We randomly split the entire dataset into train set (80%) and test set (20%).

For DNN, the model architecture we used is shown as follow:

A screenshot of a social media post

Description automatically generated

The model is trained with Adam optimizer for 200 epochs, the loss and accuracy curve are shown below.

A screenshot of a social media post

Description automatically generated

Summary of the performance for 6 different methods:

A screenshot of a cell phone

Description automatically generated

As can be shown from the above table, the six classifiers can be ranked based on the prediction accuracy as follows: SVM > Random Forest > DNN > Ada Boosting > Logistic Regression > GLMnet. SVM shows the best performance for prediction with 100% accuracy but is not perfectly fitted for the training data. Random Forest and Ada Boosting has 100% training accuracy, but show 0.3% and 1% test error respectively, this could indicate overfitting. DNN shows decent results overall. Logistic Regression and GLMnet show the worst results, this could be related to the large numbers of feature and data. Overall, SVM, Random Forest, Ada Boosting, and DNN could show good results on the Pex23 dataset.