قارئ

Qari' *(Translation: Reader)*
Arabic Handwriting AI

27 JAN 2023, Robert Golden

MAKING AN IMPACT THAT MATTERS since 1845

Outline of Presentation

Stakeholders and Use Case

Dataset and Model

Measures of Success

Stakeholders & Use Case

Three columns

Stakeholders

Professionals serving on DS&J engagements or a client's own staff of investigators and linguists.

Trained linguists can be short-staffed or unavailable depending on the engagement.

Use Case

Identifying characters in a handwritten document and outputting the Arabic characters.

The identified characters can then be output, compiled, and used for machine translation.

Value Added

The solution permits the identification of documents to be prioritized for a linguist to review, saving time and increasing efficiency.

Dataset and Model

Dataset

The *Arabic Handwritten Characters Dataset* uploaded by Mohamed Loey to Kaggle is being utilized to train the AI, consisting of 13,440 images for training and 3,360 images for testing. This means that there are 480 training images and 120 testing images of each of the 28 character isolates.

Limitations

The dataset is limited in its applicability for this project by the fact that it only provides the character isolates (e.g., [illegible]), as opposed to their initial ([illegible]), medial ([illegible]), and final ([illegible]) forms. This should permit a limited proof-of-concept, but its real-world utility would be significantly hampered as-is.

Features

The dataset is limited in scope, and no data features require exclusion for the purposes of this project.

Dataset and Model, cont'd.

Model

The model is a convolutional neural network (CNN) designed to identify images of the handwritten characters.

Results

By the completion of development, the model should be capable of recognizing and classifying all 28 character isolates of the Arabic abjad.

Measures of Success

Precision vs. Accuracy vs. F-1 Score

Reliability matters…

A stakeholder needs to be able to trust that the model will be able to return the correct character out of all the observations.

In other words, how many times does it find the right answer?

As do False Positives/Negatives…

False Positives/Negatives will add significant confusion, meaning we cannot rely on accuracy alone!

Imagine the confusion if the model could not tell the difference between l and I, or between b and d!

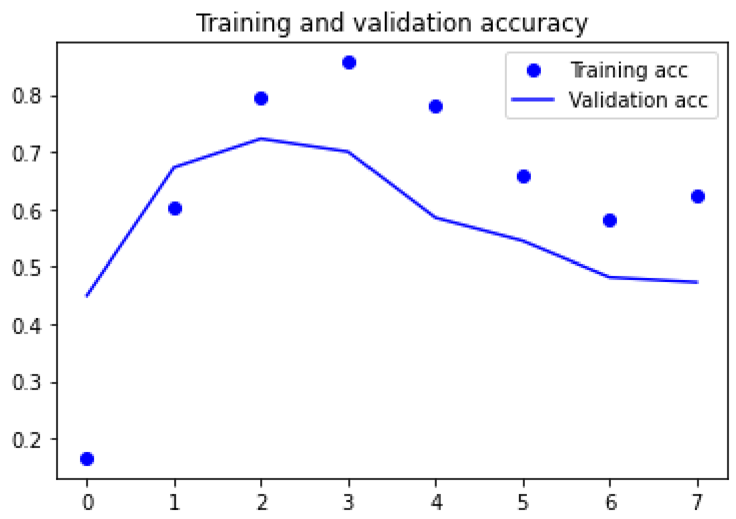
Our Measurement of Accuracy

The F-1 score is a weighted average that combines the benefits of both precision and accuracy.

Therefore, with one measure, we can evaluate how well the model is performing from 0 (very poorly) to 1 (excellent)!

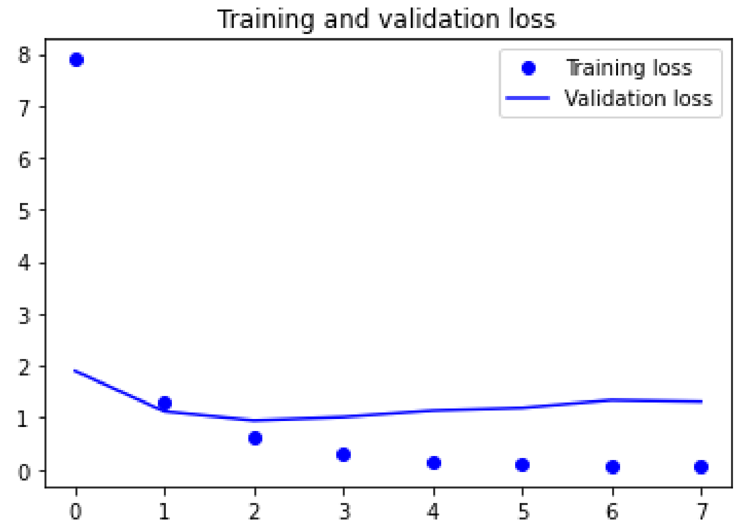
Model Evaluation

The model experiences low accuracy for the intended use case



What impact could accuracy have?

In a short, 500-character document, the model would potentially misclassify roughly 135 words.

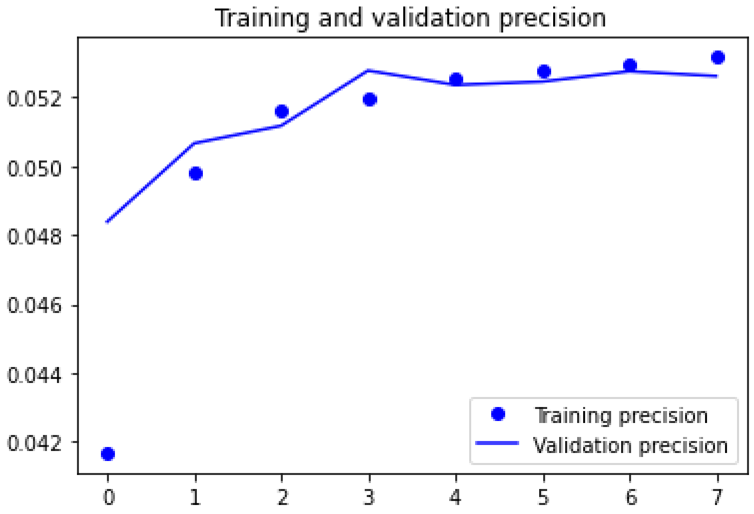


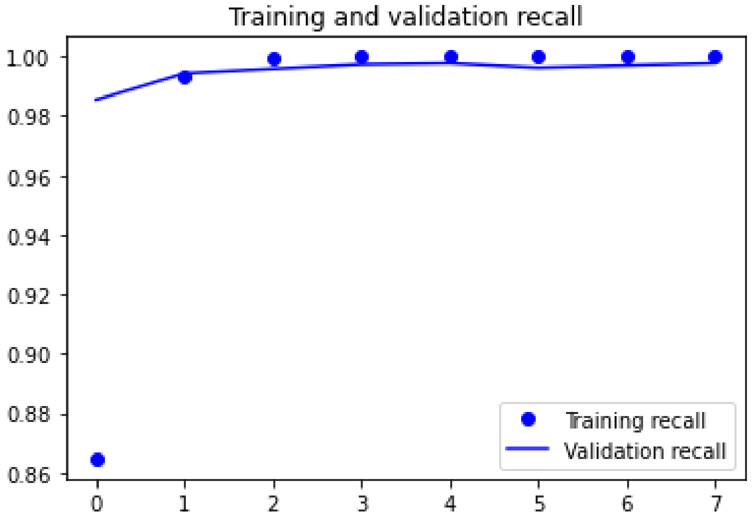
Why were only 8 epochs done?

The model terminated the training/validation process due to rising validation losses.

Model Evaluation

High recall and low precision leads to middling F1 Scores





Are the results relevant?

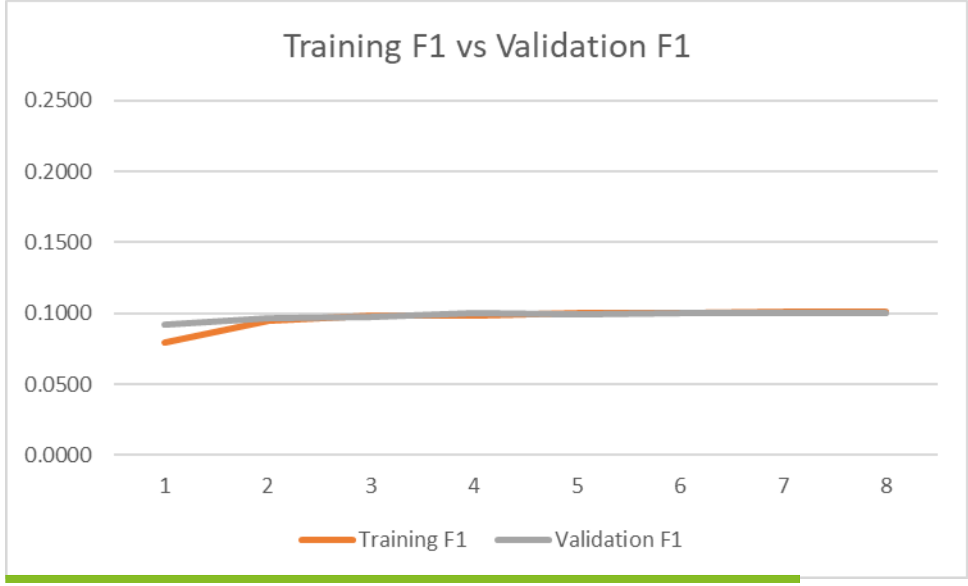
The precision scores indicate that, of all the outputs retrieved, the model is returning a lot of false positives.

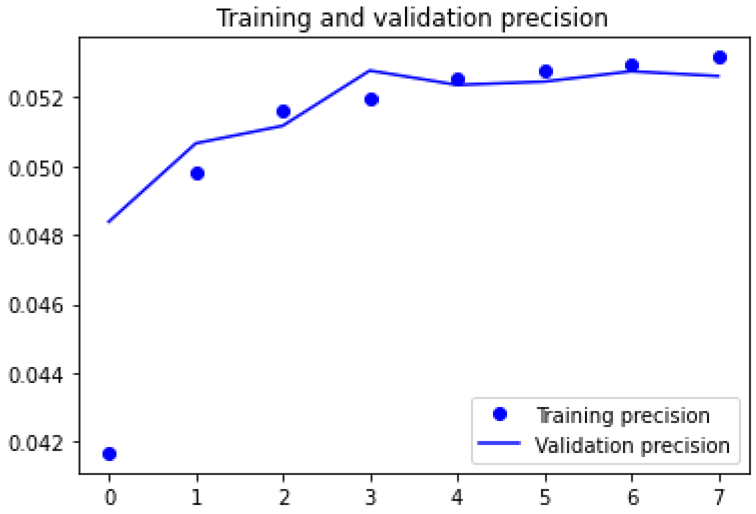
What do the recall values mean?

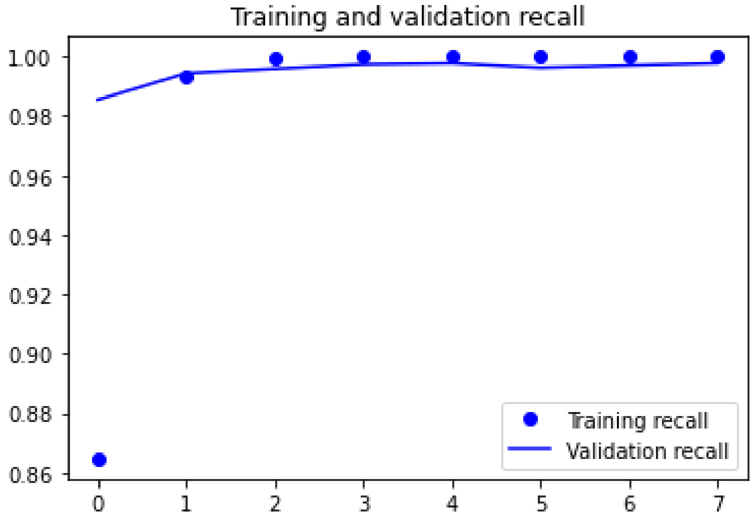
The recall values mean that this model is largely able to avoid false negatives when classifying a character.

Model Evaluation

High recall and low precision leads to middling F1 Scores







So what?

The model is unable to adequately perform at the expected level for the intended client.

Remediation and Next Steps

Short-term: Hyperparameter Tuning

Additional experimentation should be conducted to realize a model that can increase the model’s accuracy and F1 scores.

Long-term: Extend the model

Incorporate the ability to output the characters that it has identified.

Identify datasets that feature the initial, medial, final, and isolate forms of Arabic characters.

Questions?