Harshit Mishra (19BCE0799)

CSE4020 – MACHINE LEARNING

DIGITAL ASSIGNMENT 1

Question:

You have to download 6 to 8 recent journal papers from reputed journal (IEEE, Elsvier, Springer, MDPI, Hindwai etc.) belonging to same machine learning application like covid19 reproduction rate prediction. Read out the paper completely and go through the methodology used, Results, Performance metrics, pros and cons. Try to find out a core pitfall for future work from these papers.

Topic Chosen: Pneumonia Detection…

Pneumonia is an infection of lungs which makes breathing difficult for the patient inheriting it. It is can be caused due to virus, bacteria or fungi and a patients natural functioning of lungs. The topic of pneumonia detection was chosen keeping COVID19 in mind. Coronavirus infects a person's lungs and causes severe pneumonia like traits. Deaths that occurred due to COVID19 were predominantly due to progressive pneumonia in patients and hence early identification of pneumonia is necessary to avoid life threating consequences in a patient and treat one in the best possible way by the earliest.

Identification of Pneumonia is majorly proposed by chest X-ray images, but recent studies have initiated these by studying how a patient's coughing traits and affected symptoms with highly encouraging accuracies.

Harshit Mishra (19BCE0799)

Literature Review:

Sl. No.	Paper Title and Year	Method (Algorithm)	Results (performance metrics)	Advantage and Limitation
1.	Pneumonia Detection using an Improved Algorithm based on Faster R-CNN (21-Apr-2021) **Hindwai Publication**	They have employed Regions with Convolutional Neural Network approach to detect pneumonia on chest X-rays. To detect the anchor-box they have used K-mean++ algorithm and it helps them analyse target region.	They have used various evaluation metrics including ROC, Interaction over Union (IoU). They achieved an accuracy of 67.56% on test set	The very first limitation of this implementation is that u need to have a high end computing graphic card, i.e., Nvidia GeForce GTX 1080 configuration. Also, despite having these configurations the accuracy of test set was not as expected coz test accuracy was at a bare minimum of less than 70%.
2.	Automatic Detection of Acute Bacterial Pneumonia from Chest X-ray Reports . (1-Nov-2000)	They have used the accuracy of an NLP system named as **SymText** for extracting pneumonia features from chest X-ray reports. They have manually used ensemble learning of 7 different physicians with	The algorithm they used gave them an accuracy ranging from 0.86 to 0.96 as different test sets were considered. Also, the Recall, Precision, Specificity and agreement of the model was	The major backdrop of this approach is that it is not fully automated. It requires manual predictions from highly trained physicians who can clearly specify and relate the features of disease to a patient.

	Journal of the American Medical Informatics Association	different features to check majority voting.	0.94, 0.87, 0.91 and 0.84 when physicians assisted there reading. When it was fully automated, these values reduced to 0.46, 0.89, 0.95 and 0.54	Although, they have also implemented automated classification but its accuracy is as low as 46%. The next issue with their manual approach is although it gave an accuracy of 96%, they need 7 (approx.) different types of physicians. This accuracy of 96% was achieved only due to the difference in approach of each physician, the accuracy will decrease drastically if lesser number of them were to considered. Also, the paper stands out-dated as it was published in 2000.
3.	Pneumonia Detection on Chest X-Ray using Machine	They have used NIH ChestX-Ray14 dataset to train and test their classifier. They	They have achieved the best accuracy with Multilayer	The lung segmentation approach which was used in the paper is still

	Learning Paradigm. (01-Nov-2019) **Proceedings of 3rd International Conference on Computer Vision and Image Processing.**	have used 5 different classifiers. They are Multilayer Perceptron, Random Forest, Sequential Minimal Optimization, Classification via regression and Logistic Regression. They have user CXR images and have used ROI bounded lungs for feature selection.	Perceptron. Accuracy with MLP was 92.33%, with Random Forest was 90.534%, SMO was 89.806%, Classification via Regression was 91.990% and with Logistic Regression it was 91.505%. Also, the F1 score of MLP was 91.753, Sensitivity was 86.408%, Specificity was 98.058% and Precision was 97.802%.	under development. Also, it needs further enhancements for paediatric cases and pleural effusions. The method they proposed clearly ignored the medical background of the patient. This often happens with the CXR ROI bounded feature extraction that even cardiac related patients are termed to have pneumonia. This is because the ROI of a Chest X-Ray only studies the shape of lung and the lung size can be affected by lot many diseases including pneumonia and since dataset didn't deal with such cases, the accuracy seemed

				amazing, which might not be the case with heart related patients.
4.	Pneumonia Detection Using CNN based Feature Extraction. (17-Oct-2019) **Institute of Electrical and Electronics Engineering.**	They have used Densely Connected Convolutional Neural Network as classifier to predict if a given chest x-ray images is infected by pneumonia. They have divided their architecture into 3 different stages which includes the pre-processing stage, feature extraction stage and the classification stage. For Feature extraction they have used the architecture of DenseNet-169.	They have used various evaluation metrics which included C value, gamma value and AUC value. They have compared their results with similar works proposed using different algorithms which included SVM, NB, RF etc. The AUC value they achieved was 60.90%.	Similar to previous paper's limitation, their feature extraction detector detects only the boundary of lungs and hence does not consider the change in size and shape of lungs due to some other infections which are very much affected by heart related diseases. They have considered these kinds of images in their test set and hence we can see it is reflected in their accuracy as well. The accuracy stands less than 60.9% which suggests the poor performance of their model

				when past medical history of patient is not considered.
5.	A Deep Feature Learning Model for pneumonia Detection applying a Combination of mRMR Feature Selection and Machine Learning Models . (06-Nov-2019) **IRBM journal of Alliance for Engineering in Biology an Medicine Alliance pour le génie biologique et médical**	They have used pneumonia dataset which consisted of around 6000 samples. They have deployed Convolutional Neural Networks with 5-step process. This process involved, Sequential Layer, Feature Selection, Max Pooling, Flattening and Fully Connected Dense Layer. For feature extraction they have used Image Data Generator, where they flipped and scaled the image horizontally and also applied shear tension on the images within the image range. By deploying these they have made sure that they don't get into issues of lung's shape and hence minimized the	The had True Negative Value of 1270, True Positive Value of 1274, False Positive value of 10 and False negative value of 5. The accuracy of their model stands at 99% which is very overwhelming. Also, they achieved a learning rate of 0.2% which suggested with execution of every epoch they had an increment in their accuracy.	With their approach of implementation, they need 3 different kinds of feature extraction methods namely AlexNet, VGG-16 and VGG-19. These makes the algorithm very slow and training it takes more than feasible time. This difficulty is added on by the need of high-end computation Graphical Processing Unit or Online Runtime environment of Google Colab which had issues with saving runtime and hence required training of model again and again.

		error due to heart related diseases.		
6.	Differentiating novel Coronavirus performance from general pneumonia based on Machine Learning . (19-Aug-2020) **BioMedical Engineering OnLine**	The paper proposed a deep learning model named COVNet which is based on visual features from volumetric Chest Images to distinguish COVID19 from normal pneumonia. They used three-dimensional Chest X-Rays for feature identification. They applied U-net to crop the lung region as the ROI of both two dimensional and three dimensional features were extracted by COVNext based on their ROIs.	They used dataset with 40% normal pneumonia cases, 30% of COVID acquired pneumonia cases and 30% of non-pneumonia cases. They achieved a sensitivity of 90%, specificity of 96% while the CAP values stood at 87% and 92%. The accuracy value of training set was a whopping 96% and on test set it was 95%.	The feature extractor of U-net was designed on deep learning models and had millions of weighted neurons with embedded information and hence the model lacked interpretability and transparency in feature detection. Also, U-Net was serialized with ResNet-18 and fed into ResNet-23 which made the model computationally extensive and expensive process and hence was not very feasible.
7.	Pneumonia Detection using Deep Learning Approaches (25-Feb-2020)	They have used real-time dataset from Ruian People's Hospital and have categorised them into pneumonia set and non-	Their model achieved an accuracy of 94.16%, specificity of 100% and sensitivity of 88.62%. The	Their model's foundation is on COVID related pneumonia and does not identify normal pneumonia cases. The

	2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE).	pneumonia set. They have then identified 34 statistical texture of pneumonia sets and ROI region of interest declination including 16 gray-level co-occurrence matrices of GLCM features. They have deployed various classification models including Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbour (KNN) and Logistic Regression.	area under AUC curve of receiver operating characteristics was 0.99.	model hence is limited only on data acquired due to COVID acquired pneumonia systems. Also, they have used real time data from a single source. This suggests that their model is very much skewed to specific geographic location and their acquired traits. This is hence not universally applicable. The other issue with their model is that the dataset is relatively small. The accuracy they have achieved is pretty high but this could be very much due to overfitting of model and low-test case count. Hence, we cannot be sure of achieving the same results on

				any other chest x-ray that we get.

Plagiarism Report: