Class Imbalance Problem in COVID-19 Image Analysis

Arman Torikoglu - 0084452
Muhammed Esad Simitcioglu - 0084474

Introduction

In this project, we want to demonstrate the class imbalance problem in our COVID-19 dataset, and we want to try to overcome this problem in this project. Our dataset contains 3 three lung diseases with normal patient lung images. You can observe our code and dataset in the following links:

- Link: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database
- Code: https://github.com/EsadSimitcioglu/COMP-548

Weighted Loss Function

According to Aggarwal [1], It should be highlighted that most of the dataset utilized for COVID-19 diagnostic binary class or multiclass classification needs to be more balanced. Unfortunately, there is a lack of dataset problems for COVID-19 classification. This problem is not unique to COVID-19 classification, but it's a big challenge for AI researchers. It was challenging to get enough high-quality photos, especially in the early stages of the pandemic. Our dataset has 3616 COVID+ images and 10,192 COVID- images. Due to the significant imbalance between the number of COVID+ and COVID- images, our models tend to overfit by predominantly predicting COVID- for all instances in the dataset. Initially, we use a pre-trained AlexNet model (The weights coming from the ImageNet[2] dataset). In solving the class imbalance problem, there are lots of proposed techniques, such as SMOTE algorithm that we mentioned in our survey. But for this project, we implemented a weighted loss function. In a typical classification problem, all classes are treated equally, meaning each class has an equal impact on the loss calculation. However, when dealing with imbalanced datasets where some classes have significantly fewer samples than others, the model can be biased towards the majority class. To overcome this bias, the Weighted Cross Entropy Loss allows you to assign higher weights to minority classes and lower weights to majority classes. Doing so gives the loss function more importance to the underrepresented classes during the training process. This helps the model to focus on learning patterns and make accurate predictions for the minority classes, thereby reducing the impact of class imbalance.

PreTrained AlexNet With Cross Entropy Loss Function	Accuracy	Recall	Precision	F1 Score
Covid+ (class0)	0.9310	0.8029	0.9237	0.8591
Covid- (class1)	0.9310	0.9765	0.9331	0.9543

PreTrained AlexNet With Weighted Cross Entropy Loss Function	Accuracy	Recall	Precision	F1 Score
Covid+ (class0)	0.9088	0.6685	0.9758	0.7934
Covid- (class1)	0.9088	0.9941	0.8941	0.9415

Recall (Sensitivity):

- COVID+ (minority class) recall has decreased, indicating a slight reduction in the model's ability to identify positive cases correctly. The model has become more cautious in predicting COVID+ to reduce false positives.
- COVID- (majority class) recall has significantly improved, suggesting a reduction in false negatives for negative cases.

Precision:

- COVID+ precision has increased from 92% to 98%, indicating that the model is more precise in correctly identifying COVID+ cases. This implies a lower rate of false positives, where the model is making fewer incorrect predictions of COVID+ for samples that are COVID-.
- COVID- precision has slightly decreased from 93% to 89.5%, indicating a slight increase in false optimistic predictions for negative cases. However, it's essential to consider the overall trade-off between precision for different classes.

The weighted loss function has influenced the model's decision-making process by making it more conservative in predicting positive cases. This trade-off between precision and recall aims to reduce false positives and maintain a higher precision for the majority class. While the F1 score for the minority class (COVID+) has decreased, this approach can improve the overall model's performance when false positives are undesirable or costly.

Transfer Learning Architecture

The application of deep learning models for classifying CT or CXR(like in our dataset) images may have limitations due to the pre-trained weights of the most commonly used models. These weights are typically trained on the ImageNet dataset, which consists of natural images across 1000 different classes. Therefore, it is recommended to retrain a few convolutional layers to adapt the model to the specific domain of CT or CXR images when using transfer learning. To emphasize the significance of this adaptation, we conducted experiments using the same model, AlexNet, in two scenarios. In the first scenario, we employed the pre-trained version of AlexNet with weights from ImageNet [2]. In the second scenario, we utilized the model without pre-trained weights. By comparing these two scenarios, we highlighted the importance of fine-tuning the pre-trained or untrained weights when dealing with CT or CXR images.

Accuracy:

- Non-pre-trained AlexNet: Achieved an accuracy of 73.89%.
- Pretrained AlexNet: Demonstrated a higher accuracy of 93.10%.
- Transfer learning with pre-trained weights improves the model's ability to classify CT or CXR images, resulting in higher accuracy.

Recall:

- Non-pretrained AlexNet: Showed low recall (0.37%) for the minority class (COVID+) but high recall (100%) for the majority class (COVID-).
- Pretrained AlexNet: Improved recall for the minority class (80.29%) and the majority class (97.65%).
- Transfer learning enables the model to identify positive cases (COVID+) better while maintaining good performance on negative cases (COVID-).

Precision:

- Non-pre-trained AlexNet: Had a precision of 100% for the minority class but 73.86% for the majority class.
- Pretrained AlexNet: Showed high precision for both classes (92.37% for COVID+ and 93.31% for COVID-).
- Transfer learning with pre-trained weights helps reduce false positives and improves precision for both classes.

F1 Score:

- Non-pre-trained AlexNet: The F1 score for the minority class was 0.73%, reflecting an imbalance between precision and recall. The F1 score for the majority class was 84.97%.
- Pretrained AlexNet: Achieved higher F1 scores for both classes (85.91% for COVID+ and 95.43% for COVID-).
- Transfer learning enhances the model's ability to balance precision and recall, improving F1 scores.

These results highlight the importance of transfer learning using pre-trained models, such as AlexNet trained on the ImageNet[1] dataset. By leveraging prior knowledge, the pre-trained AlexNet demonstrates improved accuracy, recall, precision, and F1 scores for both positive and negative cases in the classification of CT or CXR images.

Data Augmentation

Data augmentation is a unique technique to expand the training data by modifying the existing data. The image modifications such as rotating, translating, or adding noise make the model more robust to noise and also helps to reduce the class imbalance in our context. We utilized Generative adversarial networks trained on a small training dataset to produce ~3500 COVID+ images, one problem was the model was only able to generate 112x112 images and we had to upscale to make them fit into our existing dataset that consist of 299x299 images, thus, we made it equal for both COVID+ and COVID- cases in the training set. The results showed a promising improvement in recall and minority class predictions.

Accuracy:

- Without data augmentation: The test accuracy is 93.10%
- With data augmentation: Data augmentation reduced total accuracy only slightly (92.98%).
- The quality of artificial data seems to be the main reason in the accuracy decrease, including the resolution difference in augmented dataset and original dataset and GAN accuracy.

Recall:

- Without data augmentation: The model demonstrates a satisfactory recall performance when identifying minority(80.29%) and the majority class (97.65%)
- With data augmentation: There is a significant recall performance for the COVID+ cases, more instances of smaller class made the model saturated, resulting in enhanced ability to capture positive cases (89.66%). While the majority class remains the same 97.65% (no change in majority class so it made sense)
- Data augmentation helped better identify the COVID+ cases

Precision:

- Without data augmentation: It shows a high precision for both classes (majority class is more accurate) 92.37% for COVID+ and 93.31% for COVID-
- With data augmentation: Precision accuracy for COVID- dropped when artificial images are added, the noise of those images made it harder to guess the opposite class. 98.16% for COVID+ and 87.06% for COVID-
- Transfer learning with pre-trained weights helps reduce false positives and improves precision for both classes.

F1 Score:

- Without data augmentation: Unbalanced F1 scores for the classes, 85.91% for COVID+ and 95.43% for COVID-
- With data augmentation: After augmentation, F1 scores are more even, and overall superior. 0.9372, 0.9205
- Data augmentation helped better grasp the minority data, hence, enables are more balanced accuracy as is shown in F1 score.

We observed that data augmentation strictly helps predicting the minority class, while it has no improvement in the opposite class. The overall accuracy remained about the same, while recalls and precisions for these classes were more affected from this method, thus F1 score was as expected, had a considerable increase.

Conclusion

In conclusion, there were two main methods to reduce the class imbalance, weighted loss function and data augmentation. Data augmentation was generally not advised due to the information loss. The results indicated that Data augmentation performed better in learning the minority class (in our case COVID+), while not as good as weighted loss function in predicting the majority class, the overall accuracy drops by a little amount due to this information loss. However, the F1 score in our experiments shows that data augmentation leads to a more even and fair prediction in favour of the minority class.

References

[1] Aggarwal, P., Mishra, N. K., Fatimah, B., Singh, P., Gupta, A., & Joshi, S. D. (2022). COVID-19 image classification using deep learning: Advances, challenges, and opportunities. *Computers in Biology and Medicine*, *144*, 105350. https://doi.org/10.1016/j.compbiomed.2022.105350

[2] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255).