Face Mask Detection (FMD) Project

Table of Contents

List of Figures

List of Tables

1.0 Introduction

1.1 Overview of research area and focus of the study

The broad research area of this study pertains to image processing and computer vision applications in the domain of face mask detection. Given the importance of public health measures on a global scale, face mask detection is a particularly pertinent and important subject within this large realm. For these procedures to be followed in a variety of settings, including public venues, precise face mask detection models are essential. However, getting high accuracy in face mask detection is a difficult task, particularly when dealing with unbalanced datasets. It is more challenging for the model to learn and generalize efficiently in an unbalanced dataset since there are often disproportionality more samples of one class than the other. Therefore, this study's particular focus is to find ways on how to enhance the accuracy of face mask detection models trained with unbalanced datasets.

1.2 Aims and importance

The aims and importance of this project are stated as follows:

Aims

1. Enhance Face Mask Detection Accuracy:
    a. The primary aim of this project is to greatly increase the face mask detection model accuracy. This involves cutting down on false positives and false negatives in order to identify people wearing masks with greater accuracy.
2. Mitigate Unbalanced Dataset Challenges:
    a. Address the issues related with unbalanced datasets in the context of face mask detection. This entails devising effective ways for making the model function effectively even when the quantity of examples for different classes is uneven.

Importance

1. Public Health and Safety:
    a. Accurate face mask models are critical in preserving public health and safety. The accomplishment of the project's aims can help minimizing the spread of contagious disease and protect community well-being.
2. Challenging Data Imbalance:
    a. Many real-world datasets are naturally imbalanced, and tackling this issue is critical. The strategies and techniques used in this project to handle unbalanced datasets may have a broader implications and uses.

1.3 Summarize of literatures

In this section, each article will be summarised using a review matrix.

Table 1.0
Face Mask Detection Using Deep Learning

Review Matrix	
Citation For This Paper	Kodali, R. K., & Dhanekula, R. (2021). Face mask detection using Deep Learning. *2021 International Conference on Computer Communication and Informatics (ICCCI)*. https://doi.org/10.1109/iccci50826.2021.9402670
Name(s) of the Author(s)	1. Ravi Kishore Kodali 2. Rekha Dhanekula
Title of Article, Title of Journal and Publication Year	Title of Article: Face Mask Detection Using Deep Learning Title of Journal: 2021 International Conference on Computer Communication and Informatics (ICCCI-2021) Publication Year: 2021
Context and Motivation	1. The investigation is motivated by the desire to prevent the spread of the COVID-19 virus by requiring people to wear masks. 2. Face mask detection is seen as a realistic preventive technique for saving lives and assisting frontline workers.

Methodology	1. Using well-known libraries like TensorFlow, Keras, Scikit-learn, and OpenCV, a fundamental Convolutional Neural Network (CNN) model is created to improve accuracy. 2. The task is broken down into three stages: a. Pre-processing b. CNN training c. Real-time classification
Pre-processing	1. Techniques used: a. Grayscale conversion of RGB images b. Image resizing c. Image normalization
CNN training	1. The study's core comprises training a CNN model with two output neurons to categorise faces with and without masks using Softmax activation. 2. To optimise the model's performance, categorical cross-entropy is used as the loss function. 3. The suggested model has an validation accuracy of 96%
Real-time classification	1. The model visibly labels individuals who are not wearing masks with a red rectangle and display the label "NO MASK" during real-time categorization. 2. Individuals wearing masks are distinguished by a green rectangle around their faces.

Note: The pre-processing and CNN training will be used as motivation for this FMD project, but not real-time classification since the goal of this project is to increase the model's accuracy when trained with unbalanced classes, not to deploy the project after the model has been trained.

Table 2.0

A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining

Review Matrix	
Citation For This Paper	Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of under sampling, oversampling, and smote methods for dealing with imbalanced classification in Educational Data Mining. *Information*, *14*(1), 54. https://doi.org/10.3390/info14010054
Name(s) of the Author(s)	1. Tarid Wongvorachan 2. Surina He 3. Okan Bulut
Title of Article, Title of Journal and Publication Year	Title of Article: A Comparison of Under sampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining Title of Journal: Information 2023, 14(1), 54 Publication Year: 2023
Context and Motivation	1. The paper addresses the issue of class imbalance in educational data mining, specifically when creating predictive models for various educational applications. 2. Class imbalance in educational datasets can have an impact on predictive model accuracy, especially when these models assume balanced class distributions.
Methodology	1. The research evaluates several sampling techniques for addressing the class imbalance problem in educational data mining

Sampling techniques	1. Techniques used: a. Random oversampling (ROS) b. Random under sampling (RUS) c. Hybrid resampling
Key findings	1. The study's findings indicate that random oversampling works well for moderately imbalanced data. 2. The hybrid resampling strategy (SMOTE paired with RUS) appears to function best in terms of boosting classification accuracy for very imbalanced data.

Note: For this FMD project, only ROS and RUS will be used for sample procedures.

1.4 Hypothesis and research questions

Hypothesis

In general terms, hypothesis is a statement that is utilised to conduct an experiment on the connection of two or more variables or a suggested clarification about several observation phenomena. Thus, in this report the hypothesis that will be made are stated below:

1. Balancing techniques:
    a. Oversampling and Under sampling with Data Scaling
        - The implementation of oversampling with data scaling will result in a (higher/lower) accuracy of the model trained for face mask detection
        - The implementation of under sampling with data scaling will result in a (higher/lower) accuracy of the model trained for face mask detection

    b. Oversampling and Under sampling without Data Scaling
        - The implementation of oversampling without data scaling will result in a (higher/lower) accuracy of the model trained for face mask detection
        - The implementation of under sampling without data scaling will result in a (higher/lower) accuracy of the model trained for face mask detection

    c. Data Augmentation with varying trrain-test-splits
        - The implementation of data augmentation with a train test split of 8:2 will result in a (higher/lower) accuracy of the model trained for face mask detection
        - The implementation of data augmentation with a train test split of 9:1 will result in a (higher/lower) accuracy of the model trained for face mask detection
2. Regularization:
    a. Regularization with different train-test-splits
        - The implementation of regularization with a train test split of 8:2 will result in a (higher/lower) accuracy of the model trained for face mask detection
        - The implementation of regularization with a train test split of 9:1 will result in a (higher/lower) accuracy of the model trained for face mask detection

3. Colour mode:
    a. RGB vs Grayscale colour mode
        - The implementation of RGB colour mode will result in a (higher/lower) accuracy of the model trained for face mask detection
        - The implementation of Grayscale colour mode will result in a (higher/lower) accuracy of the model trained for face mask detection
4. Activation functions:
    a. Sigmoid vs SoftMax activation
        - The implementation of sigmoid as an activation during model training will result in a (higher/lower) accuracy of the model trained for face mask detection
        - The implementation of softmax as an activation during model training will result in a (higher/lower) accuracy of the model trained for face mask detection

Research Questions

In general terms, research questions are specified questions where the research is conducted to deliver the answers to, thus normally positioned at the first step of the project. Consequently, research questions emphasize on research, dictate the methodology and hypothesis, and provide the route for all steps in analysis, inquiry, and report. Thus, in this report the research questions that will be made are stated below:

- Does the accuracy of using oversampling technique with data scaling tend to be higher than using under sampling technique with data scaling or vice versa?
- Does the accuracy of using oversampling technique without data scaling tend to be higher than using under sampling technique without data scaling or vice versa?
- Does the accuracy of using data augmentation with a train test split of 8:1 tend to be higher than the model trained with a train test split of 9:1 or vice versa?
- Does the accuracy of using regularization with a train test split of 8:1 tend to be higher than the model trained with a train test split of 9:1 or vice versa?
- Does the accuracy of using RGB colour mode tend to be higher than Grayscale colour mode or vice versa?
- Does the accuracy of using sigmoid as an activation tend to be higher than softmax as an activation or vice versa?

1.5 Approach to achieving aims

To accomplish the aim of improving the accuracy of the model trained with an unbalanced dataset, the following strategies will be employed:

1. Balancing techniques:
    a. Oversampling and Under sampling with Data Scaling
    b. Oversampling and Under sampling without Data Scaling
    c. Data Augmentation with varying train-test splits
2. Regularization:
    a. Regularization with different train-test-splits
3. Colour mode:
    a. RGB vs Grayscale colour mode
4. Activation functions:
    a. Sigmoid vs SoftMax activation

1.6 Key findings and contribution of work

Key findings

```
54/54 [==============================] - 2s 30ms/step - loss: 0.2936 - acc: 0.8823
Test Accuracy For A_train_scaled Before Data Undersampling:  88.228440284729

54/54 [==============================] - 2s 30ms/step - loss: 0.2936 - acc: 0.8823
Test Accuracy For X_train_scaled Before Data Resampling:  88.228440284729

48/48 [==============================] - 1s 30ms/step - loss: 0.2743 - acc: 0.8805
Test Accuracy For A_train_scaled After Data Undersampling:  88.05482983589172

60/60 [==============================] - 2s 30ms/step - loss: 0.2542 - acc: 0.9248
Test Accuracy For x_train_scaled After Data Resampling:  92.47764348983765

54/54 [==============================] - 2s 31ms/step - loss: 0.6934 - acc: 0.6492
Test Accuracy For A_train Before Data Undersampling:  64.91841673851013

54/54 [==============================] - 2s 33ms/step - loss: 0.6934 - acc: 0.6492
Test Accuracy For X_train Before Data Resampling:  64.91841673851013

48/48 [==============================] - 2s 34ms/step - loss: 0.6634 - acc: 0.6880
Test Accuracy For A_train After Data Undersampling:  68.79895329475403

60/60 [==============================] - 2s 33ms/step - loss: 0.5205 - acc: 0.7528
Test Accuracy For x_train After Data Resampling:  75.2761721611023

54/54 [==============================] - 2s 37ms/step - loss: 0.3833 - acc: 0.8514
Test Accuracy For X_train With Data Augmentation With Train Test Split 8:2 :  85.13985872268677

27/27 [==============================] - 1s 32ms/step - loss: 0.3108 - acc: 0.8543
Test Accuracy For X_train1 With Data Augmentation With Train Test Split 9:1 :  85.43123602867126
```

```
54/54 [==============================] - 2s 32ms/step - loss: 0.4070 - acc: 0.8765
Test Accuracy For X_train After Regularization With Train Test Split 8:2:  87.64568567276001

27/27 [==============================] - 1s 32ms/step - loss: 0.4089 - acc: 0.8636
Test Accuracy For X_train1 After Regularization With Train Test Split 9:1 :  86.36363744735718

54/54 [==============================] - 2s 32ms/step - loss: 0.2936 - acc: 0.8823
Test Accuracy For X_train_scaled With Color Mode Of RGB:  88.228440284729

54/54 [==============================] - 2s 36ms/step - loss: 0.3798 - acc: 0.8322
Test Accuracy For Xtrain_scaled With Color Mode Of GrayScale:  83.21678042411804

60/60 [==============================] - 2s 35ms/step - loss: 0.2542 - acc: 0.9248
Test Accuracy For x_train_scaled Using Sigmoid As Activation:  92.47764348983765

60/60 [==============================] - 2s 37ms/step - loss: 0.2701 - acc: 0.9169
Test Accuracy For x_train_scaled Using Softmax As Activation:  91.68858528137207
```

Figure 1.0 - Main Results

Note: Prior to any techniques applied, the benchmark for the test accuracy is 88.228440284729 The accuracy of each model will vary depending on different strategies. On the basis of different strategies, the accuracy will be stated as follows:

1. Balancing techniques:
   a. Oversampling and Under sampling with Data Scaling
      i. Researcher had concluded that data oversampling test accuracy is higher than data under sampling
      ii. Researcher had concluded that the test accuracy with data scaling is higher than the test accuracy without data scaling
   b. Oversampling and Under sampling without Data Scaling
      i. Researcher had concluded that data oversampling test accuracy is higher than data under sampling
      ii. Researcher had concluded that the test accuracy without data scaling is lower than the test accuracy with data scaling
   c. Data Augmentation with varying train-test-splits
      i. Researcher had concluded that the test accuracy of using data augmentation with a train-test-split of 9:1 is higher than using a train-test-split of 8:2

2. Regularization:
    a. Researcher had concluded that the test accuracy of using regularization with a train-test-split of 8:2 is higher than using a train-test-split of 9:1
3. Colour mode:
    a. Researcher had concluded that the test accuracy of using RGB colour mode is higher than using Grayscale colour mode
4. Activation functions:
    a. Researcher had concluded that the test accuracy of using Sigmoid as the activation is higher than using SoftMax as the activation

Contribution of work

Refer to research conducted by (Kodali & Dhanekula, 2021), Table 1.0, we can notice that the researchers used a variety of pre-processing approaches, such as grayscale conversion of RGB images, etc., and that the CNN model they trained had an accuracy of 96%. However, with reference to Figure 1.0, we find that the model trained on images that had been previously processed in RGB colour mode as opposed to grayscale had a better test accuracy.

Refer to research conducted by (Kodali & Dhanekula, 2021), Table 1.0, we can notice the researches employed softmax as the activation during CNN model training, yielding an accuracy of 96%. The model trained with sigmoid activation, on the other hand, has a greater test accuracy than the model trained with softmax activation, as shown in Figure 1.0.

Refer to research conducted by (Wongvorachan et al., 2023), Table 2.0, we can notice that the researchers asserted that random oversampling is effective for unbalanced data. With the reference to Figure 1.0 as support, we can show that this idea is sound because oversampling has a greater test accuracy than under sampling. In addition, with the help of this book's Chapter 56, which claims that oversampling performs better than under sampling (D. et al., 2023), we can conclude that oversampling is a superior method for handling imbalance datasets than under sampling.

According to Mr.Siddhardhan, data scaling is essential when training neural network models (Siddhardhan, 2023) and using Figure 1.0 as a guide, we can demonstrate that data scaling does make a difference when compared to a model trained without data scaling. The difference in test accuracy is approximately 23%.

2.0 Methodology

2.1 How was the data gathered

The dataset was collected initially from Kaggle, however it was discovered to be balanced, which did not match the project's criterion for an imbalanced dataset. To remedy this, additional images from online sources were scrapped, specifically images of persons wearing and not wearing masks. The dataset was expanded by combining images scrapped online, notably those displaying people wearing masks, with an existing dataset from Kaggle that also included mask-related images. This merging operation was carried out to create an imbalanced dataset that met the project's requirements, and as illustrated in Figure 2.0, we had created our unbalanced dataset. Refer to Appendix 1 for the web scrapping process code and Appendix 2 for the merger of scrapped and original dataset code.

```
There are 4752 images in with_mask dataset
There are 3828 images in without_mask dataset
```

∴ We can clearly see that now our dataset is **UNBALANCE** after merging the initial with_mask dataset with the scrapped WithMask dataset

Figure 2.0 - Result After Merging Scrapped and Original Dataset

2.2 Research methods selection

A research method is a systematic approach or technique for conducting scientific studies, gathering data, analysing information, and answering research questions or hypotheses (Editorial Team, 2023). There are two sorts of research methods:

1. Quantitative research methods which include:
    a. Surveys
    b. Questionaries
    c. Test
    d. Databases
    e. Organization records
2. Qualitative research methods
    a. Interviews
    b. Observation
    c. Focus groups

Since web scrapping was used for data collection in which it is not often regarded as a research method in and of itself; rather, it is a data acquisition technique utilised within quantitative research methodology (Scraping Robot, 2023).

2.2.1 Justification for method choice

As aforementioned, web scrapping is not often regarded as a research method in and of itself; rather, it is a data acquisition technique utilised within quantitative research methodology. However, the use of web scrapping in this FMD project is exclusively for achieving FMD project's aims, which is to improve the accuracy of the model trained with an **unbalanced dataset** and mitigate unbalanced dataset challenges. As a result, web scrapping is essential to make the original dataset to be unbalanced after merging, as shown in Figure 2.0.

2.2.2 How were these methods applied to analyse the research question or problem?

The section of hypotheses and research question served as the framework for the analysis in this study. They led the experimental design and offered a framework for evaluating the accuracy of various strategies and parameters in the face mask detection model trained.

Furthermore, the data collection approach, as mentioned in section 2.1, was crucial in creating the dataset required for this project. The merging of web scrapped images with an existing Kaggle dataset was a critical step in creating the imbalanced dataset required to fulfil the project's aims.

In terms of research methods, web scraping was used as a data gathering tool within the context of quantitative research methodology, as elucidated in section 2.2. It was chosen because of its compatibility. As a result, web scraping was used as a means to an end rather than a stand-alone research method.

In summary, the methods described in the hypothesis and research questions section were used to design experiments and assess the model's accuracy by using different evaluation metrics such as f1 score, recall, precision, support, test accuracy, and loss function, while the data gathering technique, which included web scraping and dataset merging, was critical to creating the required dataset.

3.0 Results

3.1 Restate the study purpose and summary of all models trained

As mentioned in section 1.1, the purpose of this study is to enhance the accuracy of face mask detection models trained with unbalanced datasets.

Table 3.0

Summary of all models trained with and without DS using different strategies

Summary					
Strategies Used		Test Accuracy		Loss Function	
DS	Without DS	DS	Without DS	DS	Without DS
Under sampling	Under sampling	88.05%	68.80%	27.43%	66.34%
Over sampling	Over sampling	92.48%	75.28%	25.42%	52.05%
Initial	Initial	88.23%	64.92%	29.36%	69.34%

Note: Prior to any techniques applied, the benchmark for the test accuracy is 88.228440284729

: DS = Data Scaling

∴ Since we can see that DS improves test accuracy while decreasing loss function, we may conclude that DS improves test accuracy while decreasing loss function as illustrated in Table 3.0. As a result, Table 4.0 will contain a summary of all models trained with DS using different strategies.

Table 4.0

Summary of all models trained with DS using different strategies

Summary		
Strategies Used	Test Accuracy	Loss Function
Data Augmentation with TTS of 8:2	85.14%	38.33%
Data Augmentation with TTS of 9:1	85.43%	31.08%
Regularization with TTS of 8:2	87.65%	40.70%
Regularization with TTS of 9:1	86.36%	40.89%
Colour mode of RGB	88.23%	29.36%
Colour mode of Grayscale	83.22%	37.98%
Sigmoid As Activation	92.48%	25.42%
SoftMax As Activation	91.69%	27.01%

Note: TTS = Train Test Split

The evaluation metrics for several models trained with different strategies are shown in Figures 3.0 to 16.0, and the evaluation metrics utilised include precision, recall, f1-score, and support.

	precision	recall	f1-score	support
0	0.87	0.86	0.87	768
1	0.89	0.90	0.89	948
accuracy			0.88	1716
macro avg	0.88	0.88	0.88	1716
weighted avg	0.88	0.88	0.88	1716

Figure 3.0 – Initial With Data Scaling (Benchmark)

	precision	recall	f1-score	support
0	0.60	0.65	0.63	768
1	0.70	0.64	0.67	948
accuracy			0.65	1716
macro avg	0.65	0.65	0.65	1716
weighted avg	0.65	0.65	0.65	1716

Figure 4.0 - Initial Without Data Scaling (Benchmark)

	precision	recall	f1-score	support
0	0.89	0.86	0.88	758
1	0.87	0.90	0.88	774
accuracy			0.88	1532
macro avg	0.88	0.88	0.88	1532
weighted avg	0.88	0.88	0.88	1532

Figure 6.0 - Under sampling with Data Scaling

	precision	recall	f1-score	support
0	0.65	0.80	0.72	758
1	0.75	0.57	0.65	774
accuracy			0.69	1532
macro avg	0.70	0.69	0.68	1532
weighted avg	0.70	0.69	0.68	1532

Figure 5.0 - Under sampling without Data Scaling

	precision	recall	f1-score	support
0	0.91	0.94	0.92	921
1	0.94	0.91	0.93	980
accuracy			0.92	1901
macro avg	0.92	0.93	0.92	1901
weighted avg	0.93	0.92	0.92	1901

Figure 7.0 - Over sampling with Data Scaling

	precision	recall	f1-score	support
0	0.70	0.86	0.77	921
1	0.83	0.65	0.73	980
accuracy			0.75	1901
macro avg	0.77	0.76	0.75	1901
weighted avg	0.77	0.75	0.75	1901

Figure 8.0 - Oversampling without Data Scaling

	precision	recall	f1-score	support
0	0.87	0.77	0.82	768
1	0.83	0.91	0.87	948
accuracy			0.85	1716
macro avg	0.85	0.84	0.84	1716
weighted avg	0.85	0.85	0.85	1716

Figure 9.0 - Data Augmentation With 8:2 Train Test Split

	precision	recall	f1-score	support
0	0.85	0.82	0.83	377
1	0.86	0.89	0.87	481
accuracy			0.86	858
macro avg	0.86	0.85	0.85	858
weighted avg	0.86	0.86	0.86	858

Figure 10.0 - Data Augmentation With 9:1 Train Test Split

	precision	recall	f1-score	support
0	0.85	0.88	0.86	768
1	0.90	0.88	0.89	948
accuracy			0.88	1716
macro avg	0.87	0.88	0.88	1716
weighted avg	0.88	0.88	0.88	1716

Figure 11.0 - Regularization With 8:2 Train Test Split

	precision	recall	f1-score	support
0	0.94	0.74	0.83	377
1	0.82	0.96	0.89	481
accuracy			0.86	858
macro avg	0.88	0.85	0.86	858
weighted avg	0.87	0.86	0.86	858

Figure 12.0 - Regularization With 9:1 Train Test Split

	precision	recall	f1-score	support
0	0.87	0.86	0.87	768
1	0.89	0.90	0.89	948
accuracy			0.88	1716
macro avg	0.88	0.88	0.88	1716
weighted avg	0.88	0.88	0.88	1716

Figure 13.0 - Colour Mode of RGB

	precision	recall	f1-score	support
0	0.83	0.78	0.81	768
1	0.83	0.87	0.85	948
accuracy			0.83	1716
macro avg	0.83	0.83	0.83	1716
weighted avg	0.83	0.83	0.83	1716

Figure 14.0 - Colour Mode of Grayscale

	precision	recall	f1-score	support
0	0.91	0.94	0.92	921
1	0.94	0.91	0.93	980
accuracy			0.92	1901
macro avg	0.92	0.93	0.92	1901
weighted avg	0.93	0.92	0.92	1901

Figure 15.0 - Using Sigmoid As Activation

	precision	recall	f1-score	support
0	0.83	0.78	0.81	768
1	0.83	0.87	0.85	948
accuracy			0.83	1716
macro avg	0.83	0.83	0.83	1716
weighted avg	0.83	0.83	0.83	1716

Figure 16.0 - Using SoftMax As Activation

3.2 Report of insignificant research findings

Based on Table 3.0, Table 4.0, and Figures 3.0 to 16.0, the statistically insignificant research findings are as follows:

1. Data augmentation with alternative train-test-splits (8:2 and 9:1) did not result in statistically significant difference in test accuracy or the classification report, which includes precision, recall, f1-score, and support. However, there is a considerable variation in the loss function, with the difference being roughly 7.25%.
2. Regularization with alternative train-test-splits (8:2 and 9:1) did not result in statistically significant difference in test accuracy, loss function, or classification report.
3. The choice of activation (Sigmoid vs. Softmax) did not result in statistically significant difference in test accuracy, or loss function. However, there is a considerable variation in the classification report, with an average difference of 19% for precision, recall, and f1-score.

3.3 Summary of key study outcomes

Among the strategies used, training neural network models using data scaling during pre-processing can significantly improve the learned model's accuracy. Second, in terms of both model trained with and without data scaling, over sampling outperforms under sampling. Third, Data Augmentation with TTS of 9:1 has somewhat better test accuracy than Data Augmentation with TTS of 8:2, however there is a substantial difference in their loss function of roughly 7.25%. Fourth, Regularisation with TTS 8:2 has somewhat better test accuracy and loss function than Regularisation with TTS 9:1. Fifth, the RGB colour mode has a higher test accuracy and loss function than the Grayscale colour mode. Finally, using sigmoid as activation has somewhat better test accuracy and loss function than using softmax as activation. To summarise, the model trained with the highest test accuracy and lowest loss function was trained using over sampling with data scaling using sigmoid as activation on an unbalanced dataset.

4.0 Discussion

4.1 Key findings interpretation

The key findings from this study can be summarized and interpret as follows:

1. Data scaling enhances model performance:
    a. The introduction of data scaling techniques enhanced the accuracy of face mask detection models while concurrently lowering the loss function. This conclusion emphasises the significance of pre-processing strategies like data scaling in increasing the model performance.
2. Over sampling outperforms Under sampling:
    a. Over sampling consistently outperformed under sampling across the investigated techniques, both with and without data scaling. This suggests that in this situation, it is more beneficial to address class inequality by boosting the representation of minority samples through over sampling.
3. Train-test split impact:
    a. Data augmentation with varied train-test-split ratios (8:2 and 9:1) produced somewhat improved test accuracy when using the train-test split of 9:1. The loss function which was about 7.25% greater in the 9:1 train-test split which showed a significant difference as compare to the test accuracy which was about 0.29%.

4. Colour mode and activation:
    a. RGB colour mode outperformed the Grayscale colour mode, showing smaller loss function and better test accuracy. Additionally, when compared to utilising softmax as an activation function, using sigmoid generated significantly higher test accuracy and a lower loss function.

Refer to Figure 1.0, Table 3.0 and Table 4.0, or Figures 3.0 to 16.0 provide the supporting evidence for the aforementioned key findings.

4.2 Comparative analysis and critical assessment

4.2.1 Compare and contrast to previous study

In comparing this study to the previous study as shown in Table 1.0 and Table 2.0, several key findings and difference emerge:

1. Image pre-processing methods
    a. In the previous study conducted by (Kodali & Dhanekula, 2021), Table 1.0, a variety of pre-processing techniques were used, including the conversion of RGB images to Grayscale. Their CNN model achieved an impressive accuracy of 96%.
    b. However, in this study as illustrated in. Figure 1.0 found that models trained on images processed in RGB colour mode outperformed those in grayscale in terms of test accuracy and loss function. This suggests that the choice of colour mode can have a significant impact on model accuracy.
2. Activation functions
    a. In the previous study conducted by (Kodali & Dhanekula, 2021), Table 1.0, researchers used softmax as the activation function during CNN model training, resulting in a 96% accuracy.
    b. In contrast, in this study as illustrated in Figure 1.0 demonstrated that a model trained with sigmoid activation achieved a higher test accuracy than one trained with softmax activation. This suggests that the choice of activation function can have a significant impact on model accuracy. According to (Arora, 2023), the researcher stated that sigmoid is chosen for binary and multi-label classification problems, whereas Softmax is used for multi-class classification problems. Hence, this project is a binary classification, sigmoid outperforms softmax.

3. Handling class imbalance
    a. In the previous study conducted by (Wongvorachan et al., 2023), Table 2.0, researchers asserted the effectiveness of random over sampling for dealing with unbalanced data, a finding supported by Figure 1.0, which showed that over sampling led to a higher test accuracy compared to under sampling.
    b. Furthermore, our study aligned with these findings, and we incorporated insights from a separate source from Chapter 56 in a book (D. et al., 2023) to conclude that over sampling is a superior method for addressing imbalanced datasets compared to under sampling.
4. Data scaling
    a. Mr.Siddhardhan (Siddhardhan, 2023) emphasized the importance of data scaling when training neural network models, a point validated in our study. With the guidance of Figure 1.0 demonstrated that data scaling made a substantial difference, resulting in an approximate 23% increase in test accuracy when compared to models trained without data scaling.

4.2.2 Strengths and limitations of this study

The following list includes this study's strengths and limitations:

Strengths

- Data scaling enhancement
    - This study effectively demonstrated the positive impact of data scaling on model accuracy.
- Comparative analysis
    - This study conducted a thorough comparative analysis of various pre-processing techniques and activation functions, providing valuable insights for researchers in the field.
- Effective handling of class imbalance
    - This study supports the notion of over sampling is a superior method for handling class imbalance in face mask detection.

Limitations

- Lack of real-world context
    - This study does not address how the model might perform in real-world scenarios, such as varying lighting conditions, mask types, or other practical challenges.
- Model Complexity
    - This study does not delve into the complexity of neural network architectures. The choice of model architecture can also influence the results.

4.2.3 Discussion of unexpected findings

Due to the fact that this study focused on confirming or validating specific hypotheses and expanding upon the body of knowledge already available in the field of face mask detection, it appears that there were no particularly unexpected findings in this study.

5.0 Conclusion

5.1 Summary of hypothesis and purpose of this study

The purpose to conduct this extensive study is to deepen the understanding on what are the different strategies that may be used to improve the model's accuracy when trained on unbalanced datasets and compare which one gives the best accuracy, and which strategy is best suited for the model's accuracy when trained on unbalanced datasets. As a result of observing section 1.4 in hypothesis and research questions and finishing training all the models using different strategies, we can conclude that the model trained using over sampling with data scaling and sigmoid as activation on an unbalanced dataset has the highest accuracy, followed by model trained using over sampling with data scaling and softmax as activation, followed by the initial model trained with data scaling and colour mode of RGB prior to any strategies applied, followed by model trained using under sampling with data scaling, followed by model trained using regularization with a train-test split of 8:2, followed by model trained using regularization with a train-test split of 9:1, followed by model trained using data augmentation during image pre-processing with a train-test split of 9:1, followed by model trained using data augmentation during image pre-processing with a train-test split of 8:2, followed by model trained using colour mode of Grayscale during image pre-processing, followed by model trained using over sampling and without data scaling, followed by model trained using under sampling and without data scaling, and lastly the initial model trained without data scaling and colour mode of RGB prior to any strategies applied.

5.2 Significance of this study

Face mask detection has grown in relevance, particularly in today's environment, when public health and safety are vital. This study is significant because it has the ability to address critical challenges and contribute to the improvement of face mask detection technologies. Several key points emphasise the significance of this research. Firstly, the public health and safety. In this day and age, accurate detection of face masks is critical for public health and safety. This study contributes to the advancement of technologies that can aid in the enforcement of mask-wearing rules in a variety of contexts, ultimately lowering the transmission of dangerous diseases. Secondly, investor confidence. This study gives insights into the practicality and usefulness of alternative tactics for stakeholders and investors in face mask detection systems. This information can increase investor trust and encourage additional investments in this technology. Lastly, real-world applications. This study's findings have direct real-world applications. They can be used to ensure compliance with face mask mandates in places including airports, healthcare institutions, schools, etc.

5.3 Unanswered questions and future enhancements

According to the research questions in section 1.4, there aren't any unanswered questions. For future enhancements, researcher will incorporate real-world data, consider using diverse and larger datasets that represent a wide range of face mask recognition issues, such as different mask kinds, lighting situations, and backgrounds, to make the findings more useful in real-world circumstances. Furthermore, multimodal detection will be developed so that the model can recognise not just face masks but also other attributes such as face shields.

6.0 References

Arora, A. (2023, May 22). *Understanding logits, sigmoid, Softmax, and cross-entropy loss in deep learning*. W&B. https://wandb.ai/amanarora/Written-Reports/reports/Understanding-Logits-Sigmoid-Softmax-and-Cross-Entropy-Loss-in-Deep-Learning--Vmlldzo0NDMzNTU3#:~:text=Softmax%20can%20also%20do%20the,to%20%E2%80%9Cpick%20a%20class%E2%80%9D.

D., S., Pal, S., Le, D.-N., & Zaman, N. (2023). *Recent trends in Computational Intelligence and its application: Proceedings of the 1st International Conference on Recent Trends in Information Technology and its application (ICRTITA, 22)*. CRC Press.

Editorial Team, I. (2023, August 1). What is research methodology? (why it's important and types). https://www.indeed.com/career-advice/career-development/research-methodology

Kodali, R. K., & Dhanekula, R. (2021). Face mask detection using Deep Learning. *2021 International Conference on Computer Communication and Informatics (ICCCI)*. https://doi.org/10.1109/iccci50826.2021.9402670

Scraping Robot. (2023, June 29). *Web scraping: Best method to collect qualitative data*. https://scrapingrobot.com/blog/methods-to-collect-qualitative-data/

Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of under sampling, oversampling, and smote methods for dealing with imbalanced classification in Educational Data Mining. *Information*, *14*(1), 54. https://doi.org/10.3390/info14010054

Siddhardhan. (2023). *Face Mask Detection using Convolutional Neural Network (CNN)*. Retrieved October 23, 2023, from https://www.youtube.com/watch?v=-6K6-qQrIWc.

7.0 Appendix

Appendix 1 – Web Scrapping

```
# Import libaries needed for web scrapping
import os
import time
import pandas as pd
import urllib.request
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
```

```
driver = webdriver.Chrome()
driver.get("https://google.com/")
```

```
search = driver.find_element("name", "q")
search.send_keys("Girls selfie with mask", Keys.ENTER)
```

```
element = driver.find_element("link text", "Images")
element.get_attribute("href")
element.click()
```

```
value = 0
for i in range(2500):
    if os.path.exists("/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/WithMask"):
        print("File exists. Stopping execution.")
        break
    driver.execute_script('scrollBy('+ str(value) +', +2500);')
    value += 2500
    time.sleep(2)
```

```
elements = driver.find_elements("xpath", '//img[contains(@class,"rg_i") and contains(@class, "Q4LuWd")]')
```

```
try:
    os.mkdir("WithMask")
except FileExistsError:
    pass
```

```
count = 0
for i in elements:
    src = i.get_attribute('src')
    try:
        if src != None:
            src  = str(src)
            count += 1
            urllib.request.urlretrieve(src, os.path.join('WithMask','image' + str(count) + '.jpg'))
            if count%10 == 0: print("downloaded", count, "images")
        else:
            raise TypeError
    except TypeError:
        pass
```

```
driver.back()
```

```
search = driver.find_element("name", "q")
search.clear()
search.send_keys("Boys selfie with mask", Keys.ENTER)
```

```
element = driver.find_element("link text", "Images")
element.get_attribute("href")
element.click()
```

```
value = 0
for i in range(2500):
    if os.path.exists("/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/WithMask"):
        print("File exists. Stopping execution.")
        break
    driver.execute_script('scrollBy('+ str(value) +', +2500);')
    value += 2500
    time.sleep(2)
```

```
elements = driver.find_elements("xpath", '//img[contains(@class,"rg_i") and contains(@class, "Q4LuWd")]')
```

```python
for i in elements:
    src = i.get_attribute('src')
    try:
        if src != None:
            src  = str(src)
            count += 1
            urllib.request.urlretrieve(src, os.path.join('WithMask','image' + str(count) + '.jpg'))
            if count%10 == 0: print("downloaded", count, "images")
        else:
            raise TypeError
    except TypeError:
        pass
```

Appendix 2 – Merge of Scrapped and Original Dataset

```python
# Import libraries needed for merging scrapped WithMask dataset with our initial with_mask dataset
# so that we will have an unbalance dataset
import os
import shutil
```

```python
source_dir1 = "/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/WithMask"
source_dir2 = "/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/with_mask"
destination_dir = "/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/Unbalance Dataset"

# Create the destination directory
os.makedirs(destination_dir, exist_ok=True)

# Copy images from the source directories to the destination directory
for filename in os.listdir(source_dir1):
    shutil.copy(os.path.join(source_dir1, filename), os.path.join(destination_dir, filename))

for filename in os.listdir(source_dir2):
    shutil.copy(os.path.join(source_dir2, filename), os.path.join(destination_dir, filename))
```

```python
# Directory where with_mask (merged) dataset is stored
df = "/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/Image Processing and Computer Vision/with_mask"

# Directory where without_mask dataset is stored
df1 = "/Users/_fangkhai/Documents/Academic/Computer Science Semester 7/Image Processing and Computer Vision/without_ma

# Use os.listdir to get a list of files in both of the directory
image_files = [f for f in os.listdir(df) if f.endswith('.jpg')]
image_file = [f for f in os.listdir(df1) if f.endswith('.jpg')]

# Count the number of image files in both of the directory
num_images = len(image_files)
num_image = len(image_file)

print(f"There are {num_images} images in with_mask dataset")
print(f"There are {num_image} images in without_mask dataset")

print("\n∴ We can clearly see that now our dataset is \033[1mUNBALANCE\033[0m after merging the initial with_mask data
```