

Model-Assisted Data Labelling: Investigating the Impact of Incrementally Trained Models on Image Annotation Time and Labelling Accuracy



University of
Staffordshire
London

Author

J Russ

Supervisors

Dr. Viraj Dawarka

Dr. Khalil Sadaat

Introduction

Creating high-quality datasets for machine learning models comes at a large cost, which is the initial time it takes to annotate enough data before machine learning models can meaningfully assist. This project investigates the effect incrementally trained models can have on reducing the amount of annotated data required for models to be useful in annotation assistance.

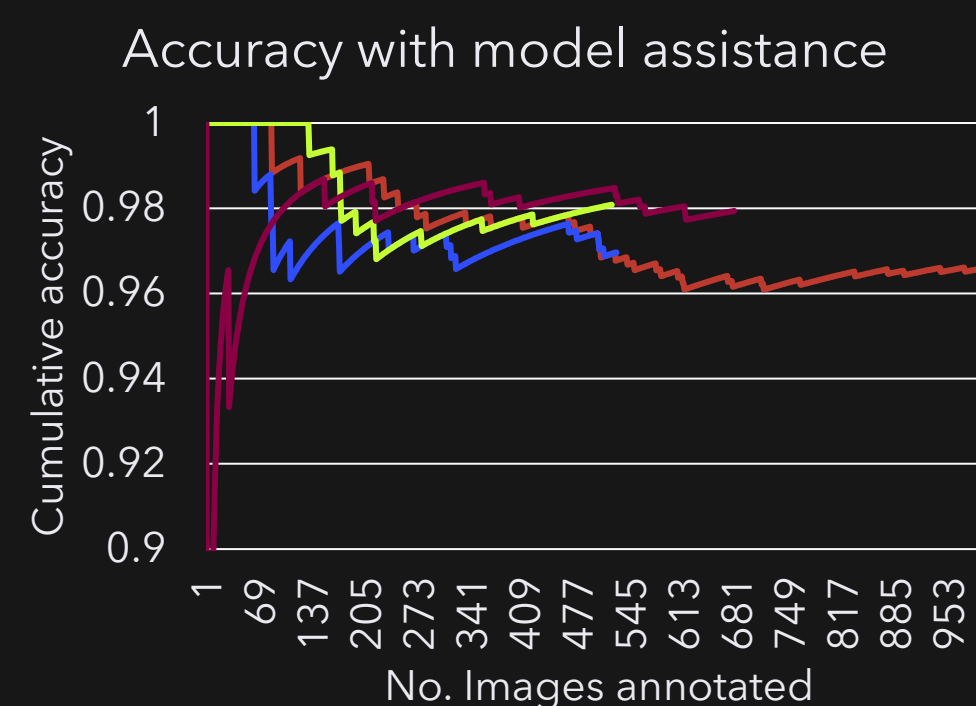
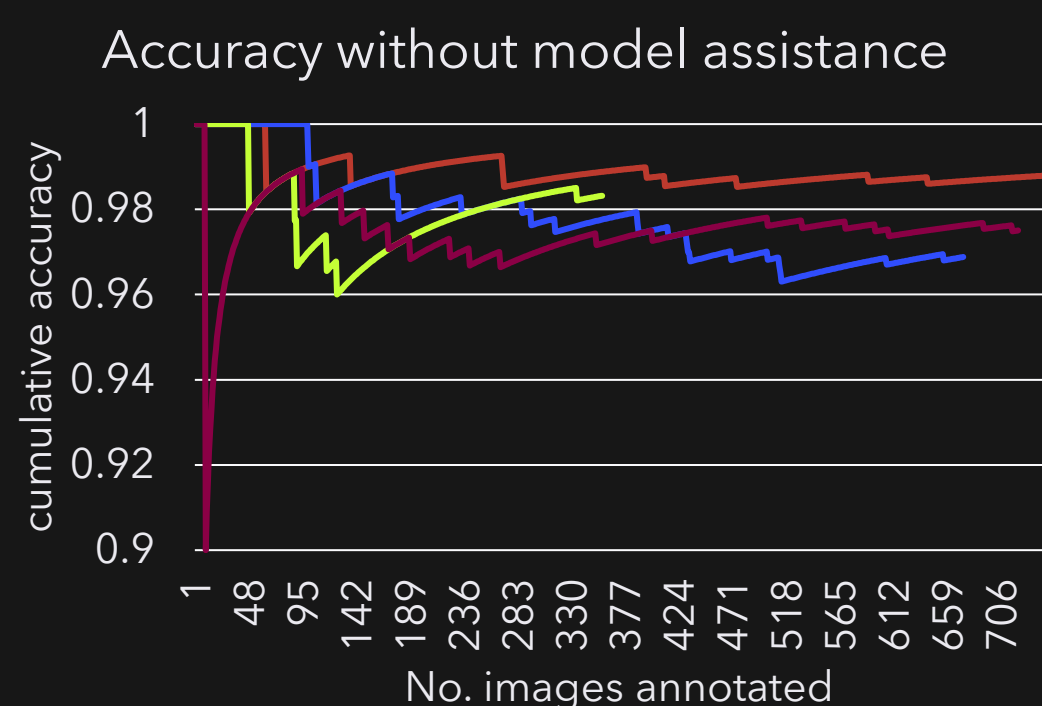
Research Gap & Project Aims

No existing annotation tools allow for training a model which improves as annotations are submitted.

- Create an annotation platform which integrates incremental model training
- Run user studies with the annotation platform
- Measure difference in annotation time with/without model assistance
- Evaluate the method overall, and when it is useful to users

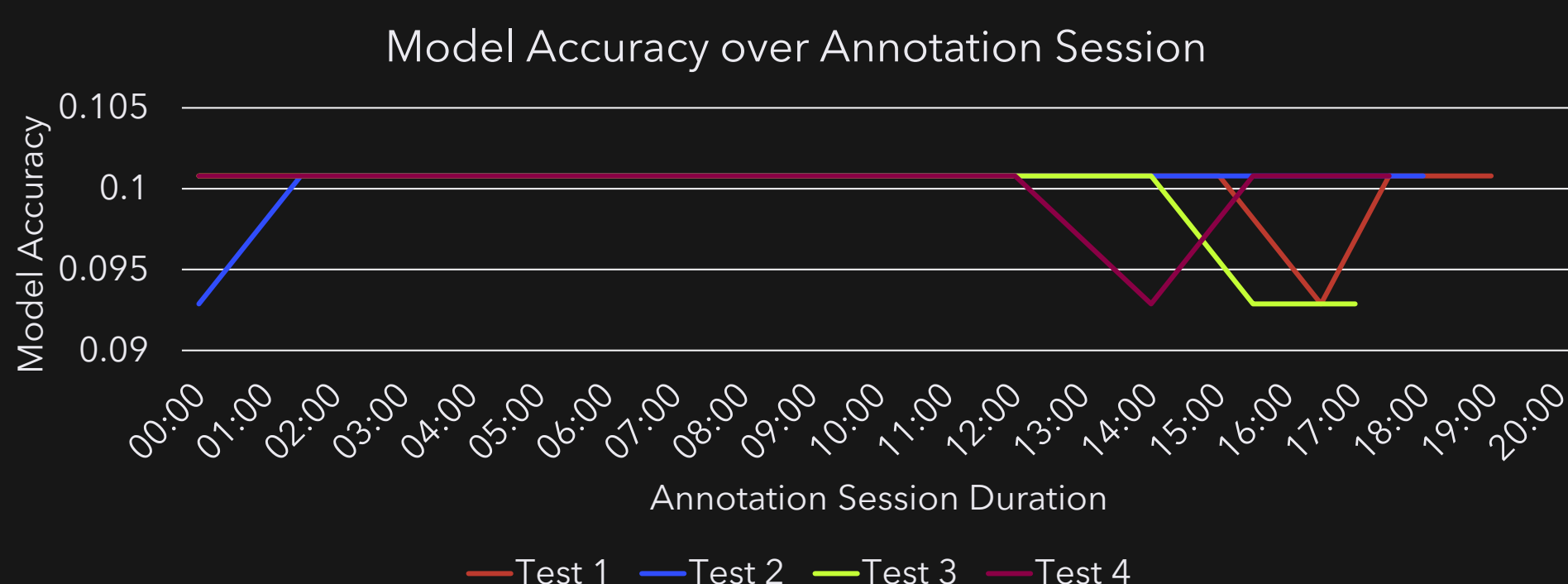
User Study

- Two 20-minute annotation sessions
 - 20 minutes as baseline
 - 20 minutes with model assistance
- Measured annotation time and accuracy
- Annotation accuracy is constant between both annotation sessions



Results

- Model accuracy didn't meaningfully improve across short annotation sessions
- 5,000 image labels required to increase model accuracy
- Annotation tool developed was quick for users to pick up and use



Tools & Platforms

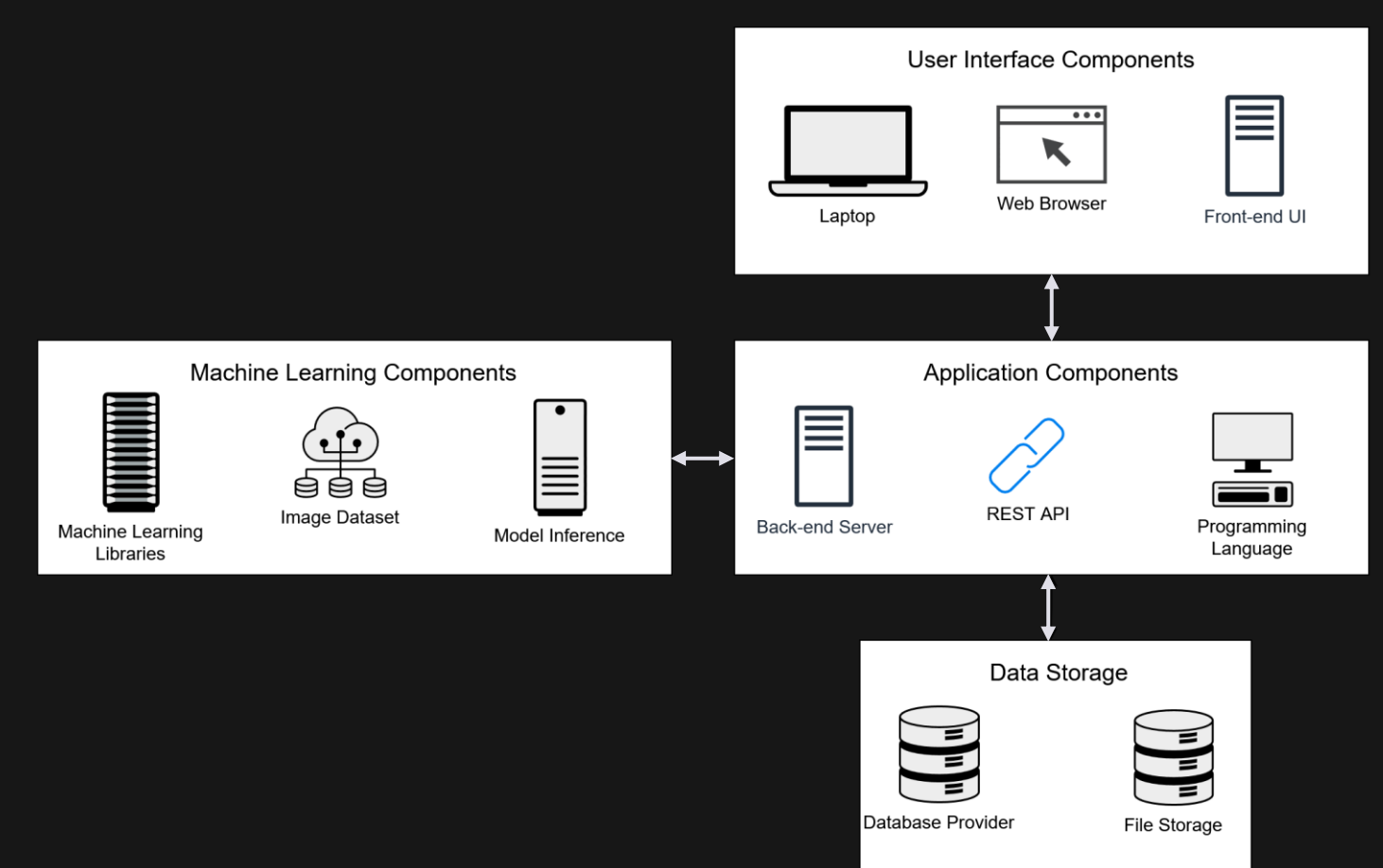


Research Question

How effectively do incrementally trained models reduce image annotation time in semi-automated labelling systems while maintaining sufficient accuracy?

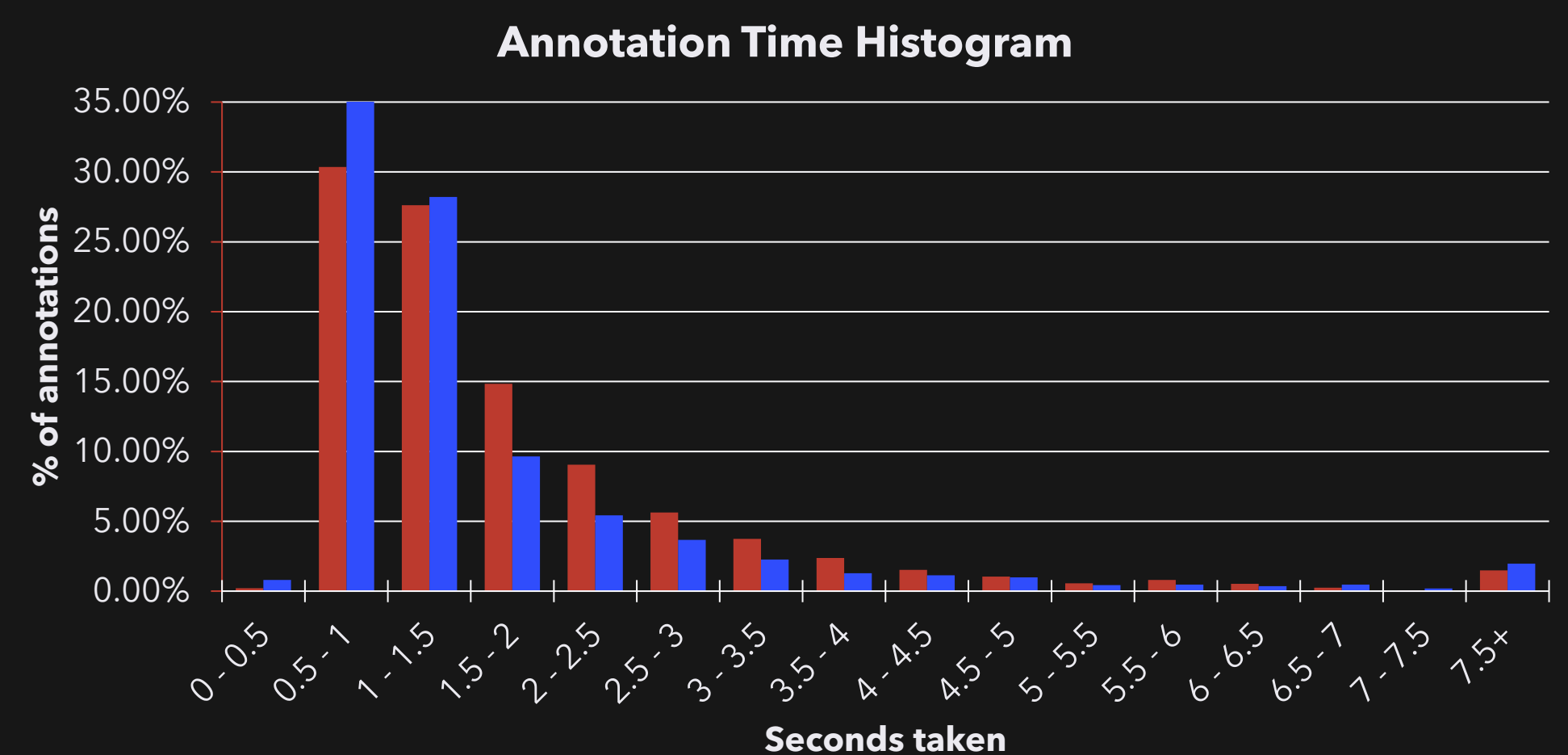
System Architecture

A high-level overview of the components required for a system to integrate incremental training into an annotation tool.



Data Analysis

- Average annotation time is reduced from 1.83s to 1.67s with model assistance
- Decrease in annotation time likely due to having practice for the model assistance session



Conclusion

This project investigated how incrementally trained models could be used to solve the cold-start problem. The method used was not effective in solving this issue, but the solution developed provided a good user experience for test participants.

Future Work

Specialized Model Architecture
Incremental training using a lightweight architecture designed for the dataset

Transfer Learning
Transfer learning from pre-trained model into a specialized dataset