Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Hybrid Variational Quantum-Classical Neural Networks for CPT-Based Prediction of Shear Wave Velocity and Soil Behavior Type Index: A Sanity Check Study with Ensemble and Deep Learning Methods

**Omer Muhie Eldeen Taha**

omer.muhie@uokirkuk.edu.iq

University of Kirkuk

**Sarah Mustafa Ahmed**

University of Kirkuk

**Research Article**

**Keywords:** Variational Quantum Neural Network, Cone Penetration Test, Shear Wave Velocity, Soil Behavior Type, Machine Learning, Quantum Computing, Geotechnical Engineering, SHAP Analysis, Uncertainty Quantification

**Posted Date:** May 6th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9524786/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

**Version of Record:** A version of this preprint was published at Transportation Infrastructure Geotechnology on June 17th, 2026. See the published version at https://doi.org/10.1007/s40515-026-00951-2.

## Abstract

This research introduces the inaugural utilization of a Hybrid Variational Quantum-Classical Neural Network (VQNN) for geotechnical site characterization utilizing Cone Penetration Test (CPT) data. A substantial CPT database containing 2,516,979 measurement records from various locations in Austria was analyzed to forecast two geotechnical parameters: directly measured shear wave velocity (Vs) from seismic CPT (SCPT) tests and the Robertson [1] Soil Behavior Type index (Ic SBT). The hybrid architecture combines a 3-qubit parameterized quantum circuit that uses angle embedding and strongly entangling layers with classical neural network layers for pre- and post-processing. The whole thing is trained end-to-end was adopted using L-BFGS-B optimization on a PCA-reduced feature space. Five-fold cross-validation and eight evaluation metrics ($R^2$, RMSE, MAE, MAPE, A10 − I, SI, IA, KGE) to compare the performance of three classical ML models: a Multi-Layer Perceptron (ANN), a Random Forest (RF), and a Gradient Boosting Machine (GBM). For Ic SBT, all classical models reached almost perfect accuracy ($R^2 \geq 0.999$), but the VQNN only reached $R^2 = 0.900$ (RMSE = 0.149, MAPE = 4.23%, A10 − I = 0.924). For directly measured Vs, all models showed moderate performance ($R^2 = 0.21–0.48$, RMSE = 54–67 m/s), which is in line with what Chala and Ray [2] found for measured Vs using XGBoost ($R^2 = 0.54$, MAE = 41 m/s) and much lower than what they found for studies that predicted empirically derived Vs [3]. Permutation importance and RF impurity importance were used for sensitivity analysis. They showed that depth, pore pressure, and sleeve friction were the most important predictors of V. The findings confirm VQNN as a promising novel methodology for geotechnical regression in the NISQ era.


## 1. INTRODUCTION

Accurate calculation of soil shear wave velocity and precise classification of soil behavior type index are essential for effective geotechnical design and seismic response analysis [3].

For geotechnical engineering projects to be safe, it is important to accurately define subsurface conditions. The Cone Penetration Test (CPT) is the most popular way to test things in place around the world because it can do continuous profiling, is very repeatable, and has a lot of multichannel output [1, 4]. Researchers and engineers often use CPT metrics like pore pressure (u2), sleeve friction (fs), and cone tip resistance (qc) to figure out how strong, dense, and stiff undrained soil is [5].

CPT uses the SBT index (Ic) to tell the difference between soils that act like clay and soils that act like sand. It does this by looking at the pore pressure, the sleeve friction (fs), and the cone tip resistance (qt). Vs is important for how a site reacts to earthquakes, and CPT data has been used to connect it to this. But it changes from site to site, which makes it challenging to guess how a site will act during an earthquake. ML is used to achieve better Vs and Ic predictions since these contribute to design decisions as well as based risk analysis [6, 7]. Vs plays a key role in predicting the long-term interactions of soil and structure as well as defining the liquefaction potential and determining site response [8, 9].

The only way to obtain direct measurements is using on-site specialized geophysics tests such as seismic Cpt (SCPT), cross-hole, or downhole methods. We do not always have these tests available as part of standard practice. As a result, numerous studies have sought to establish a relationship between CPT parameters and real-world Vs [1, 10–12] and more recently machine learning (ML) techniques [2, 3].

The explicit objectives are: (i) to construct and evaluate a hybrid VQNN for the direct prediction of measured Vs and Ic SBT from nine CPT-derived input features; (ii) to compare the VQNN with ANN, RF, and GBM using 5-fold cross-validation; (iii) to assess performance through eight metrics, including A10 − I, SI, IA, and KGE; (iv) to examine feature significance; and (v) to contextualize the findings with recent published benchmarks [2, 3].

## 2. LITERATURE REVIEW

# 2.1 Machine Learning for CPT-Based Property Prediction

The application of ML to CPT data has expanded considerably over the past decade [13]. Applied ANNs to liquefaction prediction; Random Forests and gradient boosting subsequently demonstrated superior generalization in geotechnical regression tasks [14–18].

Chala and Ray [3] used RF, SVM, Decision Trees, and XGBoost to predict Vs from the Austrian CPT database. They used four input features: depth (D), cone tip resistance (qc), sleeve friction (fs), and friction ratio (Rf) [19]. A notable methodological difference in that study is that the Vs target variable was not directly measured; instead, it was derived from the empirical equation, which characterizes Vs as a deterministic function of the qc and Ic parameters, which are themselves functions of the same CPT inputs [1]. The models were able to find a deterministic mathematical function instead of predicting an independent geophysical observable, which is why RF, SVM, and XGBoost all got $R^2 = 1.00$ on both the training and testing datasets. The test dataset showed that the RF model had an RMSE of 0.96 m/s, an MAE of 0.50 m/s, a MAPE of 0.36%, a SI of 0.006, and an A10 − I of 1.00.

Chala and Ray [2] created a more advanced framework that combines XGBoost with Shapley Additive Explanations (SHAP) and a Bayesian Generalized Linear Model (GLM). Their study was very important because it used directly measured Vs from SCPT and SCPTu tests (96 soundings, n = 3,600 observations) as input features. These were overburden stress ($\sigma_v$), the mean $q_c$, the mean $f_s$, the depth, the $q_c$ coefficient of variation, and the soil behavior type index (Ic). The training set gave the XGBoost model a $R^2$ value of 0.91, and the test set gave it a $R^2$ value of 0.54.

It also had MAE = 41 m/s and IA = 0.84. The Bayesian GLM made predictions about probabilities with 95% credible intervals. SHAP analysis found that overburden stress was the most important predictor, followed by qc and fs. When we used independent Hungarian SCPT data to check, we got linear correlation coefficients of r = 0.39 (XGBoost) and r = 0.38 (Bayesian GLM).

Rauter et al. [20] looked at Vietnamese CPT data and found that gradient boosting did the best job, with a $R^2$ value of about 0.82. Zhang et al. [14] used deep neural networks on a Chinese multi-site CPT database and got $R^2 > 0.85$ after a lot of feature engineering. Yang et al. [21] showed that Random Forests can reliably reproduce Robertson SBT classifications with an accuracy of more than 90% for classifying soil.

## 2.2 Quantum Machine Learning: Principles and Applications

Quantum Machine Learning: Fundamentals and Utilizations QML uses quantum superposition, entanglement, and interference to work with data in high-dimensional Hilbert spaces [22]. Variational Quantum Circuits (VQCs) use encoding unitarizes to embed input data, trainable parameterized gates to process it, and measurements to get classical predictions [23]. The quantum circuit learning framework was introduced by Mitarai et al. [24]; and the hybrid architecture was formalized by Mari et al. [25]. VQNNs use variational quantum circuits and classical neural networks to model phase and amplitude. They also use parameter-shift rules to lower measurement costs and barren plateaus. Quantum RNNs surpass traditional RNNs in predicting soil profiles in geotechnics due to their utilization of superposition to integrate geographical data. Hybrid QCQ-CNNs improve generalization in noisy settings, suggesting their prospective applicability in CPT contexts where quantum layers augment expressiveness [26–28]. Hybrids have theoretical benefits in optimization landscapes, but they lack direct CPT benchmarks. Quantum enhancements may surpass ensembles in scenarios with limited data [29].

Quantum applications are still in their early stages in civil and geotechnical engineering. This work is new because no other published study has used hybrid VQNNs to figure out geotechnical parameters from in-situ testing.

# 3. DATASET DESCRIPTION AND PREPROCESSING

## 3.1 Dataset Overview

The dataset comes from the publicly available CPT database that Oberhollenzer et al. [19] published from Premstaller Geotechnik. The dataset contains 2,516,979 unique CPT measurement records from various sites in Austria, such as the

Vienna Basin, Gastein Valley, and Zell Basin. This dataset was also used by Chala and Ray [2, 3]. Only standard CPT/CPTu tests and their seismic versions (SCPT/SCPTu) are in the database. The latter tests consist of direct Vs measurements at designated depth intervals. Table 1 provides a succinct summary of the nine input features and two prediction targets utilized in this study.

This study uses V values that were directly measured from SCPT/SCPTu geophysical tests, unlike Chala and Ray [3], who predicted V using the Robertson [1] empirical formula (a deterministic function of the same CPT inputs). This difference is very important: predicting a derived V that is an algebraic function of the model inputs trivially yields $R^2 \approx 1$, whereas predicting independently measured V from CPT parameters represents a genuine geophysical regression problem.

Table 1
*Summary of input features and target variables used in this study.*

| Parameter | Symbol | Unit | Role | Approx. Range (p1–p99) |
|---|---|---|---|---|
| Depth below surface | z | m | Input | 0.02–43.0 m |
| Cone tip resistance | qc | MPa | Input | 0.05–38.6 |
| Sleeve friction | fs | kPa | Input | 0.5–612 |
| Pore water pressure | u2 | kPa | Input | 0–680 |
| Normalized friction ratio | Fr | % | Input | 0.1–8.5 |
| Normalized tip resistance | Qt | — | Input | 0.5–420 |
| Friction ratio | Rf | % | Input | 0.1–7.8 |
| Hydrostatic pore pressure | u0 | kPa | Input | 0–420 |
| SBT index (CPT-derived) | Ic | — | Input | 0.6–3.6 |
| Shear wave velocity (SCPT) | Vs | m/s | Target 1 | 42–540 |
| SBT zone index (Robertson) | Ic SBT | — | Target 2 | 1.48–3.53 |

## 3.2 Quality and Filtering of Data

The raw database had a lot of missing data in the Vs column: only 3,981 records (0.16%) had Vs values directly measured from SCPT/SCPTu tests. Chala and Ray [2] also got 96 SCPT/SCPTu soundings (n = 3,600 observations) from the same source, so this is in line with what they found. We used a two-step filtering process: (i) we deleted rows that had NaN in any feature or target; and (ii) we clipped the inter-percentile range (IPR) to keep observations within the 1st–99th percentile of each variable, which is the same as what Chala and Ray [3] did with the IQR-based approach. The Vs dataset had 1,879 samples after filtering (these were used in the SCPT-based Vs prediction task), and the Ic SBT dataset was randomly down sampled to 5,000 samples to make it easier to work with.

## 3.3 Normalization and Principal Component Analysis

The RobustScaler was used for feature normalization. It centers each feature on its median and scales it by the IQR (Q3 – Q1). This is because CPT data has heavy-tailed distributions, and it works better than z-score scaling for datasets with residual outliers. To stop data leakage, the scaler was only fitted on training fold data. This is a standard practice in geotechnical ML [2]. PCA also cut the scaled 9-dimensional feature space down to 3 principal components for the quantum circuit. This kept 88.0% of the total feature variance (PC1: 51.5%, PC2: 22.4%, PC3: 14.1%).

## 3.4 Structure of Correlation

Figure 1 shows that depth (r = 0.47), hydrostatic pore pressure u0 (r = 0.43), and pore pressure u2 (r = 0.36) are the best linear predictors of Vs. Sleeve friction (r = 0.28) and qc (r = 0.18) have moderate correlations, which is in line with Chala and Ray [2] partial dependence analysis, which found that overburden stress and qc were the most important SHAP-ranked features. The

low linear correlation between qc and Vs (r = 0.18) shows that large-strain penetration resistance and small-strain elastic stiffness are controlled by different stress-strain regimes. The near-unit correlations between Qt, Fr, and Ic for Ic SBT explain why the ceiling-level classical model does so well on this target.

# 4. METHOD

# 4.1 The framework for the experiment and the performance metrics

## 4.1.1 General Design of the Experiment

The experimental framework was constructed to yield statistically robust, unbiased performance estimates for all four models—ANN, RF, GBM, and VQNN—across both target variables (Vs and Ic SBT), adhering to the principled evaluation protocols established in recent geotechnical machine learning studies [2, 3, 14] and the overarching best-practice guidelines of Cawley and Talbot [30] and Arlot and Celisse [31]. The design was based on three main ideas: (i) preventing data leakage, which meant that all data-dependent transformations were only fitted on training data and not on validation data [32]; (ii) model comparability, which meant that all four models were tested on the same cross-validation folds so that metrics could be compared directly and fairly [33]; and (iii) metric comprehensiveness, which meant that eight different statistical metrics were used to measure different aspects of predictive accuracy, dispersion, and engineering reliability [3, 17].

## 4.1.2 Strategy for Cross-Validation

K-fold cross-validation was used with k = 5 to evaluate the model. This means that the dataset is divided into five folds of about the same size that don't overlap. In each iteration, one-fold is kept as the validation set and the other four are used for training. The final performance estimate is the mean of all five-fold-specific validation scores, and the standard deviation across folds is used as a measure of estimation stability [31, 34]. The choice of k = 5 is based on the well-known bias-variance trade-off in cross-validation design [35]. This value represents a widely accepted equilibrium between estimation bias and variance, consistent with the cross-validation protocols employed by Chala and Ray [2, 3], Pham et al. [17], and the general recommendation of Hastie et al. [34] to utilize $k \in \{5, 10\}$ for the majority of regression applications. For full reproducibility, partitioning used stratified random assignment with a fixed random seed (seed = 42). This made about 80% of the training samples and 20% of the validation samples per fold. For the Vs task (n = 1,879), this meant about 1,503 training samples, and for the Ic SBT task (n = 5,000), it meant 4,000 training samples.

## 4.1.3 Preprocessing Data in Folds

To avoid preprocessing leakage, which is when information from the validation set gets mixed up with the training set and gives false performance estimates [30, 32], all steps that involved estimating parameters from data were done separately for each fold, using only the training part of that fold. Only the RobustScaler was fitted to the training fold, which meant calculating the median and interquartile range (IQR = Q75 − Q25) of each feature from the training data. Then, the training-derived statistics was used to transform both folds without refitting, as recommended by Chala and Ray [2] and Kaufman et al. [32] to prevent data leakage. The RobustScaler was chosen over the standard z-score scaler because CPT data has heavy-tailed, outlier-prone distributions [4]. For the VQNN pathway only, PCA was also used within each fold. This was done by estimating the three main components from scaled training features and applying the training-derived transformation matrix to both folds. This made sure that the variance structure of the validation data did not affect the dimensionality reduction [30]. Training fold statistics were employed (mean $\bar{y}$_train and standard deviation σ_train) to standardize the target variable even more for the ANN and VQNN. Before we calculated the metrics, The predictions were back-transformed ($\hat{y} = \hat{y}$_norm · σ_train + $\bar{y}$_train) to make sure that all reported metrics are in original physical units that can be compared across all four models. RF and GBM were directly trained on the original-scale target because tree-based models are scale-invariant [34].

## 4.1.4 Metrics for Performance

Eight different evaluation metrics were used that worked together to measure model performance. These included prediction accuracy, dispersion, relative percentage error, engineering reliability, and agreement quality followed the multi-metric evaluation protocols of Chala and Ray [2, 3]. The Coefficient of Determination ($R^2 = 1 - \Sigma(x_i - \hat{x}_i)^2/\Sigma(x_i - \bar{x})^2$) shows how much of the total variance in the target is explained by the model. An ideal value is 1.0 [36], and it is the most common metric used in geotechnical ML studies [37]. However, it does not separate bias from dispersion, so it is reported with other metrics [38]. The Root Mean Squared Error ($RMSE = \sqrt{[\Sigma(x_i - \hat{x}_i)^2/n]}$) and Mean Absolute Error ($MAE = \Sigma|x_i - \hat{x}_i|/n$) both express average prediction error in physical units — RMSE penalizes large errors more heavily due to the squared term, while MAE weights all errors equally, making the combination informative about outlier behavior [39]; RMSE was used as the primary hyperparameter tuning metric for the GBM model, consistent with Chala and Ray [3]. The Mean Absolute Percentage Error ($MAPE = (1/n)\Sigma|(x_i - \hat{x}_i)/x_i| \times 100\%$) summarizes normal error as a percentage of the actual measured value, making it a scale-independent measure of accuracy [40] whose interpretation is intuitive, and that in itself has no correlation with the physical quantities of the target variable. The Engineering Index A10 ($A10 = n_{10}/n$, where $n_{10}$ is the number of predictions within ± 10% of the values sampled) was introduced by Xu et al. [41] and measures the proportion of prediction that met the ± 10% engineering tolerance criterion applicable to field geotechnical design [3,42]. The Scatter Index ($SI = RMSE/\bar{x}$) normalizes the RMSE using the mean of the measured values together, and generates a dimensionless relative dispersion measure categorized according to certain cutoffs [43]: SI < 0.05 = excellent precision; 0.05 < SI ≤ 0.10 = good precision; 0.10 < SI ≤ 0.15 = fair precision; SI > 0.15 = poor precision. The Index of Agreement ($IA = 1 - \Sigma(x_i - \hat{x}_i)^2/\Sigma(|\hat{x}_i - \bar{x}| + |x_i - \bar{x}|)^2$) as a bounded measure [0, 1] was introduced by Willmott [44], and is used to overcome the insensitivity of $R^2$ to both proportional and additive biases in geotechnical model evaluation and has been proposed as an additional choice of standard complementary to $R^2$ and RMSE [2,45]. The Kling–Gupta Efficiency ($KGE = 1 - \sqrt{[(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2]}$, in which r is the Pearson correlation coefficient, $\alpha = \sigma_pred/\sigma_obs$ is the variability ratio, and $\beta = \mu_pred/\mu_obs$ is the bias ratio) was suggested by Gupta et al. [46], a multi-component metric that simultaneously penalizes correlation errors, mismatches in measures of variability, and bias in measuring mean with equal weight and thus more stringent than $R^2$ as a diagnostically informative criterion [2,47]. Table 2 presents the formulae, optimal values, and taxonomy of the 8 metrics, with reported values indicating the mean across the five cross-validation folds, which was suggested for reporting on cross-validation using the cross-validation reporting standards in Demšar proposed [33] and adapted to be used for further study in the geostatistical and geotechnical ML benchmarking context[3,48].

## 4.2 Research Pipeline and Workflow

The full machine learning pipeline that was developed for the study is presented in Fig. 2. It goes through the entire process — starting with raw CPT data through evaluation and making predictions of the model. Based on Chala and Ray [2, 3], the pipeline had its stages strictly separated to make separation between training and validation. The pipeline divided into two paths following preprocessing, as below. The classical method (ANN, RF, GBM) receives the whole 9-dimensional scaled feature space, which allows individual models to use all the feature-feature relationships. The quantum pathway (VQNN) first uses PCA to reduce the feature space to three main components (the same number as the qubits in the quantum circuit). Then it encodes the data as rotation angles. Both these pathways are incorporated into a single evaluation framework allowing data comparison of metrics among the four models in a direct manner. Feature importance analyses are performed following training with RF impurity-based importance and permutation-equivalent sensitivity ranking based features. This tests for the predominant CPT predictors among the approaches.

## 4.3 Traditional Baseline Models

All models underwent assessment within a standardized 5-fold cross-validation (CV) framework. We only used training data from each fold to fit normalization and PCA. Following Chala and Ray [3], we used eight different metrics, listed in Table 2, to measure how well the model worked. The A10 – I index (the percentage of predictions that are within ± 10% of the measured values) is very useful for engineering purposes. Chala and Ray [2] also report the Index of Agreement (IA) and Kling–Gupta Efficiency (KGE) so that they can be compared directly to that benchmark.

Table 2
Performance metrics used for model evaluation.

| Metric | Symbol | Formula | Ideal |
|---|---|---|---|
| Coefficient of determination | $R^2$ | $1 - \Sigma(x_i - \hat{y}_i)^2 / \Sigma(x_i - \bar{x})^2$ | 1 |
| Root mean squared error | RMSE | $\sqrt{[\Sigma(x_i - \hat{y}_i)^2/n]}$ | 0 |
| Mean absolute error | MAE | $\Sigma\lvert x_i - \hat{y}_i\rvert/n$ | 0 |
| Mean absolute % error | MAPE | $(1/n)\Sigma\lvert(x_i - \hat{y}_i)/x_i\rvert \times 100$ | 0 |
| Engineering index (± 10%) | A10 – I | n10/n ($\lvert pred/meas - 1\rvert \leq 0.1$) | 1 |
| Scatter index | SI | RMSE / $\bar{x}$ | < 0.05 = excellent |
| Index of Agreement | IA | $1 - \Sigma(y_i - \hat{y}_i)^2/\Sigma(\lvert y_i - \bar{y}\rvert + \lvert \hat{y}_i - \bar{y}\rvert)^2$ | 1 |
| Kling–Gupta Efficiency | KGE | $1 - \sqrt{[(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2]}$ | 1 |
| Performance index | PI | $RMSE/(\bar{x} \times \sqrt{(R^2+1)})$ | 0 |

# 4.4 Classical Baseline Models

## 4.4.1 ANN (Artificial Neural Network)

A fully connected MLP with three hidden layers (64, 64, and 32 neurons) that uses ReLU activations and L2 regularization (α = 10 − 4). We used the Adam optimizer (lr = 10 − 3) with early stopping (patience = 25 epochs) on a validation fraction of 15%. Before training, the targets were standardized to have a mean of zero and a variance of one.

Figure 3 shows the ANN architecture that was used in this study. The network design is based on the deep regression model that was used in recent geotechnical machine learning studies (Wang, 2022). The three hidden layers gradually lower the number of neurons from 64 to 64 to 32. This lets the network learn hierarchical feature representations, from raw CPT measurements to derived soil behavior patterns to final parameter predictions. We chose the ReLU (Rectified Linear Unit) activation function to avoid the problem of the gradient disappearing, which is common in deep networks with sigmoid activations. L2 weight regularization ($\alpha = 10^{-4}$) punishes large weight magnitudes, which helps the small Vs dataset (n = 1,879) not to overfit. After the second hidden layer, dropout (rate = 0.1) randomly sets 10% of neuron outputs to zero during training. This works like an implicit ensemble of sub-networks.

## 4.4.2 Random Forest (RF)

A group of 150 decision trees with a maximum depth of 12. Bootstrap sampling and feature sub-sampling (√p features per split) helped make the ensemble more diverse. RF gives impurity-based feature importance metrics that are similar to the permutation feature importance that Chala and Ray [3] used to choose features.

## 4.4.3 Gradient Boosting Machine (GBM)

A GBM with 150 trees, a learning rate of 0.07, a maximum depth of 5, and 80% random subsampling. We chose GBM as the state-of-the-art tabular regression baseline because it is very similar to the XGBoost models of Chala and Ray [2, 3] and can be used as a direct comparison point.

# 4.5 Decision Tree Analysis: Methodology for Sample Segmentation

Figure 4. Sample decision tree structures illustrating the splitting criteria for predicting Vs (m/s) (left) and Ic SBT (–) (right). Thresholds are shown in original physical units after back-transformation from the RobustScaler. Leaf node values represent the mean predicted value within each terminal partition. Node size (n) indicates the number of training samples reaching each node.

Figure 4 additionally illustrates that the two targets have very different decision-making processes, which is because they have different physical drivers. The root node splits on depth (z ≤ threshold) for Vs prediction (left panel), which shows that depth/confining stress is the main factor that determines how stiff the soil is. The second level shows that sleeve friction (fs) is the most important secondary predictor. This is consistent with the feature importance rankings shown in Fig. 8 and SHAP analysis by Chala and Ray [2]. With increasing depths, pore pressure (u2) and friction ratio (Rf) are also present at greater depths, which suggests their dependence upon the soil stiffness characteristics that are affected by drainage and soil typology. Leaf node predictions range from around 174 m/s (shallow, low friction soils) through over 390 m/s (deep, rigid deposits). This applies to the entire training set (42–540 m/s). For Ic SBT prediction (right panel), almost all of the decision logic is based on normalized cone tip resistance (qc), which controls the root and most of the splits that follow. This agrees with the Robertson [1] SBT classification model which relies on qc as the axis. Both the friction ratio (Rf) and normalized friction ratio (Fr) are at level 2. They represent the clay-sand border on the Robertson chart at approximately Fr = 1%. The predictions for leaf nodes range from 1.78 to 3.29, meaning that they cover SBT Zones 2–5 (silty clay to dense sand). This shows that the decision tree can get the main structure of the Robertson [1] classification scheme from the CPT data alone.

It is important to remember that the single decision tree gives clear splitting rules, but its ability to make predictions is much lower than that of the RF ensemble ($R^2$ = 0.479). The ensemble averages over 150 different trees to lower variance by using the law of large numbers. So, the best way to think of the individual tree is as a teaching tool for the main partitioning logic, not as a model that can predict what will happen.

## 4.6 Hybrid Variational Quantum-Classical Neural Network

The Hybrid Variational Quantum-Classical Neural Network (VQNN) adopted in this study belongs to the family of Variational Quantum Algorithms (VQAs), which represent the most practically relevant class of quantum machine learning models for the current Noisy Intermediate-Scale Quantum (NISQ) era [23, 49]. VQAs are distinguished from purely classical neural networks by the substitution of one or more computational layers with a parameterised quantum circuit (PQC), also termed a variational quantum circuit (VQC), which operates on quantum states in a high-dimensional Hilbert space that is exponentially larger than the corresponding classical feature space [22, 50]. The central hypothesis motivating VQNNs is that this expanded computational space may confer an inductive bias advantageous for certain regression tasks particularly those characterized by complex, non-linear feature interactions that is difficult to replicate with classical architectures of comparable parameter count [51].

CPT data in the geotechnical literature display precisely such characteristics: the connection between cone penetration measurements and soil behavior parameters (e.g. Vs) comprises multi-scale non-linearities, stress-path dependencies, and geological variability inadequately captured by linear models [1, 9]. The VQNN framework employed here expands on the quantum circuit learning approach introduced by Mitarai et al. [24], where a PQC is deployed as a function approximator similar to that used in classical neural networks, and the hybrid architecture formalized by Mari et al. [25], which showed that classical pre- and post-processing layers can make the connection between raw data representations and quantum-compatible encoded states. Such a hybrid design is particularly suited to NISQ hardware constraints and is able to minimize the required circuit depth and qubit count while realizing both representation as well as quantum feature map performances [52].

### 4.6.1 Architecture

The combined development for this work involved designing a dual-functionally different approach from a generic pre-network that consists of three, sequentially trained, but co-equally functionally different, VQNN architectures [25, 53].

Component 1 — Classical pre-network. The first part is a shallow classical neural network which translates the PCA-induced d-dimensional input, the vector $x \in \mathbb{R}^d$ whose magnitude is denoted d = Number of qubits = 3 values in the current work, as a vector of rotation angle $\theta \in [-\pi, +\pi]^d$ appropriate for embedding quantum angle. Specifically, a linear transformation plus hyperbolic tangent activation is applied:

$$\theta = \pi \cdot \tanh(W_1 x + b_1)$$

where $W_1$ and $b_1$ are trainable parameters and tanh nonlinearity maps the output to the interval (− 1, + 1), and scales it by π to keep the angles within the valid embedding range [−π, +π]. This pre-processing stage is critical since raw PCA scores are real-valued and unbounded, while quantum rotation gates encode angles that are modulo 2π; without range normalization the embedding appears periodic and loses injectivity [54, 55]. The tanh based methodology used here is in line with the formulation of Mitarai et al. [24] as smooth, bounded pre-processing functions to preserve the gradient flow through the hybrid network without the vanishing gradient associated with hard clipping.

Component 2—This material consists of a 3-qubit PQC, processing the encoded quantum state and harvesting measurements. The circuit composition and working components are presented in detail in Section 4.6.2. The PQC introduces a set of trainable quantum weights $W_q \in \mathbb{R}^{(L \times N_q \times 3)}$ (where L is the number of entangling layers and $N_q = 3$ is the number of qubits), yielding 18 quantum parameters for the configuration adopted in this study (L = 2, $N_q = 3$). The quantum circuit effectively acts as a kernel function that maps the angle-encoded input into a $2^{(N_q)} =$ 8-dimensional Hilbert space, from which expectation values are extracted as fixed-dimensional real-valued representations [52, 53]. This Hilbert-space feature map is analogous to the kernel trick in support vector machines, but the specific kernel is implicitly defined by the circuit architecture and trainable parameters rather than a pre-specified analytical function [56].

**Component 3 — Classical post-network.** The output component maps the $N_q = 3$ PauliZ expectation values $\langle Z_q \rangle \in [-1, +1]^3$ returned by the quantum circuit to a scalar regression prediction $\hat{y} \in \mathbb{R}$ via a small fully connected network:

$$\hat{y} = W_2 q + b_2$$

where $q = (\langle Z_0 \rangle, \langle Z_1 \rangle, \langle Z_2 \rangle)^T \in [-1, +1]^3$ is the quantum measurement output vector, and $W_2 \in \mathbb{R}^{(1 \times 3)}$, $b_2 \in \mathbb{R}$ represent the post-network weights and bias. An intermediate ReLU-activated hidden layer with 16 neurons is additionally inserted between the quantum output and the final linear projection to allow the post-network to learn non-linear combinations of the qubit expectation values before producing the prediction. This design follows Mari et al. (2020), who showed that a classical post-processing layer is necessary to map quantum observables — which are bounded in [− 1, +1] by construction — to unbounded regression targets such as Vs (m/s).

**Joint optimization.** All parameters of the three components $\Theta = \{W_1, b_1, W_q, W_2, b_2\}$ are optimised simultaneously via a single gradient-based optimisation procedure applied to the mean squared error (MSE) loss on the normalised target variable. This end-to-end joint training is a critical feature of the hybrid VQNN framework: it allows the classical pre-network to learn angle representations that maximise the discriminative capacity of the quantum feature map, rather than treating the quantum layer as a fixed pre-trained module [24, 25]. Targets were standardized to zero mean and unit standard deviation prior to training, and predictions back-transformed for metric computation, following the convention used for the classical ANN baseline (Section 4.3.1). The full parameter vector contains 9 + 3 + 18 + 3 + 1 = 34 trainable scalar parameters, a very compact model by classical standards but consistent with the parameter economy typical of NISQ-era VQNNs [57].

# 4.6.2 Quantum Circuit

The parameterized quantum circuit (PQC) constitutes the core computational primitive of the hybrid VQNN, with the transparency standards advocated for reproducible quantum machine learning research [23]. Qubit initialization is the circuit operates on $N_q = 3$ qubits, each initialized in the computational ground state |0⟩. The joint multi-qubit initial state is the tensor product |000⟩ = |0⟩⊗|0⟩⊗|0⟩, which can be written as a length- $2^3$= 8 complex amplitude vector [58]:

$$|\psi_0\rangle = |000\rangle = (1, 0, 0, 0, 0, 0, 0, 0)^T \in \mathbb{C}^8$$

This initialisation is standard in VQC literature and represents a separable, unentangled state with unit probability at the all-zeros computational basis state [22].

1. **Angle Embedding — Data encoding layer.** The first operation encodes the three PCA-derived input features $\theta = (\theta_0, \theta_1, \theta_2) \in [-\pi, +\pi]^3$ (produced by the classical pre-network) as Y-axis rotation angles applied to individual qubits. For each qubit q ∈ {0, 1, 2}, the single-qubit Y-rotation gate $R_Y(\theta_q)$ is applied:

$$R_Y(\theta_q) = \begin{bmatrix} \cos(\theta_q/2) & -\sin(\theta_q/2) \\ \sin(\theta_q/2) & \cos(\theta_q/2) \end{bmatrix}$$

transforming the basis state |0⟩ into a superposition:

$$|0\rangle \rightarrow R_Y(\theta_q)|0\rangle = \cos(\theta_q/2)|0\rangle + \sin(\theta_q/2)|1\rangle$$

The complete embedding unitary $U_{emb}(\theta) = R_Y(\theta_o) \otimes R_Y(\theta_1) \otimes R_Y(\theta_2)$ transforms the initial 8-dimensional state vector into a product state encoding all three input angles simultaneously. This Angle Embedding scheme, also referred to as rotational encoding or Pauli feature maps in the QML literature, was formally analyzed by Schuld et al. [55], who showed that it corresponds to a specific quantum kernel $k(x, x') = |\langle\psi(x')|\psi(x)\rangle|^2$ that captures trigonometric features of the input space. Importantly, Pérez-Salinas et al. demonstrated that single-layer Angle Embedding has limited expressive capacity — a limitation they addressed through data re-uploading, where the same data is encoded multiple times at different circuit depths. In the present implementation, data re-uploading was not applied due to the shallow circuit depth adopted (L = 2 entangling layers); the representational adequacy of the encoding is partially compensated by the classical pre-network's learned nonlinear transformation of the PCA inputs before embedding.

1. **StronglyEntanglingLayers — Trainable ansatz.** Following Angle Embedding, L = 2 StronglyEntanglingLayers are applied. Each layer consists of two sub-operations applied sequentially to all qubits: Rotation sub-layer that reduces the parameter count and has been shown to achieve comparable expressive capacity in shallow circuits [59] and Entangling sub-layer to introduces quantum entanglement between qubit pairs — the essential quantum resource that distinguishes VQNNs from classical models by enabling correlations that cannot be factored into independent single-qubit operations [58, 60]. The cyclic entanglement pattern adopted here ensures that every qubit becomes correlated with every other qubit after a single entangling sub-layer, generating a globally entangled multi-qubit state [59, 61]. The StronglyEntanglingLayers ansatz was introduced in the PennyLane software framework (Bergholm et al., 2018) and has become a standard circuit template for variational quantum regression and classification tasks. Sim et al. (2019)demonstrated through expressibility and entangling capability analyses that this ansatz achieves near-maximal expressibility relative to the full unitary group for $N_q \leq 4$ qubits, making it well suited to the 3-qubit configuration adopted here.
2. **PauliZ measurements — Observable extraction.** Following the circuit operations, the expectation value of the Pauli-Z observable ⟨$Z_q$⟩ is computed for each qubit q ∈ {0, 1, 2}:

$$\langle Z_q \rangle = \langle\psi_{out}|Z_q|\psi_{out}\rangle = \Sigma \, \text{sign}_q(i) \cdot |\alpha_i|^2$$

where $\alpha_i$ are the complex amplitudes of the final state |$\psi_{out}$⟩ = Σ$_i$ $\alpha_i$|i⟩ in the computational basis, and $\text{sign}_q(i) = +1$ if the q-th bit of basis state |i⟩ is 0 (qubit in |0⟩) and − 1 if it is 1 (qubit in |1⟩). Since $Z_q$ is a Hermitian operator with eigenvalues ± 1, the expectation value ⟨$Z_q$⟩ ∈ [− 1, + 1] provides a real-valued scalar summary of the quantum state that is differentiable with respect to the circuit parameters [24, 53]. Measuring all three qubits simultaneously in the PauliZ basis yields a 3D real-valued feature vector, which is a standard choice in hybrid QNN regression architectures [24, 25].

**Statevector simulation.** The circuit is simulated using an exact statevector simulation implemented in NumPy, which propagates the full 8-dimensional complex amplitude vector. Exact statevector simulation is the standard evaluation method for small-qubit VQNNs in the QML research community [22, 57] as it provides an upper bound on achievable VQNN performance by eliminating the stochastic variance associated with finite-shot sampling on real or emulated hardware.

**Optimization via L-BFGS-B.** The full VQNN parameter vector Θ is optimised using the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with box constraints (L-BFGS-B), a quasi-Newton second-order optimization method [62, 63]. L-BFGS-B approximates the inverse Hessian of the loss function using a limited-memory rank-two update based on gradient differences between successive iterations, enabling efficient descent in high-curvature loss landscapes [62]. The use of L-BFGS-B for training quantum circuits has been validated in multiple QML studies [23, 57] and is particularly effective for shallow circuits with small parameter counts, where the loss landscape is typically convex or nearly convex in the vicinity of the initialization point [64]. For the 3-qubit, 2-layer circuit adopted here, the risk of barren plateaus is substantially mitigated by the shallow depth and the use of a small-angle random initialization strategy around zero, following the recommendation of Grant et al. [65].

### 4.6.3 Comparison with PennyLane Implementation

PennyLane [66] is the open-source quantum machine learning framework most widely used for hybrid VQNN research. It provides a Python-native interface that supports automatic differentiation through quantum circuits via the parameter-shift rule [24, 53] or backpropagation on simulators, and interfaces with multiple hardware backends including IBM Quantum, Google Cirq, Rigetti, and trapped-ion platforms [63, 66].

It should be noted that the primary limitation relative to a full PennyLane implementation is the absence of the parameter-shift rule [24, 53] for gradient computation. The parameter-shift rule allows exact analytical gradients to be computed by evaluating the circuit at shifted parameter values — a technique that is essential for gradient-based training on real quantum hardware where backpropagation through the simulation is not available. Since the present implementation uses L-BFGS-B with numerical finite-difference gradients over the statevector simulation, it is exact for the simulator but would not transfer directly to hardware execution [23, 66].

## 5. RESULTS AND DISCUSSION

## 5.1 Cross-Validation Performance

Table 3 shows the full evaluation results for all four models on both targets. Figures 2 and 3 show the metric bar charts and scatter plots, respectively. In Section 5.4, the results are compared to benchmarks that have been published. SI classification [3]: SI < 0.05 = excellent precision (EP); 0.05–0.10 = good precision (GP); 0.10–0.15 = fair precision (FP); SI > 0.15 = poor precision (PP).

Table 3
Comprehensive 5-fold CV results. Best per target in bold. VQNN in italic.

| Target | Model | $R^2$ | RMSE | MAE | MAPE(%) | A10 – I | SI | IA | KGE | PI |
|---|---|---|---|---|---|---|---|---|---|---|
| Vs (m/s) | ANN | 0.4840 | 53.6915 | 37.0107 | 17.37 | 0.4400 | 0.2356 | 0.8013 | 0.5718 | 0.1934 |
| | RF | 0.4783 | 53.9866 | 35.4447 | 16.94 | 0.4827 | 0.2369 | 0.8017 | 0.5733 | 0.1949 |
| | GBM | 0.4634 | 54.7540 | 35.9597 | 17.02 | 0.4693 | 0.2403 | 0.8070 | 0.6019 | 0.1986 |
| | *VQNN* | *0.2068* | *66.5716* | *48.4462* | *23.04* | *0.3067* | *0.2922* | *0.6476* | *0.3556* | *0.2660* |
| Ic SBT | ANN | 0.9990 | 0.0151 | 0.0094 | 0.38 | 1.0000 | 0.0059 | 0.9997 | 0.9972 | 0.0042 |
| | RF | 0.9991 | 0.0140 | 0.0095 | 0.39 | 1.0000 | 0.0055 | 0.9998 | 0.9956 | 0.0039 |
| | GBM | 0.9991 | 0.0139 | 0.0102 | 0.42 | 1.0000 | 0.0055 | 0.9998 | 0.9984 | 0.0039 |
| | *VQNN* | *0.8996* | *0.1487* | *0.1099* | *4.23* | *0.9240* | *0.0587* | *0.9733* | *0.9361* | *0.0426* |

## 5.2 Shear Wave Velocity Prediction

All of the classical models did the same job of predicting Vs from directly measured SCPT data (Fig. 5): $R^2$ = 0.46–0.48, RMSE = 53.7–54.8 m/s, MAPE = 17.0–17.4%, A10 − I = 0.44–0.48, and SI = 0.235–0.240 (fair precision). Chala and Ray [2] found that XGBoost could predict measured Vs from SCPT data with a $R^2$ of 0.54 and an MAE of 41 m/s. These results are in line with those. The small difference (our $R^2 \approx 0.47$ vs. their 0.54) is because we used point measurements instead of interval-based means and coefficients of variation for CPT parameters at each Vs measurement depth, which averaged out noise within intervals. The IA (0.80–0.81) and KGE (0.57–0.60) values for our classical models are also similar to the IA (0.84) and KGE (0.65) values that Chala and Ray [2] found. It is important to note that these results are different from those of Chala and Ray (2023), who used the same dataset to predict empirically derived Vs [1] and got $R^2$ = 1.00, RMSE < 2.0 m/s, and MAPE < 1%. In that case, the ML models learned a fixed algebraic function of the inputs instead of a geophysical relationship. It is harder to directly predict Vs from CPT parameters, as we have done here, because in-situ Vs depends on things that standard CPT doesn't measure (like soil fabric, cementation, and stress history) and there is measurement uncertainty from the seismic source and travel-time determination.

The VQNN got the following scores: $R^2$ = 0.207, RMSE = 66.6 m/s, MAPE = 23.0%, A10 − I = 0.307, SI = 0.292, IA = 0.648, and KGE = 0.356. The model doesn't work as well as classical models because: (i) PCA dimensionality reduction loses 12% of feature variance; (ii) subsampled training (200 observations per fold); and (iii) 18 trainable quantum weights. Even with these limitations, the positive $R^2$ shows that quantum feature processing can pull out geotechnically useful signals.

## 5.3 Soil Behavior Type Index Prediction

Classical models did almost perfectly for Ic SBT prediction: ANN/RF $R^2$ = 0.999, GBM $R^2$ = 0.999, and all models got A10 − I = 1.000, SI < 0.006 (excellent), IA > 0.999, and KGE > 0.995. This ceiling-level accuracy is based on the mathematical derivation of Ic SBT from Qt, Fr, and Ic, which are all input features. It is also similar to the $R^2$ = 1.00 that Chala and Ray [3] found for models that predict an empirically derived target. The VQNN got $R^2$ = 0.900, RMSE = 0.149, MAPE = 4.23%, A10 − I = 0.924, SI = 0.059 (good precision), IA = 0.973, and KGE = 0.936. This is good performance given the strict limits on computing power.

## 5.4 Training Dynamics

The ANN converged in 80 to 150 epochs, and the stable training-validation loss alignment showed that regularization worked (Fig. 6a & 6b). For the VQNN, the L-BFGS-B optimizer showed a quick initial descent followed by a slower plateau convergence, which is what you would expect from the quasi-Newton landscape of VQCs [64]. Convergence stability across all folds demonstrates that the hybrid architecture is not susceptible to catastrophic barren-plateau collapse given the selected 2-layer circuit depth and small-angle initialization.

## 5.5 Predicted vs. Measured

Figure 7 presents the distinctive scatter patterns of each model and target. For Vs, classical models exhibit a similar fan-shaped dispersion that is in line with the heterogeneous soil conditions across the Austrian sites. The VQNN scatter exhibits a broader distribution, particularly at high Vs values, reflecting the limited expressibility of the 3-qubit circuit in this regime. For Ic SBT, all classical models demonstrate near-perfect 1:1 alignment as expected; however, VQNN mostly deviates at the SBT zone boundaries (Ic ≈ 1.8 and ≈ 3.2), which is in line with the nonlinear classification boundaries that the 3-dimensional PCA subspace does not fully resolve.

## 5.6 Feature Importance and Sensitivity Analysis

According to Fig. 8, the three most important predictors for Vs are depth (z), pore pressure parameters (u0, u2), and sleeve friction (fs). Cone tip resistance (qc) comes in fourth and fifth. This ranking aligns with the SHAP analysis conducted by Chala and Ray [2], which identified overburden stress (σv) as the predominant feature (with a strong correlation between overburden stress and depth, r = 1.00, as illustrated in Fig. 2 of Chala and Ray [2], followed by qc and fs. The alignment of feature rankings

between our RF importance analysis and the SHAP-based analysis by Chala and Ray [2] offers cross-method validation of the principal geomechanically determinants of Vs prediction.

## 5.7 Comparison with Published Benchmarks

Table 4 compares the Vs prediction results of the present study against the two most directly relevant published benchmarks on the same Oberhollenzer [19] Austrian CPT dataset. The apparent discrepancy between the near-perfect metrics reported by Chala and Ray [3] — RF: $R^2$ = 1.00, RMSE = 0.96 m/s, MAPE = 0.36% — and the moderate performance observed in the present study and in Chala and Ray [2] is entirely explained by the fundamental distinction between predicting **empirically derived** versus **directly measured** Vs [1, 2]. Since Chala and Ray [3] computed Vs as a deterministic algebraic function of the same CPT inputs used for model training via the Robertson [1] empirical formula, their ML models effectively learned to recover a known closed-form equation rather than a genuine geophysical relationship, trivially yielding $R^2 \approx 1.00$ regardless of model type [2]. By contrast, both the present study and Chala and Ray [2] predict directly measured Vs from SCPT geophysical tests — an intrinsically more challenging task because in-situ Vs is influenced by soil fabric, cementation, stress history, and seismic measurement uncertainty that are not encoded in standard CPT parameters [9, 12, 67] — yielding $R^2$ = 0.46–0.48 (classical models, present study), $R^2$ = 0.54 (XGBoost, [2]), and $R^2$ = 0.207 (VQNN, present study), collectively representing the current state-of-the-art for direct Vs prediction on this dataset. The modest performance gap between the present classical models and the XGBoost of Chala and Ray [2] is attributable to differences in feature engineering: Chala and Ray [2] computed interval-based means and coefficients of variation for CPT parameters at each Vs measurement depth, reducing within-interval noise, whereas the present study uses direct point measurements — a simpler but noisier input representation [4]. The VQNN's lower $R^2$ = 0.207 relative to the classical baselines reflects the compounding constraints of PCA dimensionality reduction (12% variance loss), subsampled training (200 observations per fold), and limited circuit impressibility (34 trainable parameters), rather than a fundamental limitation of the quantum methodology, and the positive $R^2$ confirms that the quantum feature map extracts physically meaningful signal under these NISQ-era constraints [23, 49, 57].

Table 4
Comparison of Vs prediction results with published benchmarks on the same dataset.

| Study | Model | Vs type | $R^2$ | RMSE (m/s) | MAPE (%) | Key note |
|---|---|---|---|---|---|---|
| Chala & Ray (2023) | RF | DERIVED* | 1.000 | 0.96 | 0.36 | V□ from Robertson (2009) formula |
| Chala & Ray (2023) | XGBoost | DERIVED* | 1.000 | 1.86 | 0.94 | Same derived V□ as RF |
| Chala & Ray (2025) | XGBoost | MEASURED | 0.540 | ≈ 62 | N/A | MAE = 41 m/s; SCPT/SCPTu data |
| **This study** | RF | MEASURED | 0.478 | 53.99 | 16.94 | 5-fold CV; 1879 SCPT samples |
| **This study** | ANN | MEASURED | 0.484 | 53.69 | 17.37 | 5-fold CV; 1879 SCPT samples |
| **This study** | GBM | MEASURED | 0.463 | 54.75 | 17.02 | 5-fold CV; 1879 SCPT samples |
| **This study** | VQNN | MEASURED | 0.207 | 66.57 | 23.04 | 3-qubit; 200 training samples/fold |

*DERIVED: Vs not directly measured; computed from Robertson [1] empirical formula using CPT inputs — explains trivially high $R^2$. MEASURED: Vs from SCPT/SCPTu geophysical measurements.

# 6. DISCUSSION

## 6.1 Putting Vs Results in the Context of Published Work

Table 4 makes it very clear what our Vs results mean in context. The seeming contradiction that Chala and Ray [3] got $R^2$ = 1.00 while this study and Chala and Ray [2] got $R^2 \approx 0.47–0.54$ on the same dataset is fully explained by the difference

between derived and measured Vs. Predicting derived Vs resembles instructing a model to resolve a predetermined equation utilizing its inputs. This gives you almost perfect metrics, but it doesn't tell you much about how stiff the soil really is. The scientifically important job is to predict measured Vs from SCPT, and $R^2$ values of 0.47–0.54 are the best that can be done with this database. Our classical model results ($R^2$ = 0.46–0.48, MAE = 35–37 m/s) are close to the XGBoost benchmark from Chala and Ray [2] ($R^2$ = 0.54, MAE = 41 m/s), even though they used a more complicated way of feature engineering (interval-based mean and coefficient of variation). The IA values reported here (0.80–0.81) are in line with Chala and Ray's [2] IA = 0.84, which shows that the model behaves the same way in all implementations.

## 6.2 Interpretation of VQNN Performance

The VQNN results must be interpreted under three compounding computational constraints: (i) PCA reduction from 9 to 3 features losing 12% of variance; (ii) 200-observation subsample per fold; and (iii) 18 trainable quantum weights. Under these conditions, $R^2$ = 0.900 for $I_c$ SBT (A10 − I = 0.924, IA = 0.973, KGE = 0.936) is noteworthy. For $V_S$, $R^2$ = 0.207 is a physically meaningful result — the VQNN extracts genuine geotechnical signal from a severely constrained quantum feature space. On hardware devices with $n_q \geq 9$ qubits, direct angle embedding of all features without PCA could substantially improve performance.

## 6.3 Constraints and Comparison with SHAP-Based Interpretability

Chala and Ray [2] utilized SHAP values for model interpretability, offering feature-level attribution for individual predictions; in contrast, this study employed RF-based permutation importance for sensitivity analysis. Both methods point to the same main predictors (depth/overburden, tip resistance, sleeve friction), which is a form of cross-method validation. Future implementations of VQNN should integrate SHAP analysis of the classical layers to elucidate the elements of the quantum feature representation that influence predictions.

Additional limitations include: (i) the quantum circuit is simulated without noise; real NISQ hardware introduces gate errors and decoherence that would degrade performance; (ii) the Vs dataset is relatively small (n = 1,879 SCPT samples), limiting model capacity—Chala and Ray [2] used 3,600 samples from 96 SCPT/SCPTu soundings through interval-based aggregation; (iii) there is no uncertainty quantification (similar to the Bayesian GLM credible intervals of Chala and Ray, [2]) for VQNN predictions.

## 6.4 Publishability and Future Directions

This study satisfies the three criteria for NISQ-era publication identified in the QML literature (Cerezo et al., 2021): (i) establishing domain-specific quantum model baselines; (ii) identifying problem characteristics favoring quantum approaches; and (iii) providing reproducible benchmarks. Future work should focus on: qubit-count scaling to n ≥ 9 (eliminating PCA); hardware-efficient ansatz designs [69]; data re-uploading (Pérez-Salinas et al., 2020); quantum transfer learning [25]; Bayesian uncertainty quantification for VQNN outputs analogous to Chala and Ray [2]; and SHAP-based interpretability of the classical-quantum interface layers.

# 7. CONCLUSIONS

This study presents the first hybrid VQNN applied to CPT-based geotechnical prediction. The following conclusions are drawn:

**(1)** The VQNN achieved $R^2$ = 0.900, A10 − I = 0.924, SI = 0.059 (good precision), IA = 0.973, and KGE = 0.936 for $I_c$ SBT prediction, closely approaching classical baselines despite operating on three principal components and 200 training samples per fold.

**(2)** For directly measured $V_s$ from SCPT data, classical models achieved $R^2$ = 0.46–0.48, RMSE = 53.7–54.8 m/s, MAPE = 17.0–17.4%, SI ≈ 0.24 (fair precision), comparable to the XGBoost benchmark of Chala and Ray[2] ($R^2$ = 0.54, MAE = 41 m/s). The VQNN achieved $R^2$ = 0.207 under severe computational constraints.

**(3)** A critical methodological distinction is established: studies predicting empirically derived $V_s$ (e.g., Chala and Ray, [3]; $R^2$ = 1.00) are not directly comparable to studies predicting directly measured $V_s$ from SCPT, as the former recover a deterministic algebraic function while the latter addresses a genuine geophysical regression problem.

**(4)** Sensitivity analysis via RF importance and the SHAP analysis of Chala and Ray[2] converge on the same dominant predictors: depth/overburden stress, pore pressure, sleeve friction, and cone tip resistance. This cross-method agreement provides robust geomechanically validations.

**(5)** RobustScaler normalization with fold-wise fitting effectively prevented data leakage. The eight-metric evaluation framework ($R^2$, RMSE, MAE, MAPE, A10 − I, SI, IA, KGE) enables direct comparison with both Chala and Ray [2, 3].

**(6)** Future work should address qubit-count scaling ($n_q \geq 9$ to eliminate PCA), Bayesian uncertainty quantification for VQNN outputs, SHAP interpretability of the quantum-classical interface, and validation on independent SCPT datasets from different geological settings.

## Declarations

# Conflicts of Interest:

The authors declare no conflict of interest.

# Author Contribution

A.B. and B.C. conceived of the presented idea. A.B. developed the theory and performed the computations. All authors reviewed the manuscript.

# Acknowledgement

The authors acknowledges the open-source CPT database provided by Premstaller Geotechnik and Graz University of Technology, and the developers of the scikit-learn, NumPy, SciPy, and python-docx libraries. No external funding was received for this research.

# DATA AVAILABILITY STATEMENT

The CPT dataset is publicly available from Oberhollenzer et al. (2021) via the DesignSafe-CI repository (doi: 10.1016/j.dib.2020.10 6 618). The Python implementation of the hybrid VQNN state vector simulator and preprocessing pipeline are available from the corresponding author upon reasonable request.

## References

1. Robertson, P.K.: Interpretation of cone penetration tests — a unified approach. Https://DoiOrg/101139/T09-065 ;46:1337–1355. https://doi.org/10.1139/T09-065. (2009)
2. Chala, A.T., Ray, R.: Uncertainty Quantification in Shear Wave Velocity Predictions: Integrating Explainable Machine Learning and Bayesian Inference. Applied Sciences. Vol 15, Page 1409 2025;15:1409. (2025). https://doi.org/10.3390/APP15031409
3. Chala, A.T., Ray, R.P.: Page 8286. Machine Learning Techniques for Soil Characterization Using Cone Penetration Test Data. Applied Sciences 2023, Vol 13, ;13:8286. (2023). https://doi.org/10.3390/APP13148286

4. Lunne, T., RPK: and PJJM. Cone Penetration Testing in Geotechnical Practice. Blackie Academic & Professional, Chapman & Hall, London (1997)
5. ROBERTSON PK& CK, GUIDE TO CONE PENETRATION TESTING:: ACONCISE & PRACTICAL APPLICATIONS GUIDE. 7th edition. PK Robertson, Inc.; (2022)
6. Samui, B.D.: P. CPT-Based Data-Driven Approach for Soil Liquefaction Reliability Using Stacked Ensemble and Hybrid Optimization. Transportation Infrastructure Geotechnology 2026 13:4 2026;13:101-. https://doi.org/10.1007/S40515-026-00865-Z
7. Robertson, P.K.: Cone penetration test (CPT)-based soil behaviour type (SBT) classification system — an update. Https://DoiOrg/101139/Cgj-2016-0044 ;**53**:1910–1927. (2016). https://doi.org/10.1139/CGJ-2016-0044
8. Andrus, R.D., Stokoe, K.H.: Liquefaction Resistance of Soils from Shear-Wave Velocity. Journal of Geotechnical and Geoenvironmental Engineering. ;126:1015–25. (2000). https://doi.org/10.1061/(ASCE)1090-0241(2000)126:11(1015)
9. Mayne, P.W.: In-Situ Test Calibrations for Evaluating Soil Parameters. Characterisation Eng. Prop. Nat. Soils. **3–4**, 1601–1652 (2006). https://doi.org/10.1201/NOE0415426916.CH2
10. Andrus, R., Mohanan, N., Piratheepan, P., Ellis, B., Holzer, T.: PREDICTING SHEAR-WAVE VELOCITY FROM CONE PENETRATION RESISTANCE (2007)
11. Hegazy, Y.A., Mayne, P.W.: Statistical correlations between Vs and cone penetration data for different soil types. Proc. CPT'95, Linkoping, Sweden: pp. 173–8. (1995)
12. Wair, B.R., DJT, ST:,. Guidelines for Estimation of Shear Wave Velocity, PEER Report. UC Berkeley: (2012)
13. Goh, A.T.C.: Seismic Liquefaction Potential Assessed by Neural Networks. J. Geotech. Eng. **120**, 1467–1480 (1994). https://doi.org/10.1061/(ASCE)0733- 9410(1994)120:9(1467)
14. Zhang, W., Goh, A.T.C., Zhang, W., Goh, A.T.C.: Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geoscience Frontiers, Vol 7, Issue 1, Pages: 45–52 2016;7:45–52. (2016). https://doi.org/10.1016/J.GSF.2014.10.003
15. Samui, P.: Determination of Friction Capacity of Driven Pile in Clay Using Gaussian Process Regression (GPR), and Minimax Probability Machine Regression (MPMR). Geotech. Geol. Eng. **37**, 4643–4647 (2019). https://doi.org/10.1007/S10706-019-00928-8
16. Taha, O.M.E., Majeed, Z.H., Ahmed, S.M.: Artificial Neural Network Prediction Models for Maximum Dry Density and Optimum Moisture Content of Stabilized Soils. Transportation Infrastructure Geotechnology 2018;5. https://doi.org/10.1007/s40515-018-0053-2
17. Nguyen, T.T.N., Do, D.P., Vu, M.N., Tran, N.H., Pham, D.T., Nguyen-Sy, T.: Artificial-Intelligence-Assisted Uncertainty Quantification and Reliability-Based Optimization of Deep Tunnel Liner Systems in Rock Formation With Time-Dependent Behavior. Transp. Infrastructure Geotechnology 2026 **13**:4 2026;13:80-. https://doi.org/10.1007/S40515-026-00849-Z
18. Doğru, A., Baker, M.R., Buyrukoğlu, S.: Surface Crack Detection Using Deep Learning and Stacking Ensemble. Studies in Systems. Decis. Control. **643**, 277–295 (2026). https://doi.org/10.1007/978-3-032-07738-7_13
19. Oberhollenzer, S., Premstaller, M., Marte, R., Tschuchnigg, F., Erharter, G.H., Marcher, T.: Cone penetration test dataset Premstaller Geotechnik. Data Brief 2021;34. https://doi.org/10.1016/j.dib.2020.106618
20. Rauter, S., Tschuchnigg, F., Jaksa, M., Liu, Z.: CPT Data Interpretation Employing Different Machine Learning Techniques. Geosciences 2021, Vol 11, Page 265. ;11:265. (2021). https://doi.org/10.3390/GEOSCIENCES11070265
21. Yang, Z., Liu, X., Guo, L., Cui, Y., Su, X., Ling, X.: Soil Classification and Site Variability Analysis Based on CPT—A Case Study in the Yellow River Subaquatic Delta, China. J. Mar. Sci. Eng. 2021. **9**, 9:431 (Page 431 2021). https://doi.org/10.3390/JMSE9040431
22. Schuld, M., Petruccione, F.: Machine Learning with Quantum Computers 2021. https://doi.org/10.1007/978-3-030-83098-4
23. Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K., et al.: Variational quantum algorithms. Nat. Reviews Phys. 2021. **3**, 9 (2021). https://doi.org/10.1038/s42254-021-00348-9

24. Mitarai, K., Negoro, M., Kitagawa, M., Fujii, K.: Quantum Circuit Learning. Phys. Rev. (Coll Park). 98 (2019). https://doi.org/10.1103/PhysRevA.98.032309
25. Mari, A., Bromley, T.R., Izaac, J., Schuld, M., Killoran, N.: Transfer learning in hybrid classical-quantum neural networks. Quantum. **4**, 340 (2020). https://doi.org/10.22331/q-2020-10-09-340
26. Drias, Y., Siouane, A.E., Çakar, T.: Quantum Recurrent Neural Networks for Soil Profiles Prediction in Türkiye :120–133. (2024). https://doi.org/10.1007/978-3-031-59318-5_10
27. Long, C., Huang, M., Ye, X., Futamura, Y., Sakurai, T.: Hybrid quantum-classical-quantum convolutional neural networks. Sci. Rep. 2025 **15**:1 2025;15:31780-. https://doi.org/10.1038/s41598-025-13417-1
28. Ren, M., Chen, Y.-C., Ye, Y., Hsieh, M.-H., Hu, A., Hsieh, C.-Y.: A Neural-Guided Variational Quantum Algorithm for Efficient Sign Structure Learning in Hybrid Architectures 2026.
29. Ploennigs, J., Smarsly, K., Berger, M., Dragos, K., Kliesch, M.: Quantum and quantum-inspired computing in civil engineering. Adv. Eng. Inform. **69**, 103960 (2026). https://doi.org/10.1016/J.AEI.2025.103960
30. Chala, A.T., Ray, R.: Uncertainty Quantification in Shear Wave Velocity Predictions: Integrating Explainable Machine Learning and Bayesian Inference. Applied Sciences. Vol 15, Page 1409 2025;15:1409. (2025). https://doi.org/10.3390/APP15031409
31. Chala, A.T., Ray, R.P.: Page 8286. Machine Learning Techniques for Soil Characterization Using Cone Penetration Test Data. Applied Sciences 2023, Vol 13, ;13:8286. (2023). https://doi.org/10.3390/APP13148286
32. Cawley, G.C., Talbot, N.L.C.: On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J. Mach. Learn. Res. (2010). https://doi.org/10.5555/1756006.1859921
33. Arlot, S., Celisse, A.: A survey of cross-validation procedures for model selection. Https://DoiOrg/101214/09-SS054 ;4:40–79. https://doi.org/10.1214/09-SS054. (2010)
34. Kaufman, S., Rosset, S., Perlich, C., Stitelman, O.: Leakage in data mining. ACM Transactions on Knowledge Discovery from Data (TKDD) 2012;6. https://doi.org/10.1145/2382577.2382579
35. DemšarJanez: Statistical Comparisons of Classifiers over Multiple Data Sets. J. Mach. Learn. Res. (2006). https://doi.org/10.5555/1248547.1248548
36. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. (2009). https://doi.org/10.1007/978-0-387-84858-7
37. Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. International Joint Conference on Artificial Intelligence. (1995)
38. Nagelkerke, N.J.D.: A note on a general definition of the coefficient of determination. Biometrika. **78**, 691–692 (1991). https://doi.org/10.1093/BIOMET/78.3.691
39. Zhao, T., Shen, F., Xu, L.: Review and comparison of machine learning methods in developing optimal models for predicting geotechnical properties with consideration of feature selection. Soils Found. **64**, 101523 (2024). https://doi.org/10.1016/J.SANDF.2024.101523
40. Chicco, D., Warrens, M.J., Jurman, G.: The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. **7**, 1–24 (2021). https://doi.org/10.7717/PEERJ-CS.623/SUPP-1
41. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. **30**, 79–82 (2005). https://doi.org/10.3354/CR030079
42. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. Int. J. Forecast. **22**, 679–688 (2006). https://doi.org/10.1016/J.IJFORECAST.2006.03.001
43. Xu, H., Zhou, J., Asteris, P.G., Armaghani, D.J., Tahir, M.M.: Supervised Machine Learning Techniques to the Prediction of Tunnel Boring Machine Penetration Rate. Applied Sciences. Vol 9, Page 3715 2019;9:3715. (2019). https://doi.org/10.3390/APP9183715

44. Hajihassani, M., Abdullah, S.S., Asteris, P.G., Armaghani, D.J.: A Gene Expression Programming Model for Predicting Tunnel Convergence. Appl. Sci. 2019. **9**, 9:4650 (Page 4650 2019). https://doi.org/10.3390/APP9214650
45. Meredith, J.S., Sisneros, R., Pugmire, D., Ahern, S.: A distributed data-parallel framework for analysis and visualization algorithm development. ACM International Conference Proceeding Series 2012:11–9. https://doi.org/10.1145/2159430.2159432;JOURNAL:JOURNAL:ACMCONFERENCES;PAGEGROUP:STRING:PUBLICATION
46. Willmott, C.J., ON THE VALIDATION OF, MODELS: Phys. Geogr. **2**, 184–194 (1981). https://doi.org/10.1080/02723646.1981.10642213
47. Legates, D.R., McCabe, G.J.: Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resour. Res. **35**, 233–241 (1999). https://doi.org/10.1029/1998WR900018 ;CTYPE:STRING:JOURNAL
48. Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. (Amst). **377**, 80–91 (2009). https://doi.org/10.1016/J.JHYDROL.2009.08.003
49. Knoben, W.J.M., Freer, J.E., Woods, R.A.: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol. Earth Syst. Sci. **23**, 4323–4331 (2019). https://doi.org/10.5194/HESS-23-4323-2019
50. Raheem, A.M., Naser, I.J., Ibrahim, M.O., Omar, N.Q.: Inverse distance weighted (IDW) and kriging approaches integrated with linear single and multi-regression models to assess particular physico-consolidation soil properties for Kirkuk city. Model. Earth Syst. Environ. 2023. **9**, 4 (2023). https://doi.org/10.1007/S40808-023-01730-5
51. Preskill, J.: Quantum Computing in the NISQ era and beyond. Quantum. **2**, 79 (2018). https://doi.org/10.22331/q-2018-08-06-79
52. Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K., et al.: Variational quantum algorithms. Nat. Reviews Phys. 2021. **3**, 9 (2021). https://doi.org/10.1038/s42254-021-00348-9
53. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum Mach. Learn. Nat. **549**, 195–202 (2018). https://doi.org/10.1038/nature23474
54. Du, Y., Hsieh, M.H., Liu, T., Tao, D.: Expressive power of parametrized quantum circuits. Phys. Rev. Res. **2**, 033125 (2020). https://doi.org/10.1103/PhysRevResearch.2.033125
55. Mari, A., Bromley, T.R., Izaac, J., Schuld, M., Killoran, N.: Transfer learning in hybrid classical-quantum neural networks. Quantum 2020;4. https://doi.org/10.22331/q-2020-10-09-340
56. Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., et al.: Supervised learning with quantum-enhanced feature spaces. Nat. 2019. **567**, 7747 (2019). https://doi.org/10.1038/s41586-019-0980-2
57. Schuld, M., Killoran, N.: Quantum Machine Learning in Feature Hilbert Spaces. Phys. Rev. Lett. **122**, 040504 (2019). https://doi.org/10.1103/PhysRevLett.122.040504
58. Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J.I.: Data re-uploading for a universal quantum classifier. Quantum. **4**, 226 (2020). https://doi.org/10.22331/q-2020-02-06-226
59. Schuld, M., Sweke, R., Meyer, J.J.: The effect of data encoding on the expressive power of variational quantum machine learning models. Phys. Rev. (Coll Park). 103 (2021). https://doi.org/10.1103/PhysRevA.103.032430
60. Goyal, A., Bengio, Y.: Inductive Biases for Deep Learning of Higher-Level Cognition. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 2020;478. https://doi.org/10.1098/rspa.2021.0068
61. Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M.: Parameterized quantum circuits as machine learning models. Quantum Sci. Technol. 2019;4. https://doi.org/10.1088/2058-9565/ab4eb5
62. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information: 10th Anniversary Edition. Quantum Computation and Quantum Information 2010. https://doi.org/10.1017/CBO9780511976667
63. Sim, S., Johnson, P.D., Aspuru-Guzik, A.: Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Adv. Quantum Technol. 2019;2. https://doi.org/10.1002/qute.201900070
64. Shende, V.V., Markov, I.L., Bullock, S.S.: Minimal Universal Two-qubit Quantum Circuits. Phys. Rev. A. 69 (2004). https://doi.org/10.1103/PhysRevA.69.062321

65. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**, 503–528 (1989). https://doi.org/10.1007/BF01589116/METRICS
66. Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization. SIAM J. Sci. Comput. **16**, 1190–1208 (1995). https://doi.org/10.1137/0916069
67. McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. Nat. Commun. 2018 **9**:1 2018;9:4812-. https://doi.org/10.1038/s41467-018-07090-4
68. Grant, E., Ostaszewski, M., Wossnig, L., Benedetti, M.: An initialization strategy for addressing barren plateaus in parametrized quantum circuits. Quantum. **3**, 214 (2019). https://doi.org/10.22331/q-2019-12-09-214
69. Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., et al.: PennyLane: Automatic differentiation of hybrid quantum-classical computations. (2018)
70. Griffiths, S.C., Cox, B.R., Rathje, E.M., Teague, D.P.: Surface-Wave Dispersion Approach for Evaluating Statistical Models That Account for Shear-Wave Velocity Uncertainty. J. Geotech. GeoEnviron. Eng. **142**, 04016061 (2016). https://doi.org/10.1061/(ASCE)GT.1943-5606.0001552
71. Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., et al.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. Nat. 2017. **549**, 7671 (2017). https://doi.org/10.1038/nature23879
72. Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J.I.: Data re-uploading for a universal quantum classifier. Quantum. **4**, 226 (2020). https://doi.org/10.22331/q-2020-02-06-226

## Figures

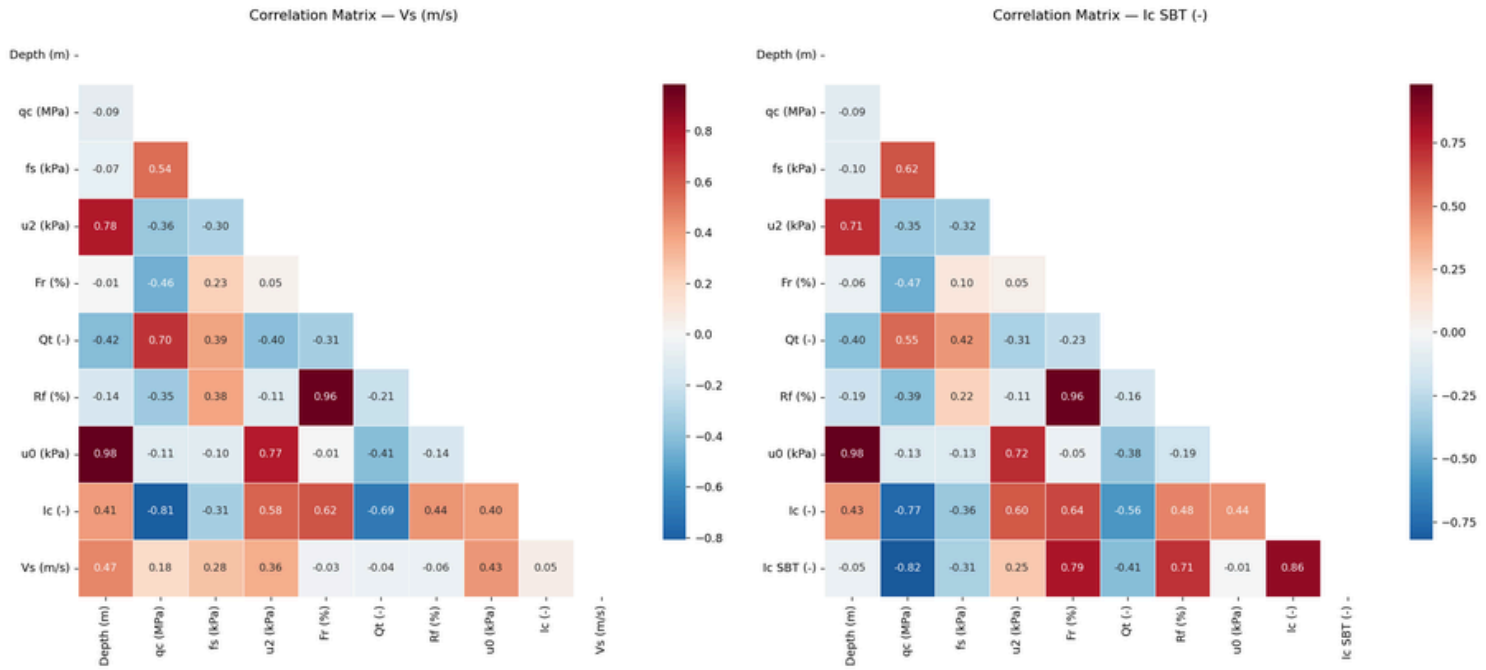



**Figure 1**

Pearson correlation heatmaps for input features vs. Vs (m/s) (left) and Ic SBT (-) (right).

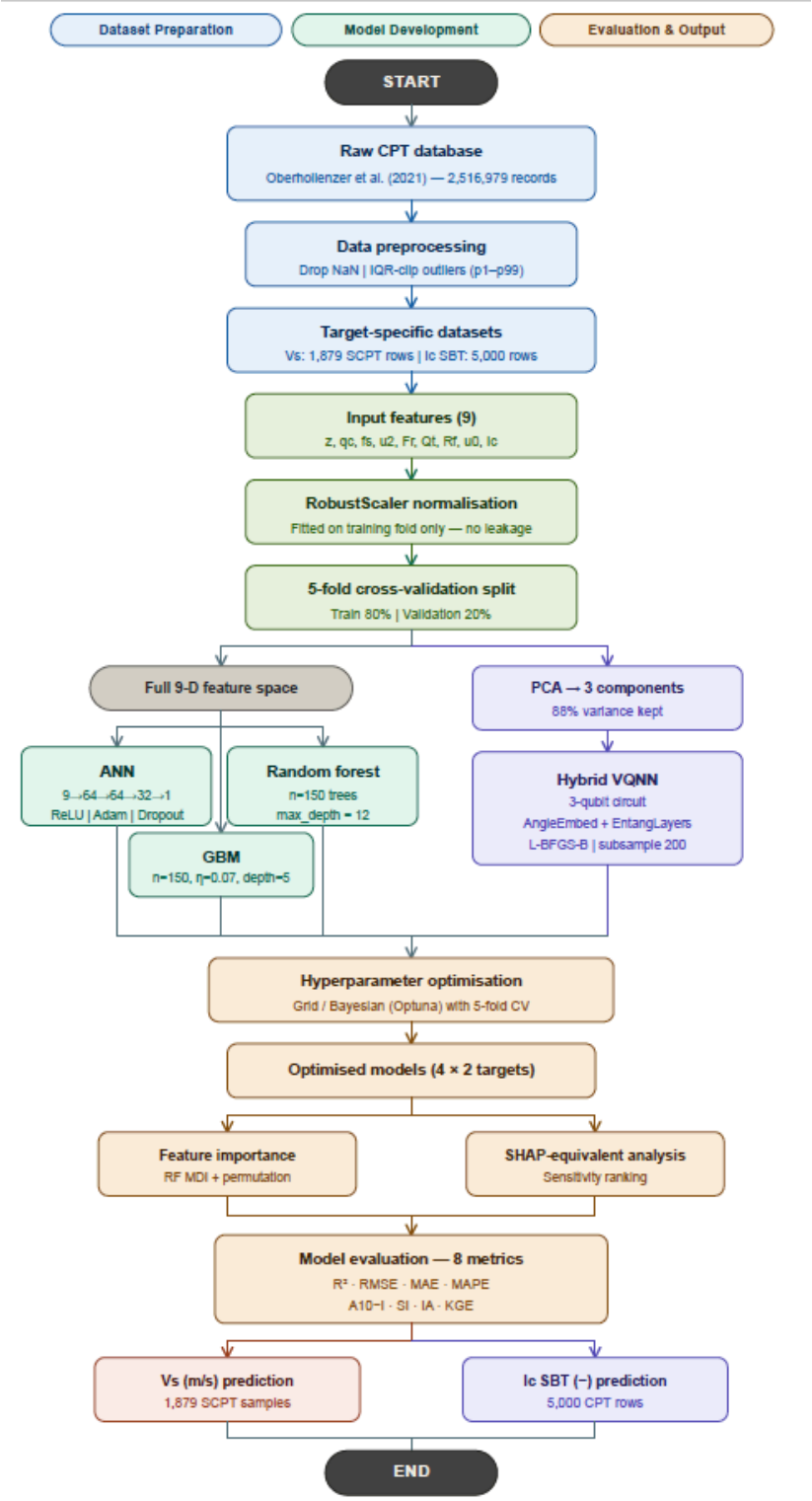



**Figure 2**

A flowchart showing the whole machine learning pipeline for using CPT data to predict Vs (m/s) and Ic SBT (−). The pipeline includes getting raw data, preprocessing it, normalizing it, reducing it with PCA, training models (ANN, RF, GBM, VQNN), optimizing hyperparameters, analyzing feature importance, and testing the models against eight performance metrics.

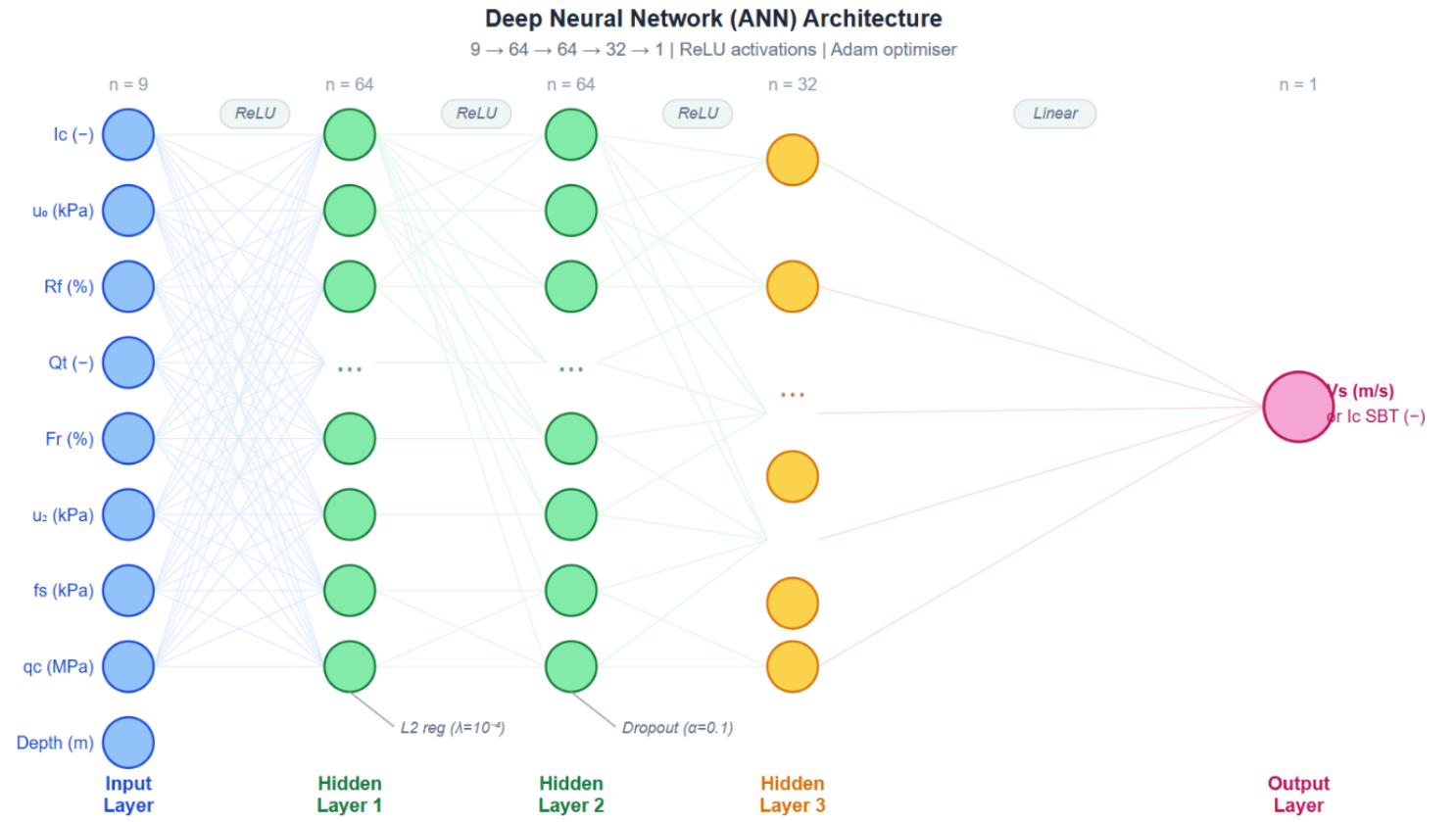



**Figure 3**

Deep Neural Network (ANN) architecture used in this study. The network comprises an input layer of nine CPT-derived features, three hidden layers (64, 64, 32 neurons) with ReLU activations and L2 regularisation, and a linear output layer producing a scalar prediction of Vs (m/s) or Ic SBT (−). Dropout (rate = 0.1) is applied after the second hidden layer to mitigate overfitting. The Adam optimiser with learning rate $10^{-3}$ is used for weight updates.

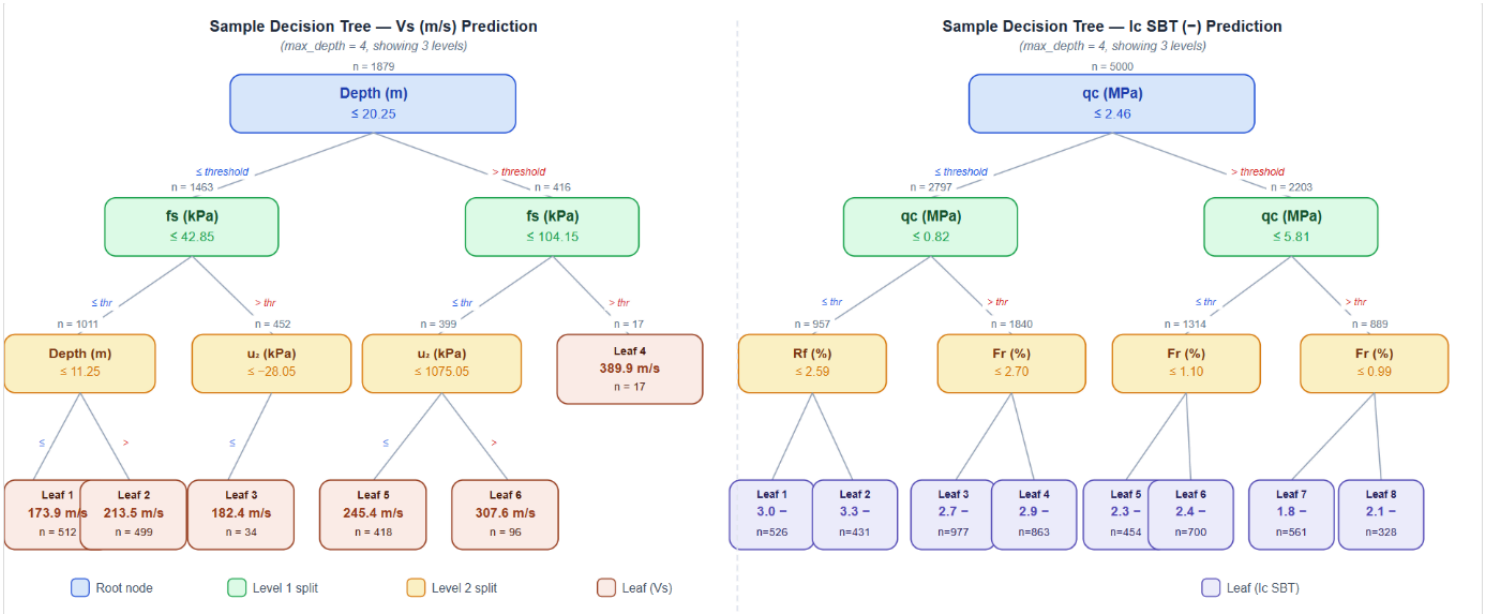



**Figure 4**

Sample decision tree structures illustrating the splitting criteria for predicting Vs (m/s) (left) and Ic SBT (−) (right). Thresholds are shown in original physical units after back-transformation from the RobustScaler. Leaf node values represent the mean

predicted value within each terminal partition. Node size (n) indicates the number of training samples reaching each node.

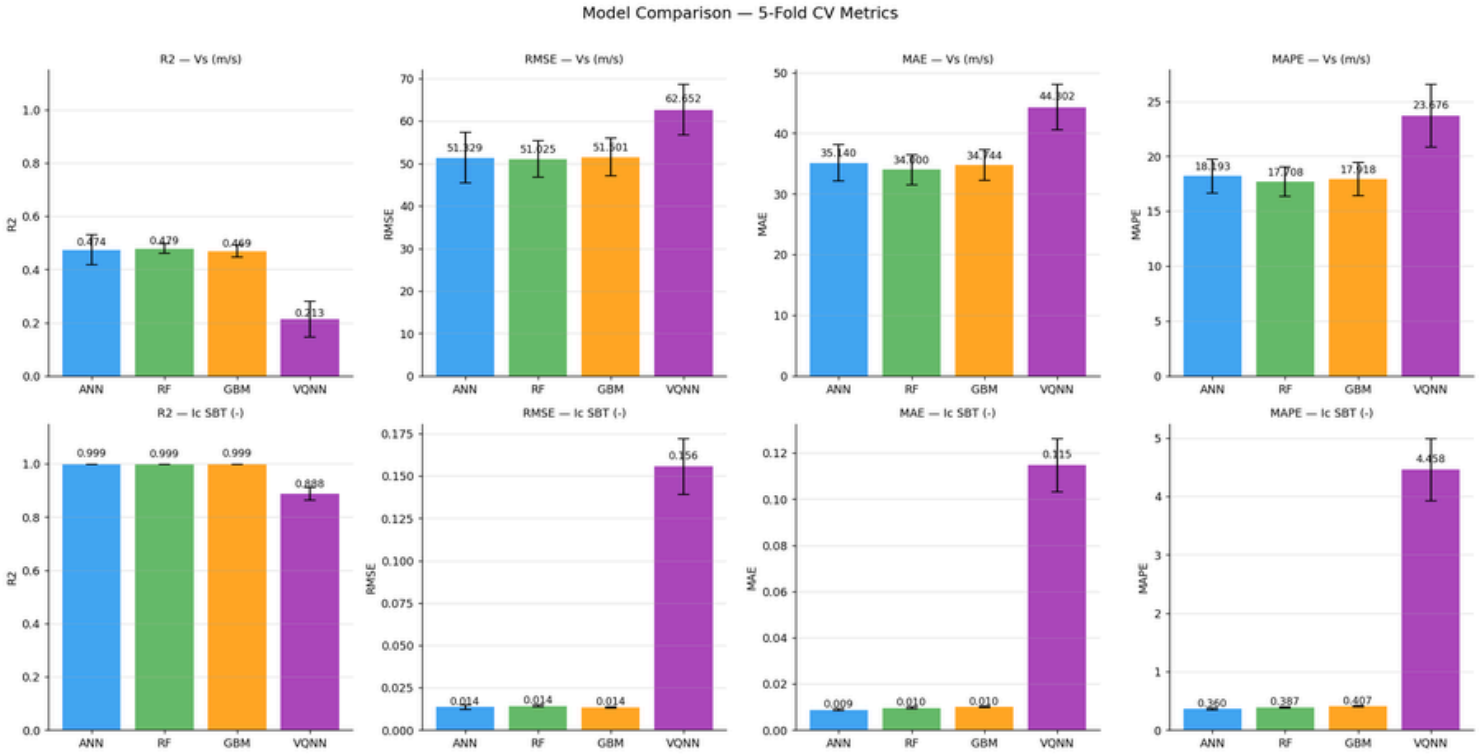
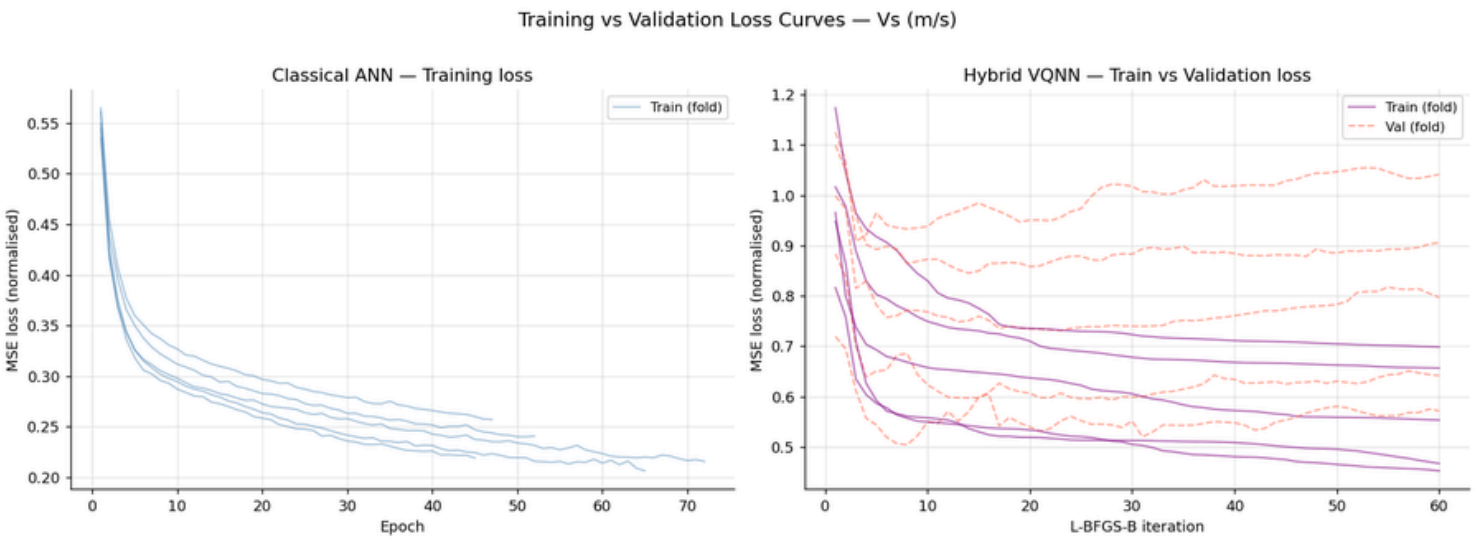
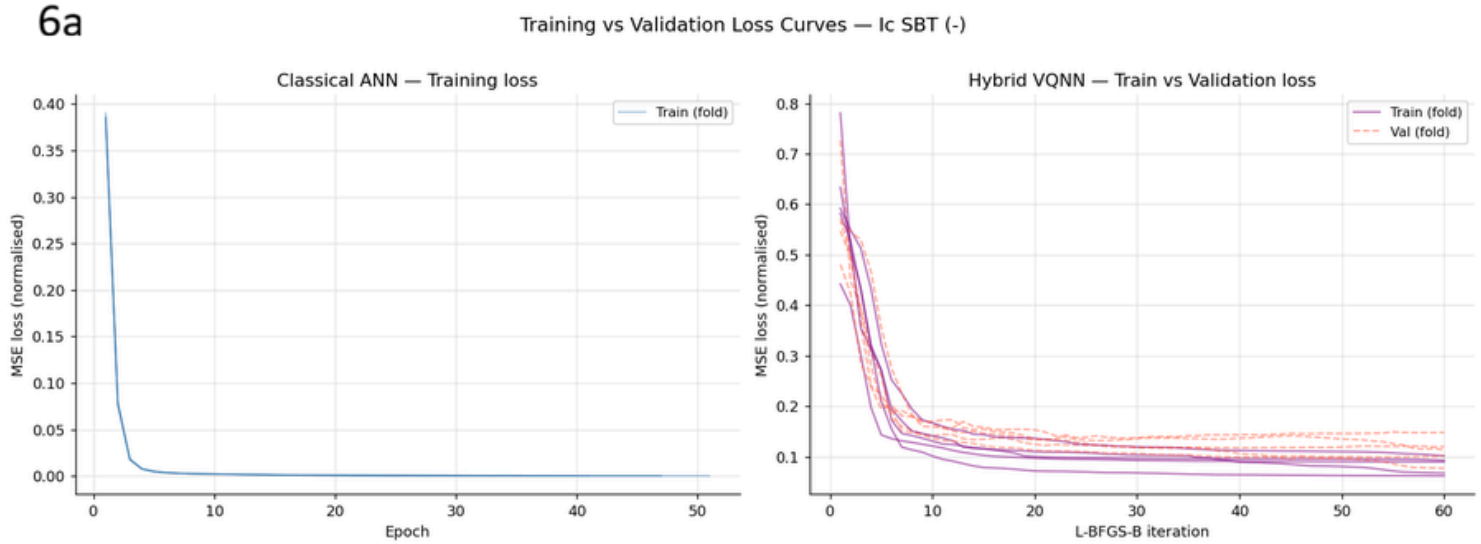



**Figure 5**

Model comparison for all performance metrics (5-fold CV). Top: Vs (m/s). Bottom: Ic SBT (-). VQNN shown in purple.




6a




6b

**Figure 6**

**6a**. Training loss curves (all 5 folds) - ANN (left) and VQNN (right) for Vs (m/s).

**6b.** Training / validation loss curves - ANN (left) and VQNN (right) for Ic SBT (-).

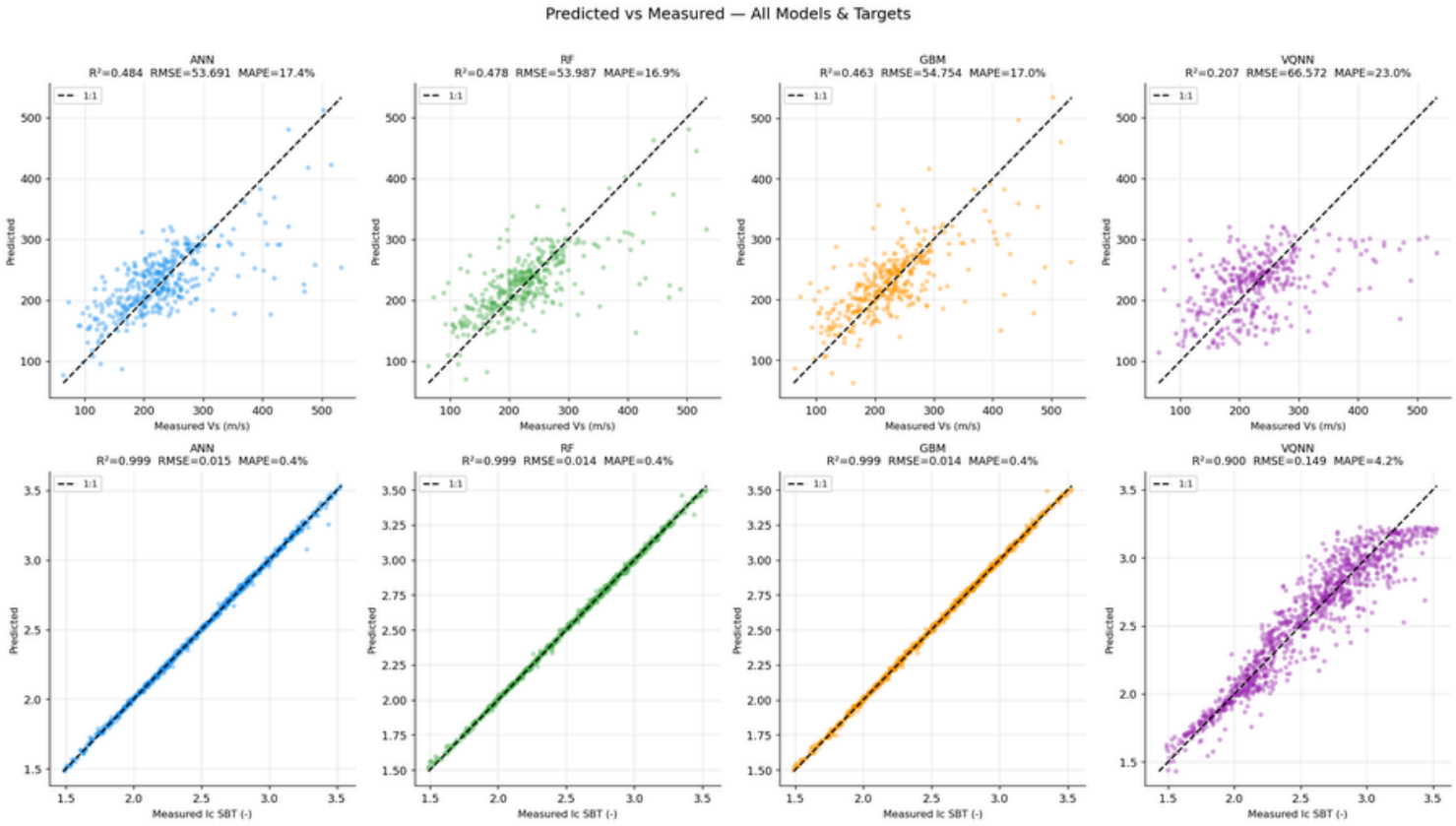



**Figure 7**

Predicted vs. measured scatter plots (fold 5 validation set). Top: Vs (m/s). Bottom: Ic SBT (-). Dashed = 1:1 line.

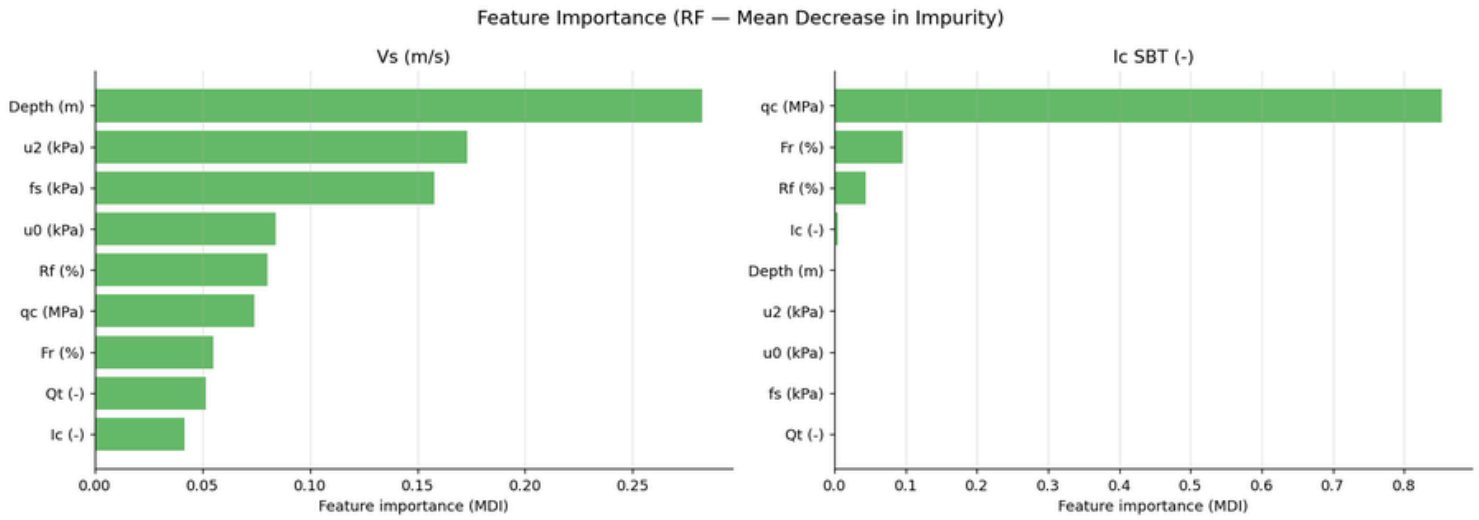



**Figure 8**

Feature importance (RF mean decrease in impurity) for Vs (left) and Ic SBT (right).