Measuring Performance of Audio Augmentation Techniques

Objective:

- To research and implement methods to measure the performance of the audio augmentation.

What I have done:

- I evaluate the performance using Cross-Validation and Validation Split method.
- I evaluate it again by augmenting it.
- I visualize using a bar chart to compare the accuracies between both methods based on my audio datasets.
- Creating a documentation to explain everything.

Datasets:

For the dataset I use animal and environmental audio which are located in different folders. I use Librosa and Numpy for the libraries.

Link to Dataset (make sure to use your deakin email to access this or else you wouldn't be able to access it): Audio_Dataset

Augmentation Technique:

For the augmentation I use audiomentations for the library and they are:

- Gaussian Noises
- Time Stretching
- Pitch Shifting
- Audio Shift

Model:

The model that I use in this project is the 1D CNN which can classify the audio into 2 categories which are Animal sounds and Environmental sounds.

Performance Measurement:

I use both Cross-Validation and Validation Split method.

Cross-Validation Method:

- I split the dataset into k folds with 5 splits.
- The model is trained and validate on different subset of data.
- Compute the accuracy and gives the results of the mean accuracy and standard deviation.

Validation Split Method:

- I split the dataset into training and validation which is 80/20 split
- Compute the results after training.

Results:

The results are a comparison of both the original and augmented data.

Original:

Cross Validation: The results that I have gotten which is on the code as well is:

- Mean accuracy: 0.72
- Standard Deviation: 0.06

Validation Split: The results that I have gotten which is on the code as well is:

- Accuracy: 0.77

Augmented:

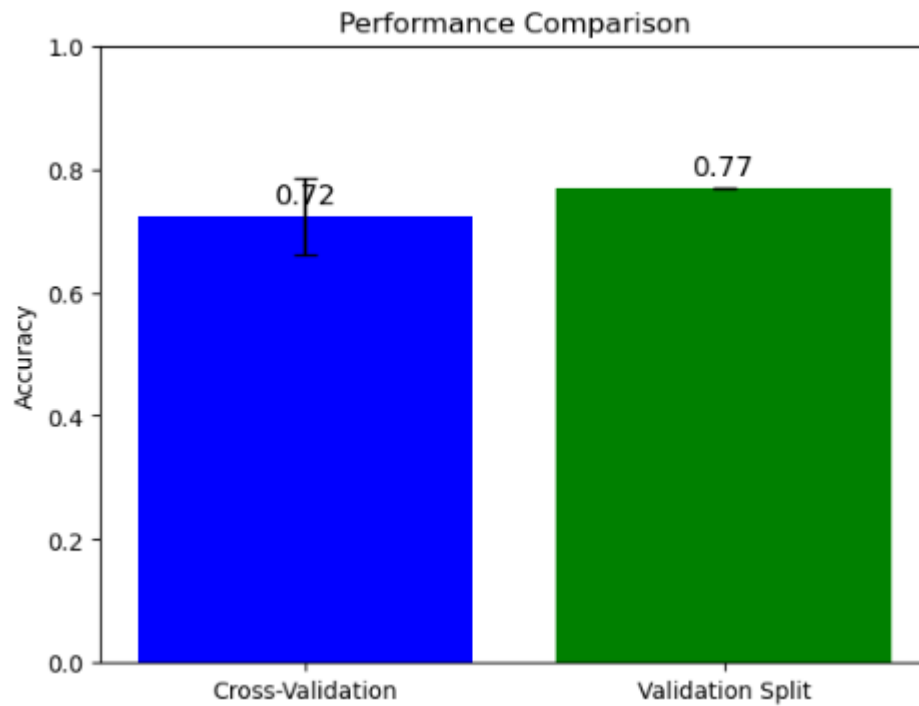Cross Validation: The results that I have gotten which is on the code as well is:

- Mean accuracy: 0.55
- Standard Deviation: 0.06

Validation Split: The results that I have gotten which is on the code as well is:
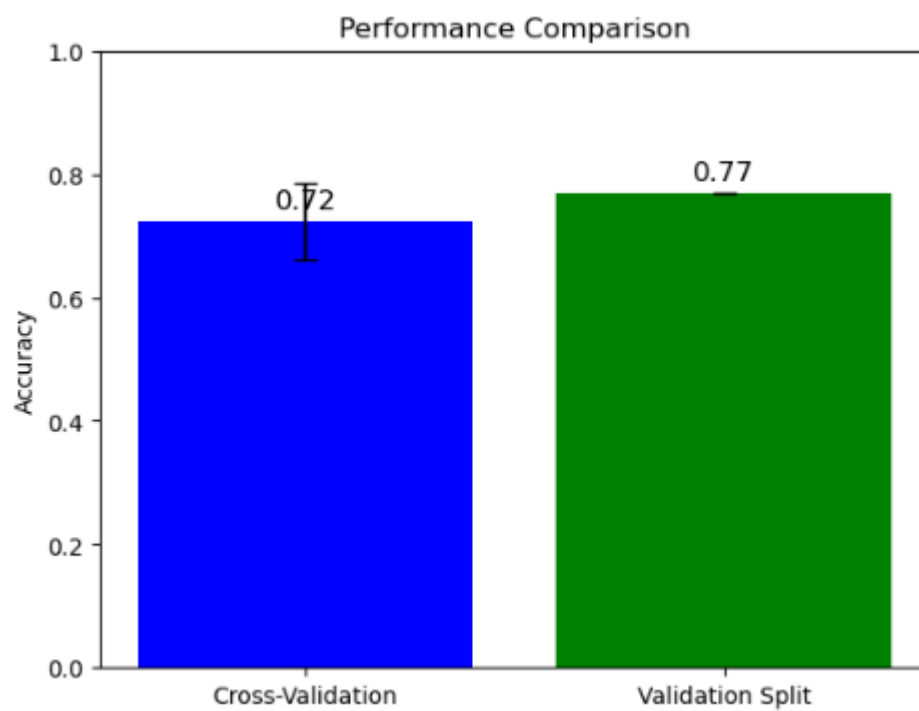
- Accuracy: 0.69

Visualisation:

Original Comparison:

Augmented Comparison:

Analysis:

- Original data proves to have more higher performance when compared to the augmented data in this project.
- The Cross-Validation appears to offer more robust evaluation with lower variability.
- The validation split seems to be less reliable for smaller dataset.

Areas to Improve:

1. **Data Augmentation:** Try different augmentation techniques and adjust it with the characteristic of the dataset.
2. **Dataset:** The current dataset doesn't have that many audios and diversity which causes the variability in cross-validation results.
3. **The model:** Currently I am implement the model using 1D CNN which might be too simple to be able to take on different complex patterns.
4. **Cross-Validation:** The variability in the method is probably because of insufficient folds or imbalance splits.