

A comparative study of machine learning approaches for lithostratigraphic interpretations of CPT data

Tom Van Haren^{*1}, Pieter Jan Kerstens², Filip Schouwenaars², Hossein Ghorbanfekr³, Katrijn Dirix¹, Prof. Dr. Dirk Valkenborg⁴

¹VITO [Environmental Intelligence, tom.vanharen@vito.be]

²VITO [Data & Analytics]

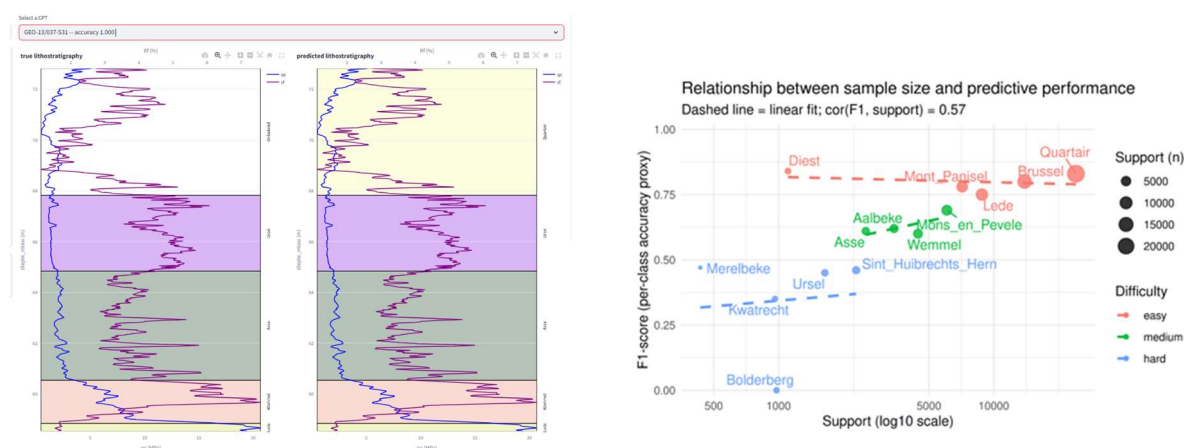
³IMEC

⁴Data Science Institute, UHasselt

Cone penetration tests (CPTs) provide quick and detailed information about the subsurface and are widely used in geotechnical engineering. They also help to provide detailed insight into lithostratigraphic correlation and interpretation for the development of urban subsurface models. However, lithostratigraphic interpretation of CPT data has remained largely a manual process, making it a time-consuming and subjective process. Automated and reproducible interpretation workflows are clearly needed to help put more detail into shallow urban subsurface models.

Our study explores the potential of machine learning models to support geologists and geo engineers with lithostratigraphic interpretation of CPT data on a large scale. It focuses on the development, comparison and integration of complementary approaches. A reference set of around 200 CPT's from the northern periphery of Brussels was manually interpreted by expert geologists. These CPT signals were subsequently enriched with a wide range of derived features and expert-informed constraints focusing on geological layer properties and sequences.

The presentation of this study will focus on the comparison and combination of different approaches with promising accuracies (>70%), demonstrating that this is an interesting way of working with supervised learning approaches to effectively capture lithostratigraphic patterns in CPT data. We will present different solution strategies encountered along the way and how they failed. We then present the latest results, with particular attention to model interpretability, and the role of expert feedback in assisted interpretation workflows.



Dashboard assisted CPT lithostratigraphy (left) and the effects of lithostratigraphy and sample size on predictive performance (right)