

Detection of Covid-19 From an Imbalanced Chest X-ray Image Data Set

Sharath Madhu Manjunath¹[0000-0002-1894-5331], Manju Gurjar²[0000-0002-6774-1087], Neil O’Kane³[0000-0003-3506-1324], Andrew McCarren⁴[0000-0002-7297-0984], and Leonardo Gualano⁵[0000-0002-7543-4394]

¹ Dublin City University, Dublin, Ireland

² sharath.madhumanjunath2@mail.dcu.ie

Abstract. The Covid-19 pandemic has spread quickly, making identification of the virus critically important in assisting overburdened health-care systems. Numerous techniques have been used to identify Covid-19, of which the Polymerase chain reaction (PCR) test is the most common. However, obtaining results from the PCR test can take up to two days. An alternative is to use X-ray images of the subject’s chest area as inputs to a deep learning neural networks algorithm. The two problems with this approach are the choice of architecture and the method used to deal with the imbalanced data. In this study a comparative analysis of a standard convolutional neural network (CNN) and a number of transfer learning algorithms with a range of imbalanced data techniques was conducted to detect Covid-19 from a data set of chest x-ray images. This data set was an amalgamation of two data sets extracted from the Kaggle Covid-19 open source data repository and non-Covid illnesses taken from the National Institute of Health. The resulting data set was had over 115k records and 15 different type of findings ranging from no-illness to illnesses such as Covid-19, emphysema and lung cancer. This study addresses the problem of class imbalance on the largest data set used for x-ray detection of Covid-19 by combining undersampling and oversampling methods. The results showed that a CNN model in conjunction with these random over and under sampling methods outperformed all other candidates when identifying Covid-19 with a F1-score of 93%, a precision of 90% and a recall of 91%.

Keywords: Covid-19 · Oversampling · Undersampling · CNN · transfer learning · chest x-ray

1 Introduction

The novel coronavirus illness of 2019 has created a substantial and immediate threat to global health. The outbreak has paralyzed the world economy and created widespread disruptions to people’s lives. According to the World Health Organization (WHO), the coronavirus pandemic is putting unprecedented and escalating strain on healthcare systems around the world.

Covid-19 has been diagnosed using a range of medical imaging modalities, blood tests (CBCs) and PCR. According to the WHO, all diagnoses of Covid-19 must be validated by reverse-transcription polymerase chain reaction tests. PCR techniques are performed manually and have a high degree of inaccuracy with a positivity rate of only 63% [1]. Also, PCR techniques require a lot of time and resources. Patients must apply to health centers in batches due to the virus’s fast-spreading nature [2]. These factors ensure that the necessity for quick diagnostic procedures is of critical importance at this point in time. Another technique for monitoring and diagnosing SARS-CoV-2 infections is visually, using radiological images such as chest x-rays or computed tomography (CT). Covid-19 creates abnormalities, evident on chest x-rays and CT scans, in the form of ground-glass opacities. It is foreseeable that a diagnostic could be made with radiological imaging. This would have certain advantages over PCR testing in terms of the detection rate in the early phases of Covid-19. One way would be to have professionals manually analyse the images. However, an alternative is an Artificial Intelligence based approach, which could provide a more timely and accurate means of diagnosing Covid-19.

Deep learning may be used to make predictions based on medical images by extracting distinctive characteristics such as shape and spatial rotation from the images [3]. CNNs have been critical in extracting features and learning patterns that enable prediction [4].

For this study, a comparative analysis of a standard convolution neural network (CNN) and number of transfer learning algorithms with a range of imbalanced data techniques was conducted to detect Covid-19 from a data set of chest x-ray images. Each algorithm and imbalance combination will be assessed using the F1-score, Precision, Recall and the area under the curve (AUC).

In section 2, past research is examined that employed deep neural networks to identify Covid-19 and other chest illnesses. The data sets used are described in section 3 and the methodology is detailed in section 4. Finally, our findings and analysis based on the model and data set are presented in section 5.

2 Literature Review

Researchers have developed deep learning techniques to detect and diagnose Covid-19 based on x-ray images of the chest. This literature review, describes the development of deep learning techniques, to detect Covid-19 and the methods used to handle an imbalanced data set.

2.1 Deep Neural Networks for detecting Covid-19 cases using chest x-ray images:

A Generative Adversarial Network, is introduced by Loey et al., for pre-processing chest x-ray images [5]. The scheme used three pre-trained models: AlexNet, GoogleNet, and RestNet18. The data set included 69 Covid-19 images, 79 pneumonia bacteria images, 79 pneumonia virus images, and 79 normal images.

GoogleNet was selected as a main deep learning technique with a 80.6% test accuracy containing four class scenarios, AlexNet with a 85.2% test accuracy with three class scenarios and again GoogleNet with 99.9% scenario with two class scenarios. This shows how results can be seen to improve when the number of classes is reduced.

A combination of a CNN and a Long Short-Term Memory (LSTM), was used by Islam et al., to detect Covid-19 using x-ray images [6]. In this system, a CNN is used for feature extraction and LSTM for detection using the extracted feature. The data set includes 613 Covid-19 images, 1525 pneumonia images, and 1525 normal cases images. The experimental results showed that their system achieved an accuracy near 99% with an AUC of 99.9%, a specificity of 99.2%, a sensitivity of 99.3%, and a F1-score of 98.9%.

An inception model was proposed by Das et al., to screen Covid-19 from other non-Covid cases like pneumonia, tuberculosis, and healthy cases [7]. The data set included 162 Covid-19 images, 1583 pneumonia images and 342 tuberculosis images. The proposed model achieved an AUC of 100% in classifying Covid-19 from a combined pneumonia and healthy cases data set. Similarly, it achieved an AUC of 99% in classifying Covid-19 cases from a combined non-covid and healthy diseases data set. This shows that it is easier to classify images when there are fewer disease types in the data set. The fact that our data set contains 14 different illnesses alongside one for healthy cases is a reason for lower scores when compared to those that contained only Covid-19 and healthy cases.

A CNN was implemented by Wang et al., to train the model over 13,975 chest x-ray images to detect Covid-19 [8]. Their model was evaluated to a classification accuracy of 98.9%.

A concatenated network was developed by Rahhimzadeh et al., with Xception and ResNet50V2 models to classify Covid-19 chest x-ray images [9]. This developed system contains 180 images of Covid-19, 6054 images of pneumonia, and 8851 of healthy images. A total of 633 images were considered for the training phase, with an outcome of 99.56% of accuracy, 100% specificity, and 93.3% sensitivity.

A Deep Bayes-SqueezeNet model is proposed by Ucar et al., to detect Covid-19 images [10]. The data set includes three classes consisting of 76 Covid-19 images, 4290 pneumonia images, and 1583 normal images. The proposed model achieved 98.3% accuracy for the Covid-19 class.

A deep neural network was designed by Khan et al., to automatically detect and diagnose Covid-19 images [11]. The data set consists of 284 images of Covid-19, 330 images of pneumonia-bacterial, 327 images of pneumonia-viral, and 310 images of healthy cases. Overall, the proposed system obtained 89.5% accuracy, 97% precision, and 100% recall for Covid-19 cases.

The system proposed by T. Ozturk et al., includes seven pre-trained models: VGG19, MobileNetV2, InceptionV3, ResNetV2, DenseNet201, Xception, and InceptionResNetV2 [12]. The data set consists of 25 images of Covid-19 and 25 images of non-Covid cases. Overall, VGG19 and DenseNet201 were selected as

the main model which achieved the highest values of 90% accuracy and 83% precision.

A VGG16 deep transfer learning model was considered by Singh et al., to detect Covid-19 from CT scans [13]. Principal Component Analysis was used for feature extraction and trained for four different classes using ensemble method and an SVM classifier resulting in 95.7% accuracy, 95.8% precision, and 95.3% F1-score.

A review of the application of Machine-Learning and Deep-Learning in the diagnosis of Covid-19 through x-ray and CT images was conducted by Mohammad-Rahimi et al [14]. 105 studies that used ML or DL methods to identify chest images of Covid-19 patients were reviewed. All of these studies were conducted on data sets that were much smaller than used in our study, in the range of 1000 - 10,000 images, with the largest study using a total of 60,000 images. Several studies employed CNNs, of which, the highest accuracy, precision, sensitivity and specificity were obtained by Chowdhury et al, 99.7%, 99.7%, 99.7% and 99.5% respectively [15]. Butt et al., using a balanced data set of 618 images received an AUC score of 99.6%, a sensitivity of 98.2% and a specificity of 92.2% [16]. Hassantabar et al., using a balanced data set of 682 images, received an accuracy of 93.2% a sensitivity of 96.1% and a specificity of 99.71% [17]. All of these studies employing CNNs have relatively small balanced data sets, alongside a low number of competing diseases: one or two diseases including Covid-19 (compared to 15 in ours). Their higher results can be attributed to the fact that they are undertaking an easier classification problem.

2.2 Techniques used for handling the imbalance problem:

During the pre-processing step, Lin et al., used two undersampling techniques that involved clustering [18]. The first technique, used the cluster centres to represent the majority class, and the second technique, used the nearest neighbours of the cluster centres. The experimental results revealed that the second strategy outperformed the first.

Sun et al, proposed an ensemble method to convert the imbalanced data sets into multiple balanced data sets and then built classifiers on them using a classification algorithm [19]. The proposed model was compared with other sampling methods, such as conventional sampling, a cost-sensitive learning method and bagging and boosting based ensemble methods. The proposed method was successful in resolving the imbalance to a greater degree.

Nanni et al., proposed a neighborhood balanced bagging technique to modify samples according to the class distribution in their neighborhood using two approaches [20]. Firstly, they kept a large sample size through oversampling, and secondly, they reduced the sample size using stronger undersampling.

3 Methodology

In this section, an overview of the approaches utilised in this study is given.

Table 1. Chest X-ray Image Data Set

Class	No. Instances	Total
Covid	3616	3616
Non-Covid	Atelectasis - 11559	112,120
	Cardiomegaly - 2776	
	Consolidation - 4667	
	Edema - 2303	
	Effusion - 13317	
	Emphysema - 2516	
	Fibrosis - 1686	
	Infiltration - 19894	
	Mass - 5782	
	Nodule - 6331	
	Pleural thickening - 3385	
	Pneumonia - 1431	
	Pneumothorax - 5302	
No findings - 60353		
Total		115736

3.1 Data Set

For the identification and categorization of the Covid-19 illness, two types of chest x-ray image data sets are used in our study. The Covid-19 images are taken from the Covid-19 chest x-ray image data set which is freely available on Kaggle [21]. The second set of x-ray images are taken from the National Institute of Health (NIH) chest x-ray data set [22]. The chest x-ray is the most commonly used imaging technique, due to its cost effectiveness, however, it poses more of a challenge clinically than alternatives such as chest CT imaging. In reflection of this, the model proposed in this study has been positioned to deliver clinically meaningful diagnostics using a combination of chest x-ray imagery and publicly available data sets that contain a large number of cases. The data set is relatively large and complex owing to the many different disease types in the database.

From Table 1, it can be seen that just over half the observations have been classified under the 'healthy' or 'No findings' label. The data set allows for multiple label peer samples, however, a 'No findings' can only occur on its own. A little over half of the observations in the general data set represent the absence of disease and 3616 subjects had Covid-19. This amalgamated data set can be found in a Github repository [23].

Each picture sample is accompanied by numerous fields in each of the two data sets. These variables contain patient information such as gender, ID and age, as well as information on the images perspective (PA or AP). The number of images per class is presented in Table 1.

3.2 Image Pre-processing

The x-ray images used were taken at different orientations, positions, there was some variation in size. The following pre-processing steps were taken: resize the images to equal length and width; Gaussian Blur was used to remove the noise from the sample image; smooth the edges of the images to allow easy detection; segmentation was applied (histogram equalization), to normalize the brightness and contrast.

3.3 Resampling Methods

In general, for machine learning or deep learning imbalanced learning problems are dealt with by using resampling [4]. Resampling decouples the imbalance problem from classification techniques allowing users to employ any standard algorithm after the resampling pre-processing phase has been completed.

In this section, we present the different resampling methods for handling imbalance in image data classification.

Oversampling Oversampling is a strategy that adds more samples to an imbalance class. Many studies have used this strategy to solve classification challenges for extremely imbalanced, image data sets.

There are many oversampling methods available which are designed as variations of the oversampling technique SMOTE such as: Borderline-SMOTE, Synthetic Sampling (ADASYN), SVM-SMOTE [18]. SMOTE selects examples from the feature space by drawing the line between the examples and by generating samples along that line.

Instead of smaller particular decision areas, synthetic methods cause a classifier to produce bigger and less specific decision regions [20]. Thus, positive examples are learned in more general regions rather than being impacted by negative instances around the sample. SMOTE, however, runs into the problem of over-generalization. It ignores the majority class and arbitrarily generalizes a minority class region [19]. In the event of a highly skewed class distribution this results in a higher likelihood of class mixing.

Undersampling Random undersampling in the training data set removes certain data points from the majority class. This results in a decreased number of instances in the majority class of the training data set. This method is repeated until each class has the same number of examples. This approach is suitable for dealing with the imbalance when there are a sufficient number of samples in the minority class that holds all the necessary information.

Undersampling approaches, may, on the other hand, result in a loss of information, when data from the majority class are removed, leading to under-fitting. No method exists to recognize or preserve information-rich instances from the majority class, since examples are eliminated at random [18].

There are different approaches available making use of undersampling e.g. TomekLinks. These approaches are used to solve problems with imbalanced data sets; however, they are confined to moderate or low skewed data sets and are not recommended for highly imbalanced data sets [20].

In order to avoid a low level of recall due to undersampling, this study applied a combination of oversampling and undersampling, similar to the work by Sun et al [19]. This process oversamples the minority class, and then undersamples the majority class by 70%. It was found that making the ratio 70% minimised the recall loss while reducing the imbalance problem.

3.4 Classification Models

Convolutional Neural Network In this study, a convolutional neural network architecture was constructed to detect Covid-19 from chest x-ray images. The network has 14 layers: six convolutional layers, four pooling layers, four dropout layers, one fully connected layer, and one output layer. Each convolution block is made up of two 2-D CNNs and one pooling layer, followed by a dropout layer with a 25% dropout rate. The 5 x 5 kernel is used for feature extraction with an ReLU activation function at the convolutional layer. The 3 x 3 kernel is used by the max pooling layer to reduce the size of the input image. Finally, after extracting the features, it is connected to fully connected layers for making the prediction.

Transfer Learning Transfer learning is a deep learning approach that involves training a neural network model on a problem that is similar to the one being solved [4]. The learned model's layers are then utilized in a new model that is trained on the problem of interest.

In addition to the developed CNN model, four pre-trained models for classification have been used with applied transfer learning: ResNet50, NasNetLarge, Xception and InceptionV3. ResNet50 is made up of convolutional layers with an average pooling layer at the end. It uses residual learning as a building component. CIFAR10 and ImageNet are two types of architectures that are part of NasNetLarge [4]. The SoftMax function is implemented in the CIFAR10 architecture, which includes N number of regular cells and one reduction cell repeating after each other. ImageNet has the same architecture as CIFAR10, with two strides of convolutional layers with a 3x3 kernel size at the start, followed by two reduction cells [4]. In Xception there is one entry flow, eight intermediate flows, and one exit flow [14]. Convolutional and max-pooling layers make up the entry flow with a ReLU serving as the activation function.

Convolutional layers with a ReLU activation function are used in the intermediate flow on their own. Convolutional, max pooling and global average pooling layers, with a ReLU activation function, make up the exit flow at the end of the architecture [4]. The InceptionV3 model's fundamental architecture comprises convolutional and pooling layers, as well as three inception designs. For the final layer, logistic and SoftMax functions are used.

To implement this structure, models were derived from the Python Keras package and initialized with a shape of (80,80,1). The transfer learning models are capable of categorizing a maximum of 1000 classes, however, for our purposes all that was needed was two. Therefore the transfer learning models were modified by substituting the fully connected layer with one that has only outputs.

3.5 Evaluation Measures

The CNNs performances were assessed by using the following metrics: sensitivity, recall, precision and F1-Score (F-Measure). All of these metrics may be evaluated upon testing the model. The different outcomes for recall and precision is discussed below [4]:

- High Precision & High Recall: the model performs well with the classification.
- High Precision & Low Recall: the model is unable to correctly categorize the data points of a certain class or may overfit them.
- Low Precision & High Recall: the model correctly identifies data points from a certain class, but incorrectly labels a large number of data points from other classes.
- Low Precision & Low Recall: the model performed poorly in handling the classification.

The Receiver Operating Characteristic (ROC) curve expresses the performance of the classification CNN model used in this study. A graphical approach is used that works at all classification thresholds. It is able to achieve this by graphing the True Positive Rate and False Positive Rate. The metric gives a summary of the performance of a classifier over all possible thresholds. Similar to the ROC is the Area Under the ROC curve (AUC) which examines the entire two-dimensional area underneath the entire ROC curve covering from (0,0) to (1,1). This metric is effective at checking the proper or wellness and the quality of our model's prediction performance [3].

4 Results

The CNN and pre-trained models developed in this study, receive pre-processed grayscale images as their input and binary classification is performed to distinguish between Covid and non-Covid cases. Meanwhile, the model is trained for 100 epochs and a batch size of 64 for each experiment.

To handle the imbalance problem, we tested the proposed model with a combination of oversampling and undersampling as described in Section 3.3. The new population of data was used for training the model. The obtained results are displayed in Table 2. The CNN model that uses mixed sampling outperforms the other combination of algorithms tested.

In machine learning, learning curves are a common diagnostic tool for algorithms that learn progressively from a training data set. After each update during training, the model can be assessed on the training data set and a hold

Table 2. Results obtained for proposed models when mixed sampling was done with Random oversampling and Undersampling

	Random Oversampling & Undersampling			
Models	F1-score(%)	Precision(%)	Recall (%)	AUC(%)
CNN	93	90	91	94.5
ResNet50	74	68	86	68
InceptionV3	69	64	67	63
Xception	71	67	65	66
NasNetLarge	77	73	69	72
	SMOTE and Random Undersampling			
Models	F1-score(%)	Precision(%)	Recall (%)	AUC(%)
CNN	90	86	88	89
ResNet50	69	67	65	67
InceptionV3	63	61	62	61
Xception	65	64	64	65
NasNetLarge	71	70	68	70.5

out validation data set. Graphs of the measured performance can be generated to display the learning curves. In this study, the training performances for both SMOTE and Random Oversampling (ROS) were tested. A comparison of a learning curve is shown in Fig.4. Looking at these graphs, mixed sampling with ROS performs well and there is no sign of overfitting or underfitting. On the other hand, the mixed sampling with SMOTE graph, suggests that the model often fluctuates at different epochs.

The classification evaluation metrics are summarised in the confusion matrix which results in the True positive, True Negative, False Positive and False Negative shown in Fig. 2.

The CNN model outperforms mixed sampling with random oversampling approach by achieving 90% precision, 91% recall (this avoids unnecessary false alarms) and 93% F1-score which measures the misclassified cases.

In the second phase of testing, pre-trained models were trained on different combinations of resampling techniques. The results from this analysis were not very significant. The large number of classes and continual overfitting of the models were the main factors contributing to these uninformative results.

5 Conclusion and Future Work

Medical image analysis and feature extraction, which are used to diagnose a wide range of chest illnesses, have benefited greatly from deep learning. CNN architectures are well-known for their capacity to learn and predict mid-level and high-level visual representations.

In this paper, a combination of over and under sampling methods has been used to address the problem of class imbalance on the largest data set used for x-ray detection of Covid-19. Image pre-processing was performed as described

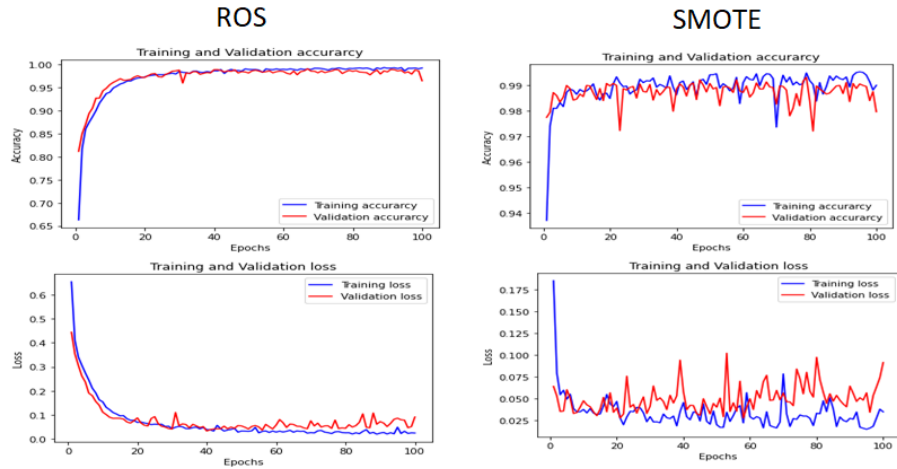


Fig. 1. Accuracy and loss over 100 epochs of training and validation data set for ROS and SMOTE of CNN model

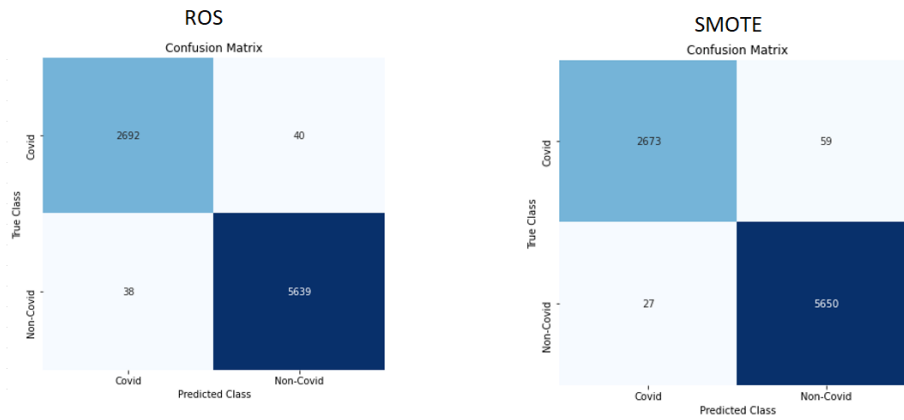


Fig. 2. Confusion matrix for both ROS and SMOTE for CNN model

in Section 3.3 to remove unnecessary details and each model was trained with all the combinations of the resampled data. The proposed modelling analysis outperformed many of the existing methods.

The data set used here, was comparatively large relative to other research conducted into x-ray classification of Covid-19. The presence of many different illnesses in the data set poses a higher challenge in classification compared to those works that used fewer disease types. Satisfactory results of F1-score, precision and recall have been obtained when compared to those obtained by previous works. For the purposes of reproducibility and experimental validation, both the data set and the code have been made publicly available.

Due to the exponential growth in Covid-19 reported cases and treatments throughout the world, the content of Covid-19 databases are continuously growing by significant amounts. As a result, future research could focus on improving the architecture of the deep-learning model proposed in this study, as well as testing its robustness on even larger data sets.

In future work, the study could be expanded to explore the use of One-Class-Classification algorithms used for imbalanced data sets, in conjunction with the techniques that have been proposed in this paper.

References

1. Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., Tan, W.: Detection of SARS-CoV-2 in different types of clinical specimens. *Jama*, 323(18), pp.1843-1844., (2020)
2. Yang, T., Wang, Y.C., Shen, C.F., Cheng, C.M.: Point-of-care RNA-based diagnostic device for COVID-19. *Diagnostics*, 165, (2020)
3. Oyelade, O.N., Ezugwu, A.E.: Deep Learning Model for Improving the Characterization of Coronavirus on Chest X-ray Images Using CNN. *medRxiv*, (2020)
4. Albahli, S., Yar, G.N.A.H.: Fast and Accurate Detection of Covid-19 Along With 14 Other Chest Pathologies Using a Multi-Level Classification: Algorithm Development and Validation Study. *Journal of Medical Internet Research*, 23(2), p.e23693., (2021)
5. Loey, M., Smarandache, F., M Khalifa, N.E.: Within the Lack of Chest Covid-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning. *Symmetry*, 12(4), p.651., (2020)
6. Islam, M.Z., Islam, M.M., Asraf, A.: A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in Medicine Unlocked*, 20, p.100412., (2020)
7. Das, D., Santosh, K.C. and Pal, U.: Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Physical and engineering sciences in medicine*, 43(3), pp.915-925., (2020)
8. 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. *Scientific Reports*, 10(1), pp.1-12., (2020)
9. Rahimzadeh, M., Attar, A.: A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Informatics in Medicine Unlocked*, 19, p.100360., (2020)

10. Ucar, F., Korkmaz, D.: COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical Hypotheses*, 140, p.109761., (2020)
11. Khan, A.I., Shah, J.L., Bhat, M.M.: CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, 196, p.105581., (2020)
12. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine*, 121, p.103792., (2020)
13. Singh, M., Bansal, S., Ahuja, S., Dubey, R.K., Panigrahi, B.K., Dey, N.: Transfer learning-based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data. *Medical biological engineering computing*, 59(4), pp.825-839., (2020)
14. H. Mohammad-Rahimi, M. Nadimi, A. Gjalyanchi-Langeroudi, M. Taheri, S. Ghafouri-Fard: Application of Machine Learning in Diagnosis of Covid-19 Through X-Ray and CT Images: A Scoping Review. *Frontiers in cardiovascular medicine*, 8, p.185., (2021)
15. M. Chowdhury, A. Khandakar, M. Kadir, Z. Bin Mahbub, K. Islam, M.Khan, A. Iqbal, M. Reaz, M. Islam: Can AI Help in Screening Viral and Covid-19 Pneumonia?. *IEEE Access*, 8, pp.132665-132676., (2020)
16. Butt, C., Gill, J., Chun, D., Babu, B. A.: Deep learning system to screen coronavirus disease pneumonia. *Applied Intelligence*, (2020)
17. S. Hassantabar, M. Ahmadi, A. Sharifi.: Diagnosis and detection of infected tissue of Covid-19 patients based on lung x-ray image using convolutional neural network approaches. *Chaos, Solitons Fractals*, 140, p.110170., (2020)
18. Lin, W.C., Tsai, C.F., Hu, Y.H., Jhang, J.S.: Clustering-based undersampling in class-imbalanced data. *Information Sciences*, 409, pp.17-26., (2017)
19. Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B., Zhou, Y.: A novel ensemble method for classifying imbalanced data. *Pattern Recognition*. 48(5), pp.1623-1637., (2015)
20. Nanni, L., Fantozzi, C., Lazzarini, N.: Coupling different methods for overcoming the class imbalance problem. *Neurocomputing*, 158, pp.48-61., (2015)
21. Chowdhury, M.E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al Emadi, N., Reaz, M.B.I.: Can AI help in screening viral and COVID-19 pneumonia?. *IEEE Access*, 8, pp.132665-132676., (2020)
22. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In *IEEE CVPR (Vol. 7)*., (2019)
23. Github, <https://github.com/sharath30/Covid>. Last accessed 15 Oct 2021