1 Score	Accuracy
Xception	1.0	0.944	0.973
InceptionResNetV2	0.995	0.968	0.952
ResNet50	0.995	0.902	0.904
Stacked Model	**0.995**	**0.988**	**0.998**

Table 2. *Binary classification between normal and pneumonia results*

The stacked model outperformed all constituent classifiers - Xception, InceptionResNetV2, and ResNet50 in three-class and binary classification with an accuracy of 0.954 and 0.998, respectively.

Model (Three-class)	Precision	Recall	Accuracy
Xception	0.922	0.939	0.942
InceptionResNetV2	0.901	0.940	0.944
ResNet50	0.884	0.878	0.890
Stacked Model	**0.911**	**0.953**	**0.954**

Table 3. *Three-class classification between normal, non-COVID-19 pneumonia, and COVID-19 pneumonia results*

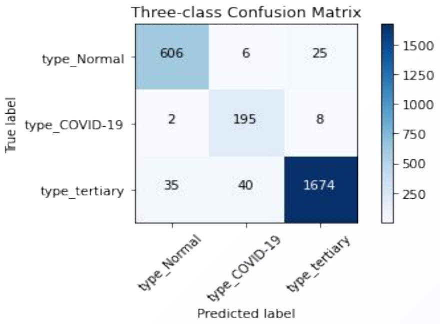


Figure 12. *Three-class Unnormalized Confusion Matrix*

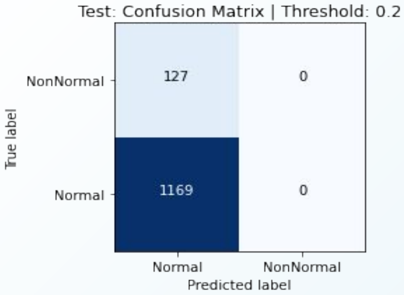


Figure 13. *Two-class Unnormalized Confusion Matrix*

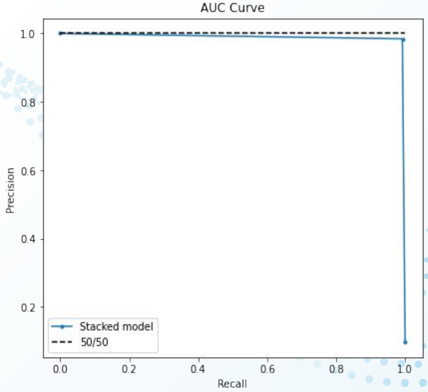


Figure 14. *Binary Classification Precision-Recall Curve*

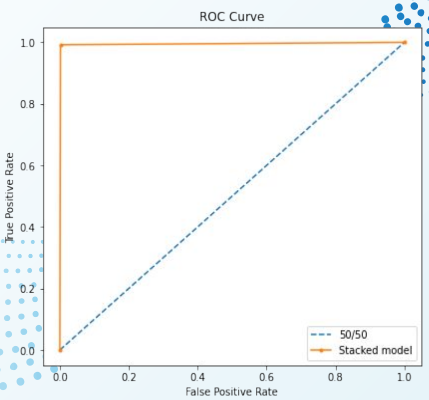


Figure 15. *Binary Classification ROC Curve*

Conclusion

- *PneumoStack* performed significantly better than any of its constituent models, as well as existing models used for pneumonia binary and multiclass classification
- PneumoStack also performed better than individual models investigated in other studies (Table 4)
- Opens doors to higher performance in automated medical imaging and possibly higher performance in applications of computer vision tasks in healthcare and medicine
- **Limitations:**
    - Trained on public dataset - clinical data may present varying characteristics.
- **Future work**:
    - Use SMOTE to counter class imbalance in dataset
    - Investigate clinical usability by validating on clinical data
    - Apply to other medical imaging tasks ex. MRI analysis for the early detection of neurodegenerative diseases
    - Apply to other CNN tasks in medicine ex. differential gene analysis and biomarker identification to investigate if superior performance projects to various computer vision applications

Study	Data Type	Model	Classes	Accuracy (%)
Wang et al. [16]	X-ray	COVID-Net	3	93.3
Apostolopoulos et al. [17]	X-ray	VGG-19	3	87
Umer et al. [18]	X-ray	COVINet	3	89.9
Nishio et al. [19]	X-ray	VGG-16	3	83.68
Singh et al. [20]	X-ray	MADE-based CNN	2	92.55
Zhang et al. [21]	X-ray	CAAD	2	95.18
Sahinbas and Catak [22]	X-ray	VGG16, VGG19, ResNet, DenseNet, InceptionV3	2	80
Mehdi et al. [23]	X-ray	Deep CNN	2	93
Narin et. al. [24]	X-ray	InceptionV3, ResNet50, Inception-ResNetV2	2	98
PneumoStack (proposed)	**X-ray**	**Xception, InceptionResNetV2, ResNet50 stacked model**	**3**	**95.4**
PneumoStack (proposed)	**X-ray**	**Xception, InceptionResNetV2, ResNet50 stacked model**	**2**	**99.8**

References

1. Pneumonia. (2019). Retrieved 12 February 2021, from https://www.who.int/news-room/fact-sheets/detail/pneumonia
2. Pneumonia | NHLBI, NIH. (2021). Retrieved 12 February 2021, from https://www.nhlbi.nih.gov/health-topics/pneumonia
3. A. Ghaderi and V. Athitsos, "Selective unsupervised feature learning with Convolutional Neural Network (S-CNN)," *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, 2016, pp. 2486-2490, doi: 10.1109/ICPR.2016.7900009.
4. Bai, H., Hsieh, B., Xiong, Z., Halsey, K., Choi, J., & Tran, T. et al. (2020). Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. *Radiology*, *296*(2), E46-E54. doi: 10.1148/radiol.2020200823
5. Saha, S. (2021). *A Comprehensive Guide to Convolutional Neural Networks—the ELI5 way*. Towards Data Science. Retrieved 12 February 2021, from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164
6. Zhang, J. (1999). Developing robust non-linear models through bootstrap aggregated neural networks. Neurocomputing, 25(1-3), 93-113. doi: 10.1016/s0925-2312(99)00054-5
7. Ensemble Transfer Learning Framework for Vessel Size Estimation from 2D Images - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Ensemble-transfer-learning-using-pretrained-CNN-model-initialized-with-weights-trained-on_fig4_333619654 [accessed 28 Feb, 2021]
8. Brownlee, J. (2017). A Gentle Introduction to Transfer Learning for Deep Learning. Retrieved 28 February 2021, from https://machinelearningmastery.com/transfer-learning-for-deep-learning/
9. Wolpert, D. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241-259. doi: 10.1016/s0893-6080(05)80023-1
10. Detecting Behavioral Microsleeps from EEG Power Spectra - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Schematic-diagram-of-stacked-generalization_fig2_5898833 [accessed 28 Feb, 2021]
11. Cohen, Joseph, Paul Morrison, and Lan Dao. "COVID-19 Image Data Collection: Prospective Predictions Are The Future." *arXiv 2006 11988* (2021): n. pag. Web 25 Feb 2021
12. Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. Retrieved 31 January 2021, from https://arxiv.org/abs/1610.02357
13. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. Retrieved 31 January 2021, from https://arxiv.org/abs/1512.03385
14. Alemi. Alex. (2021). Improving Inception and Image Classification in TensorFlow. Retrieved 28 February 2021, from https://ai.googleblog.com/2016/08/improving-inception-and-image.html
15. ImageNet Winning CNN Architectures (ILSVRC) | Data Science and Machine Learning. (2021). Retrieved 28 February 2021, from https://www.kaggle.com/getting-started/149448
16. Wang, L., Lin, Z.Q. & Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep* 10, 19549 (2020). https://doi.org/10.1038/s41598-020-76550-z
17. Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and engineering sciences in medicine,* 43(2), 635–640. https://doi.org/10.1007/s13246-020-00865-4
18. Umer, M., Ashraf, I., Ullah, S. et al. COVINet: a convolutional neural network approach for predicting COVID-19 from chest X-ray images. *J Ambient Intell Human Comput* (2021). https://doi.org/10.1007/s12652-021-02917-3
19. Nishio, M., Noguchi, S., Matsuo, H., & Murakami, T. (2020). Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Scientific Reports*, *10*(1). doi: 10.1038/s41598-020-74539-2
20. Singh D, Kumar V, Yadav V, and Kaur M. Deep Convolutional Neural Networks based Classification model for COVID-19 Infected Patients using Chest Xray Images. *International Journal of Pattern Recognition and Artificial Intelligence*, https://doi.org/10.1142/S0218001421510046, 2020.
21. J. Zhang et al., "Viral Pneumonia Screening on Chest X-rays Using Confidence-Aware Anomaly Detection," in *IEEE Transactions on Medical Imaging*, doi: 10.1109/TMI.2020.3040950.
22. Sahinbas K, and Catak FO. Transfer Learning Based Convolutional Neural Network for COVID-19 Detection with X-Ray Images. https://www.ozgurcatak.org/files/papers/covid19-deep-learning.pdf, 2020
23. Jamil M, and Hussain I. Automatic Detection of COVID-19 Infection from Chest X-ray using Deep Learning. medRxiv, https://doi.org/10.1101/2020.05.10.20097063, 2020.
24. Narin A, Kaya C, and Pamuk Z. Automatic Detection of Coronavirus Disease (COVID19) Using X-ray Images and Deep Convolutional Neural Networks. arXiv:2003.10849v1, 2020
25. Osterburg, Stephen. “Implementation of Xception Model.” *Coding*, stephan-osterburg.gitbook.io/coding/coding/ml-dl/tensorfow/ch3-xception/implementation-of-xception-model.
26. Aguas, Kenneth. (2020). A guide to transfer learning with Keras using ResNet50. Retrieved 5 February 2021, from https://medium.com/@kenneth.ca95/a-guide-to-transfer-learning-with-keras-using-resnet50-a81a4a28084b
27. Brownlee, J. (2018). Stacking Ensemble for Deep Learning Neural Networks in Python. Retrieved 5 February 2021, from https://machinelearningmastery.com/stacking-ensemble-for-deep-learning-neural-networks/