Machine Learning Applications

Case study - CPT Interpretation

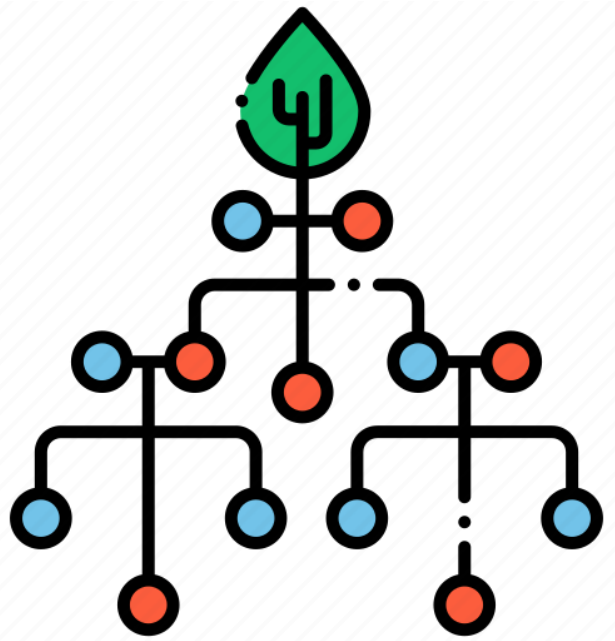
Nicholas Wong

Alex Wong

Content

Objective of the study

- A large portion of geotechnical engineering is about working with data
- Robust approaches usually found on solid engineering theories and judgement
- Albeit importance of engineering judgements, machine learning (ML) could aid engineering studies and applications
- Proof of concept with a trial on Cone Penetration Test (CPT) interpretation



Introduction to Cone Penetration Test (CPT)

- Determine properties of soils
- Suitable for soft ground

A relatively consistent tool

- ✓ Tidy dataset
- ✓ Relatively easy for machine learning

3 key parameters measured

- Cone resistance
- Porewater pressure
- Sleeve friction

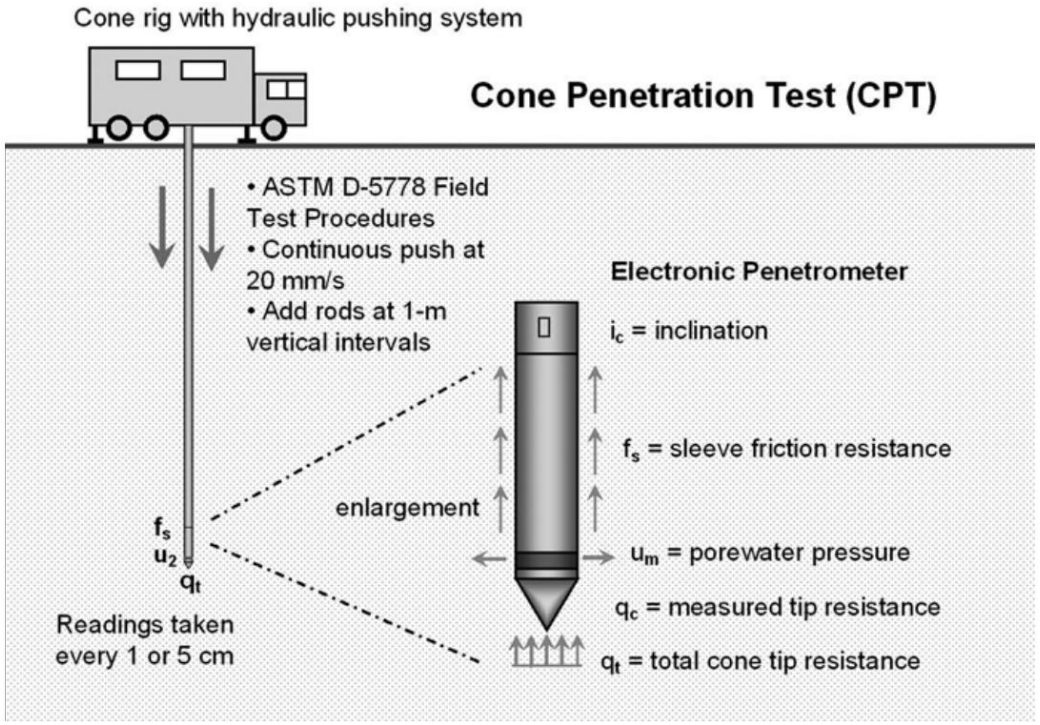


FIGURE 1 Overview of the cone penetration test per ASTM D 5778 procedures.

Traditional approach of interpretation

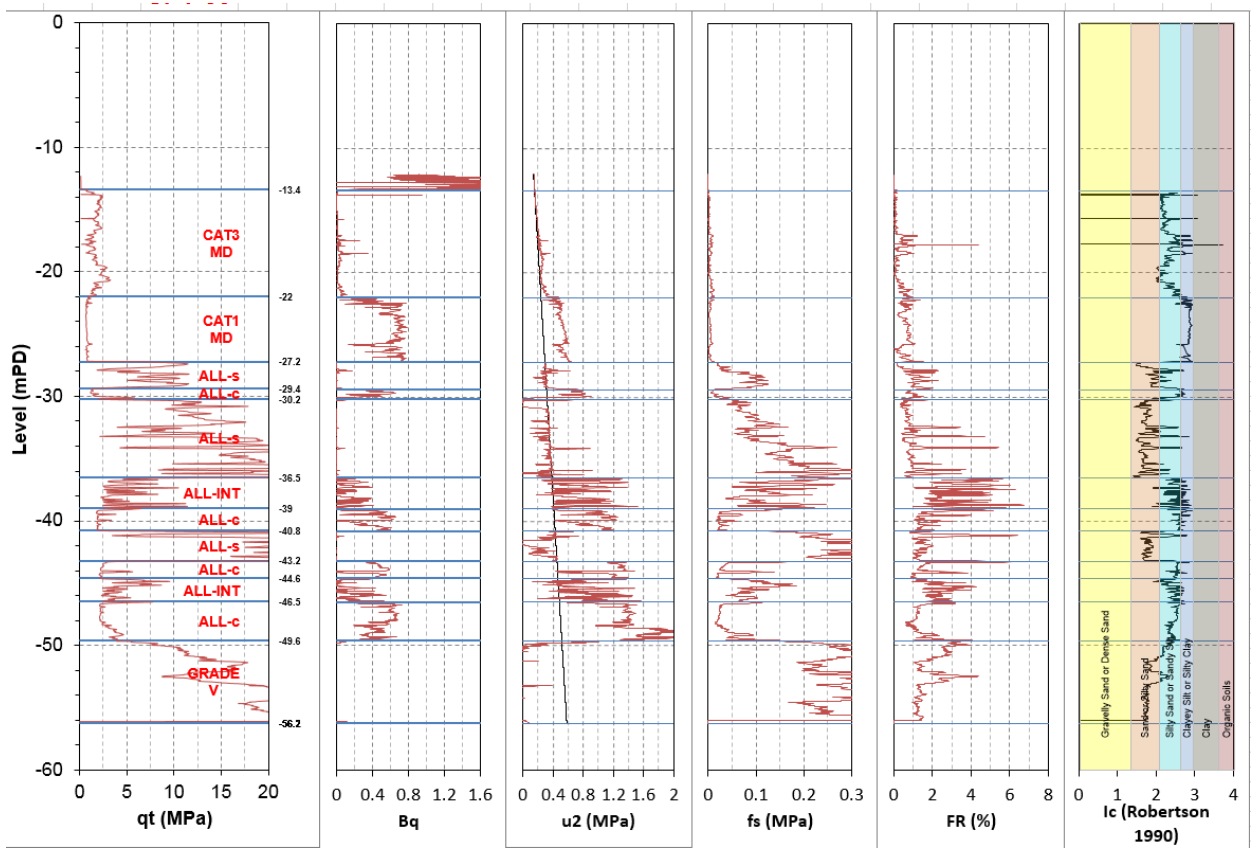
- Derived parameters from the three measurements
- Empirical approaches
- Engineering judgements

Issue

- Time-consuming
- Inconsistence

Machine learning may assist

- Speed up by automating the interpretation
- Improve consistency (NOT accuracy !)



Machine learning approach

Trial on 700 CPTs dataset in a HK project

- 600nos. CPTs completed by the trained model
- Compare with engineer's interpretation

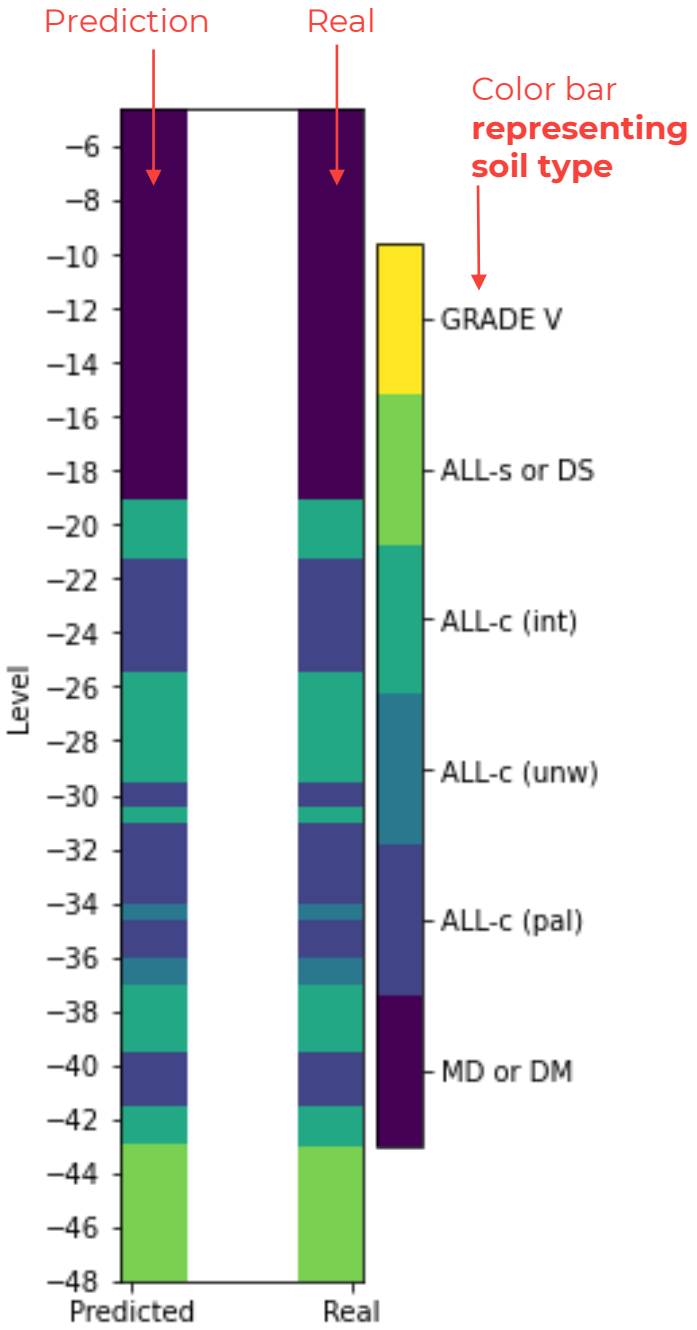
- 100nos. CPTs interpreted by engineer
- Used for training the model

- Python (free programming language)
- Baseline 76% accuracy with neural network
- 4 inputs: depth + 3 measurements
- 1 prediction: soil type

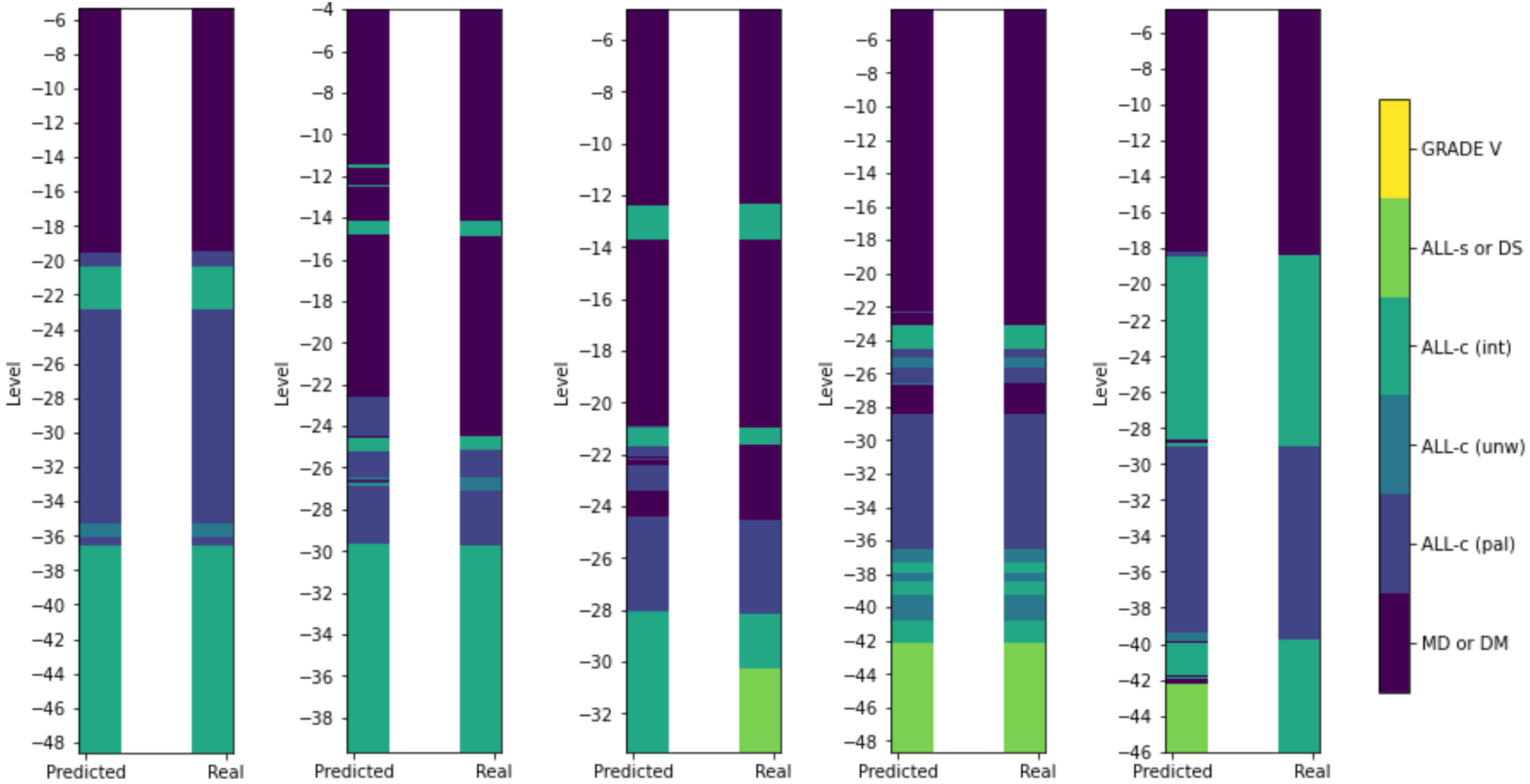
Results

- Satisfactory accuracy
- Further improvement possible
  - More sophisticated models
  - Manipulate input dataset

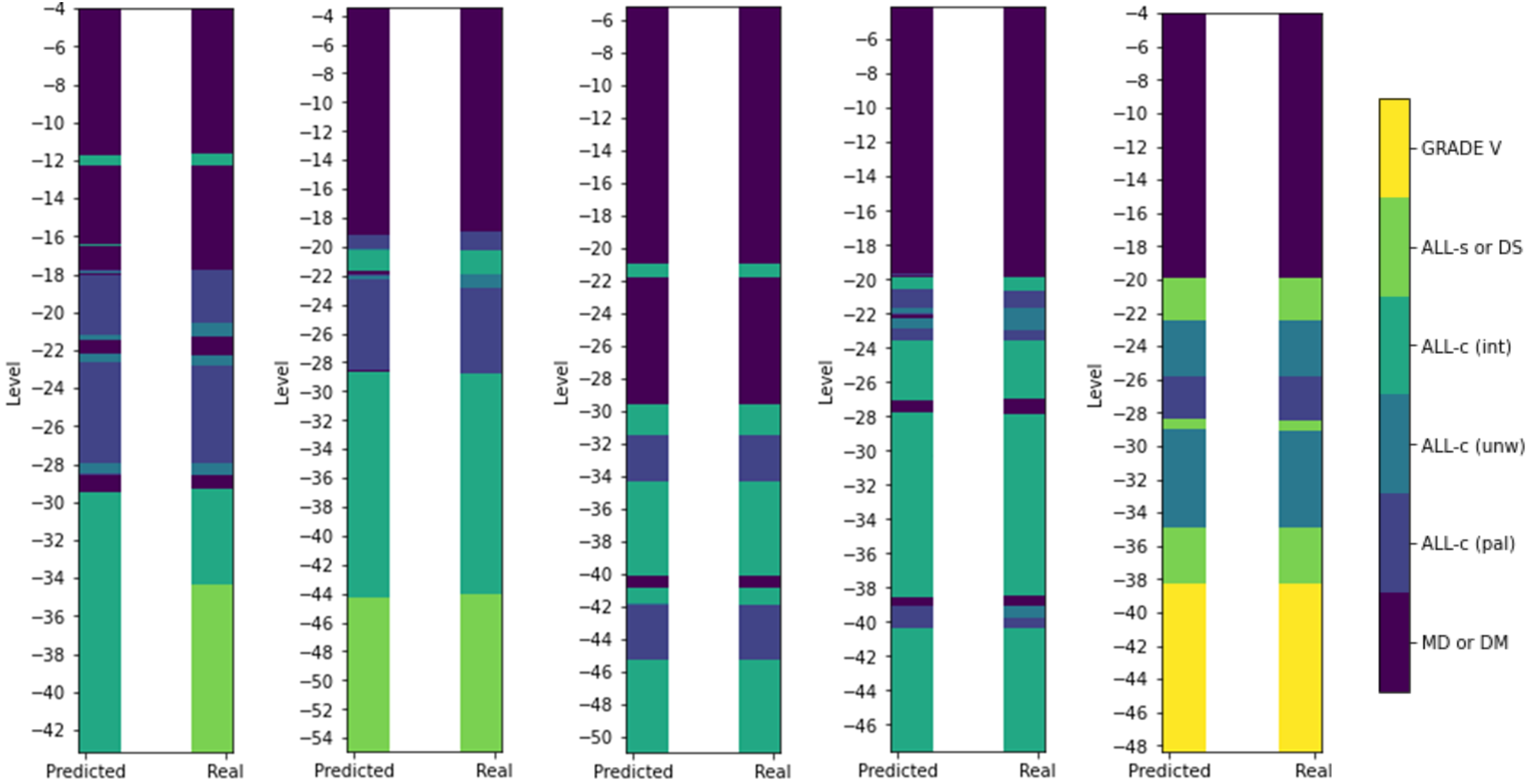
Overall Accuracy
90%



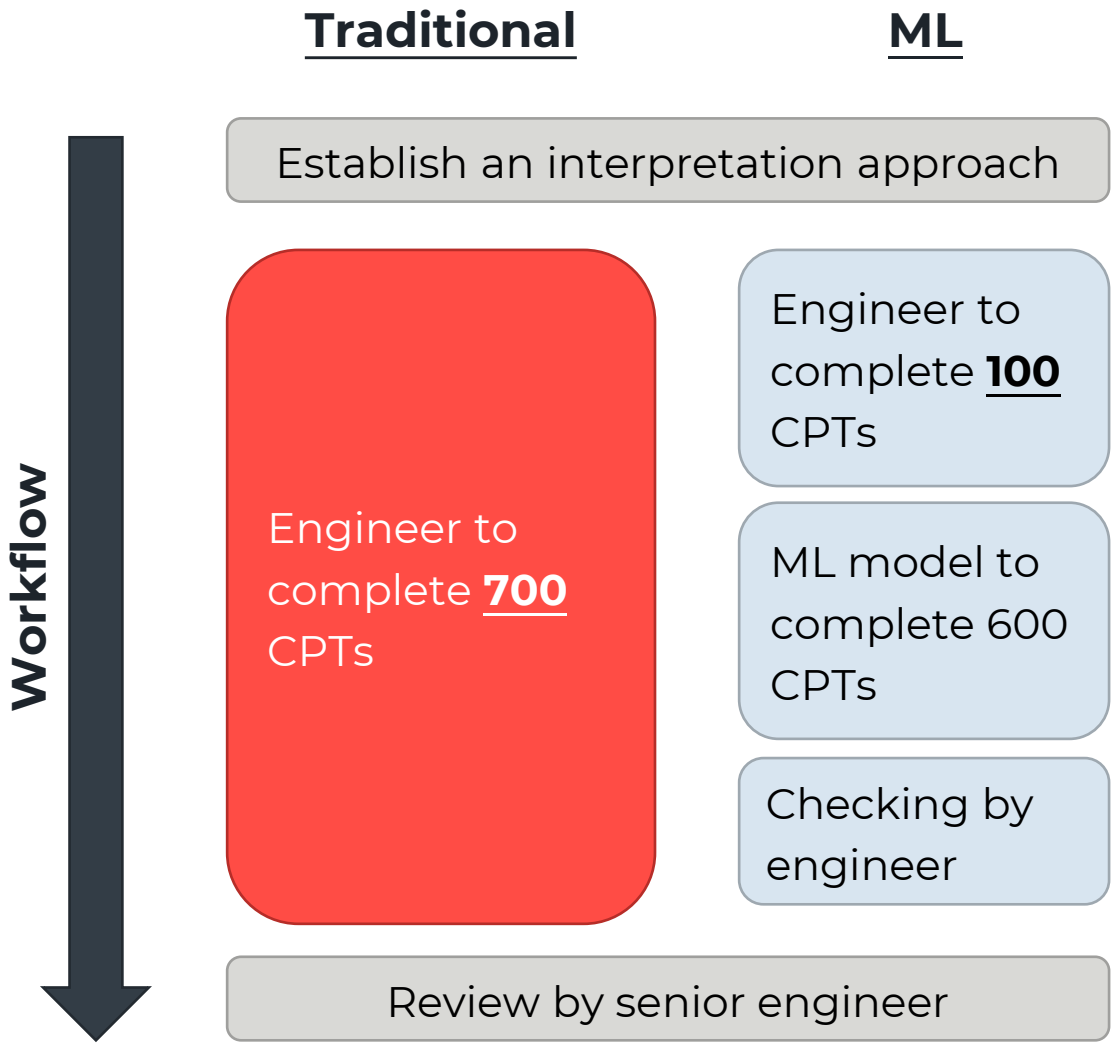
Results – 1/2


Level
Predicted
Real
GRADE V
ALL-s or DS
ALL-c (int)
ALL-c (unw)
ALL-c (pal)
MD or DM

Results – 2/2


Level
Predicted
Real
GRADE V
ALL-s or DS
ALL-c (int)
ALL-c (unw)
ALL-c (pal)
MD or DM

ML approach saves time and labour cost



✓Estimated 80% saving in time with ML

Summary and future works

Further improvements for CPT interpretation

- Use more sophisticated ML model
- Manipulate input dataset
- Smoothen the results

Key take-aways

- CPT interpretation exercise as a proof of concept
- ML streamline workflow saving time and labour cost
- Great potential to apply ML to other applications
  - Boreholes and laboratory tests
  - Text and photo recognition
  - Forecasting of time-series dataset

IMPORTANT: ML model is bounded by training data. Engineering judgement remains crucial.

Appendix – Technical Details

Baseline Score and Performance

Neural Network

Initial approach

- Established a model based on NN
- Baseline score at 70%
- NN achieved 76%

Random Forest

Polished Method

- Set base score at 76%
- RF achieved 90%
- Higher than expectation
- Significant result

Model is performing significantly better

Back to page 6

Testing Method

Random Forest

Minimum sample size for optimum result = 100

Achieved

90%~90.5%

Accuracy

Model is trained specifically with 100, 200 and 400 data sets respectively

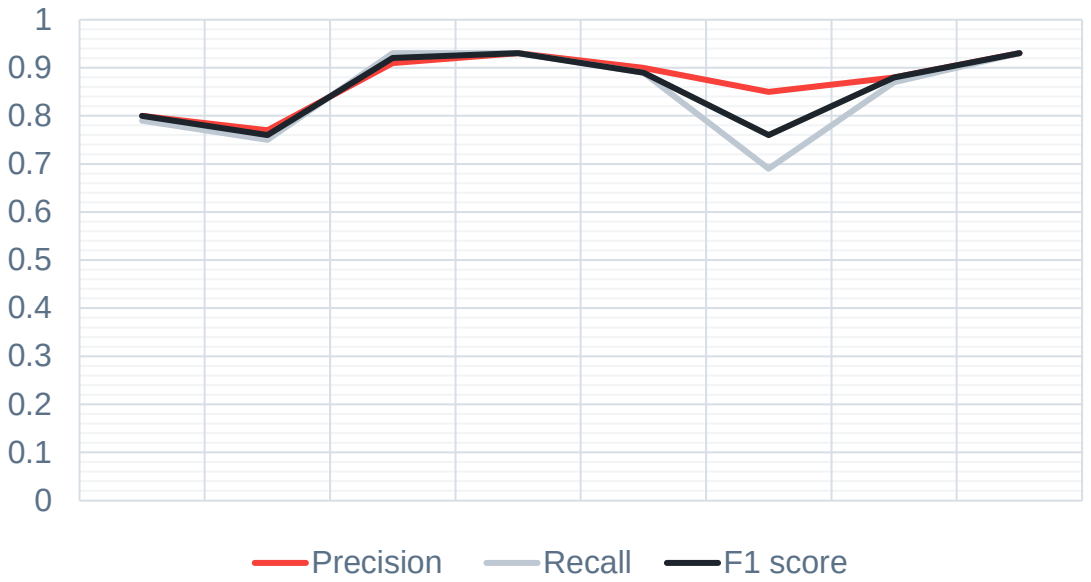
It is categorized into group with randomly selected data sets

While using different range of data, the achieved prediction results are close to 90%

Higher than the baseline score of the hypothesis of our study, 76% based on NN

Precision vs Recall

Random Forest



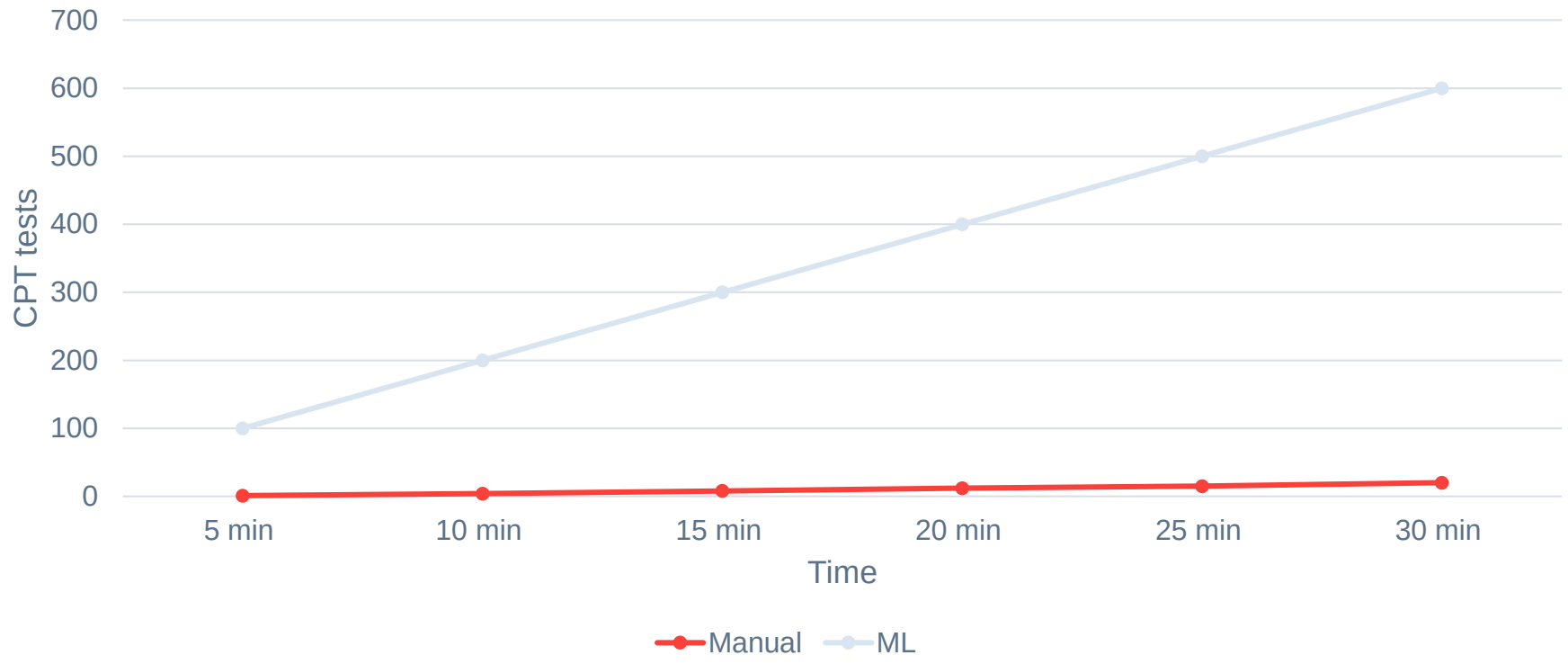
Type of Soil Layer	Precision	Recall	F1 score
All – c int	0.80	0.79	0.80
All – c pal	0.77	0.75	0.76
All – c unw	0.91	0.93	0.92
All – s	0.93	0.93	0.93
DM	0.90	0.89	0.89
DS	0.85	0.69	0.76
GRADE V	0.88	0.87	0.88
MD	0.93	0.93	0.93

Precision and Recall are generally consistent

Further improvement

- Aim for Recall > Precision

Work time productivity
Manual vs ML



Thank you

wsp.com