

City University of Hong Kong

EE4211 Computer Vision Project - COVID-19 Segmentation

Abstract—Coronavirus disease 2019 (COVID-19) is a pandemic that is threatening the world. To indicate whether the patient is infected by COVID-19 or not, medical imaging such as computed tomography (CT) Scanning, X-Ray, plays an important role. To further indicate the infection of COVID-19, Artificial intelligence (AI) can help medical specialists to classify and segment the appearance of COVID-19. In this project report, we will cover the segmentation process pipeline of medical imaging and analysis involved with COVID-19.

I. INTRODUCTION

Coronavirus disease (COVID-19) is a critical public health emergency issue which is threatening the world from last year. It is the biggest pandemic disease in the recent decade. COVID-19 is caused by a new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case was identified in Wuhan in December 2019, and it spreads around the world at a rapid rate. The World Health Organization (WHO) has stated that the pandemic starts in March 2020. Until December 2020, 67 million cases have been confirmed with around 1.54 million deaths in the world.^[1]

In traditional, the computed tomography (CT) plays a critical role in observing disease. Medical specialists identify whether the patient is infected by reading their CT scans. If the characteristic of COVID-19 can be observed, it can help to increase the accuracy of identifying infection of COVID-19. This project aims to apply the imaging segmentation method with artificial intelligence (AI) to the CT scans of human lungs. With the help of imaging segmentation, medical specialists will be able to identify and locate the infection situation. For a normal CT lung scanning, the lungs should appear in black and clean; For a COVID CT lung scanning, the lung should appear some gray patches and ground-glass opacities, which only appears when the patient is sick.^[2]

II. METHODOLOGY

The workload of medical specialists to give the judgment whether the patient is infected by COVID-19 is heavy. Therefore, applying imaging segmentation with AI deep learning model to locate the infected tissue is one of the ways to reduce the workload of medical specialists.

A. Environment Setup

The program for segmentation of CT scan is complied with following setup^[3]:

- Operation System: Ubuntu 18.04
- Programming Language: Python 3.6

- Python package:
 - miscnn 0.34
 - numpy 1.18.2
 - scikit-misc 0.1.1
 - SimpleITK 1.0.0
 - pandas 0.25.3
 - plotnine 0.6.0
 - tensorflow 2.2.1
 - tqdm 4.35.0

B. Dataset

The dataset consists of 20 annotated COVID-19 CT lung scans and publicly available from Zenodo^[4]. The CT images are labelled by junior annotators at first, then fine-tuned by radiologists with more than 5 years experience. The annotations are finally verified by the senior radiologists.

The dataset consist of 4 classes:

- Background
- Left lung
- Right lung
- COVID-19 Infection



Fig. 1. Labelled CT scan of Lungs infected by COVID-19 (Left Lung – Red, Right Lung – Yellow, Infected Area – Green)

C. Deep Learning Model

There are several deep learning architectures, for example, Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), Recurrent Neural Networks (RNN), etc. These deep learning architectures are applied to different fields including computer vision, voice recognition, data prediction, etc. As this project aims to segment the COVID-19 infected

area in the CT scan, the Medical Image Segmentation with Convolutional Neural Networks (MIScnn) model is selected as the deep learning model.^[5]

CNN is one of the main deep learning architectures for image classification and segmentation. It takes the input data, processes, classifies, and segmenting the data under certain labels. The CNN model is formed by the input layer, hidden layers, and the output layer. Each input image will be processed through the hidden layers and apply corresponding activation functions to categorize the object with the probabilities (between 0 and 1).^[6]

The hidden layers are combined with different layers, such as multiple convolutional layers, pooling layers, fully connected layers, and normalization layers. The purpose of these layers are listed below:

- **Convolutional Layer:** It aims to extract the features in the input image. The layers will obtain a convolved map by applying different mathematical operations and studying the features of the input image.
- **Pooling Layer:** It helps to reduce the number of parameters when the input image is too massive. This process is called sub-sampling. This layer reduces the spatial dimension but keeping the important information in the image.
- **Normalization Layer:** It is used to increase the speed and stabilize the artificial neural network by re-centering and scaling the input layer.
- **Fully Connected layer:** It connects all the layers. The principle of fully connected layer is the same as the multi-layer perceptron neural network (MLP). The image passes through the layer and it is classified with an activation function.

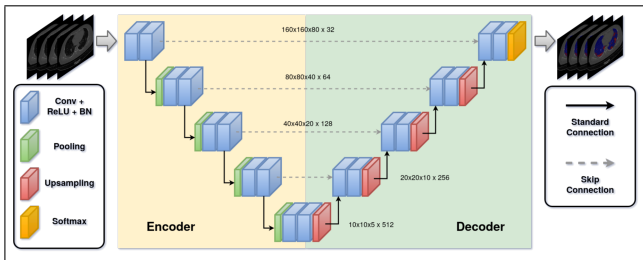


Fig. 2. MIScnn Model Pipeline^[3]

As the deep learning model needs to be implemented with medical data, Medical Image Segmentation with Convolutional Neural Networks (MIScnn) is used to utilize the Neuroimaging Informatics Technology Initiative (NIfTI) data to develop the deep learning model.^[7] As the medical image contains 3D data (Z-axis), 3D U-Net has a better performance compared to using 2D U-Net. The standard 3D U-Net is used to encode and decode the image to label objects in CT scans.^[8] It has reserved the characteristic of 2D U-Net, which are contracting path and expansive path. However, 3D U-Net will perform an element-wise summation in which the 2D U-Net will concatenate the feature maps.^[9] The 3D U-Net uses different 3D operations, such as 3D convolutions, 3D max

pooling, batch normalization (BN) in 3D up-convolutional layers, to increase the speed of convergence.

D. Procedures

As the limited size of the dataset, the CT scans are varied to increase the database. Also, the k-fold cross-validation method is used in the model training process. In this project, 5-fold cross-validation is used, and the number of classes of the CT scans is 4. The 20 CT scans are split in the 5-fold cross-validation during the process of training the deep learning model. The training and predicting the unlabelled images are process in each fold.

E. Statistic Analysis

The result of the CT image segmentation is evaluated by 4 benchmarks:

- **Dice Similarity Coefficient (DSC):** It is a statistic that shows the similarity of two labels. It compares the labeled area of the ground truth image (denoted by T), and the prediction image (denoted by P).^[10] The higher value implies the higher accuracy for the prediction image. The equation is listed below^[11] :

$$\frac{2 |T \cap P|}{|T| + |P|} \quad (1)$$

- **Sensitivity:** It measures the proportion of positives that are correctly identified (i.e. the percentage of sick people who are correctly identified and actually having some illness).^[12] The equation is listed below^[11] :

$$\frac{|T \cap P|}{|T|} \quad (2)$$

- **Specificity** is the ability of a test to correctly identify people without the disease.^[12] There are 3 cases:

- True positive: The patient has the disease and the test is positive.
- True negative: The patient does not have the disease and the test is negative.
- False positive: The patient does not have the disease and the test is positive.

The equation is listed below^[11] :

$$\frac{|\sim T \cap \sim P|}{|\sim T|} \quad (3)$$

- **Accuracy:** It is the index that shows the accuracy of the prediction image compared to ground truth. The equation is listed below^[11] :

$$\frac{|T \cap P| + |\sim T \cap \sim P|}{S} \quad (4)$$

(Where S = The size of the whole image)

Moreover, as the deep learning model is implemented with k-fold cross-validation, it splits the training set into 5 subsets. When one fold is tested in the model, other folds are used as the training set in the model. The errors of the output in each fold are evaluated by the following loss function:

- **Tversky Loss (Tversky Index):** It is the generalization of Dice's coefficient. It adds the weight to false positives

(FP) and false negatives (FN). The equations is listed below^[11] :

$$\frac{|T \cap P|}{|T \cap P| + \alpha|P - T| + \beta|T - P|} \quad (5)$$

(When $\alpha = \beta = 0.5$, Tversky Loss is equal to Dice's coefficient)

- Soft Dice Loss: It is the difference function of DSC. The range is between 0 and 1. The equation is listed below^[11] :

$$1 - \frac{2|T \cap P|}{|T| + |P|} \quad (6)$$

III. RESULTS

The MIScnn model is trained in Google CoLab. Google CoLab provides free GPU support for training deep learning model. The project model has been trained by 5-fold, each fold is trained for 500 epochs with 500 iterations. The training result is listed in the following section:

A. Training Result

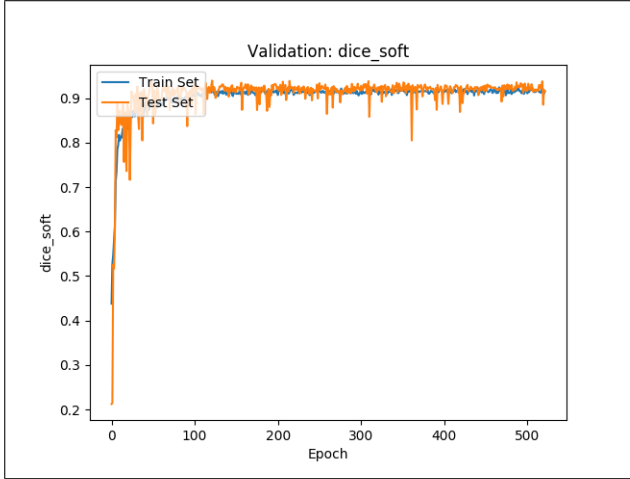


Fig. 3. Dice Soft of MIScnn

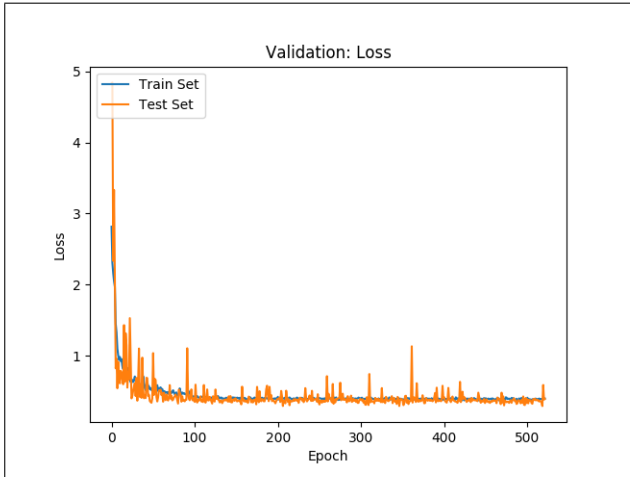


Fig. 4. Validation Loss of MIScnn

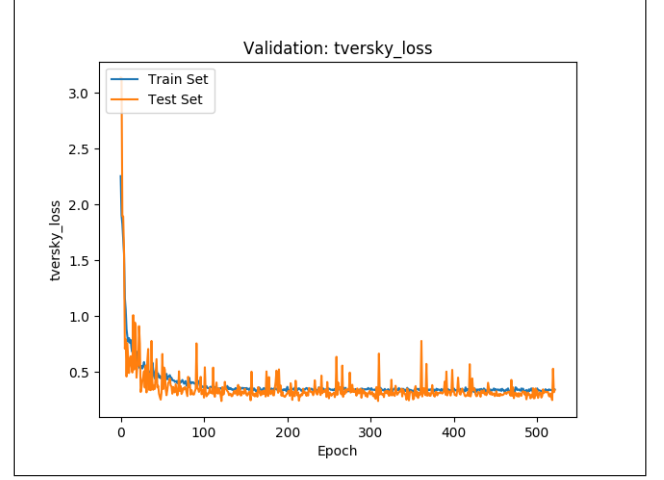


Fig. 5. Tversky Loss of MIScnn

In figure 3, it shows the validation rate of the trained MIScnn model. At epoch 200 the training set validation accuracy is about 0.92, where the testing set validation accuracy is also around 0.92. Both training and testing sets are close, thus no overfitting was observed.

For figure 4 and 5, it shows the fitting curve during the training process. The performance is settled at a loss of around 0.383.

As the training result shows no signs of overfitting and the validation accuracy is about 0.9. It has high efficiency in training the MIScnn model with a limited dataset.

B. Data Score

index	score	background	lung_L	lung_R	infection
coronacases_001	DSC	0.9989167189	0.95143868	0.9355691499	0.6291268221
coronacases_002	DSC	0.9976269727	0.9746728436	0.9723960113	0.8088044233
coronacases_003	DSC	0.9978762294	0.9197886829	0.9490495498	0.8900930757
coronacases_004	DSC	0.9978299508	0.9771784521	0.9873073138	0.7402623877
coronacases_005	DSC	0.9978684189	0.9843618058	0.9829764565	0.8550023343
coronacases_006	DSC	0.9986786513	0.9864858944	0.9806692084	0.8811962648
coronacases_007	DSC	0.9977458987	0.9687488979	0.9736739864	0.7933063338
coronacases_008	DSC	0.9985296308	0.973033297	0.9715749885	0.8339351712
coronacases_009	DSC	0.9983232485	0.980962082	0.9694835149	0.7285354305
coronacases_010	DSC	0.998353132	0.9223728831	0.9416363913	0.8051043803
radiopaedia_10_85902_1	DSC	0.9974726519	0.9797347876	0.9740179239	0.842354301
radiopaedia_10_85902_3	DSC	0.998163339	0.9799051203	0.9706993616	0.8553118555
radiopaedia_14_85914_0	DSC	0.9787556099	0.7608056288	0.7309364997	0.5311049529
radiopaedia_27_86410_0	DSC	0.9958390418	0.9396853548	0.934545652	0.6827588252
radiopaedia_29_86490_1	DSC	0.9981854524	0.8743897492	0.957089031	0.001852383555
radiopaedia_29_86491_1	DSC	0.9978159498	0.9817343062	0.9845226708	0.8639565801
radiopaedia_36_86526_0	DSC	0.9976030102	0.9694675394	0.9729975829	0.637170583
radiopaedia_40_86625_0	DSC	0.9973245992	0.7089791363	0.8035671873	0.8239760047
radiopaedia_4_85506_1	DSC	0.9964661314	0.8830461874	0.9355977241	0.5803657347
radiopaedia_7_85703_0	DSC	0.9979316316	0.9782011365	0.9355879327	0.8023814647

Fig. 6. DSC of CT Scans

In figure 6, it is a table of the DSC of the 20 cases in the dataset. DSC measures the similarity between the ground truth and the prediction image. The larger the DSC value, the higher the accuracy of predicted image segmentation is.^[10] In the table, the DSC values of background, left lung, right lung, and the COVID-19 infection area are around 0.6 to 0.9, which means that the overall prediction performance of CT scans is quite well.

index	score	background	lung_L	lung_R	infection
coronacases_001	Sens	0.9989735883	0.9726244491	0.9786939086	0.4770661471
coronacases_002	Sens	0.9972204214	0.9734102686	0.966270351	0.9377727464
coronacases_003	Sens	0.9985071337	0.8964603043	0.9465482175	0.8942973681
coronacases_004	Sens	0.9971819015	0.979746917	0.9924896432	0.7455710682
coronacases_005	Sens	0.9971660778	0.9890976957	0.9863614445	0.8827255354
coronacases_006	Sens	0.9985961257	0.9883585845	0.9819350424	0.8591195305
coronacases_007	Sens	0.9970570456	0.9814529365	0.9701919392	0.8673789268
coronacases_008	Sens	0.9989756602	0.969349529	0.9750625472	0.7830841952
coronacases_009	Sens	0.9986810016	0.980559429	0.9688901906	0.6860615441
coronacases_010	Sens	0.9989825734	0.894999252	0.9413699876	0.8093574304
radiopaedia_10_85902_1	Sens	0.9989186744	0.9708843114	0.9607336331	0.8129608004
radiopaedia_10_85902_3	Sens	0.9979366817	0.9748039278	0.9816227247	0.8302337461
radiopaedia_14_85914_0	Sens	0.9965243235	0.7850137995	0.6800353479	0.3704381695
radiopaedia_27_86410_0	Sens	0.9972880318	0.9523812238	0.9121641731	0.652317415
radiopaedia_29_86490_1	Sens	0.9989794479	0.7897896516	0.9283198982	0.6134969325
radiopaedia_29_86491_1	Sens	0.996748819	0.9934738299	0.988104125	0.8724947403
radiopaedia_36_86526_0	Sens	0.9991724057	0.9509206145	0.9658594226	0.6914556015
radiopaedia_40_86625_0	Sens	0.9986959801	0.7758411036	0.8187501582	0.725817771
radiopaedia_4_85506_1	Sens	0.9981023062	0.9621936793	0.9701704512	0.416164701
radiopaedia_7_85703_0	Sens	0.9982716506	0.9845263116	0.9829397186	0.7020391939

Fig. 7. Sensitivity Score of CT Scans

In figure 7, it is a table of the sensitivity score of the 20 cases in the dataset. Sensitivity score is used to illustrate the proportion of true positive cases. The larger the sensitivity score is, the high proportion of positives that it is correctly identified.^[12] In the table, most of the sensitivity scores are above 0.5, which indicates that the segmentation method performed by the MIScnn model is able to locate the background, lungs, and the COVID-19 infected correctly.

index	score	background	lung_L	lung_R	infection
coronacases_001	Spec	0.9833339567	0.9980599748	0.9965499945	0.9996704793
coronacases_002	Spec	0.9863678378	0.998556396	0.998555271	0.9979988674
coronacases_003	Spec	0.969215231	0.9987084503	0.9983029706	0.9971308317
coronacases_004	Spec	0.990457219	0.9982936266	0.9985879409	0.9994811339
coronacases_005	Spec	0.9907645008	0.9986941457	0.9984162043	0.9995765466
coronacases_006	Spec	0.9877455772	0.9992780378	0.9990397756	0.9997561373
coronacases_007	Spec	0.9838235624	0.9981088925	0.9989379811	0.9991141639
coronacases_008	Spec	0.9811453835	0.9990728724	0.998426706	0.9993305423
coronacases_009	Spec	0.977892344	0.9992401871	0.9986972707	0.9994555254
coronacases_010	Spec	0.9654266517	0.9989922268	0.9982489695	0.997772256
radiopaedia_10_85902_1	Spec	0.966673468	0.9993951205	0.9993545201	0.9995637972
radiopaedia_10_85902_3	Spec	0.9863741543	0.9991982333	0.9978137615	0.9996184327
radiopaedia_14_85914_0	Spec	0.7905470888	0.9845759696	0.9871203641	0.998960921
radiopaedia_27_86410_0	Spec	0.9533181968	0.9968752923	0.9976227296	0.9970912139
radiopaedia_29_86490_1	Spec	0.9638919748	0.9994986421	0.9995395917	0.9935387915
radiopaedia_29_86491_1	Spec	0.9920632868	0.9981192825	0.9986968926	0.9998402569
radiopaedia_36_86526_0	Spec	0.9743411074	0.9992920913	0.998558435	0.9982018913
radiopaedia_40_86625_0	Spec	0.9462730733	0.9956882975	0.9959414415	0.9955331447
radiopaedia_4_85506_1	Spec	0.9645882215	0.9894982484	0.9940001058	0.999504378
radiopaedia_7_85703_0	Spec	0.9814149877	0.9984030613	0.9950530559	0.9989530895

Fig. 8. Specificity Score of CT Scans

In figure 8, it is a table of the specificity scores of the 20 cases in the dataset. Specificity score is used to assess whether the test is able to correctly identify patients are infected or not. There are 3 testing cases, which are true positive, true negative, and false positive. The larger the specificity score is, the more it is truly positive.^[12] In the table, most of the specificity scores are above 0.9, which indicates that the MIScnn model is able to identify and segment the true positive area and the true negative area.

C. Data Visualization

In order to visualize the training result, the MIScnn model will generate the images which show the segmentation result of the COVID-19 infection area. The image is combined with 3 parts, which are the original CT Scan, the Ground Truth, and the model prediction.

In figure 9 and figure 10, both images are examples of the data visualization generated by the MIScnn model, which are coronacase_005 and coronacase_010. All the generated

cases, it has been divided into 300 slices, and the slices are combined into a Graphics Interchange Format (GIF) file. In figure 9, the original CT scan is a gray scanned image of the patient's lungs, the Ground Truth is the standard answer of the segmentation result, and the prediction is the result of imaging segmentation performed by the MIScnn model. The Ground Truth images show the COVID-19 infected area by the CT scanning, it labels the lung tissue in red and labels the infected area in green. For the model prediction, it labels the lung tissue and the infected area in the same as the Ground Truth. The model prediction of coronacase_005 and coronacase_010 are very similar to the corresponding Ground Truth, indicating that the model has high accuracy in the prediction.

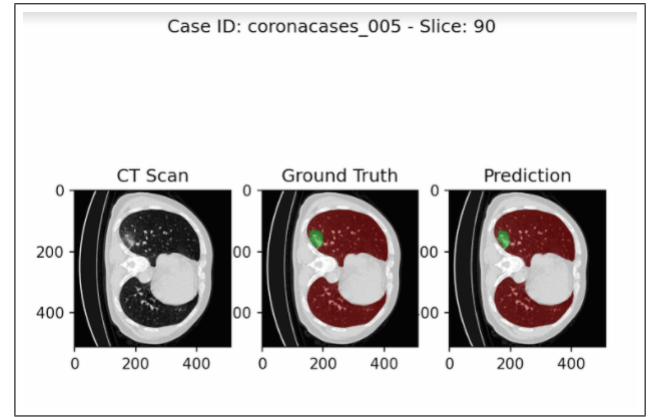


Fig. 9. Visualization of Case coronacase_005

When comparing coronacase_005 with coronacase_010, coronacase_010 is more serious as the infected area is larger, and more infected area is segmented.

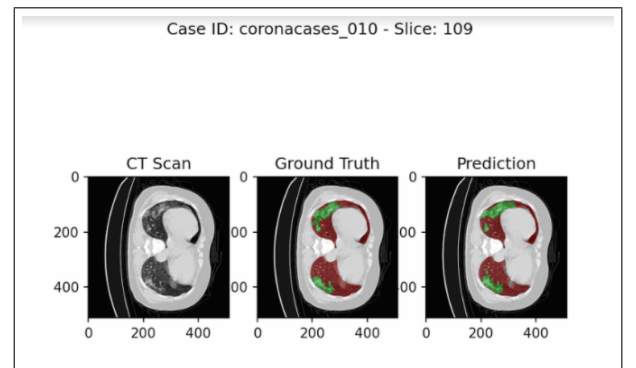


Fig. 10. Visualization of coronacase_010

IV. Limitations

A. Size of Sample Size

As it is only a short period since the outbreak of COVID-19, the public data sets for citizens are in small numbers. The data set contains 20 samples only, which leads to higher variance estimation and may alter the accuracy of the result. Variance is a measure of the contribution to the error of deviations from the central tendency. It occurs when the

model is highly sensitive to the changes in the independent variables. The model tries to fit every target, causing the model to be too general and learn noise, which then may lead to the result of over-fitting^[13]. A higher variance estimation as a result lowers the amount by which the prediction, over one training set, differs from the expected value over all the training sets. With the poor approximation, the accuracy of the training result would be affected significantly.

B. Computational Power

The MIScnn model requires a large amount of computational power. To run this model for training, prediction and evaluation of the performance, a high-end graphics processing unit (GPU) and a central processing unit (CPU) is required. Otherwise, the computer will not be responding when executing the program. To relieve the problem, the Google CoLab is used to execute the program in a cloud platform as it provides high-performance CPU and GPU which is suitable for machine learning. Thus, the requirement of personal computer's performance can be reduced. However, as the deep learning program and the CT image files are complicated, the training and prediction processes are still time-consuming. At least a day is required to generate the results even only 20 CT files are used in this project.

C. Complexity of Data

NIfTI is a common medical file format for neuroimaging. The format specifies 348-bytes of header data and uncompressed image data. The header and image data can be saved as separate files using the file extensions '.hdr' and '.img', or as a single file using the '.nii' extension, with the first 348 bytes devoted to the header and the image data typically beginning at byte 352. Images can have up to 7 dimensions: three spatial dimensions, time, and then other dimensions such as diffusion gradient direction^[14]. The file type is complicated and it takes time to understand and implement the file type to the project.

V. CONCLUSIONS

The MIScnn model demonstrates very high accuracy in the prediction of COVID-19 infection area when compared with the Ground Truth image. This model is trained with the k-fold cross-validation, which helps to relieve the problem of the insufficient data sample. Moreover, the model follows the batch normalization layer and the softmax activation function to produce the probabilities between 0 to 1. The evaluated score shows that the MIScnn model has a great performance on the segmentation of COVID-19 infected area on the CT scans.

REFERENCES

- [1] WHO. (7 December 2020). Weekly Operational Update on COVID-19. [Online]. Available: <https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19—7-december-2020>. [Accessed: 7-Dec-2020].
- [2] Groth, L. (2020, June 29). What Do “Ground Glass Opacities” Mean in Lung Scans of COVID-19 Patients? [Online]. Available: <https://www.health.com/condition/infectious-diseases/coronavirus/ground-glass-opacities-covid-19>. [Accessed: 7-Dec-2020].
- [3] Dominik Müller, Iñaki Soto Rey, & Frank Kramer. (2020). Automated Chest CT Image Segmentation of COVID-19 Lung Infection based on 3D U-Net (Version 1.0) [Data set]. Zenodo. Available: <http://doi.org/10.5281/zenodo.3902293> [Accessed: 7-Dec-2020].
- [4] Ma Jun, Ge Cheng, Wang Yixin, An Xingle, Gao Jiantao, Yu Ziqi, ... He Jian. (2020). COVID-19 CT Lung and Infection Segmentation Dataset (Version Version 1.0) [Data set]. Zenodo. Available: <http://doi.org/10.5281/zenodo.3757476> [Accessed: 7-Dec-2020].
- [5] S. Labs, “Understanding Deep Learning: DNN, RNN, LSTM, CNN and R-CNN,” Medium, 21-Mar-2019. [Online]. Available: <https://medium.com/@sprhlab/understanding-deep-learning-dnn-rnn-lstm-cnn-and-r-cnn-6602ed94dbff>. [Accessed: 7-Dec-2020].
- [6] Prabhu, “Understanding of Convolutional Neural Network (CNN) - Deep Learning,” Medium, 21-Nov-2019. [Online]. Available: <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>. [Accessed: 7-Dec-2020].
- [7] Dominik Müller and Frank Kramer. (2019) MIScnn: A Framework for Medical Image Segmentation with Convolutional Neural Networks and Deep Learning. ArXiv. Available: <https://arxiv.org/abs/1910.09308>. [Accessed: 7-Dec-2020].
- [8] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI.
- [9] Çiçek, Ö., Abdulkadir, A., Lienkamp, S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. ArXiv, abs/1606.06650.
- [10] Lee, Sumin & Kim, Hojin & Ji, Yunseo & Cho, Byungchul & Kim, Su & Jung, Jinhong & Park, Jin-hong & Lee, Sang-wook & Kim, Jong-Hoon & Yoon, Sang. (2018). Evaluation of Hepatic Toxicity after Repeated Stereotactic Body Radiation Therapy for Recurrent Hepatocellular Carcinoma using Deformable Image Registration. Scientific Reports. 8. 10.1038/s41598-018-34676-1.
- [11] Jadon, S. (2020, September 3). A survey of loss functions for semantic segmentation. [Online]. Available: <https://arxiv.org/pdf/2006.14822.pdf>
- [12] Swift A, Heale R, Twycross A What are sensitivity and specificity? Evidence-Based Nursing 2020; 23:2-4. [Accessed: 8-Dec-2020].
- [13] Alpaydin, Ethem. Introduction to Machine Learning, MIT Press, 2014. ProQuest Ebook Central. [Online]. Available: <http://ebookcentral.proquest.com/lib/cityuhk/detail.action?docID=3339851>. [Accessed: 8-Dec-2020].
- [14] Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016, May 1). The First Step for Neuroimaging Data Analysis: DICOM to NIfTI conversion. [Online]. Available: [https://discovery.ucl.ac.uk/id/eprint/1495621/1/Li et al The first step for neuroimaging data analysis - DICOM to NIfTI conversion](https://discovery.ucl.ac.uk/id/eprint/1495621/1/Li%20et%20al%20The%20first%20step%20for%20neuroimaging%20data%20analysis%20-%20DICOM%20to%20NIfTI%20conversion) [Accessed: 8-Dec-2020].