Ethical Evaluation

John Robson (w19036980)

# Bias Analysis

Having an awareness of potential bias that can be present in data is the first step of recognizing the problem that AI models face. Biases can appear through the entire process from the collection of data, the pre-processing, analysis of the data and even during model creation. (Table 1).

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Collection** | **Pre-Processing** | **Analysis Issues** | **During Modelling** |
| Data Misrepresenting | Over Filtered | Misleading Representation | Class Imbalance |
| Incorrect Readings | Missing Values | Forced Justification | Variance & Bias |
| Inaccurate/False Data | Outliers |  |  |

Table 1 - Bias that can appear during the model creation process

During data collection, it is important to try and collect data in a manner that doesn’t jeopardise whom the data is intended to represent by singling to a small group of the audience rather than the overall collective. Tools or machinery used are required to be calibrated and functional, this is to ensure the correct readings are being observed and extracted. In cases when data is being collected from people, checking for inaccurate or falsified data is a must, as this could cause bias-related issues further down the process.

When pre-processing data, identifying the occurrence of over filtering is important as removing variables from a dataset can cause it to change the meaning and can alter the output of the system later. Missing values require sanitising from datasets as well as outliers as their presence can cause changes to the model that can decrease accuracy. When analysing the data, it’s necessary to avoid trying to make data speak the point you want it too, this is known as forced justification which leads to failures to create accurate conclusions from the data. Also, outputting your results in a manner that misleads a reader is poor practice as this misrepresents what the data concluded.

When modelling the data, checking for class imbalance held by the data should be a consideration, by making sure that fair distributions of classes are given, to avoid the classification model from becoming bias to the majority class. Variance and Bias are a balance that needs to be kept in check, as overfitting and underfitting of data need to be avoided during this stage.

The data used in this study is from Kaggle, which contains x-ray images of lungs present with and without pneumonia. This data has no other information besides the label which classifies the images with or without pneumonia. Due to this, there is no way of knowing if there is an issue of bias contained during this stage.

# Ethical Discussion

This study’s final product will attempt to produce a model capable of classifying x-ray images of lungs, which have pneumonia present or not present in them, and because of this, there are potential ethical implications that need to be addressed.

Several consequences can occur from the result of the model. Incorrect diagnoses of patients could result in failure to receive needed treatment or alternatively, sending patients for unneeded treatment. This could cause a loss of trust for the organisation. The cost of health care can increase from a false-positive result, as the patient will be required to pay for further un-needed tests or treatments. This could also impact the cost of health insurance policies from false-positive results, costing a wider range of patience as the count of treatments increase. Overall, this could lead to mental health impacts, as all previous points could potentially cause stress to patients.

It is a legal requirement to ensure any personal information is kept safe and secure as covered by GDPR. The process of our system can be used as a pipeline for classifying images and requires no personal information to complete its task. Therefore, the requirements of GDPR would not be required for this system. Regarding the data used in this study, the Kaggle dataset is publicly available and stripped of identifiable information and as this is the case, no considerations are required regarding GDPR.

# Technical Solutions

Following the issue of having no personal information on the participants of the Kaggle set, there is no method to review the collection process to identify a bias that occurred during this stage. On inspection of the Kaggle dataset, it was noted to be small which make creating an accurate model difficult, and with no way to inspect the dataset for bias during collection, it was decided to create new samples using data augmentation to create a larger dataset of data for the training and testing sets. This was performed by creating new copies of previous samples and performing manipulations to the images. The list of parameters used for this process are displayed in Table 2.

|  |
| --- |
| Rescale = 1./255.0  rotation\_range = 6  zoom\_range = 0.3  brightness\_range = [0.8, 1.3]  horizontal\_flip = True  vertical\_flip = True |

Table 2 - Variables used to augment data

The parameters chosen were to ensure that the images stayed to a consistent standard with the Kaggle dataset, by only manipulating features that could occur with X-ray images like scale, rotation, and brightness. The use of sheer would sometimes cut out parts of the image that could have the pneumonia present and could have an impact on the overall accuracy, so it was decided to not use this parameter. Through this process the control over positive and negative pneumonia images was achievable, giving the ability to reduce the chances of class imbalances from occurring, ensuring the classes are fairly represented for the model.

# Solution Verification

A test was conducted using the Kaggle dataset without any data augmentation applied (Figure 1) and another with data augmentation (Figure 2) applied. Without utilizing the augmentation options, the model has a high accuracy rating against the test set from Kaggle, but the validation accuracy fluctuates heavily from a small loss to a huge loss of accuracy. This is possibly a sign of the model being overfitted, and a bias occurring as it too closely related to the data set, which would explain why the validation set performed so poorly.

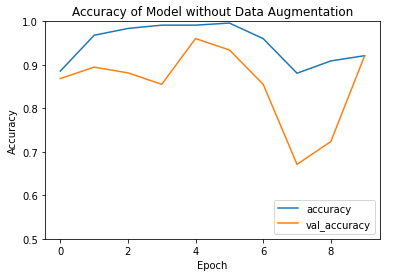


Figure 1 - CNN Accuracy Without Data Augmentation

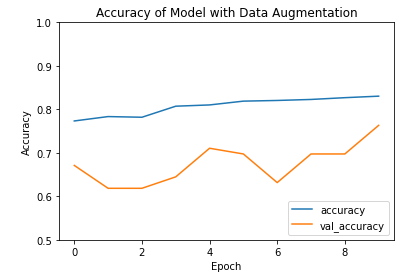


Figure 2 - CNN Accuracy with Data Augmentation

By utilizing augmentation changes to the Kaggle dataset, the testing performance of the model falls to around 80% with a steady increase of accuracy over increasing epochs. More importantly, the validation set follows a similar trend of accuracy to the test, with less dramatic increases and decreases of accuracy. By using data augmentation, the CNN model appears to show less bias to its train set but suffered from a loss of accuracy overall with the test and validation set.

Seeing the loss of accuracy as an issue, ensembling was suggested between several different models. By doing this, multiple models can vote in the classification of a sample, providing a more accurate result and reducing the chances of misclassification. After testing this theory, the system was able to produce an accuracy of 93% against the test set and 96% with the optional validation set.