1. Introduction

In this assignment, I classified self-collected audio data of car and tram pass-bys using a machine learning model. The training and validation utilized data collected by other students.

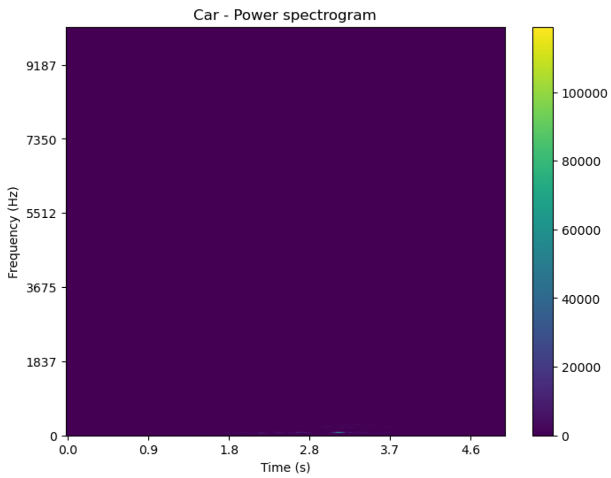
2. Data Description

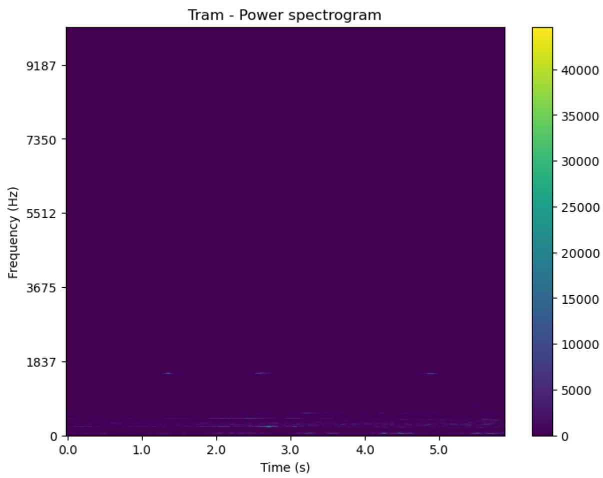
Data consists of two classes, Car and Tram, recorded near the Opiskelijankatu intersection in Hervanta, Tampere, Finland. The data collection spanned two days with similar weather conditions. Each data segment is 5-6 seconds long, with 24 samples for Car and 20 for Tram.

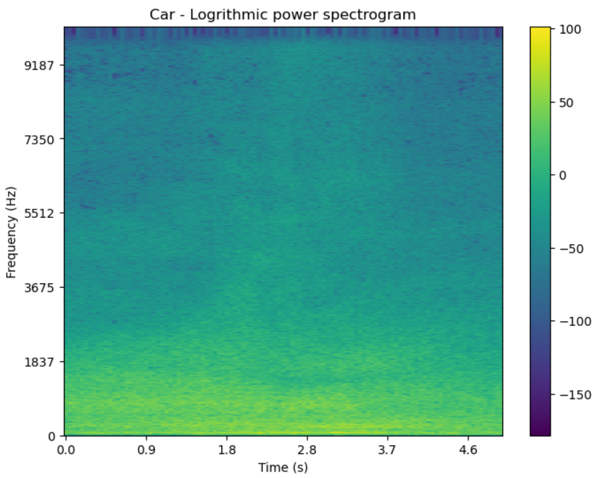
3. Feature Extraction and Selection

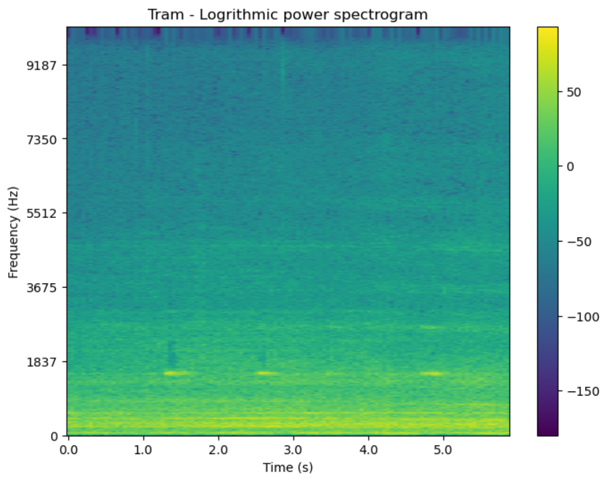
In this study, five features were analyzed: power spectrogram, log power spectrogram, mel power spectrogram, log power mel spectrogram, and MFCC. The objective was to identify features that clearly distinguish between Car and Tram sounds. Normalized audio data was used in all of the following work.

Throughout all these processes, attempts were made to verify and visualize differences in Car and Tram using multiple data points for each. For explanatory purposes, plots using one data point from each class will be presented. Starting with plotting a power spectrogram representing the frequency components of audio signals over time and a log power spectrogram, displaying the amplitude in a logarithmic scale. The log transformation compresses large value ranges, making subtle differences more visually distinct.



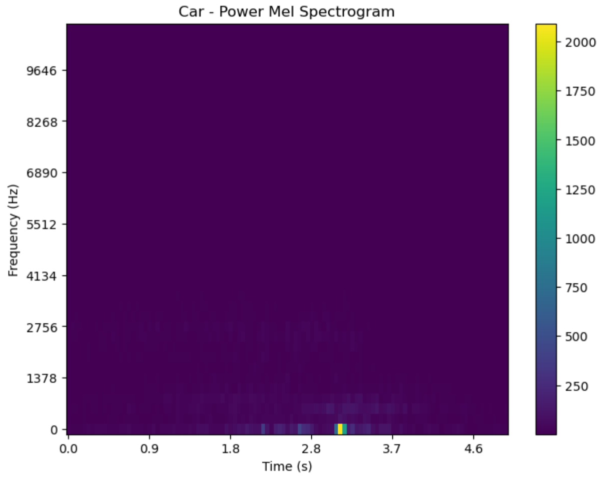


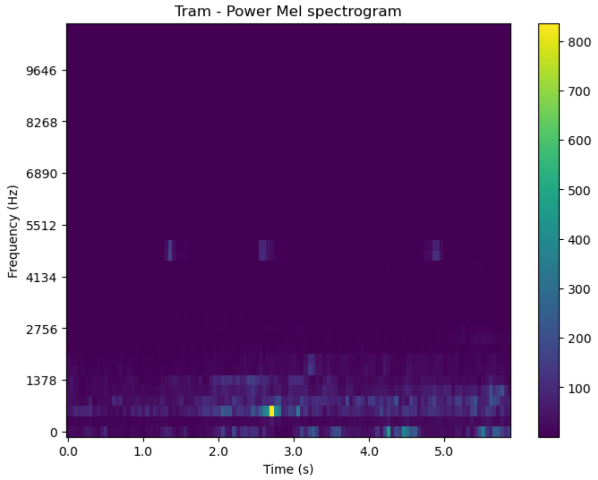




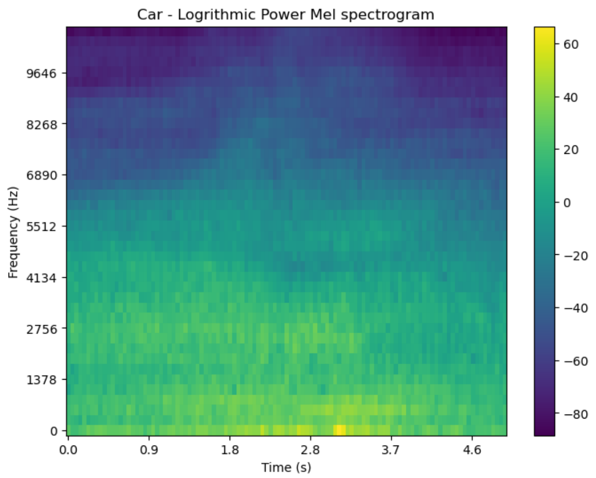
In the former, minimal differences between Car and Tram data were observed. In the latter, some distinctions between Car and Tram data were revealed, but the disparities were not particularly pronounced. This result suggests that the two classes lack the frequency characteristics to be classified.

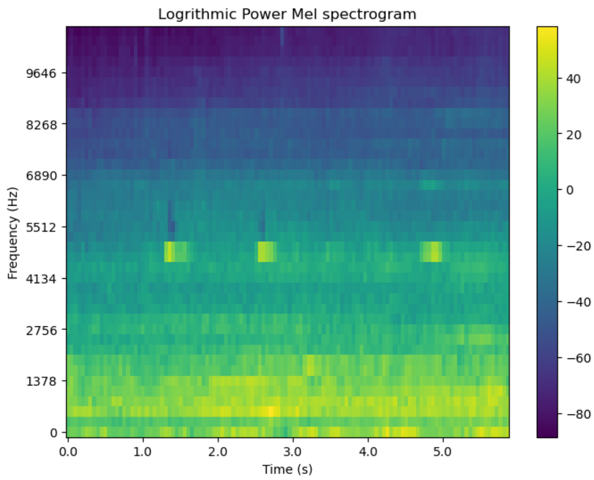
Next, we generated Mel power spectrograms by transforming the frequency axis into the Mel scale, which is based on human auditory perception. This scale is designed to mirror the way humans perceive differences in sound, providing higher resolution in the low-frequency range and lower resolution in the high-frequency range.





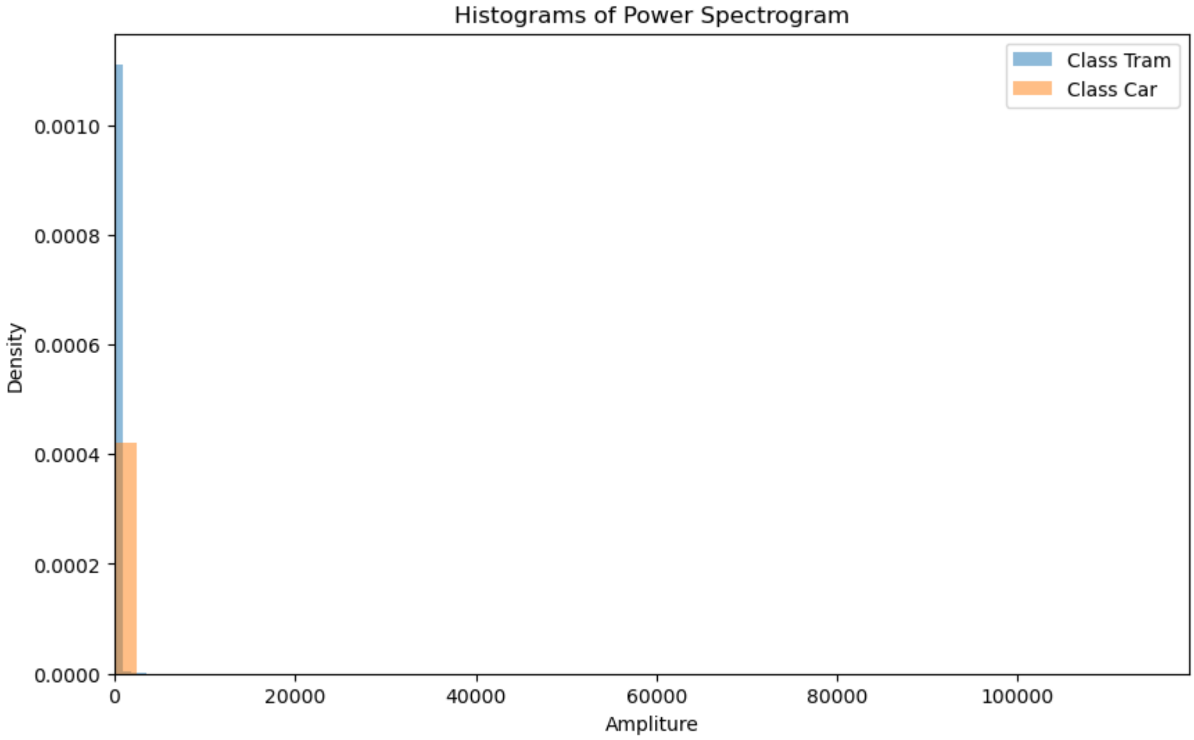
When the data was plotted, clearer differences between Car and Tram data emerged compared to power spectrograms and log power spectrograms. To emphasize these differences further, we plotted logarithmic Mel spectrograms, involving a logarithmic transformation of the Mel scale. The logarithmic transformation, as mentioned earlier, aids in reducing the impact of large amplitude differences, making detailed structures more visually discernible.





The results of this plot showcased the most pronounced disparities between Car and Tram data observed thus far.

To validate these analyses, histograms were plotted.



The log mel-spectrogram data exhibited the most significant differences between the two classes. Therefore, it can be considered as one of suitable choices for classification purposes.

4. Model Selection and Calculation Steps

For this study, the chosen model is logistic regression, a statistical model widely used for binary classification problems. It models the probability by passing the linear combination of input variables through a sigmoid function. Logistic regression, known for its simplicity and interpretability, has demonstrated strong performance across various classification tasks. Therefore, it can be considered as one of suitable choices for this investigation.

The output of logistic regression is the estimated probability of belonging to one of the two classes. Typically, a threshold of 0.5 is applied, classifying instances with a probability of 0.5 or higher into class 1 and those below 0.5 into class 0. In this study, Car is designated as class 0, and Tram as class 1.

The training data utilized in this study were collected by other students. In total, the data sets consisted of 106 and 103 samples for the Car and Tram classes, respectively. Each was split into training and validation sets. The following is a list of data providers and datasets used in this study.

Table 1: Data set for Training and Validation (Downloaded Data)

Provider	Data set for Car	The number of data set for Car	Data set for Tram	The number of data set for Tram

ali.abdelsalam	36709	32	36722	31
eetut	36831	21	36830	21
ulasanilgunes	36812	29	26813	22
koirankarva84581682	37128	24	37129	29
	Total	106	Total	103

Table 2: Data set for Testing (Uploaded Data)

Provider	**Data set for Car**	**The number of data set for Car**	**Data set for Tram**	**The number of data set for Tram**
wowaka	36709	24	39645	20

The specific steps are outlined below:

(1) Preprocessing of Training Data: Each class's log power mel-spectrogram data were subjected to padding with zeros based on the maximum length and flattened into one-dimensional vectors. (2) Labeling of Training Data:Labels of 0 were assigned to Car data, and labels of 1 to Tram data. (3) Concatenation of Training Data: The training data and labels were combined to create the training dataset (X and y). The validation set was configured with a test_size of 0.2, representing a 20% proportion of the dataset. (4) Training the Logistic Regression Model: A LogisticRegression model was constructed, and the training data were used to train the model. (5) Model Evaluation (Validation and Test Sets): Following training, the model underwent evaluation on the validation set (X_val and y_val) and the test set (X_test and y_test). This assessment provided insights into the model's performance on both the validation and test sets, with respective accuracies reported.

This approach ensures the reproducibility of experiments by utilizing the train_test_split function to partition the training data into training and validation sets. Setting a random seed (42 in this case) maintains consistent conditions, allowing the generation of the same training and evaluation data under similar circumstances. This contributes to the reliability of the results across different executions and environments.

5. Results and Analysis

Below is the table of accuracy obtained from the evaluation:

Table 3: Model Accuracy

	Validation Data	Test Data
Model Accuracy	0.9762	0.9545

The accuracy for the validation data is 0.9762, and for the test data, it is 0.9545. This indicates that the model is not overfitting to the training data and can generalize well to new data.

6. Conclusion

In this study, data from four students were used as training data. While the performance on the test data was satisfactory, achieving very high accuracy on both validation and test sets raises the question of whether obtaining a more diverse training dataset could have been beneficial. Additionally, the classification was conducted using a single feature, and there is potential for improved accuracy by combining it with other features.