**AIN SHAMS UNIVERSITY FACULTY OF ENGINEERING**

**ICHEP – Senior 2 Level – CESS**

**CSE485 - Deep Learning**

**Fall 2023**

Final Project

CSE 485: Deep Learning

Leaf Classification

Submitted to:

Dr Mahmoud Khalil

Eng Mahmoud Soheil

Mohamed Obia 19p5170

Mohamed Fathy 19p4707

Ziad Assem 19p6363

Karim Shalaby 19p6044

Contents

[1.0 Problem 2](#_Toc154341043)

[2.0 Part 1 3](#_Toc154341044)

[3.0 Part 2 5](#_Toc154341045)

# 1.0 Problem

There are estimated to be nearly half a million species of plant in the world. Classification of species has been historically problematic and often results in duplicate identifications.

The objective of this project is to use extracted features, including shape, margin & texture, to accurately identify 99 species of plants. Leaves, due to their volume, prevalence, and unique characteristics, are an effective means of differentiating plant species. They also provide a fun introduction to applying techniques that involve image-based features.

# 2.0 Part 1

The data

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The training data consists of id column which is a column that has each leaf image id. It has a species column which us the label of each leaf. And finally, the feature columns which consists of 3 features which are margin, shape, and texture of each leaf. Each feature consists of 64 different values that represent this feature.

Understanding the data shapes

A black and grey rectangular object

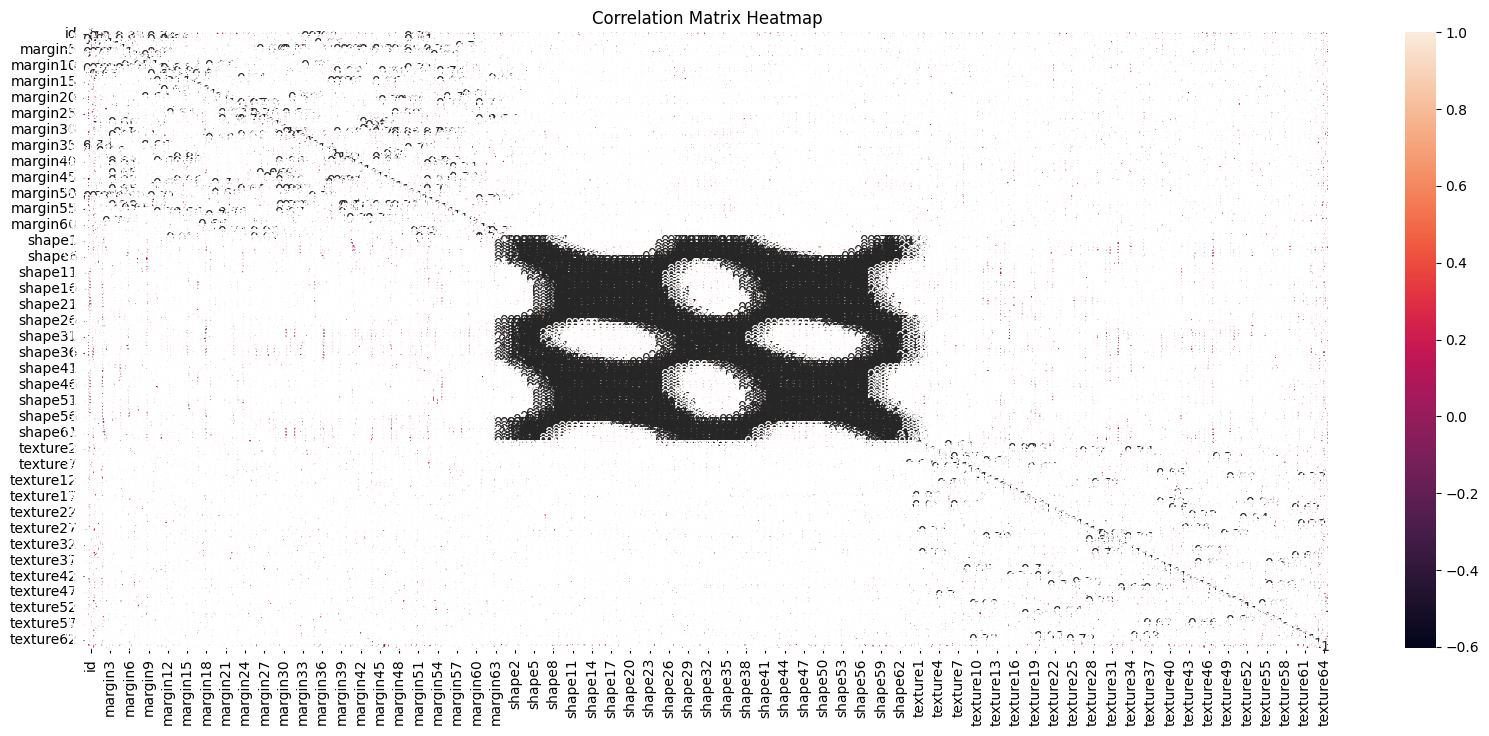
Description automatically generated

Plotting histogram of the species distribution

A red and white lines

Description automatically generated

Correlation matrix



The correlation matrix doesn’t give a good indication due to the huge amounts of data that we have, so we plotted the correlation heatmap between the highest and lowest features.

A graph showing a line

Description automatically generated with medium confidence

A graph with red and blue squares

Description automatically generated

# 3.0 Part 2

Encoding the data: removing the unnecessary columns and encoding the labels to numerical values

A screen shot of a computer

Description automatically generated

Splitting the data into training and validation with 80:20 ratio.

A screen shot of a computer

Description automatically generated

Changing the 2d vector to 3d vector so that the model focuses on each feature at a time.

A screenshot of a computer

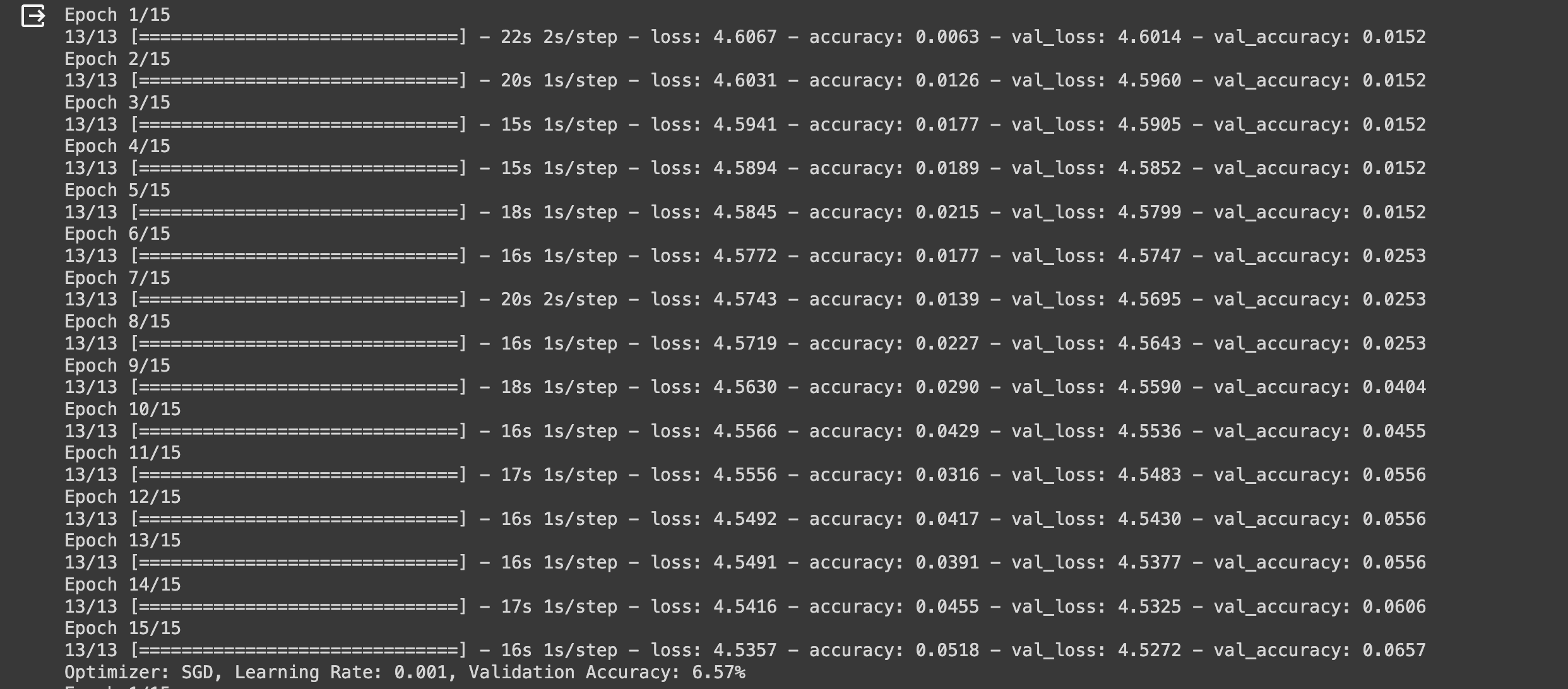
Description automatically generated

Creating a model with 512 kernels. Each kernel is 1d. the model focuses on each feature of the three previously mentioned features, and we tune the hyper parameters of the model in terms of optimizers and learning rates.

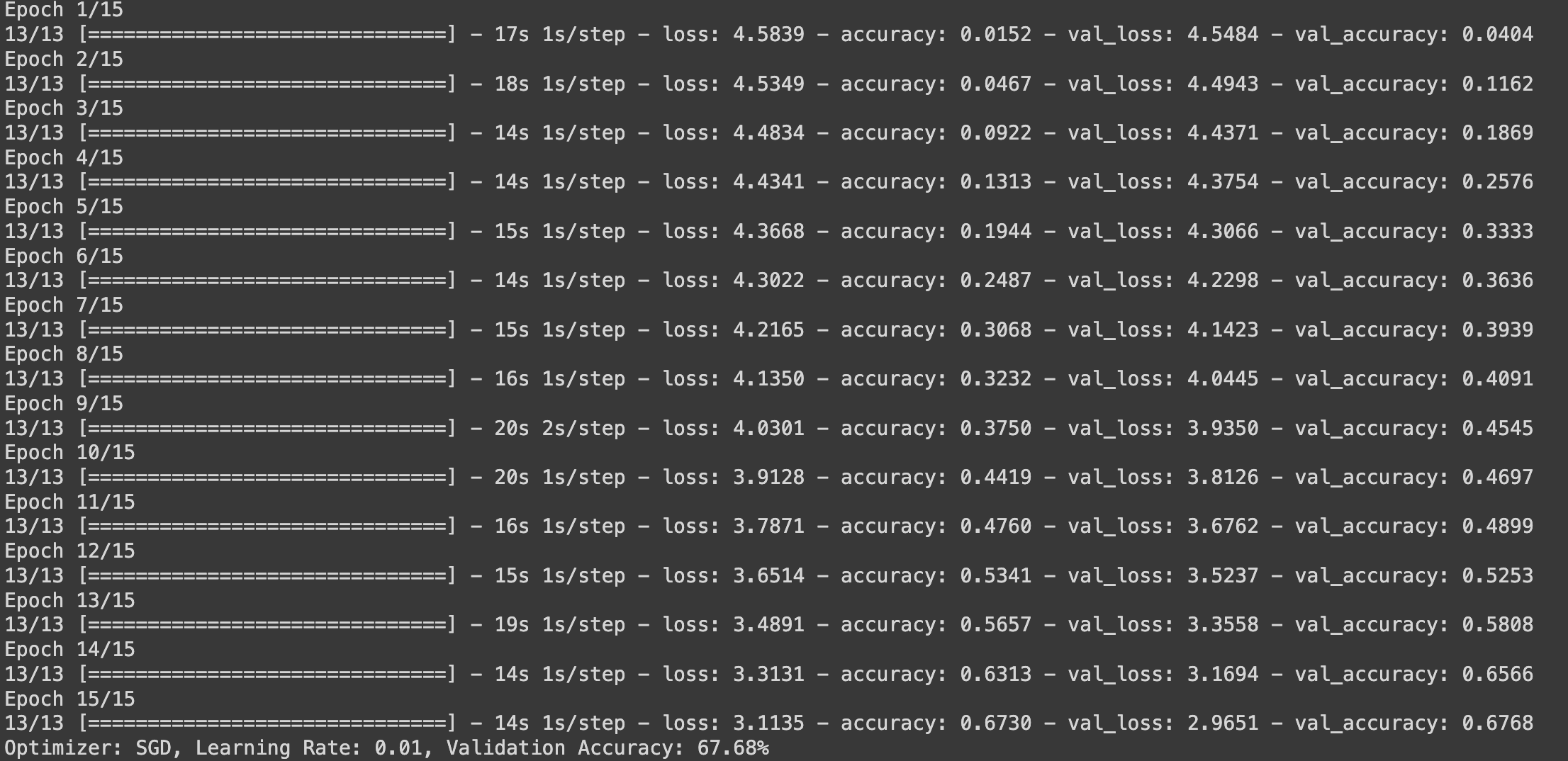
A screen shot of a computer program

Description automatically generated

This is the output using sgd with learning rate 0.001. 6% is a very bad validation accuracy.



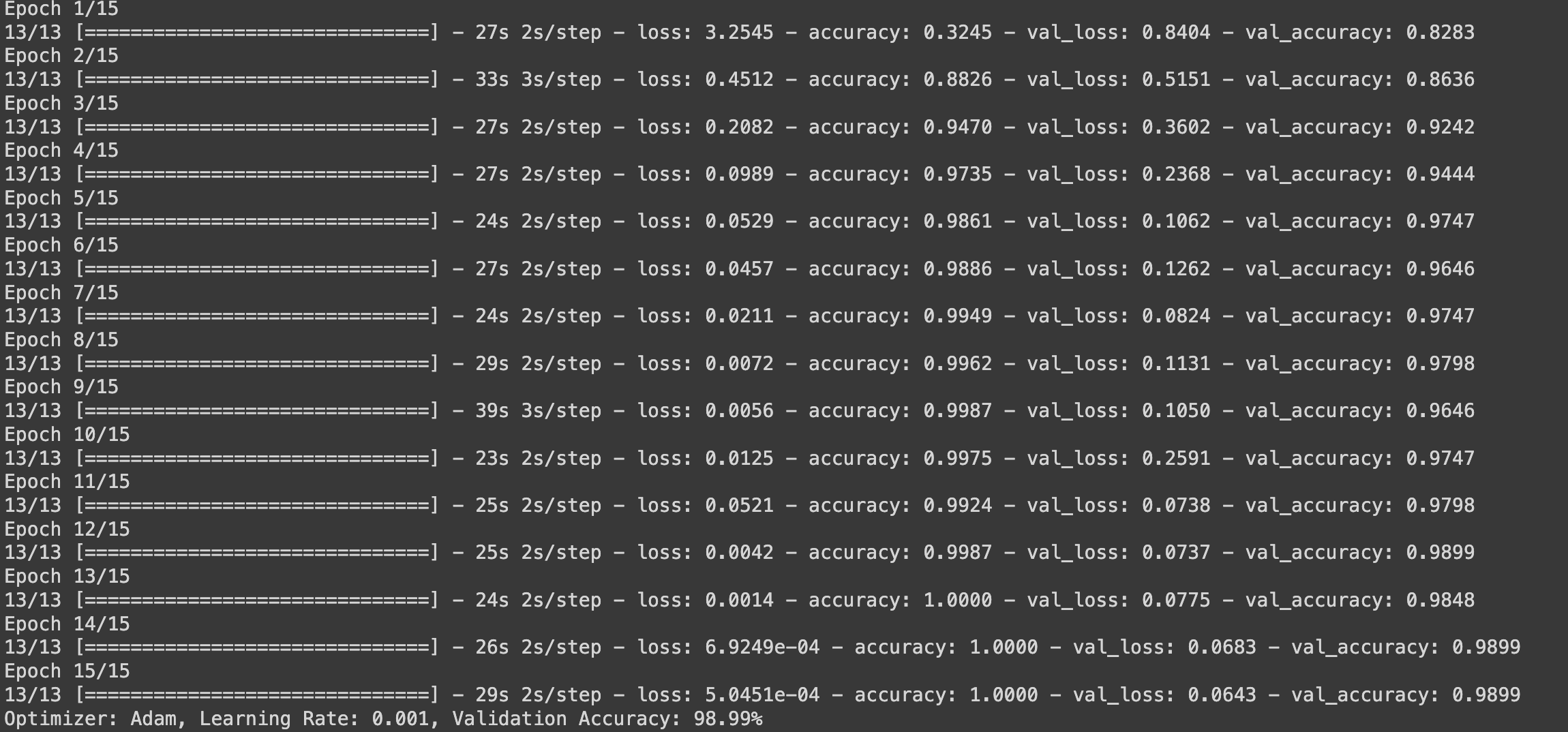
Increasing the learning rate increased the model’s validation accuracy by more than 10-fold.



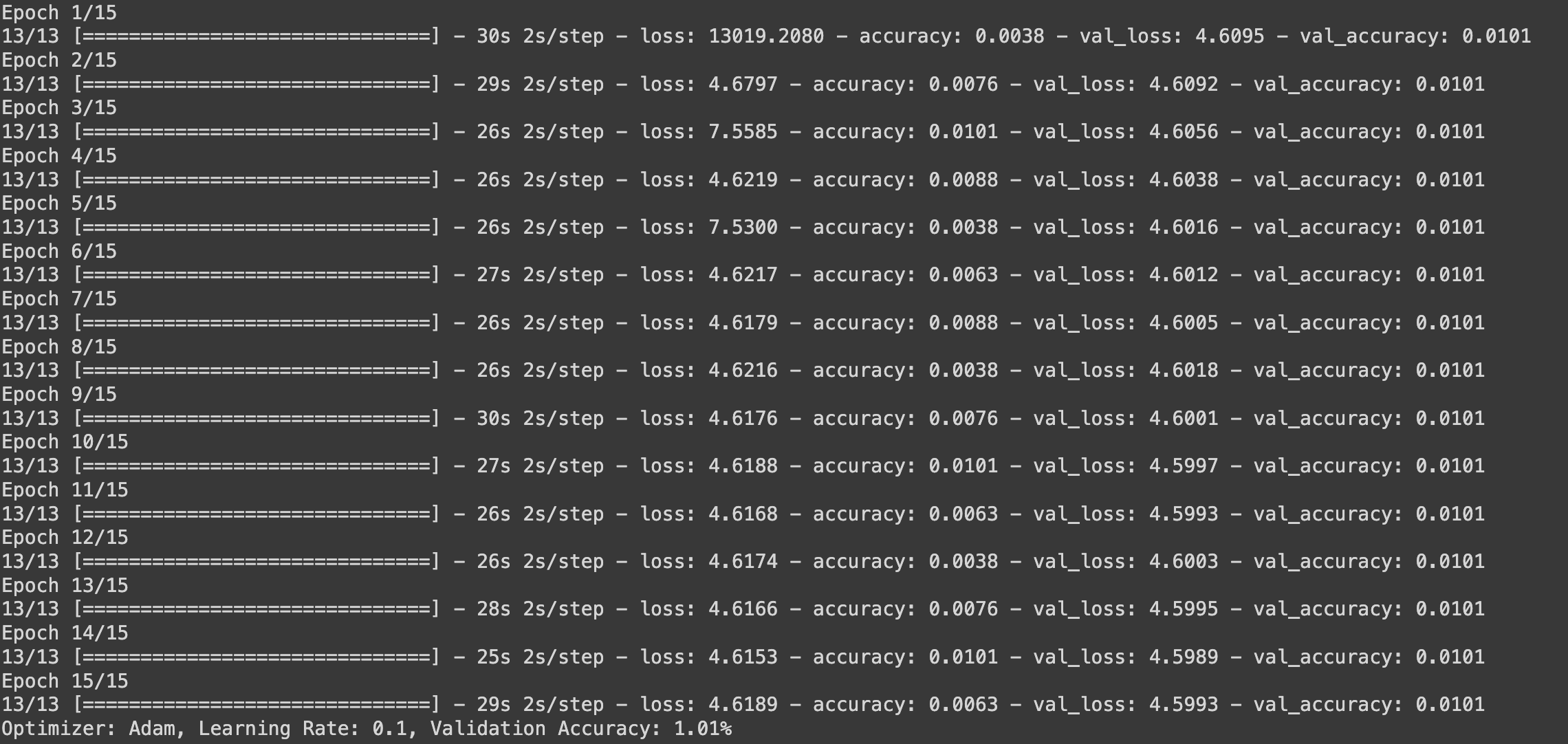
Reaching a maximum validation accuracy for the sgd optimizer



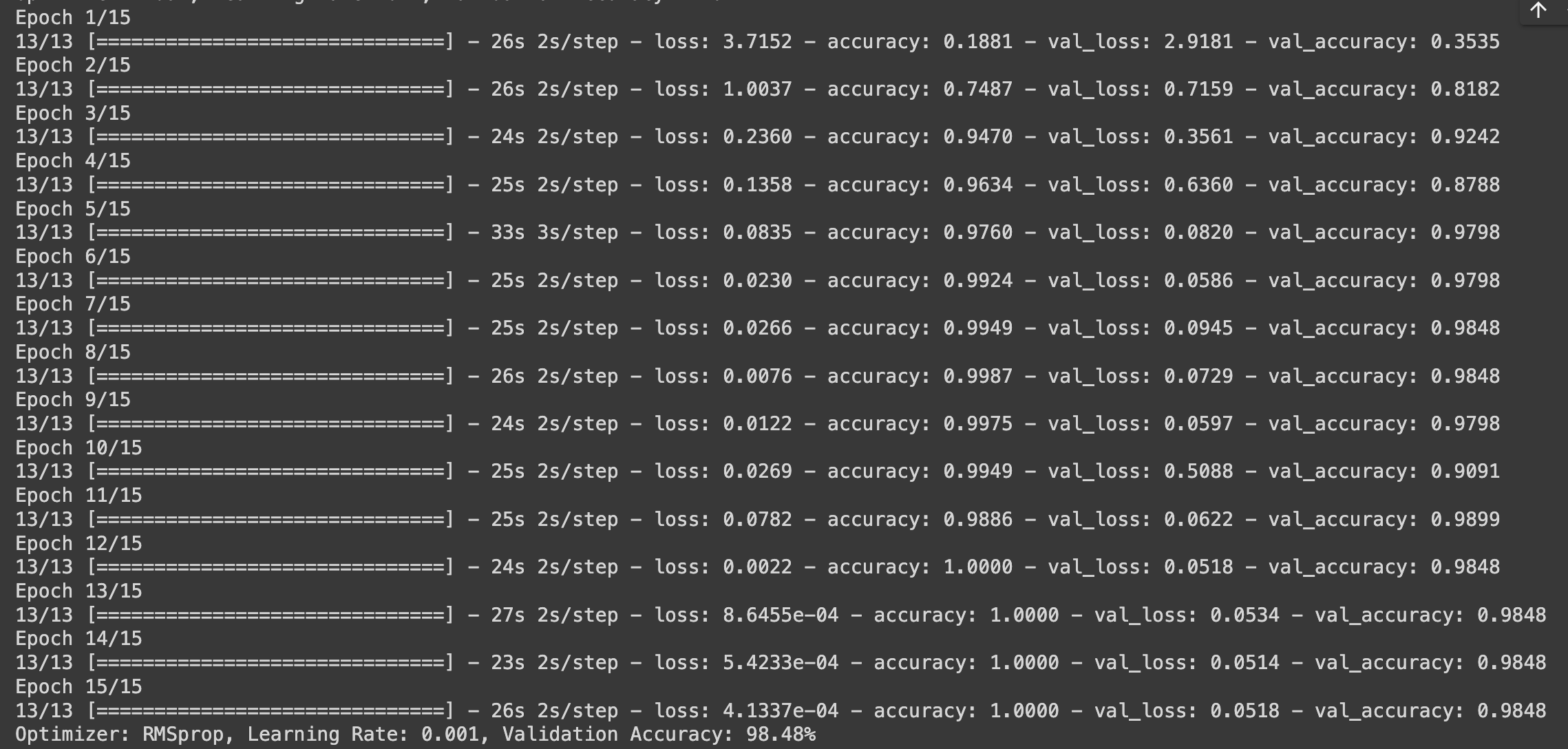
Adam optimizer with 0.001 learning rate was our best optimizer.



Here the choice of the learning rate 0.1 with Adam optimizer made the model not learn.

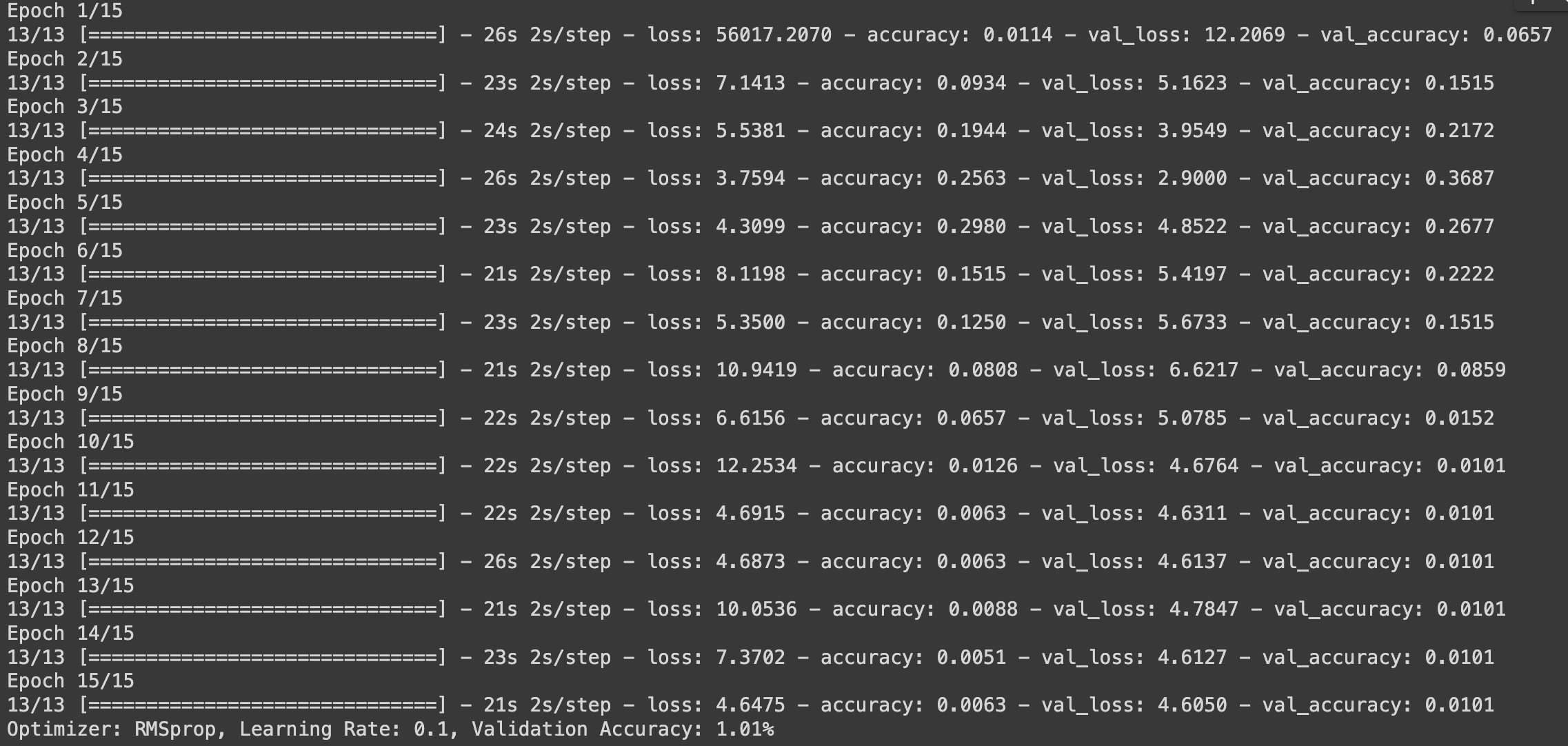


These are the vlaues for RMSprop with differnet learning rates



A screen shot of a computer

Description automatically generated

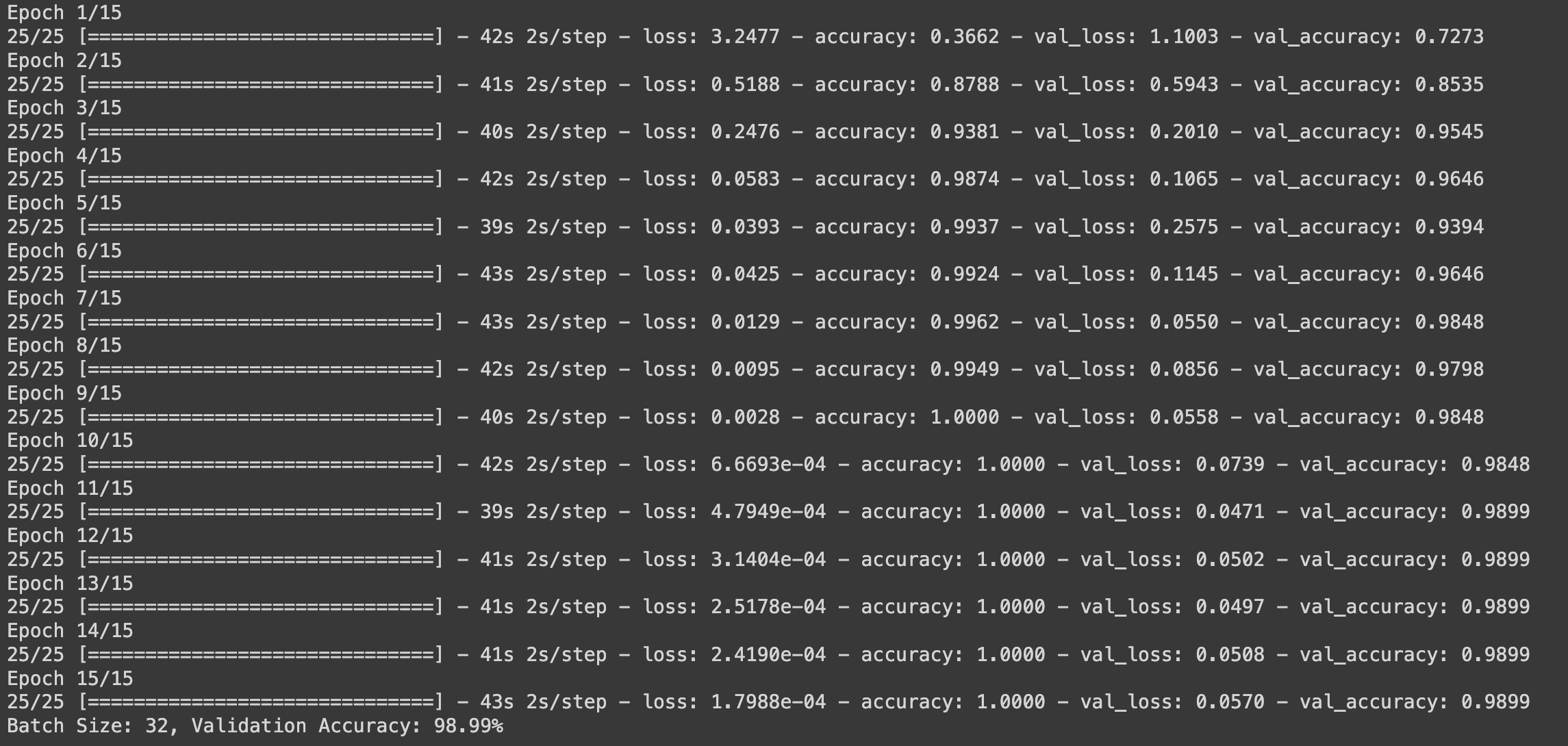


We can see that increasing the learning rates with the RMSprop optimizer reduces the socre of the model untill it stops the learning process.

So we choose our best model which was Adam with learning rate 0.001 and tried to change the batch size and see the results.

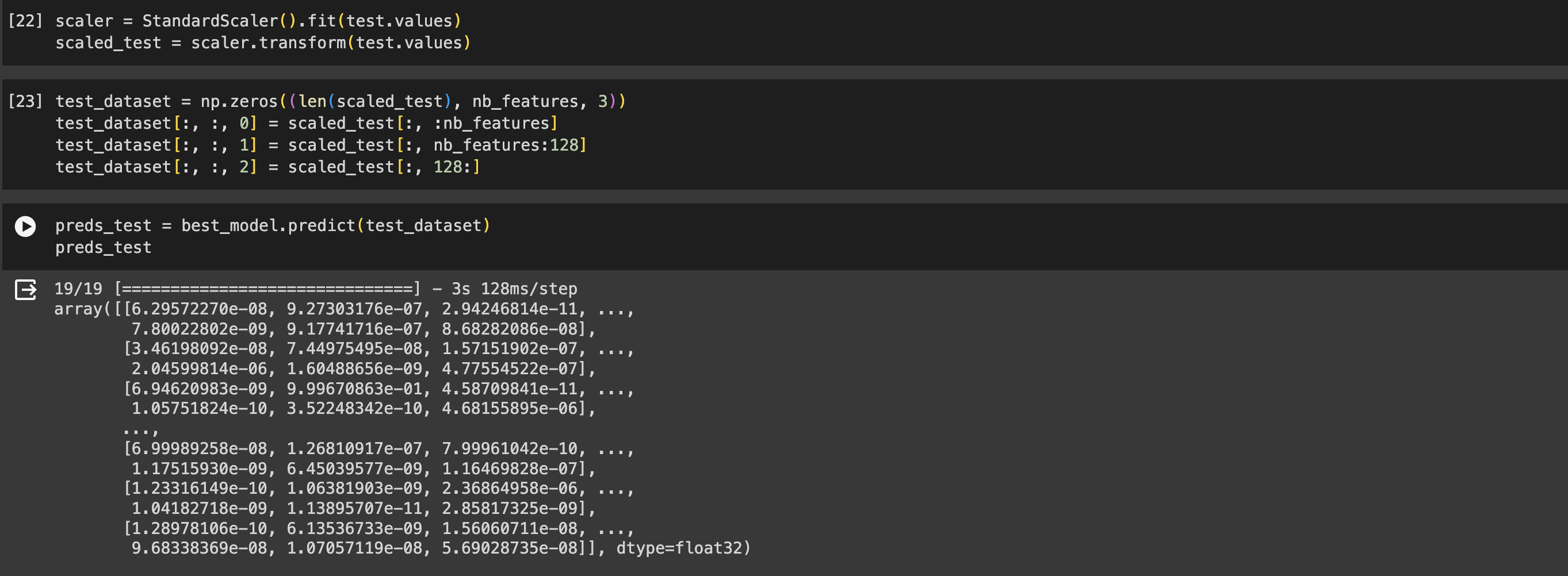
A screen shot of a computer

Description automatically generated



We achieved the same results as the 64 batch size

We then used the test set to predict the output of species



We then plotted a histogram of the species predicted and their counts.

A bar code with text

Description automatically generated

And this is the classification report of the accuracy of the model on the validation data.

A black background with white numbers

Description automatically generated