COSC 2673/2793 - Assigment 4

Vikas Virani(s3715555)

I. Literature Review

Medical image analysis and computer-aided diagnosis (CAD) in radiology plays an important role in clinical treatment and diagnosis and things like x-ray image classification is an active research area. X-ray machines are used to scan the affected body such as tumors , pneumonia, lung infections, fractures and bone dislocations. CT scanning is kind of an advanced X-ray system which examines the soft structure of the body part and clearer images of the inner soft tissues and organs. Using X-ray is a cheaper, easier, faster and less harmful method than CT. If COVID-19 pneumonia is not identified and treated promptly, it may lead to increased mortality. So, it is very crucial to build a model which can early detect COVID-19 so that treatments or precautions can be done easily.

Clinical image analysis has been going on in medical area since a log time ago for example, in 2009, Caicedo JC, Cruz A, Gonzalez FA [1] used support vector machines (SVM) classifiers along with SIFT (Scale-invariant feature transformation) as a feature descriptor to detect ad classify clinical images and got the precision of 67%, highest of that time. But SIFT was patented by University of British Columbia, so Rublee et al. [2] proposed an efficient alternative to sift, oriented fast and rotated binary robust independent elementary features (ORB) – faster feature descriptor, which has a greater than or equal performance than SIFT depending on use case. SVM was also widely used classification algorithms in medical image analysis tasks which gives an excellent performance [3,4]. In 2011, Parveen and Sathik [5] published a study to identify Pneumonia from x-ray images. They have described feature extraction techniques like DWT, WFT and WPT to extract important features from images. Unsupervised fuzzy c-means algorithm was used to classify pneumonia infection. This approach generated better results compared to other methods. Paredes et al. [6] used k-nearest neighbour (KNN) and small pieces of clinical images as local features and performed classification of the whole medical image, which achieved start-of-art accuracy during that time.

The good thing about these approaches are it generates good results in their respective time, it provides a basis to expand this research area more, but all these researches are based on traditional machine learning approaches, which requires feature generation or feature extraction in one way or another, which is a hard task to do manually. It requires a lot of research and configuration changes in the machine learning process.

After CNN-based deep neural networks score higher on ImageNet challenge [7] in recent years, these systems are widely been used in clinical image analysis tasks instead of traditional approaches. In modern research on clinical image analysis of x-ray and Computer Tomography (CT), variations of Neural networks, more specifically Convolutional neural networks (CNN) and Deep learning are used as they don't require explicit feature generation, all of the processing happens inside network itself. This can avoid complex and expensive feature engineering task from the process. However, CNN requires large amount of data to train an adequate model. Qing et al. [8] proposed an updated CNN with shallow Convolutional layers to categorize image pieces of lung disease. Furthermore, in separate research, it was found that CNN based models can be trained with big dataset of chest x-ray (CXR) with high accuracy and sensitivity for example Stanford Normal Diagnostic Data has approximately 4,00,000 images of chest x-ray. Also, training a CNN model with less data makes it hard to get an accurate model. So, CNN with transfer learning widely is used in clinical image analysis tasks. Moreover, Kermany et al. [9] proposed a transfer learning system to categorize 1,08,000 Optical coherence tomography (OCT) images and it was found that the weighted average error is of the result is equivalent to the average performance of 6 human experts.

However, there is an inherent problem with medical domain due to the recency of this domain in CNN models, i.e. clinical images are hard to find. The best chest x-ray image dataset published so far is still far smaller than the general image dataset available- ImageNet, which has more than 14 million images as of 2010. Even if the datasets are combined, it needs to be labelled manually, which requires enormous amount of time for large dataset and opinions from expert of this industry. If it is labelled already than it needs to be checked for authenticity because this dataset will play an important role in training a model and therefore labels need to be identified correctly. It also poses privacy issues when collecting images from different sources. There can be 2 solutions to this, 1st is to gather more data using some techniques like crowdsourcing or looking into the data of existing medical reports to make a big dataset. 2nd solution to this problem is to find out how to increase the model accuracy on a small dataset. There can be many strategies found to increase the model performance of CNN when using small dataset, like data augmentation to change images and feed it to CNN model. Wang and Perez [10] explored the impact of data augmentation in image classification task. They have found that the transformation-based data augmentation has more accurate results than GAN and other neural network-based methods. Other strategy to increase the model performance is, as stated earlier, transfer learning. Kermany et al. [9] obtained 92% accuracy on a small x-ray image dataset of pneumonia using transfer learning. Third strategy is to use the capsule network. Sabour et al. [11] proposed a new neural network structure called capsule network which obtained the best results on small datasets. Afshar et al. [12] have used capsule network and achieved 86.56% accuracy in identifying brain tumors. But there re some limitations in above approach, Kermany's approach uses InceptionV3 model and it does not retrain convolutional layer of InceptionV3 due to the problem of overfitting. Also, Afshar et al. [12] has not compared his results of capsule network with the results of other methods.

CNN with transfer learning is more accurate compared to other methods for a small dataset. S.S. Yadav and S.M. Jadhav [13] stated that, for transfer learning, it is crucial to retrain specific features on a new target dataset to improve performance. Second most important thing for transfer learning is a network complexity that matches the scale of the dataset [13]. They have done research on the above stated limitations and compared Kermany's approach with retrained convolutional layer, capsule network and CNN with transfer learning, along with the analysis of the impact of data augmentation, overfitting, network complexity and tuned convolutional layer to see which one performs better among these 3 approaches. They have found that configured transfer learning method along with data augmentation applied efficiently reduces overfitting and generates better result than rest of the two models: capsule networks which are trained from scratch and a transfer learning method with only last convolutional layer retrained [13].

Hence, it was identified that for x-ray image analysis/classification on small dataset, CNN with transfer learning and further configuration for data augmentation, network complexity etc. gives higher results compared to all other methods. From these conclusions, the exploration was done on such approaches which are state-of-the-art and handling the task like our original task of COVID-19 patients' classification. Some of the Chest X-ray Image classification approaches are given in the table [1] in the annexure section.

One of the first methods for identifying coronavirus is Reverse Transcription Polymerase Chain Reaction (RT-PCR). It is performed on respiratory samples, and the results are obtained in few hours to two days. Second method to detect COVID-19 is using Antibodies, where blood samples are used to detect the virus. Bu, medical professionals use chest x-ray (CXR) scans occasionally to specify lung pathology. In Wuhan, a research was done on computerized tomography (CT) image reports, and it was found that the sensitivity of CT images for the COVID-19 infection rate was about 98% while for RT-PCR, It was only 71% [19]. As stated earlier, CT images are an advanced x-ray system, but they are hard to find for a new disease like COVID-19. So, instead x-ray images are used more commonly which are cheaper, faster, and easier to find than CT images.

Out of all approaches of x-ray image classification, 3 different state-of-the art approaches in relation to COVID-19 as discussed here. Ali Narin et al. [14] proposed 3 different convolutions neural network-based models named ResNet50, InceptionV3 and Inception-ResNetV2 for the detection of COVID pneumonia using CXR images. They have used 5-fold cross validation to analyse ROC analyses and confusion matrices. They have found that ResNet50 model obtains highest accuracy of 98% among all 3 models. ResNet(Residual neural network) is an improved version of CNN. It adds shortcuts between layers to solve a problem which stops the distortion that occurs as the network gets deeper and complexity is increased. The advantage of this research was, as it used x-ray images to detect COVID-19 instead of Computer Tomography (CT), images can be obtained easily comparatively. As it is a CNN based approach, no feature extraction is required as well as it provides high accuracy. But they have used only 100 images while training a mode, 50 for COVID-19 pneumonia and 50 for healthy patients. However, it was a binary classification.

Kabid Hassan Shibly et al. [16] proposed a different approach known as Faster Regions with Convolutional Neural Networks (Faster R-CNN) framework which is based on Deep Neural Network (DNN). They have used 5450 images in total for model training and obtained an accuracy of 97.56%. However, both above described approaches are denoting an accuracy for binary classification i.e. it predicts if a person has COVID-19 or not. For our scenario, we are interested in classification between healthy people, ill with pneumonia and those who have COVID-19. So, other similar studies have been explored to in that setting which uses multi-class classification.

Ozturk T. et al. [15] proposed deep learning model for both binary (COVID-19 vs. no findings) and multi-class (COVID-19 vs. no findings vs. Pneumonia) classification known as DarkCovidNet which was built by choosing Darknet-19 model as a starting point, which form the basis of you only loon once (YOLO); a real-time object detection system. It was trained on a small number of images, 1125 in total and gave an accuracy of 98.08% and 87% accuracy for binary and three classes, respectively. As COVID-19 pneumonia is a subset of pneumonia diseases, the accuracy of model decreases to 87% when used as 3 class classification, because it generates more false positives.

With the use of transfer learning, the Deep model of Ioannis et al. [17] reached a success rate of 97.82% for two classes and 93.48% for three classes for COVID-19 detection. It uses 1427 images in total for model training. Which shows that CNN model with transfer learning works best on detecting diseases from medical CXR images in general as well as for detecting COVID-19 for a small dataset.

As COVID-19 is a recent disease, there are less number labelled of x-ray images that can be collected. Hence, we will also use pre-trained CNN model with transfer learning as a solution to our problem as our data set has only 200 images of COVID-19 and 300 images each for other 2 classes. Pre-trained CNN models are trained on large number of images; hence their convolutional layers provide local level features. This features along with data augmentation, customized network complexity and customized convolutional layers retrained on our new dataset will give more accurate results than other approaches.

Apart from above mentioned challenges, Algorithmic interpretability is a crucial challenge in machine medical image analysis when machine learning model are used. Grad-CAM [20] heat map approach can be used to visualize how model is making decision, so that experts can interpret the working of algorithm and find out how accurate it is. Other key challenge is that evaluation metrics may not reflect medical applicability. We will discuss this in technical issues section.

II. Proposed Methodology

A. Project plan

The complete framework will be built and evaluated in python programming language using Keras deep learning library with TensorFlow backend. It can be run on a local machine (for a small dataset) or on a cloud services like Google GCP or Amazon AWS. We have 200 COVID-19 images and 300 images of each other category. Which is a small dataset, so we will be performing some data pre-processing steps as stated earlier like data augmentation, normalization etc. As the images will have different size and aspect ratios, we will be processing image by cropping and resizing them to a pre-defined standard size of for example 224x224 pixels. We will also do data augmentation to make our model more robust to changes in x-ray images. Images are applied 0.3 - shear range, horizontal and vertical flips, 0.05 - zoom, 0.05 of height and width shift range, rotation of images etc. This will provide variation in training data ad make our model more accurate. This parameter values are randomly set up here, it can be customized based on which works well and which will be more accurate. Other augmentation techniques like affine transformations can also be checked on dataset.

Essentially, 2 CNN models will be considered in the approach of finding the model with highest performance; ResNet50 and InceptionV3. ResNet50 is a 50-layer network which is trained on ImageNet dataset. ImageNet is an image database with more than 14 million images classified into to more than 20,000 categories. It was created for image recognition competitions [18]. While Inception V3 is a CNN model which consists of several convolutional layers and maximum pooling steps. These are pre-trained with random initialization weights using Adam optimizer. Different parameter values like batch size, number of epochs, learning rate etc. can be randomly initialised and then Grid Search/Random search can be applied to find the best combination of parameters among all. For example, we will be setting batch size, number of epochs and learning rate to 2, 100 an 1e-4, respectively. Momentum of 0.5 for optimizer and the activation function ReLu will be used. There are some packages/libraries in python for hyperparameter tuning of CNN model. But we can leverage the Grid Search method provided by sklearn package to find out best parameters for our model from the combination space. The final model will be trained based on the best hyper parameter combination selected from the grid search method. This hyperparameter optimization can also be done using a n open source Future Gadget Laboratory [22]. Other approaches that can be considered are genetic algorithms and Bayesian optimization can be used efficiently for parameter tuning. Whole dataset will be randomly divided into 80% and 20% with former being training data and later being test data. We will be using 5-fold cross validation method to validate or model accuracy. So, the average value of these 5 folds in each epoch will be counted as final value of model accuracy. Here, cross-validation is necessary as it will reduce overfitting because it takes average value of accuracy from 5 different train test split on the same subset of data. All these processes will be applied to both our model ResNet50 and InceptionV3 and whichever has higher accuracy and promising results, will be selected as final model for our task.

Hyperparameter tuning can be done by a single person or it can be divided in groups as well. One group or person can fin-tune learning rate while other can find best batch size and epoch. As we are using pre-trained model, we will not be looking at configuring filter size, kernel size, maxpooling, dropout rate etc. of all layers. Also, we will apply transfer learning technique that was realized when using ImageNet data to makeup for both insufficient data as well as training time. The schematic representation of the models for the prediction of COVID-19 is as per [2] in annexure obtained from [14 - Figure-4]. As ResNet has large number of layers in the network, it has high time complexity, which can be reduced using a bottleneck design [21]. Skip connection resolves the problem of vanishing gradient, so that features are not lost in a network with high complexity. To reduce overfitting and improve convergence, mini-batch of stochastic gradient descent (mSGD) can be used to minimise the objective function with categorical-across-entropy loss. We can change the last layers configuration and hyperparameters tuning which we have added to change the output of pre-trained model to give the output classes of our new dataset.

B. Evaluation framework

Choosing right evaluation metrics is a crucial task when dealing with medical images because the final model tested on right evaluation metrics will actually be of use in live scenarios instead of bad model with some evaluation metrics that identify it as good model. Evaluation will be done in 2 parts. 1) technical evaluation, 2) Manual evaluation by experts in this field.

For technical evaluation, we will test our model on 20% test data that we have split at the start of model training process. It is important to note that only accuracy Is not a good measure as we want to remove False negatives as much as possible so that COVID-19 spread can be prevented. So, we will be looking for some other evaluation measures as well here. We will also be looking at AUC, Precision, Recall, f-1 score as well as sensitivity and specificity of each class. Each of these values can be monitored for each fold of k-fold cross validation and average values of all fold can be compared to see it how it performs on unknown test data.

For manual evaluation, prediction outputs of this model can be manually interpreted with the help of expert radiologist, so they can compare how model is performing. Also, mislabeled x-ray images and actual label for those images should be provided to these experts for them to understand where the model is lacking, so they can provide some guidance on what can be improved. To visualize how model is predicting/making decision, Grad-CAM [20] heat map approach can be used described in [15]. This will also be provided to experts to understand which area does our model focus on to make decisions. This manual evaluation is done with experts of different healthcare systems to ensure there is no bias in decision made by the experts. Because experts from different regions or different area will give their opinion about model, there will be no bias in the final verdict for a model, as it will be done by combined decision of all.

This model can act as a second opinion for radiologist working in health industry. It can significantly reduce the work of medical experts and help them make more accurate diagnosis in their work. As proposed model can predict fast, diagnostic process becomes fast than regular approach, which can save time of experts so they can focus on more critical cases.

C. Technical issues

There can be some technical issues as well while executing this project, as every algorithm can potentially suffer from different shortcomings of their own.

One important issue is the dataset shift as the input data evolves over time due to shifting population of patients and data that was used in model training may not be a true sample from whole population anymore. For that, the performance of the model should be carefully observed, and some drift measures should be set to identify drift in a model so that it can be updated/retrained to predict more accurately. Other solution can be to prospective studies instead of retrospective studies for CXR images [23].

Other challenge is that evaluation metrics may not reflect medical applicability. These metrics are not easily understandable by medical health professionals. Even though multiple evaluation measures are provided none of them will accurately reflect what is most important to patients, i.e. whether the outcome of model be of actual benefit to the patient care. There are some approaches proposed like decision curve analysis, which aims to resolve the benefit of using a trained model to guide subsequent actions [24].

Another challenge is when providing algorithm with images to train, it will use whatever pattern it may find in images to provide more accurate result, but by doing so, it can sometimes exploit unknown cofounding variables which are not reliable, which ultimately weakens the ability of algorithm to generalize on the new dataset. Ongoing work is required regarding this to understand the specific features learned by machine learning CNN model.

Another issue can be of the model's ability to generalization to new population. It is hard due to difference in input data because of variations in equipment and different administrative practices in local clinics. There should be some way provided to used standardized techniques while curating data across industry and site-specific training can be given to adapt new population with global standardization while maintaining those.

D. Biases

Model bias is somewhat intertwined with model's ability to generalization.

As the historical data is used for training, we may not know if this data comes from a specific age-group or a specific gender or a location (i.e. Australia). If that is the case or if there is some other bias present when training a model than it is a bias model and may not generalize well in real world settings. For example, for different ethnicity, hospital mortality prediction algorithm has varying accuracy [25].

As stated earlier, if images CNN extracts whatever features it can from images to predict with high accuracy, so if x-ray images contain same health care facility name/device name etc. it may prone to use it as a feature as well, which will be biased towards only CXR images from that facility and may not give that accurate results on other x-ray images.

Apart from biases, there are also some privacy & ethical considerations needed to keep in mind. Healthcare systems are concerned with how user's data is used in machine learning systems. Users expect that their healthcare providers are following safety measures to protect their right to privacy of their personal information or data. The data should be anonymized to mitigate privacy beaches and malicious use of data should be prevented. For user-centric application like healthcare, ensuing ethical use of data is foremost important. One important consideration is to understand how data collection can harm a person's dignity and well-being. To ensure fair and ethical working of machine learning systems, it is crucial to have a clear understanding of machine learning system in ambiguous and complex situations.

E. List of Abbreviations

CAD - Computer-aided diagnostics

CXR - chest x-ray

CT - Computer Tomography

SIFT - Scale-invariant feature transformation

SVM - support vector machines

ORB - oriented fast and rotated binary

DWT - discrete wavelet transform

WFT - wavelet frame transform

WPT - wavelet packet transform

KNN - k-nearest neighbour

CNN - Convolutional neural networks

OCT - Optical coherence tomography

GAN - generative adversarial network

ResNet - Residual neural network

R-CNN - Regions with Convolutional Neural Networks

DNN - Deep Neural Network

YOLO - you only look once

References

[1] Caicedo JC, Cruz A, Gonzalez FA. Histopathology image classification using bag of features and kernel functions. In: Conference on artificial intelligence in medicine in Europe. Berlin: Springer; 2009. p. 126–35

[2] Rublee E, Rabaud V, Konolige K, Bradski GR. Orb: an efficient alternative to sift or surf. Citeseer. 2011;11:2.

[3] Mueen A, Baba S, Zainuddin R. Multilevel feature extraction and X-ray image classification. J Appl Sci. 2007;7(8):1224–9.

[4] Yuan X, Yang Z, Zouridakis G, Mullani N. Svm-based texture classification and application to early melanoma detection. In: 2006 international conference of the IEEE engineering in medicine and biology society. IEEE. 2006. p. 4775–8.

[5] Parveen N, Sathik MM. Detection of pneumonia in chest X-ray images. J X-ray Sci Technol. 2011;19(4):423–8

[6] Paredes R, Keysers D, Lehmann TM, Wein B, Ney H, Vidal E. Classification of medical images using local representations. In: Bildverarbeitung für die Medizin 2002. Berlin: Springer. 2002. p. 71–4.

[7] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. Imagenet large scale visual recognition challenge. Int J Comput Vision. 2015;115(3):211–52.

[8] Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M. Medical image classification with convolutional neural network. In: 2014 13th international conference on

control automation robotics & vision (ICARCV). IEEE; 2014. p. 844–8.

[9] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172(5):1122–31.

[10] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. 2017. arXiv preprint arXiv:1712.04621.

[11] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. In: Advances in neural information processing systems. 2017. p. 3856–66.

[12] Afshar P, Mohammadi A, Plataniotis KN. Brain tumor type classification via capsule networks. In: 2018 25th IEEE international conference on image processing (ICIP). IEEE. 2018. p. 3129–33.

[13] Yadav, S.S., Jadhav, S.M., Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6, 113 (2019).J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed.

[14] Ali Narin, Ceren Kaya, Ziynet Pamuk, Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks

[15] Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, 103792. Advance online publication.

[16] Kabid Hassan Shibly, Samrat Kumar Dey, Md. Tahzib -Ul-Islam, and Md. Mahbubur Rahman, COVID Faster R-CNN: A Novel Framework to Diagnose Novel Coronavirus Disease (COVID19) in X-Ray Images *In Review*

[17] Ioannis D. Apostolopoulos1, Tzani Bessiana, COVID-19: Automatic detection from X-ray images utilizing Transfer Learning with Convolutional Neural Networks, arXiv:2003.11617

[18] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115: 211-252, 2015.

[19] Y. Fang, H. Zhang, J. Xie, M. Lin et al., "Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR", Available at: https://pubs.rsna.org/doi/full/10.1148/radiol.2020200432#panepcw-references

[20] Selvaraju R.R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Proceedings of the IEEE International Conference on Computer Vision. 2017. Grad-cam: visual explanations from deep networks via gradient-based localization; pp. 618–626.

[21] Wu, Z., Shen, C., and Van Den Hengel, A. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. Pattern Recognition, 90: 119-133, 2019.

[22] https://github.com/Kaixhin/FGLab

[23] Kelly, C.J., Karthikesalingam, A., Suleyman, M. et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 17, 195 (2019). https://doi.org/10.1186/s12916-019-1426-2

[24] Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008; 8:53.

[25] Chen IY, Johansson FD, Sontag D. Why Is My Classifier Discriminatory? In: 32nd Conference on Neural Information Processing Systems (NeurIPS). 2018.

ANNEXURE

1.

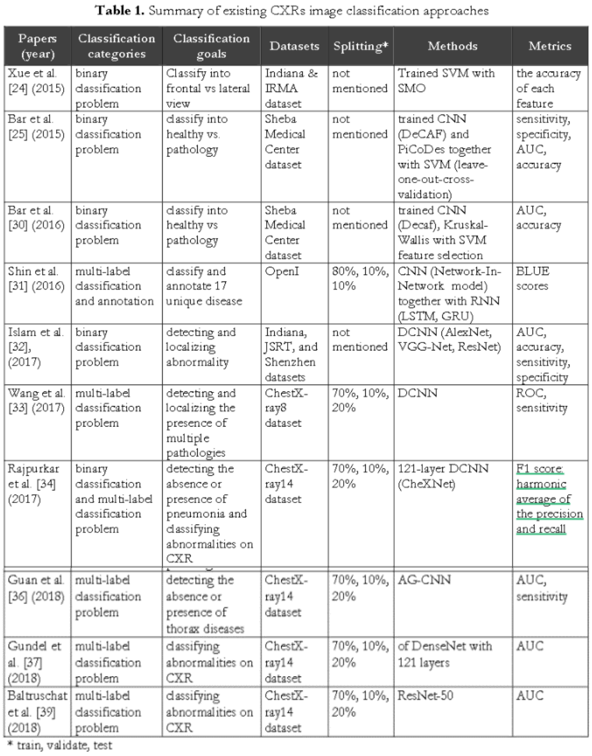
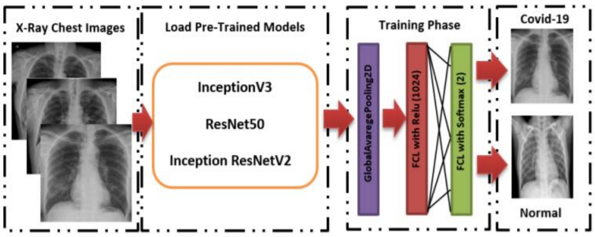


Table 1. Summary of existing CXRs image classification approaches

Papers (year)	Classification categories	Classification goals	Datasets	Splitting*	Methods	Metrics
Xue et al. [24] (2015)	binary classification problem	Classify into frontal vs lateral view	Indiana & IRMA dataset	not mentioned	Trained SVM with SMO	the accuracy of each feature
Bar et al. [25] (2015)	binary classification problem	classify into healthy vs. pathology	Sheba Medical Center dataset	not mentioned	trained CNN (DeCAF) and PiCoDes together with SVM (leave-one-out-cross-validation)	sensitivity, specificity, AUC, accuracy
Bar et al. [30] (2016)	binary classification problem	classify into healthy vs pathology	Sheba Medical Center dataset	not mentioned	trained CNN (Decaf), Kruskal-Wallis with SVM feature selection	AUC, accuracy
Shin et al. [31] (2016)	multi-label classification and annotation	classify and annotate 17 unique disease	OpenI	80%, 10%, 10%	CNN (Network-In-Network model) together with RNN (LSTM, GRU)	BLUE scores
Islam et al. [32], (2017)	binary classification problem	detecting and localizing abnormality	Indiana, JSRT, and Shenzhen datasets	not mentioned	DCNN (AlexNet, VGG-Net, ResNet)	AUC, accuracy, sensitivity, specificity
Wang et al. [33] (2017)	multi-label classification problem	detecting and localizing the presence of multiple pathologies	ChestX-ray8 dataset	70%, 10%, 20%	DCNN	ROC, sensitivity
Rajpurkar et al. [34] (2017)	binary classification and multi-label classification problem	detecting the absence or presence of pneumonia and classifying abnormalities on CXR	ChestX-ray14 dataset	70%, 10%, 20%	121-layer DCNN (CheXNet)	F1 score: harmonic average of the precision and recall
Guan et al. [36] (2018)	multi-label classification problem	detecting the absence or presence of thorax diseases	ChestX-ray14 dataset	70%, 10%, 20%	AG-CNN	AUC, sensitivity
Gundel et al. [37] (2018)	multi-label classification problem	classifying abnormalities on CXR	ChestX-ray14 dataset	70%, 10%, 20%	of DenseNet with 121 layers	AUC
Baltruschat et al. [39] (2018)	multi-label classification problem	classifying abnormalities on CXR	ChestX-ray14 dataset	70%, 10%, 20%	ResNet-50	AUC

* train, validate, test

2.


X-Ray Chest Images
Load Pre-Trained Models
InceptionV3
ResNet50
Inception ResNetV2
Training Phase
GlobalAvaregePooling2D
FCL with Relu (1024)
FCL with Softmax (2)
Covid-19
Normal