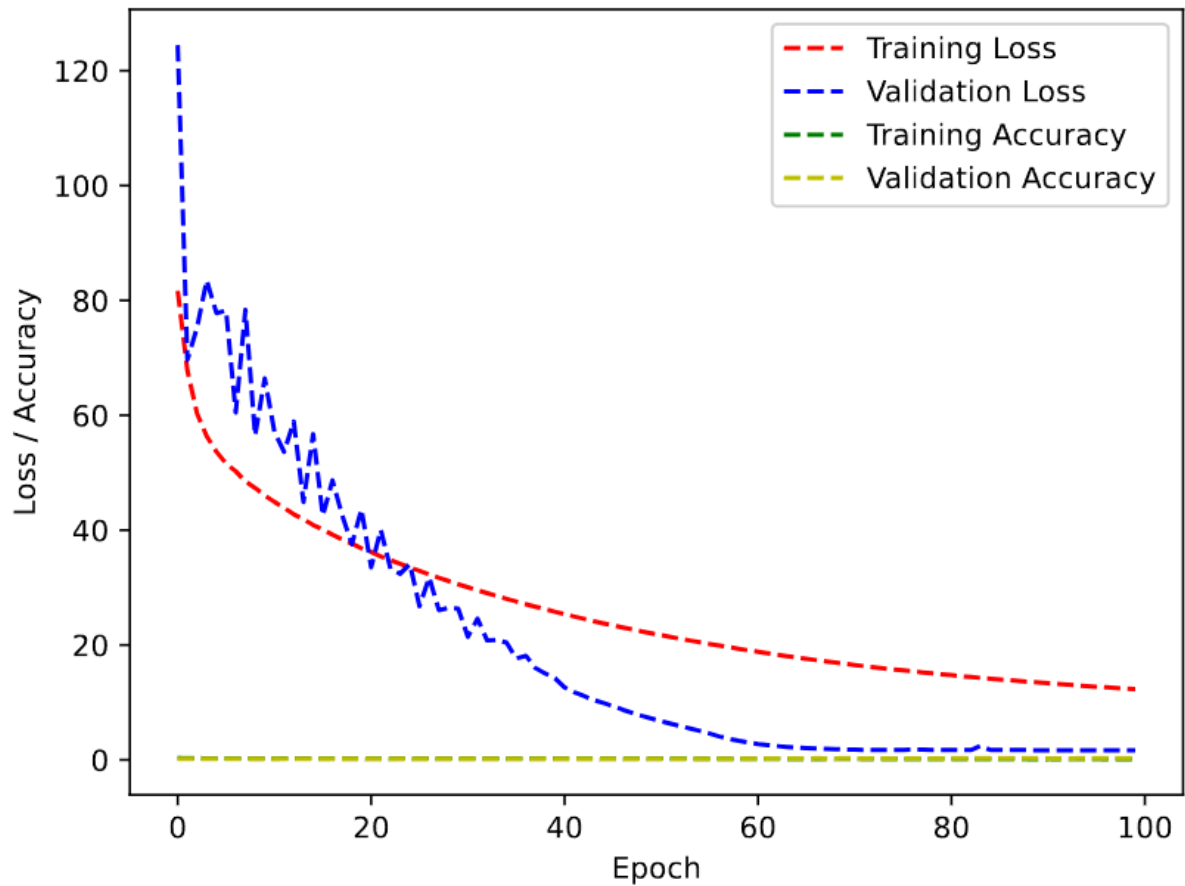


Task 2 Data And Analysis

Genre	f1-score
blues	0.0
country	0.0
metal	0.5455
pop	0.0
rap	0.3571
rock	0.0

Overall F1 Score: 0.1504



Analysis:

As shown in the table, only the rap and metal genres got correct guesses, although, it is more likely that the model just was more biased into predicting any song from the test set as either metal or rap. This raises concerns about the model's ability to distinguish between different genres.

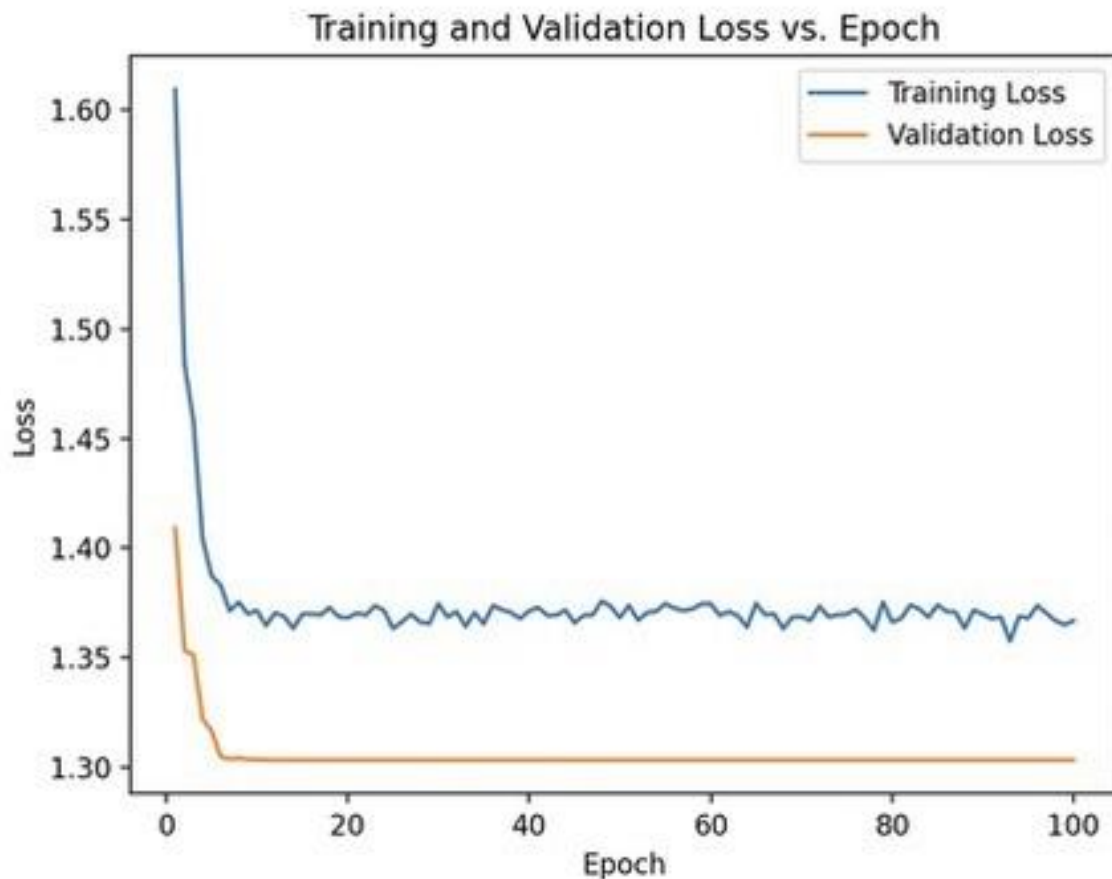
The main reason why the model performance is so mediocre is probably because of the features extracted from the songs which could be considered really basic, like the number of words for the songs, and the number of unique words in the song. These features are not sufficient enough as shown in the table to predict the genre of a specific song.

Now for the epoch diagram, the model may be overfitting as the training loss decreases while the validation loss increases, and after some more epochs, both, the validation and training loss decrease, indicating that the model is learning.

Task 3 Data And Analysis

Overall f1 score: 0.5244107744107743

Genre	F1 Score Per Genre
Blues	0.5
Country	0.545455
Metal	0.666667
Pop	0.545455
Rap	0.888889
Rock	0



Analysis:

Overall, this model seems to be pretty mediocre. The overall f1 score is .52, having predicted almost every genre at least once except Rock. Interestingly, this model never guessed Rock, which explains its f1 score of 0. As for the highest f1 score per genre, it's Rap with a score of .89, meaning that it guessed every Rap song correctly, which could mean we had very good rap song data.

From this data, it seems like a strength of this model would be predicting Rap songs, and a weakness would be predicting Rock songs. Perhaps our data for Rock songs wasn't effective at teaching our model how to predict that genre. However, even considering its efficiency at guessing rap, the model still struggled to achieve a high overall f1 score.

The training vs validation epoch loss diagram shows training loss in blue and validation loss in red. The validation loss curve shows some sudden spikes, which could cause some overfitting. As for the training loss curve, it shows a little “rockiness”, but the gradual decrease in loss shows that it was able to train somewhat, although, I’m not sure what this “rockiness” could indicate. Overall, I should look into a more efficient process of tuning hyperparameters and increasing performance.