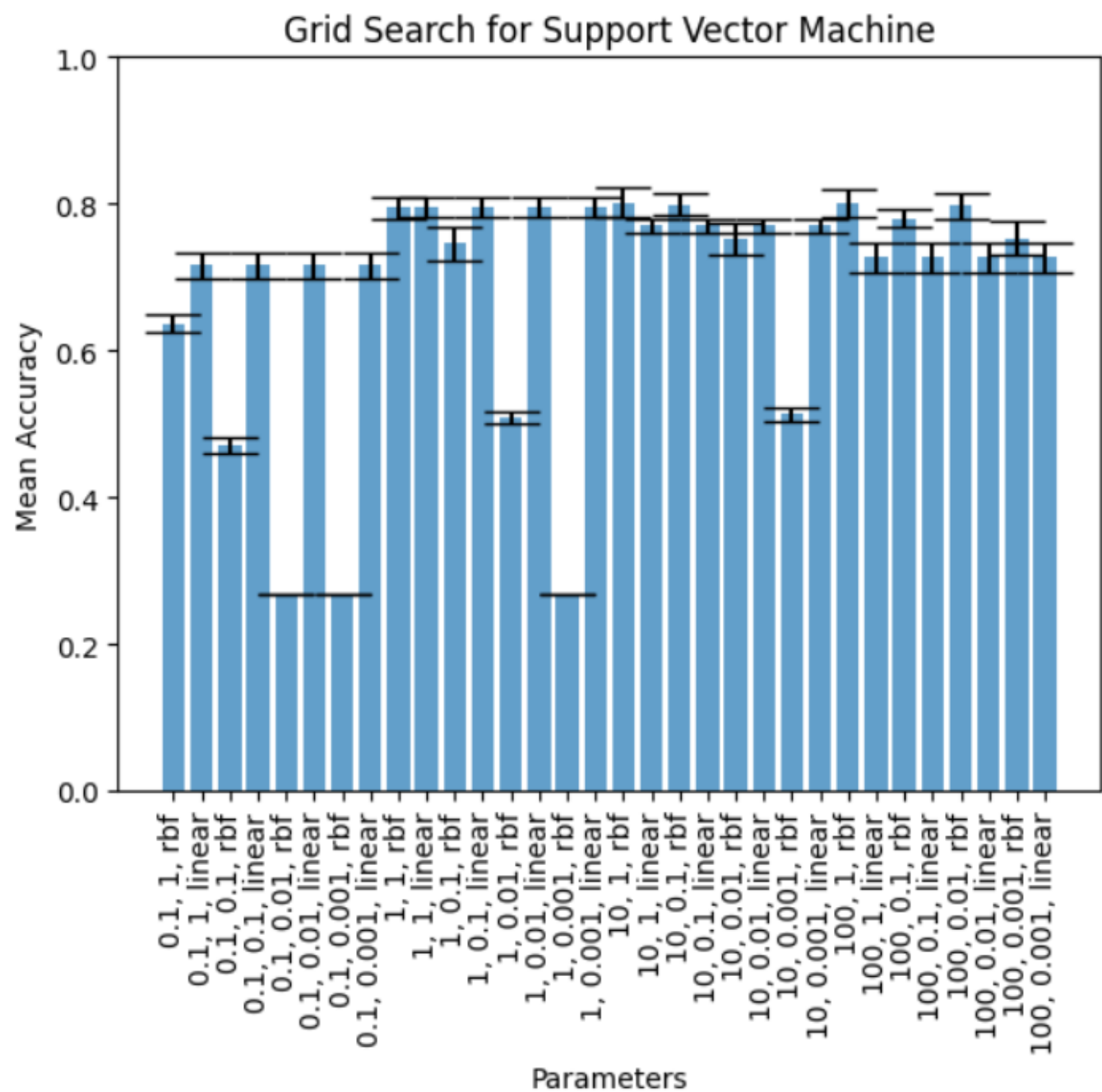


Classification of Swiss language dialects is an intriguing task that demands intricate analysis and processing of text data. In my work, I employed various machine learning methods and experimented with neural networks to achieve optimal results.

Using the Bag of Words approach initially, I trained a logistic regression model, which has shown and continues to demonstrate the best results compared to other methods.

Logistic regression

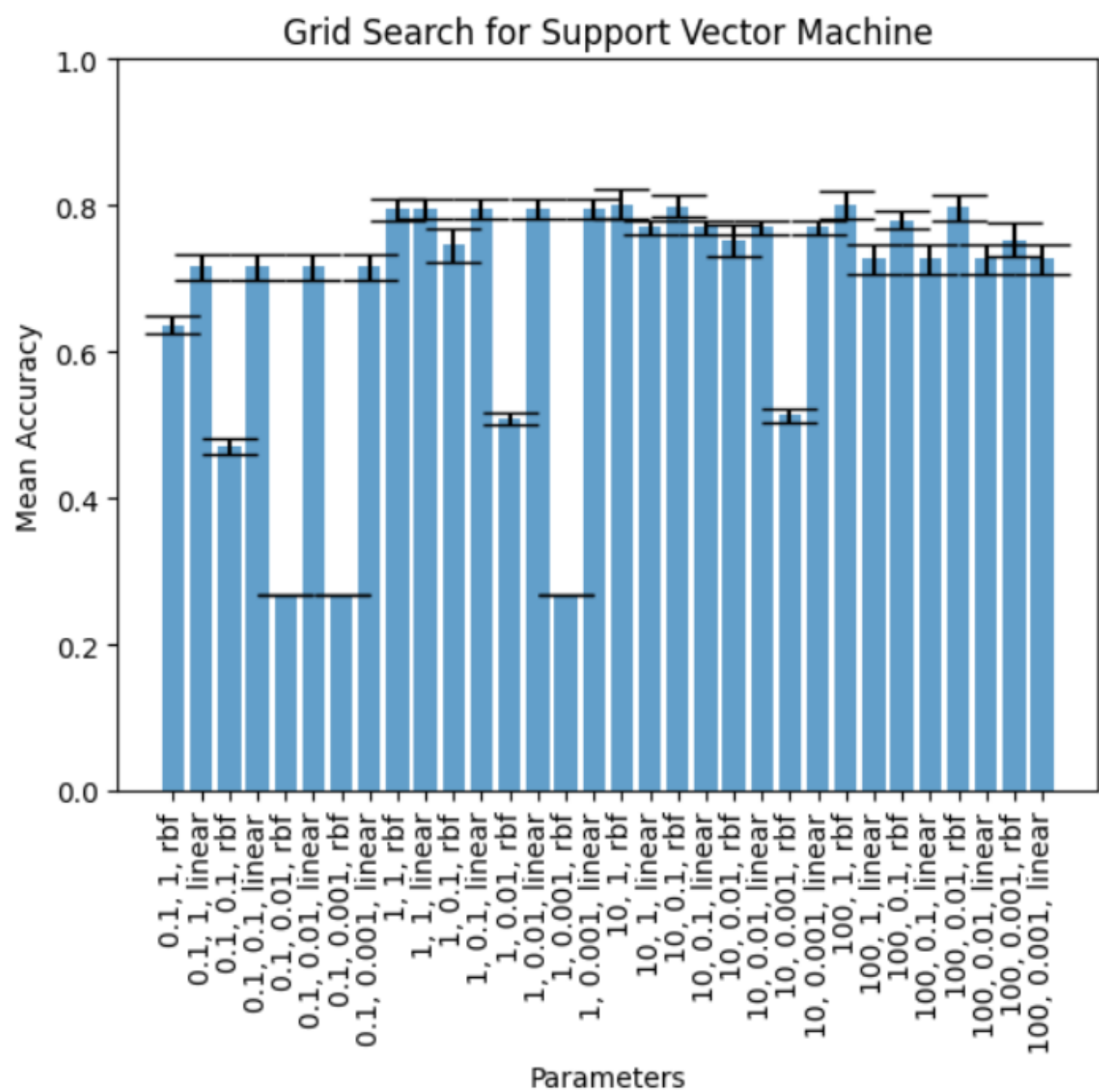


Test Accuracy: 0.808
Train Accuracy: 0.9964139344262295

The best results for lr are as follows:

Test Accuracy: 0.8092, Train Accuracy: 0.9583760245901639.

Following that, we explored a decision tree model, which did not boast promising results.



Test Accuracy: 0.808
Train Accuracy: 0.9964139344262295

Similarly, I attempted the naive Bayes method, which yielded poor results (accuracy below 50%).

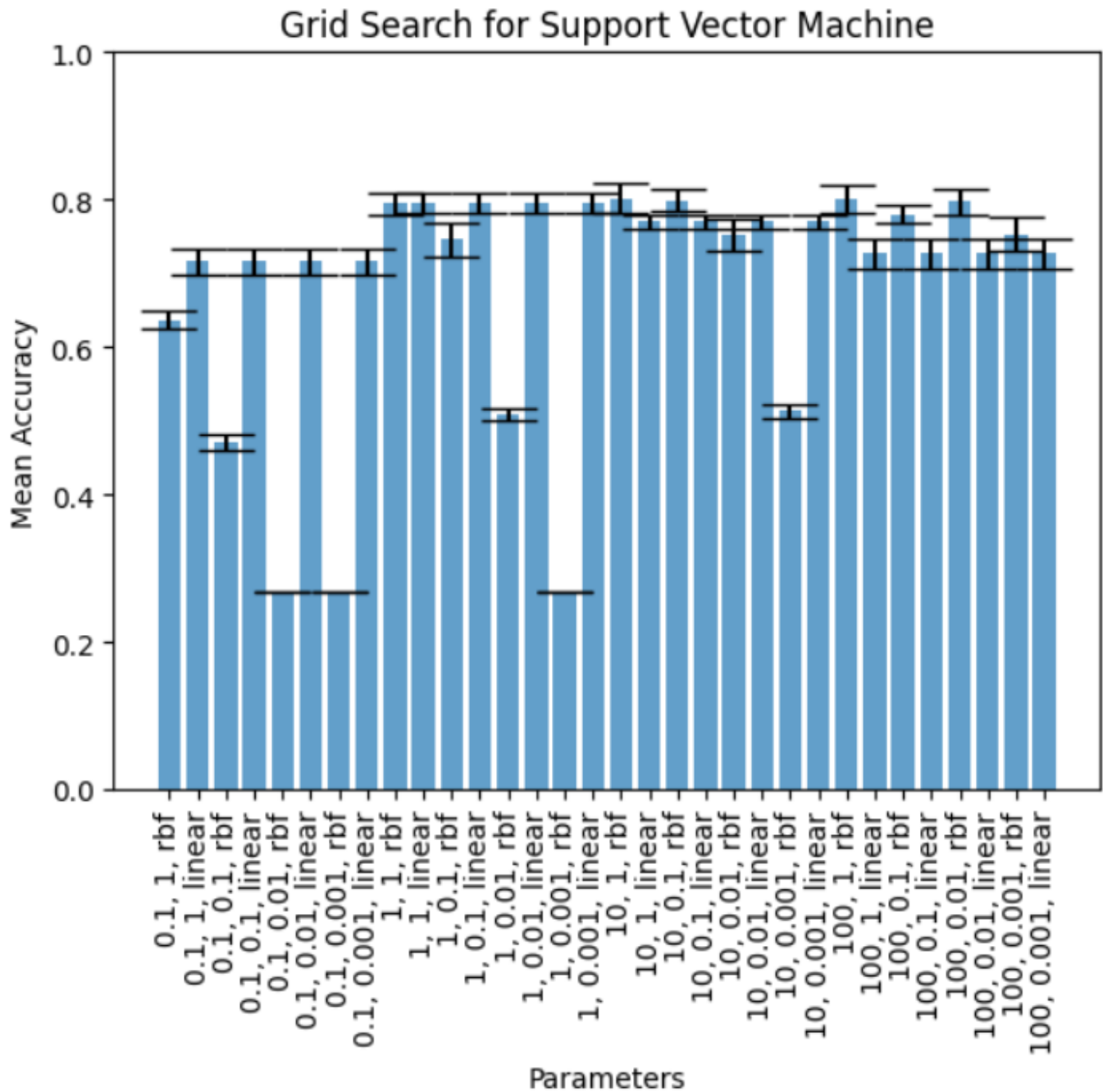
I also tried using TfidfTransformer, which overall worsened the models' performance.

However, our top candidate only improved its metrics: the best results from logistic regression were as follows:

Test Accuracy: 0.8232 Train Accuracy: 0.9829661885245902

The metrics for the decision tree also improved:

Next, I experimented with Support Vector Classifier (SVC), which yielded quite promising results due to parameter optimization.



Test Accuracy: 0.808

Train Accuracy: 0.9964139344262295

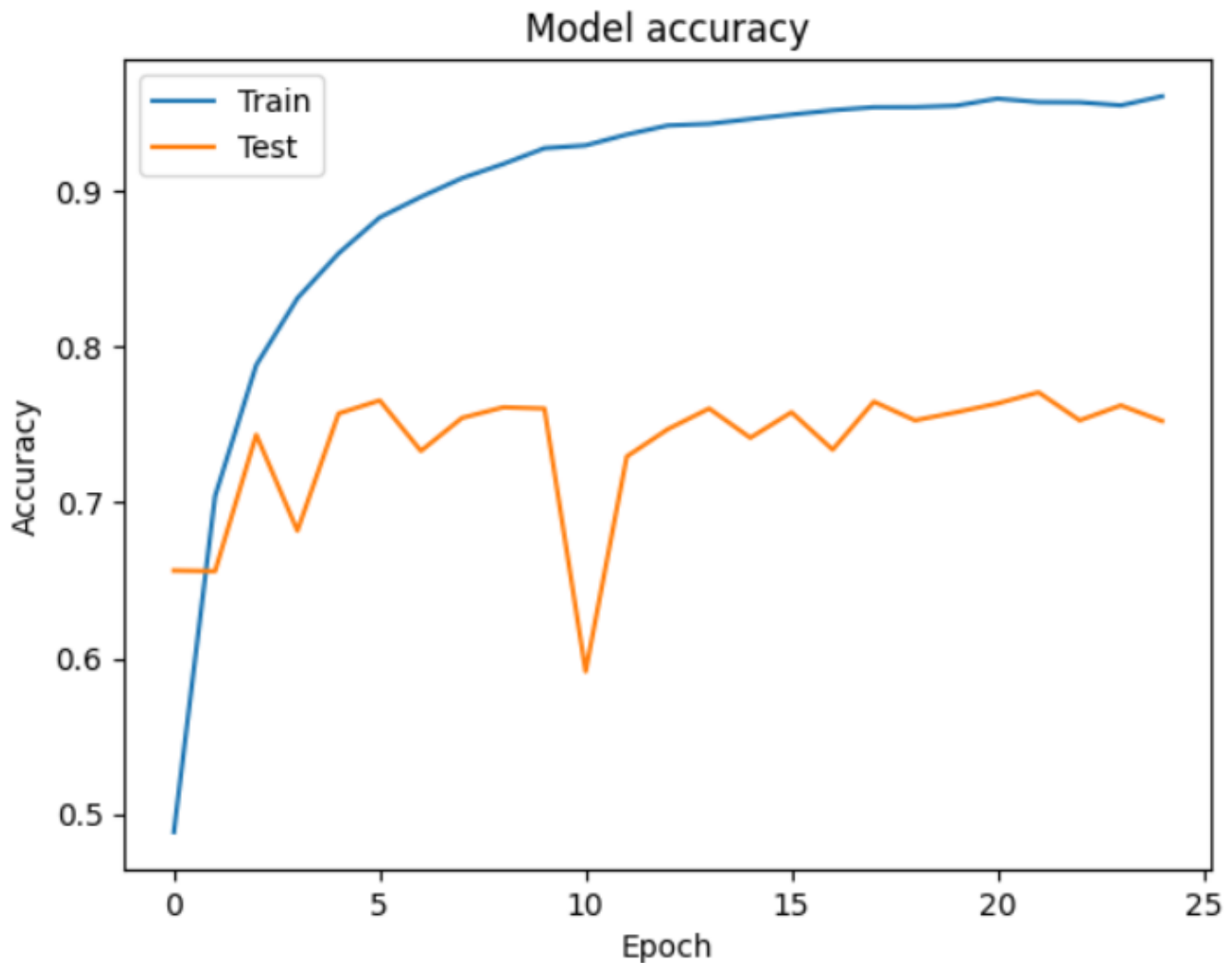
After this, I concluded my attempts to solve the task using machine learning methods and transitioned to neural networks.

I manually tokenized and processed the textual data and used it for further experiments

My first neural network model was a SimpleRNN model,

which demonstrated an accuracy of 0.9608 and a validation accuracy of 0.7524 as the best result.

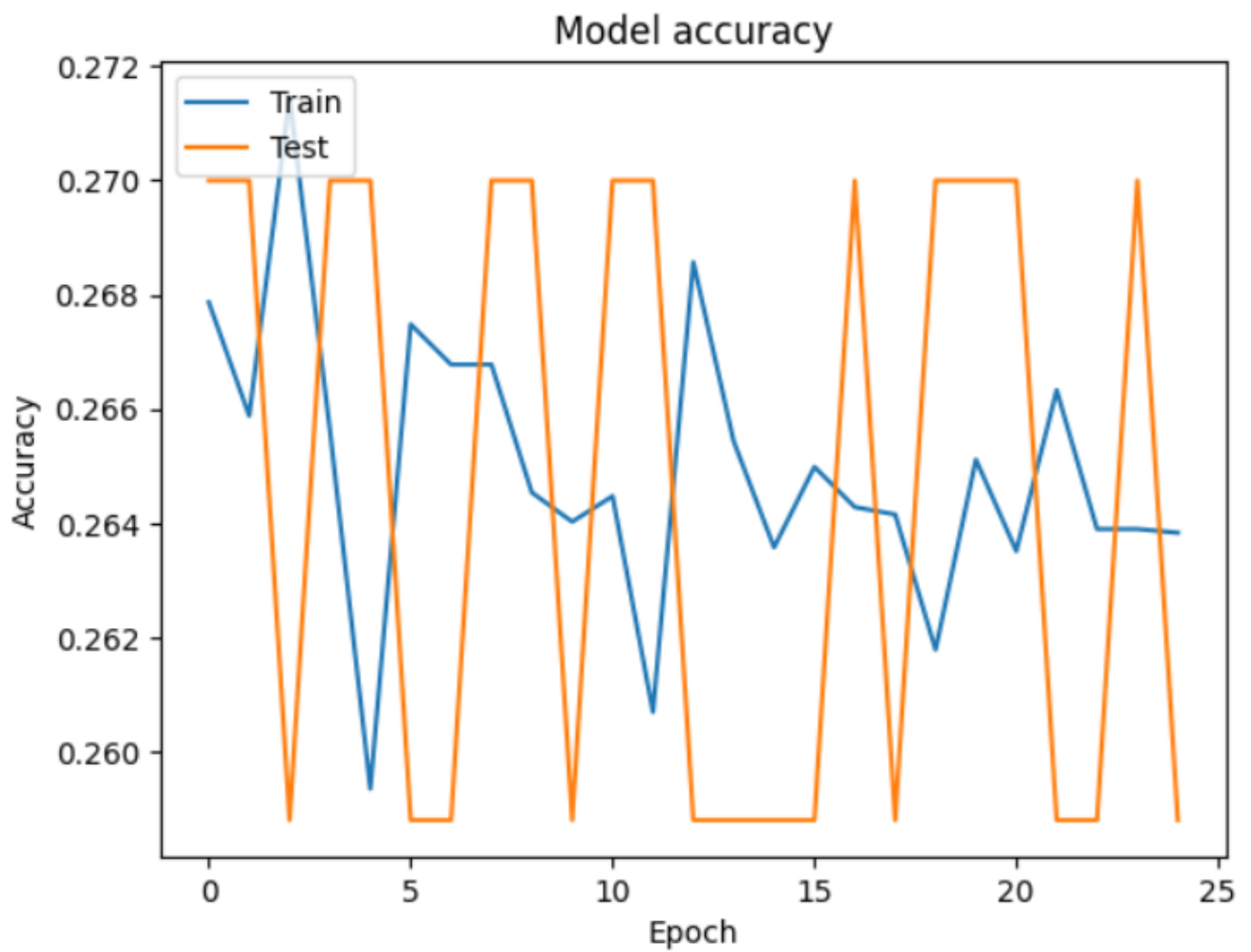
Fully Simple RNN



I experimented with an LSTM model, which did not yield promising results:

the accuracy was 0.2686, the validation loss was 1.3850, and the validation accuracy was 0.2588 as the best outcome.

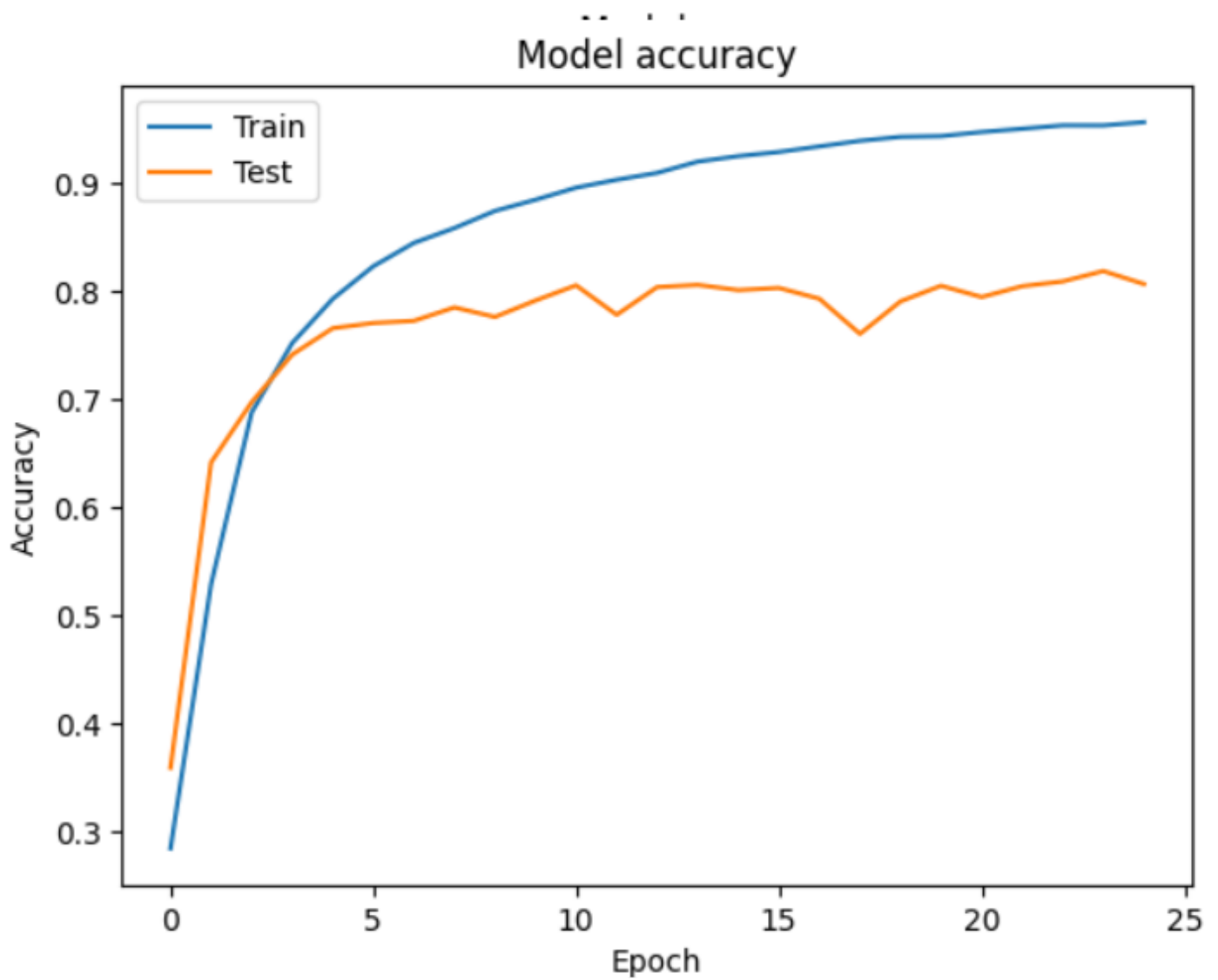
LSTM



The next model I utilized was a simple fully connected neural network, which surprisingly demonstrated decent results:

Accuracy: 0.9529 Validation Accuracy: 0.8184

Fully connected NN

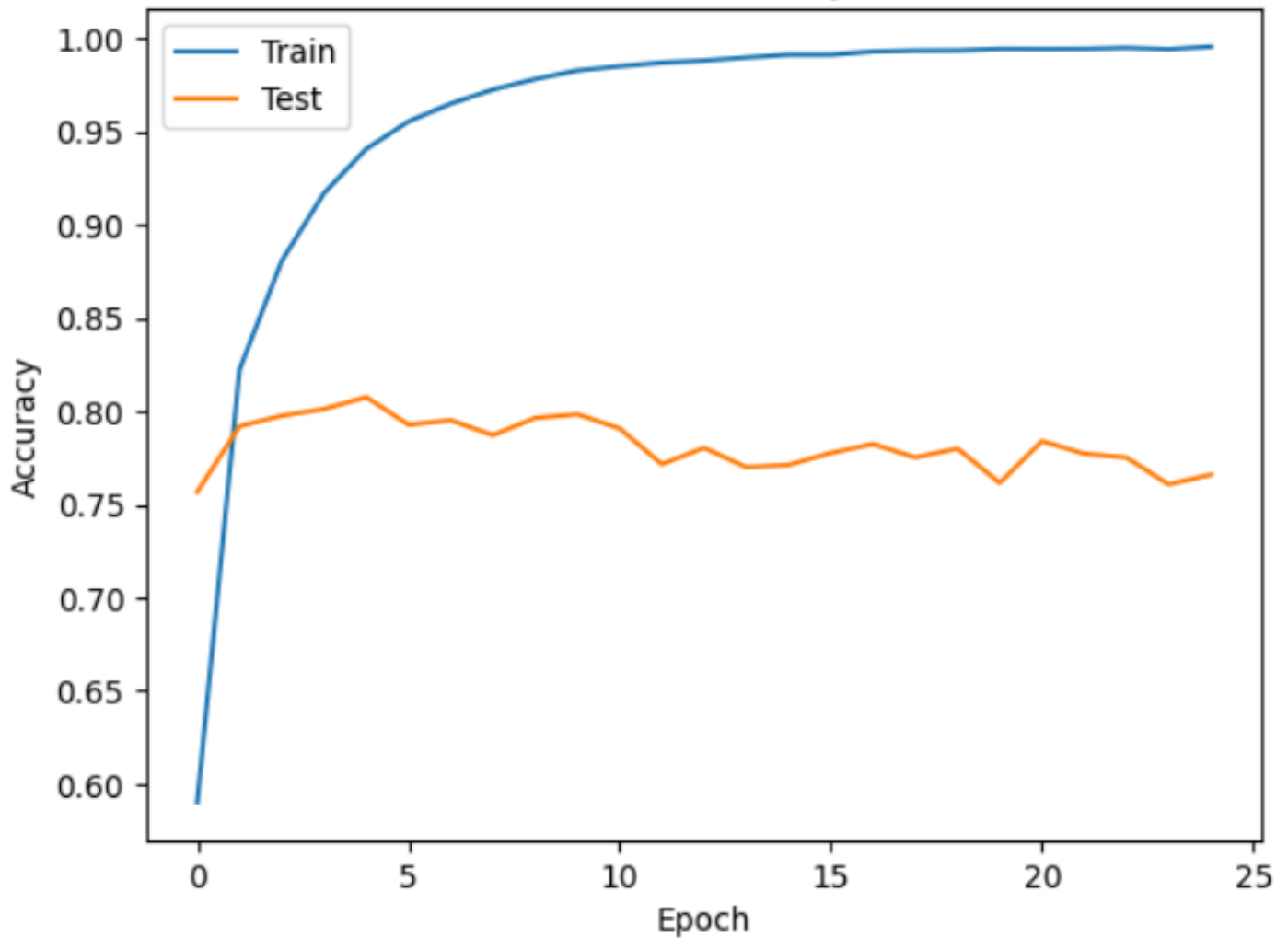


The convolutional network also presented a very decent result:

Accuracy: 0.9944 Validation Accuracy: 0.7616

Convolutional NN

Model accuracy

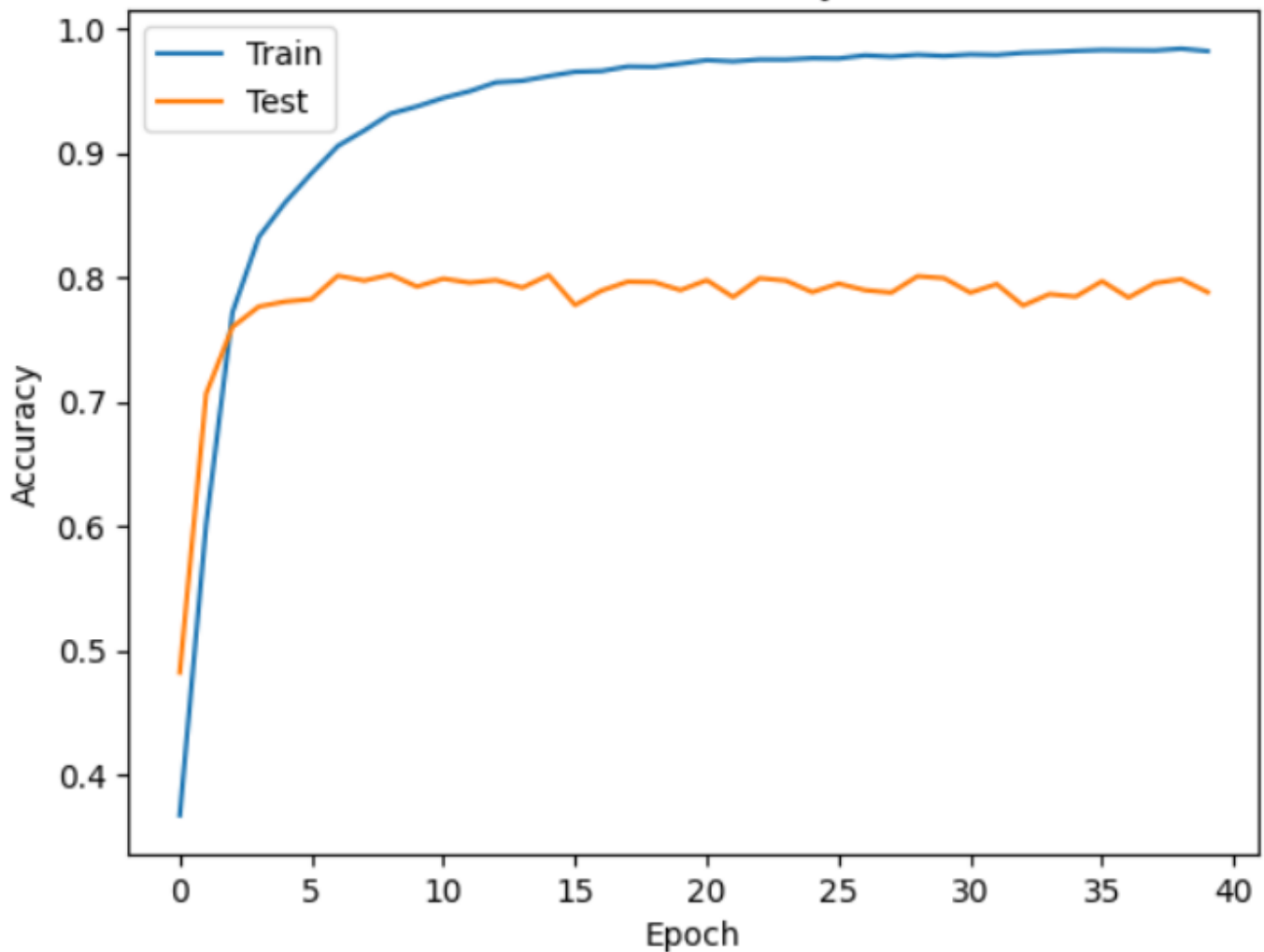


Next, I attempted to build my custom neural network, which combined convolutional layers after embedding and an RNN layer. Although it produced good results, it was significantly more complex.

Best Result: Accuracy: 0.9836 Validation Accuracy: 0.7984

My Try

Model accuracy



Conclusion

Within the scope of this task, I improved my machine learning skills and applied them to this problem. Additionally, I experimented with various neural network architectures to tackle the classification of Swiss language dialects. This experience has deepened my understanding of model selection and feature engineering, contributing to a more comprehensive understanding of the complexities involved in natural language processing tasks.

Метод	Accuracy	Validation Accuracy
Logistic Regression	0.8232	0.9829
Decision Tree	0.602	0.9829
Naive Bayes	<0.5	<0.5
Support Vector Classifier (SVC)	<0.5	<0.5
SimpleRNN	0.9608	0.7524
LSTM	0.2686	0.2588
Fully Connected NN	0.9529	0.8184
Convolutional NN	0.9944	0.7616
Custom Model	0.9836	0.7984