# SEnsembleNet: A Squeeze and Excitation based Ensemble Network for COVID-19 Infection Percentage Estimation from CT-Scans

Talha Anwar[1][0000−0002−7343−8500]

Independent Researcher, Pakistan
chtalhaanwar@gmail.com

**Abstract.** Coronavirus (COVID-19) is a contagious disease caused by SARS-CoV-2 virus. Usually, COVID-19 is diagnosed by PCR test, which requires less human expertise, but this test's false-negative ratio is high. COVID can also be diagnosed from radiographs such as CT-scan and X-ray, but it requires expert radiologists. So there is a need for an automated way to interpret chest radiographs using artificial intelligence. Several labelled datasets and deep learning algorithms are available to diagnose corona patients using radiographs. These algorithms classify the images into predefined categories such as healthy or infected. But there is no way to know how much area of chest radiograph is infected by COVID. This paper proposed an ensemble network to predict COVID-19 percentage infection from a chest CT scan. The proposed ensemble network used squeeze and excitation bock to learn individual models' weights during the training process. On validation data and test data, the proposed approach obtained a mean absolute error of 4.469 and 3.64, respectively. Implementation is publicly available at https://github.com/talhaanwarch/Covid-Infection-Estimation

**Keywords:** COVID-19 · Deep Learning · Ensemble network · CT-scans · infection estimation

## 1 Introduction

Since the end of 2019, the world has faced a severe pandemic, the Coronavirus, which has affected hundreds of millions of people and caused the death of more than 5 million people worldwide. Though Corona vaccine is available, its efficacy declines with time, and a booster dose is needed. Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) tests are usually used to diagnose COVID, but these tests have false-negative rates and are time-consuming [13]. As COVID is a respiratory disease, chest CT-scan and X-ray can be used as additional tools to reduce the chances of false-negative reports. Only expert radiologists and pulmonologists can diagnose from CT scan or XRay. However, with the exponential increase in COVID cases, an automated way of diagnosis from radiographs is required that can reduce the burden from the clinicians and enhance the level of confidence [17, 8].

Hundreds of research papers have been published to provide better and better solutions to automate the diagnosis process via radiographs[16, 1]. EfficientNet model with reduce on learning rate policy using CT scan slices achieved an F1 score of 90% [3]. Choosing proper deep learning models and the other hyper-parameters to achieve a good result is time-consuming, so AutoML models are also proposed. T.Anwar used the AutoML technique on a CT-scan dataset comprised of 1560, 374 and 3455 3D scans for training, validation, and testing. Model is first trained on 2D slices and using rule based technique predictions from 2D slices are assigned to 3D scans. A macro F1-score of 89% and 88% on validation and test data is obtained, respectively [2]. Bougourzi *et. al.* also classify 3D scans in two stages. In stage 1, four models are trained on 2D slices and using the pre-trained models in stage II with XGBoost 3D CT scans are classified [5]. These studies are based on images only, IU. Khan *et. al.* incorporated different vitals features such as temperature, pulse, respiratory rate, blood pressure, cough, etc. Comorbidities-based questions include whether the patient has diabetes, hypertension, cardiac, kidney, or any other disease are also incorporated. With clinical and images data combined 98% F1-score is obtained [12]. These studies tried to solve the classification problem of whether a person has COVID.

None of these studies tried to estimate the percentage of infection from the radiograph. F. Bougourzi *et. al.* proposed a dataset and a baseline approach to overcome this limitation that resulted in a 5.10 mean absolute error [6].

In this paper, we tried to improve the F. Bougourzi *et. al.* work and proposed a deep ensemble network that learnt the weighted average for each model in the ensemble network.

## 2 Materials and Methods

### 2.1 Dataset

The dataset is comprised of three sets: train, validation and test set [6]. Train set consists of 132 CT-scans among 128 are labelled as COVID infected and 4 CT-scans are infection-free. The validation dataset has 57 CT scans, among which 55 scans are COVID infected. The labelling is performed based on reverse transcription-polymerase chain reaction (RT-PCR). All these 3D CT scans are converted to 2D slices and two experience radiologist manually labels each slice according to the slice area infected by the COVID. The healthy CT slices are assigned the label zero because there is no infection in those slices. Visually the test data images looked different than validation or train data. Various image statistics such as average, standard deviation and root mean square are applied to verify this.

### 2.2 Deep learning architectures

Deep learning is a branch of artificial intelligence that inputs raw data. Deep learning extracts features automatically from this data to perform specific tasks.

Deep learning requires a large amount of (preferably) labelled data and substantial computation. There are multiple deep learning models; each requires a different amount of time and computation power to achieve a certain performance. In the early era of deep learning, it was believed that stacking layers increased performance. ResNet model proposed in December 2015 rejected this concept and proved that adding layers helped to some extent [10]. ResNet architecture proposed a concept known as identity function from shallow network to deep network, which acted as skip connection and skipped specific layers in between. If adding more layers reduces the performance, the model prefers to skip these layers by learning the identity function; otherwise, keep the layers. So in both cases, the model performance increased. Six ResNet family members from TIMM library [19] pre-trained model on ImageNet data are used in this study. The list is as follow.

**Table 1.** Pre-trained model paper publishing year, total parameters counts, top 1 and 5 accuracies

| **Model Name** | **Year** | **Parameters** | **top1** | **top5** |
|---|---|---|---|---|
| resnest50d [20] | 2020 | 27 Million | 80.96% | 95.38% |
| resnetrs50 [4] | 2021 | 35 Million | 79.89% | 94.96% |
| seresnext50_32x4d [11] | 2018 | 28 Million | 81.27% | 95.62% |
| ecaresnet50t [18] | 2020 | 25 Million | 82.35% | 96.13% |
| skresnext50_32x4d [14] | 2019 | 27 Million | 80.15% | 94.64% |
| seresnet50 [11] | 2018 | 28 Million | 80.26% | 95.07% |

After the last layer in each architecture, three more layers are added with the number of neurons as 500, 250 and 1. The batch size is set to 72 except for the skresnext50_32x4d where it is reduced to 48 because of memory issues as this model take more space and time. We used AdamW optimizer with a learning rate of 1e-3.

### 2.3 Data Transformation

All the images are resized to 256x256 and then center cropped to 224x224. The augmentation techniques applied are random horizontal flipping and shift scale rotation. Images are normalized and converted to tensors. For transformation Albumentations package is used [7].

### 2.4 Learning rate schedulers

Learning rate is an essential hyper-parameter for the training of deep learning models. The learning rate value should not be too large or too small, causing non-convergence or prolonged training. With AdamW optimizer learning rate of 1e-3 usually work well, but this can be improved by using a dynamic learning rate. The learning rate changes during the iteration process, resulting in faster

training and better convergence in the dynamic learning rate. For dynamic learning rates, learning rate schedulers are used. In this study , cosine annealing and cosine annealing warm restarts learning rate schedulers are used [15] along with constant learning rate. The cosine annealing learning rate function has a large learning rate in the start that decreased to a lower learning rate value and then raised to a large learning rate. This process is repeated during the training process. High learning rates prevent the model from stucking in local minima, and low learning rates help converge in true global minima. In the warm restart, the learning rate is initialized to some value before starting to decrease.

### 2.5 Loss Function

Smooth L1 loss [9] is used as a loss function. It acts as an mean square error (MSE) loss if the absolute element-wise error is below $\beta$; otherwise, it work as mean absolute error (MAE) loss. The value of $\beta$ is set to 1. Eq 1 shows the smooth L1 loss equation.

$$\begin{cases} 0.5 * (x_n - y_n)^2/\beta & if|x_n - y_n|\beta \\ |x_n - y_n| - 0.5 * \beta & otherwise \end{cases} \tag{1}$$

### 2.6 Cross validation

Five fold cross-validation is applied on the subject level to avoid over-fitting and it provides an accurate picture of model performance. First, all six architectures for all three learning rate schedulers are trained with Group K fold cross-validation where the group is the subject ID. In each fold, the validation data and test data are predicted, and the predictions of 5 folds are averaged. Figure 1 shows the cross validation pipeline.

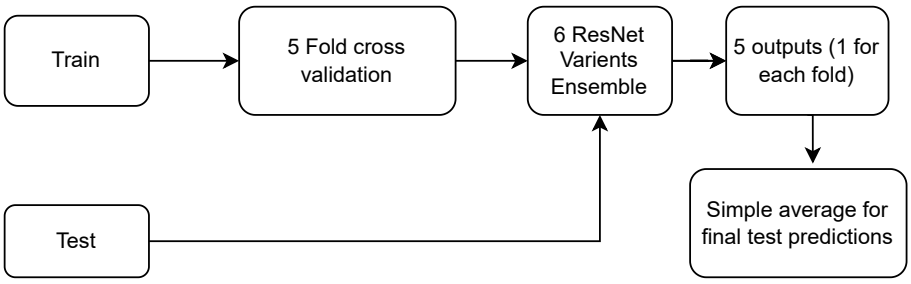


**Fig. 1.** Cross validation pipeline flow diagram

### 2.7 Model ensemble technique

**Average Ensemble** Once all six models are trained, layers of all models are frozen and the last three linear layers are removed. The input is passed through

all six frozen models' layers, and the output is stacked and averaged. This output is passed to new three layers with the number of neurons as 500, 250 and 1. These new layers are re-trained. The remaining process is the same as in individual model training. Here five-fold cross-validation is also applied such that the splitting is not randomized to make sure there is no data leakage. The output from 5 fold on the validation data and test data is finally averaged.

**Weighted Ensemble** After the model stacking, the data is passed to the squeeze and excitation block (SE block) before the averaging. In 2D convolution, a feature map is generated from each input channel to produce the output channels where equal weight is given to all input channels information of output channels. SE block adds a content-aware mechanism that weight each channel adaptively. In simple words, SE block acts as an attention mechanism that gives more weight to essential features in the feature map. The SE block is comprised of two components, squeeze and excitation. The squeezing part captures the global information by using the global average pooling layer from each input channel. The excitation part comprises two fully connected layers that serve the purpose of adaptive calibration by generating weights for each input feature map [11]. We utilize this mechanism, but instead of feeding convolution features map, we fed features map from fully connected layers of each network. So the SE block assigns the weightage to each pre-trained network instead of equally weighting them. Figure 2 shows the weightage ensemble process pipeline.

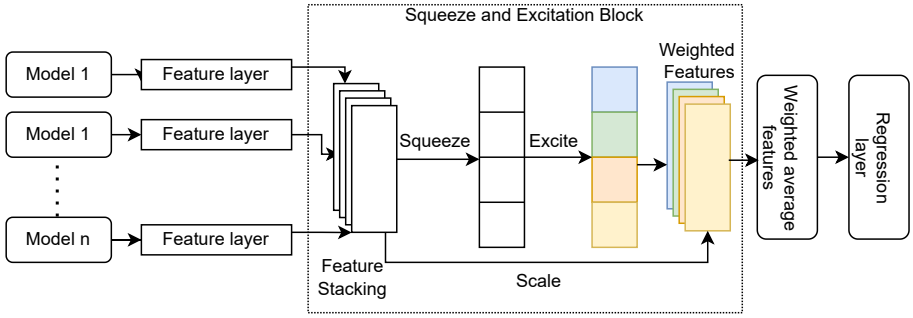


**Fig. 2.** Weighted ensemble network flowchart

## 3 Results

In this study, COVID-19 percentage is estimated from CT-scan slices using different deep learning models. Mean, standard deviation and root mean square of the data is calculated to check their distribution. Table 2 shows the data distribution and the distribution of test data is different that train and validation data.

**Table 2.** Data distribution in different sets

| | Average | Standard deviation | Root Mean Square |
|---|---|---|---|
| Train | 126.08 | 94.26 | 157.56 |
| Validation | 124.95 | 94.15 | 156.55 |
| Test | 72.65 | 63.32 | 96.91 |

The results are obtained by averaging the predictions during five-fold cross-validation. In the case of a single model, ResNest50d got the best result. It received 5.072, 4.1, and 4.056 mean absolute error with constant learning rate, cosine annealing, and warm restarts learning rate policy. Figure 3 shows the mean absolute error for five fold cross validation for individual deep learning models and ensemble models.

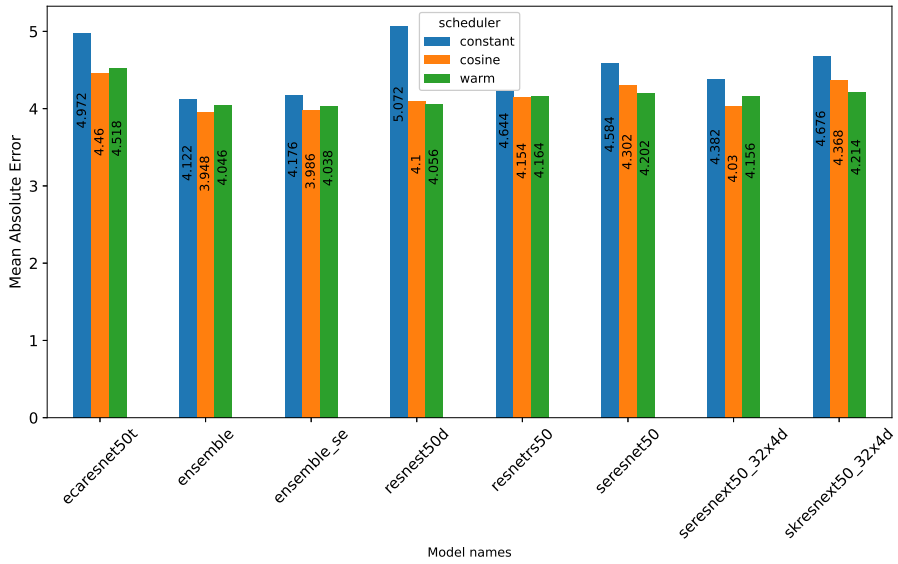


**Fig. 3.** Average 5 folds MAE using different models and schedulers

Table 3 shows the mean absolute error on validation data. The lowest MAE score of 4.46 is obtained by ensembling the model with SE block. We did not test the performance of models on validation with constant learning rate as constant learning rate performed poor in 5 fold cross-validation. Our proposed approach reduced the error rate by 0.64 from baseline [6].

For final test predictions we got mean absolute error of 3.64 with SE Ensemble network with cosine learning rate scheduler.

**Table 3.** MAE on validation data

| **Technique** | **Scheduler** | **MAE** |
|---|---|---|
| Ensemble | Cosine | 4.504109 |
| Ensemble | Warm | 4.480184 |
| SE ensemble | Cosine | 4.469321 |
| SE Ensemble | Warm | 4.476814 |

## 4 Conclusion

This paper discuss the COVID-19 percentage estimation from CT-scan slices. Different deep learning architectures with varying learning rate schedulers are used and compared. Models are ensembled by simple or weightage averaging technique. Squeeze and excitation (SE) block is used as an attention mechanism to give weights to individual models used in the ensemble. With cosine annealing learning SE Ensemble model obtained the lowest mean absolute error of 4.46 on the validation data. This block reduced the error by 0.04 with very little extra computation. The test data used has different distribution and the error rate obtained is 3.64.

## Acknowledgement

I am thankful to Fares Bougourzi (ORCID:0000-0001-5077-4862) for organizing a competition "Covid-19 Infection Percentage Estimation " and evaluating the validation and test score mentioned in this paper. The competition can be viewed at https://competitions.codalab.org/competitions/35575

## References

1. Alghamdi, H., 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). https://doi.org/10.1109/ACCESS.2021.3054484
2. Anwar, T.: Covid19 diagnosis using automl from 3d ct scans. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp. 503–507 (2021). https://doi.org/10.1109/ICCVW54120.2021.00061
3. Anwar, T., Zakir, S.: Deep learning based diagnosis of covid-19 using chest ct-scan images. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). pp. 1–5 (2020). https://doi.org/10.1109/INMIC50486.2020.9318212
4. Bello, I., Fedus, W., Du, X., Cubuk, E.D., Srinivas, A., Lin, T.Y., Shlens, J., Zoph, B.: Revisiting resnets: Improved training and scaling strategies. Advances in Neural Information Processing Systems (34) (2021)
5. Bougourzi, F., Contino, R., Distante, C., Taleb-Ahmed, A.: Recognition of covid-19 from ct scans using two-stage deep-learning-based approach: Cnr-iemn. Sensors (21)(17), 5878 (2021). https://doi.org/10.3390/s21175878
6. Bougourzi, F., Distante, C., Ouafi, A., Dornaika, F., Hadid, A., Taleb-Ahmed, A.: Per-covid-19: A benchmark dataset for covid-19 percentage estimation from ct-scans. Journal of Imaging (7)(9) (2021). https://doi.org/10.3390/jimaging7090189

7. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. Information (11)(2), 125 (2020). https://doi.org/10.3390/info11020125
8. Ghaderzadeh, M., Asadi, F.: Deep learning in the detection and diagnosis of covid-19 using radiology modalities: a systematic review. Journal of healthcare engineering (2021). https://doi.org/10.1155/2021/6677314
9. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015). https://doi.org/10.1109/ICCV.2015.169
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
11. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018). https://doi.org/10.1109/CVPR.2018.00745
12. Khan, I.U., Aslam, N., Anwar, T., Alsaif, H.S., Chrouf, S.M.B., Alzahrani, N.A., Alamoudi, F.A., Kamaleldin, M.M.A., Awary, K.B.: Using a deep learning model to explore the impact of clinical data on covid-19 diagnosis using chest x-ray. Sensors (22)(2) (2022). https://doi.org/10.3390/s22020669
13. Kucirka, L.M., Lauer, S.A., Laeyendecker, O., Boon, D., Lessler, J.: Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based sars-cov-2 tests by time since exposure. Annals of internal medicine (173)(4), 262–267 (2020). https://doi.org/10.7326/M20-1495
14. Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 510–519 (2019). https://doi.org/10.1109/CVPR.2019.00060
15. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016). https://doi.org/10.48550/arXiv.1608.03983
16. Serena Low, W.C., Chuah, J.H., Tee, C.A.T., Anis, S., Shoaib, M.A., Faisal, A., Khalil, A., Lai, K.W.: An overview of deep learning techniques on chest x-ray and ct scan identification of covid-19. Computational and Mathematical Methods in Medicine (2021). https://doi.org/10.1155/2021/5528144
17. Vantaggiato, E., Paladini, E., Bougourzi, F., Distante, C., Hadid, A., Taleb-Ahmed, A.: Covid-19 recognition using ensemble-cnns in two new chest x-ray databases. Sensors (21)(5), 1742 (2021). https://doi.org/10.3390/s21051742
18. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11531–11539 (2020). https://doi.org/10.1109/CVPR42600.2020.01155
19. Wightman, R.: Pytorch image models. https://github.com/rwightman/pytorch-image-models (2019). https://doi.org/10.5281/zenodo.4414861
20. Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., et al.: Resnest: Split-attention networks. arXiv preprint arXiv:2004.08955 (2020). https://doi.org/10.48550/arXiv.2004.08955