**Author Names and Introduction:** ~0.5-1 page

1. A title and abstract are not needed. This should save you some pace.
2. The author names should be written in alphabetical order of the surnames.
3. The introduction should identify and outline related work from the literature.
4. The introduction should briefly compare and contrast the algorithms presented here in the context of the literature.

**Technical Contributions:** 1-1.5 pages per team member

1. This section should describe the classifiers used in the study and any processing done to the images before they were used for classification.
2. Key steps should be justified.
3. Each team member should write about at least one of the classifiers (the one they implemented). Use subsections and label each subsection with the name of the person implementing it ("Graph Method - Ash").
4. **Introduction**

Pneumonia is a prevalent infectious disease with considerable morbidity and mortality. It is characterised by inflammation in the alveoli of the lungs, and most commonly diagnosed through a physical examination of a chest radiograph [1]. While this diagnosis is often done manually by a clinician, there is an ongoing need to automate this process to assist clinicians in their decision-making processes.

Machine learning models are being used in various fields [2], and present the opportunity to classify pneumonia in a patient through x-ray image analysis. This report implements and compares two machine learning methods: decision trees and convolutional neural networks (CNNs).

The benefits of CNNs in the context of image classification

* Pre-processing is minimal
* Highly specialised – through further training, fine-tuning…

The drawbacks of CNNs include:

Similarly, DL models have also been assessed for this type of application and is illustrated through the works of

Similarly, numerous papers have explored the success of DL models, specifically neural networks, for this type of application. Convolutional neural networks (CNNs) are a specialized subset of neural networks which are built specifically for feature extraction by identifying visual patterns in images with minimal preprocessing [3] It is particularly effective for analyzing large amounts of data from their unprocessed state [4]. CNNs involve a network of layers, where each layer takes the output of each preceding layer as input. Each layer applies a linear or non-linear operation on the image, with the extracted features becoming more refined over each layer [5]. A common method of implementation involves transfer learning, where a pretrained model is used as the starting point for this task. This enables generalized features to be identified without needing to undergo redundant training, which increases computational efficiency. Fine-tuning can then be implemented for the specific data being trained [2]. Wang et al. compared the outcomes of different CNN architectures on a prominent chest x-ray dataset (ChestX-Ray14), consisting of 1,315 labelled pneumonia images and 6,041 normal images [6]. The four architectures ResNet-50, AlexNet, GoogLeNet, and VGG-16 achieved area under the receiver operating characteristic (AUROC) scores were 0.63, 0.55, 0.60, and 0.51 respectively, where 1 indicates perfect classification. These scores indicate that deeper architectures such as ResNet-50 have greater success compared to the shallower VGG-16 model. Guendal et al. [7] studied the success of DenseNet-121 (with 121 layers), another deep architecture, which achieved an AUROC score of 0.76 on the same ChestX-Ray14 dataset. While it may be deduced that deeper models are more successful across all datasets, it has been suggested that the success of the model is influenced by the dataset, and specifically the size and split of the data. This has been confirmed by Rajamaran et al. who implemented a VGG-16 model on a different dataset and achieved an AUROC score of 0.99 and classification accuracy of 96.2%. Thus, while there are numerous CNN architectures available for use, their success is dependent on a range of factors outside of the specific layers being implemented. The primary benefit of deep learning is the fact that computational performance increases as the scale of data increases, with a model consistently learning as more information is provided. However, this process often takes longer to train, though it is reduced through transfer learning. Particularly for large-scale datasets, this is far more beneficial than traditional machine learning models which are restricted by the parameters set during training. Therefore, CNNs pose great potential in effectively classifying pneumonia.

~~CNN’s have become a dominant method in image analysis and classification, and as a result are often applied in radiology. They operate by adaptively learning spatial features and patterns of images [5]. Through the implementation of specific convolution layers, pooling layers, and dense layers, an algorithm can effectively identify the desired features.~~

1. **Convolutional Neural Networks – Erica**

From the literature, it is evident that CNNs have great potential in radiology with improved efficiency and ability to obtain higher training and validation accuracy [2]. This CNN implementation applies transfer learning with both ResNet50 and VGG-16 architectures, two of the most successful networks identified in the literature.

The premise ResNet50 architecture is a model which is based around ‘shortcut connections’ which avoids the vanishing gradient that occurs when a model is trained on too many layers, meaning the model can be trained on up to 3000 layers. In comparison, VGG-16 is a far shallower model which only uses convolutions, demonstrating the potential simplicity of a classification model. This model reduces the number of trainable variables to encourage faster learning and more robustness to overfitting.

Transfer learning was chosen as it allows the model to learn faster by starting with a higher accuracy. As full training is such a computationally expensive task and because the dataset was relatively small (which is often the case in medical imaging), implementing a pre-trained model requires less computational effort, data, and training time. The motivation of fine-tuning was based on the observation that while shallower layers are often more generic, such as edges, deeper layers can be trained for a specific purpose [5].

The CNN was implemented as a binary classifier using the Tensorflow Keras libraries on the same small dataset collected by Kermany et al.

1. Pre-processing:

Extensive pre-processing was not required as CNNs have the benefit of being able to receive raw image data as input [2]. The classifier first loads the train and test directories for both normal and pneumonia classes. Both the normal and pneumonia classes in the train directory were split in an 80/20 ratio to create separate training and validation subsets. Once these subsets were defined, images were resized to a standardised size of (224, 224), the recommended dimensions for the both the ResNet and VGG models [2]. This also ensures all images were given the same weighting, and reduce background noise [8]. Augmentation of each image also involved normalising each pixel by dividing each value by 255.

1. Models

The CNN models were then implemented. The fine-tuning and parameters were based on a seminal study by He et al. [9] which explores the optimisation of a residual network (ResNet) in comparison with other architectures including VGG.

After the pretrained ResNet-50 model was implemented, the first convolutional layer was removed, and fine-tuning was implemented for modification. This involved implementing a global average pooling layer, a Dense layer with a ReLU activation, and then another Dense layer with a softmax activation. These decisions were based on the work of Misra et al. which achieved

This same process was applied on the VGG model. The following layers were implemented: Dense Dropout, Dense, Dropout, Dense. The choice of layers were based on

The model was then compiled.

The pretrained ResGen layers were frozen as retraining these layers is redundant, and so training the

As the layers in ResGen50 are pretrained, these layers were frozen and only the additional layers were trained. The model was then compiled.

This process was repeated on the VGG model,

1. Training

For both the ResNet and VGG models once the original and fine-tuned layers had been compiled,

Both the ResNet and VGG models were then trained on the training dataset, comprised of 56% of the original dataset images. This training used 10 epochs, a batch size of 32, and learning rate of 0.001. These values were chosen as they have proven successful values in a study by Hashmi et al., in which pneumonia classification using a DenseNet model achieved an accuracy of 98.43% [2]. They were deemed an appropriate trade-off between achieving an accurately trained model while maintaining some computational efficiency.

1. Evaluation

The ResNet and VGG models were then evaluated on the validation and test datasets using the Model.evaluate() method. This produced loss and accuracy values for the datasets. To visualise the improvement of the model during training, the accuracy and loss were plotted against each epoch for both the train and validation data. A confusion matrix was then plotted, and values for precision, recall, and f1-score were calculated to evaluate the performance of the classification models.

1. **Results**

Results

20 epochs: loss: 1.4084 - accuracy: 0.8072

High loss and high accuracy: suggests big errors on a small amount of data

However, these metrics were not deemed reliable for comparison as the datasets were unbalanced.

Discussion:

* Overtraining
* Plot of loss vs accuracy

NORMAL

0: 112

1: 122

PNEUMONIA

0: 3

1: 387

|  |  |  |
| --- | --- | --- |
|  | ACTUAL: Pneumonia | ACTUAL: Normal |
| PRED: Pneumonia | TP: 387 | FP: 122 |
| PRED: Normal | FN: 3 | TN: 112 |

TN FP

FN TP

* Currently,

Train loss:

Validation loss:

The outcomes of the CNN classifiers showed similar success true positive cases, identifying pneumonia, with ResNet correctly identifying 334 cases and VGG correctly identifying 274 cases. This said, both models also produced a high number of false positives, incorrectly classifying normal images as pneumonia 208 and 171 times respectively, of a total of 234 normal cases. From these values, it is evident that both CNN models have significant bias towards pneumonia.

While this model is not perfect, the bias towards pneumonia is preferrable as false negative outcomes have a far greater cost [13]. That is, if a patient was to have pneumonia but was classified as normal, this may lead to ineffective treatments which can have detrimental consequences.

The likely reason for the skew in results (with high counts of true positives and false positives) is because of the imbalance of data. Using a normal loss function (which was the case in this model) was likely to result in a model which was biased towards the dominating class. It is therefore unsurprising that the model was skewed towards classifying pneumonia since the training set contained 1,982 pneumonia images compared with 672 normal images. Training of a future model may apply a weighted loss function to balance the data.

SLIDES

\*\* is it worth commenting on the fact that more TPs and FPs is better/ lower risk???

* Future

[1] T. P. Htun, Y. Sun, H. L. Chua, and J. Pang, “Clinical features for diagnosis of pneumonia among adults in primary care setting: A systematic and meta-review,” *Scientific reports,* vol. 9, no. 1, pp. 7600-7600, 2019.

[2] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, “Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning,” *Diagnostics (Basel, Switzerland),* vol. 10, no. 6, pp. 417, 2020.

[3] N. M. Elshennawy, and D. M. Ibrahim, “Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images,” *Diagnostics (Basel, Switzerland),* vol. 10, no. 9, pp. 649, 2020.

[4] Z. Yue, L. Ma, and R. Zhang, “Comparison and Validation of Deep Learning Models for the Diagnosis of Pneumonia,” *Computational Intelligence and Neuroscience,* vol. 2020, pp. 1-8, 09/18, 2020.

[5] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights into Imaging,* vol. 9, no. 4, pp. 611-629, 2018/08/01, 2018.

[6] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases." pp. 3462-3471.

[7] S. Guendel, S. Grbic, B. Georgescu, S. K. Zhou, L. Ritschl, A. Meier, and D. Comaniciu, “Learning to recognize Abnormalities in Chest X-Rays with Location-Aware Dense Networks,” 03/12, 2018.

[8] R. Almamlook, S. Chen, and H. Bzizi, *Investigation of the performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images*, 2020.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." pp. 770-778.