Marine Geophysical Research (2026) 47:5
https://doi.org/10.1007/s11001-025-09602-6

ORIGINAL RESEARCH PAPER

Interpretable XGBoost-based predictions of shear wave velocity from CPTu data

Héctor Marín-Moreno[1] · James Willis[1] · Yuting Zhang[2] · Susan Gourvenec[2]

Received: 11 August 2025 / Accepted: 1 December 2025

Abstract

Accurate estimation of shear wave velocity (V_s) is critical for offshore geotechnical design, yet direct measurements remain sparse due to cost and logistical constraints. Empirical correlations of V_s from cone penetration test data are derived for specific conditions, thus introducing uncertainty when applied more generally. Machine learning (ML) based correlations have become popular, yet to date have prioritised accuracy over interpretability. To address this gap and enhance transparency, this study integrates the computationally efficient XGBoost technique with SHapley Additive exPlanations (SHAP) to predict V_s and attribute prediction contributions to individual input features. A combined open-source dataset of 7485 paired cone penetration test data with pore pressure measurement (CPTu) and V_s measurements was integrated and used for training and validation. SHAP analysis on a testing dataset of 1526 samples shows that depth, corrected cone resistance (q_t), and sleeve friction (f_s) are the most influential features, with depth increasing in importance when the V_s predicted deviates from the mean V_s in the training dataset. Compared to a widely used empirical correlation, the ML approach demonstrated superior accuracy across most cases, while also offering insight into decision-making logic. This study highlights the value of interpretable ML in offshore site investigations, with this specific CPTu-V_s interpretable ML model particularly relevant to bottom-fixed foundation designs for offshore wind developments, which are governed by stiffness criteria, in V_s data-limited projects.

Keywords Seismic and standard piezocone tests · Shear wave velocity · Machine learning interpretability · XGBoost · Offshore wind

Introduction

Overview

The rapid expansion of offshore wind infrastructure and other marine-based developments has intensified the demand for efficient, scalable, and cost-effective site investigation techniques (Fischer et al. 2019; Gourvenec 2024). Current offshore investigation campaigns typically involve a combination of geophysical seismic reflection data and in-situ geotechnical tests, mainly borehole sampling and cone penetration testing with pore pressure measurement (CPTu). CPTu provides continuous depth profiles of several key parameters, including cone tip resistance (q_t), sleeve friction (f_s), and pore water pressure (u_2). Direct shear wave velocity (V_s) measurements, such as those obtained through seismic (S)CPTu, are only performed at a small subset of CPTu locations. However, in geotechnical engineering design of offshore wind foundations, particularly of bottom-fixed solutions, V_s is directly related to the soil's small-strain stiffness, which is essential for the analysis of foundation response, cyclic loading, and dynamic behaviour (Trafford et al. 2022).

The sparse data coverage of V_s reflects the logistical and overall financial constraints associated with SCPTu tests (e.g. Masters et al. 2024), particularly over large offshore windfarm zones where hundreds of potential foundation locations need characterizing. As a result, it is common practice to estimate V_s indirectly from CPTu data using empirical correlations (e.g., L'Heureux and Long 2017;

✉ Héctor Marín-Moreno
h.marin-moreno@soton.ac.uk

[1] School of Ocean and Earth Science, University of Southampton, Southampton, UK

[2] School of Engineering, University of Southampton, Southampton, UK

Published online: 20 December 2025

Wair et al. 2012). While empirical correlations are attractive for their simplicity and low cost, they have notable limitations. Most have been developed for onshore sites and are calibrated on relatively narrow datasets that do not necessarily capture the diversity of soil types, geological histories, or stress conditions present in marine environments. These correlations often assume a specific soil behaviour type or fixed stress state, which restricts their applicability and introduces considerable uncertainty in cases that deviate from calibration conditions (Chala and Ray 2023a; Entezari et al. 2022; Stuyts et al. 2024). Additionally, with offshore developments increasingly targeting deeper waters and more complex stratigraphy, reliance on generic empirical correlations is becoming increasingly problematic (Hu and Jeng 2025). Advances in data-driven modelling and increased access to publicly available offshore soil investigation data have created new opportunities to extract more information from CPTu data than empirical correlations allow. This has prompted a growing interest in the scientific community in applying machine learning (ML) methods to enhance prediction accuracy for key soil parameters such as V_s, especially when large volumes of in-situ CPTu data are available but direct measurements of V_s are scarce or incomplete (Entezari et al. 2022; Stuyts and Suryasentana 2023; Jin et al. 2025; Chala and Ray 2023a).

The role of V_s in geotechnical design

Shear wave velocity is directly related to the soil's small-strain stiffness G_{max} (Eq. 1), which governs how soils respond to cyclic and dynamic loads.

$$G_{\max} = \rho {V_s}^2 \tag{1}$$

where ρ is the bulk density of the soil.

This relationship holds under small strain levels (typically < 0.001%), which are relevant for many serviceability limit state conditions in geotechnical and offshore engineering (Zorzi et al. 2019). These include operational deflections, dynamic amplification and resonance avoidance. These aspects are especially important for infrastructure like offshore wind turbines which are affected by millions of wave and wind loading cycles over their operational life (Damgaard 2011).

Underestimating shear stiffness can lead to overly conservative designs and increased material costs, whereas overestimating it can result in a structure's natural frequency matching wave or wind loading frequencies, increasing the risk of resonance (Abdelghani et al. 2024). Discrepancies between designed and observed structural behaviour (e.g., natural frequency) in several offshore wind projects have been related to inaccurate foundation stiffness assumptions (Hucker et al. 2019; Abdelghani et al. 2024). In response, geotechnical practice should embrace ML-based approaches which are cost and time efficient to estimate $G_{\max}$ and its degradation with strain (e.g. Gourvenec 2024; Charles et al. 2023).

While direct measurements of V_s through SCPTu or downhole geophysics are available, they are costly and sparse, and laboratory experiments, such as those using bender elements, often underestimate field stiffness by 30–50% due to sample disturbance and scale effects (Gomez and Stuyts 2022). Consequently, engineers often rely on empirical correlations between CPTu cone resistance (q_t) and either V_s or $G_{\max}$. However, such correlations can exhibit large prediction errors and are sensitive to factors including effective stress, void ratio, overconsolidation, and soil type, with accuracy tending to break down in layered, cemented, or unusual depositional environments (Stuyts et al. 2024). Therefore, as offshore developments expand into increasingly complex geological settings, particularly in the North Sea and Atlantic margins, there is a clear need to enhance our estimation of V_s from CPTu data.

Why empirical approaches fall short

Many empirical approaches for estimating V_s from CPTu data are based on simplified fixed-form regression models that relate V_s to q_t, f_s, and additional derived CPTu parameters such as the soil behaviour type index (I_c). These models are widely adopted in practice due to their simplicity, accessibility, and small data requirements. However, such correlations are typically calibrated on generalised onshore soil behaviour assumptions. They often presume homogeneity, isotropy, and predictable drainage conditions that are frequently violated in offshore and glaciomarine environments, such as the North Sea (Stuyts et al. 2024). Importantly, the issue lies not with the input parameters themselves but with the static nature of these regression-based models. Such models cannot easily adapt to local variations or account for more complex interactions between variables. Studies have reported that while empirical models may perform reasonably in clean sands (±30%), errors can exceed ±100% in sensitive clays, silty units, or stratified profiles (Entezari et al. 2022). These large errors in sensitive clays and silty units may partly result from the partly drained or undrained behaviour of these soils, which makes the correlation between CPTu data and V_s more complicated and dependent on the CPTu penetration rate.

These challenges have prompted a shift in focus towards approaches that can move beyond fixed empirical assumptions. ML offers a different modelling framework, one that can learn complex, nonlinear relationships directly from data without requiring pre-defined equations or site-specific calibration. Rather than assuming a fixed form, ML techniques optimise their structure based on the available data, making them well-suited for applications where soil behaviour is highly variable (Khawaja et al. 2024). For this reason, ML presents a compelling alternative for improving V_s predictions from CPTu data in offshore settings where traditional empirical correlations are insufficient.

Similar uses of machine learning

A growing number of studies have explored the use of ML techniques to predict V_s from cone penetration test data, particularly CPTu. A variety of ML techniques have been trialled in this context, ranging from ensemble models like Random Forest and Gradient Boosting to kernel-based methods such as Support Vector Machine, and Artificial Neural Network, including deep learning architectures (i.e. DNN) (Entezari et al. 2022 and 2024; Chala and Ray 2023a, 2023b; Marin-Moreno et al. 2024).

However, no single technique has yet emerged as universally superior (Entezari et al. 2024). Instead, suitability of a technique tends to be highly context-dependent and shaped by the volume and quality of available training data, the geological variability of the sites, and the balance needed between predictive performance and model transparency (Koldasbayeva et al. 2024). Ensemble-based tree models, such as Random Forest and Extreme Gradient Boosting (XGBoost), have gained traction for their capabilities in capturing complex, nonlinear relationships while maintaining relatively low sensitivity to noise and multicollinearity (Chala and Ray 2023a, 2023b). Compared to simple regression models or more opaque deep learning architectures, these ensemble methods strike a trade-off between accuracy and interpretability (Zhang et al. 2023).

Interpretability remains a broader challenge across many high-performing ML algorithms. For instance, neural networks offer exceptional flexibility in modelling nonlinear patterns and have shown strong results in studies involving diverse soil types (e.g. Marin-Moreno et al. 2024). Yet their 'black box' nature, characterised by dense, abstract internal representations, makes them harder to interpret in practice. By attributing prediction contributions to individual input features, explainable artificial intelligence techniques such as SHapley Additive exPlanations (SHAP), enable a transparent interrogation of the model's internal logic, an important asset in geotechnical contexts where model outputs may influence design or safety-related decisions (Chala and Ray 2025). While these techniques can be applied to DNNs, they are often more computationally intensive than tree-based models like XGBoost or Random Forest (Hamilton and Papadopoulous 2023). The XGBoost technique offers a good balance between predictive accuracy and ability to assess complex feature interactions (Chen et al. 2024) and importance, e.g. by quantifying how frequently each feature contributes to decision trees splits, and its computationally cheaper than Random Forest (Bentejac et al. 2021), which allows for quicker training. This makes it particularly suitable for computability with robust post hoc interpretability frameworks like SHAP.

Despite advances in ML techniques, the quality and preprocessing of input data remain foundational to their application for soil property characterisation. The presence of outliers (defined as datapoints that lie outside from most of the data space), missing features, training data leakage, or resolution mismatches between V_s measurements and CPTu data, can significantly degrade model reliability if not properly addressed. It has been consistently observed that the best-performing models are those built on well-curated, representative datasets, highlighting the ongoing importance of data cleaning, feature engineering, and thoughtful model evaluation (Entezari et al. 2022; Chala and Ray 2023a, 2023b). Furthermore, because V_s behaviour is inherently influenced by site-specific factors like sediment age, overconsolidation, and mineralogy, even the most sophisticated ML techniques can require site-specific recalibration before being applied to a different geological setting (Entezari et al. 2022). Even with robust data pre-processing and validation, ML techniques are influenced by both aleatoric and epistemic uncertainty. In our application, aleatoric uncertainty mainly arises from the potential measurement error in CPTu and V_s data (caused by factors such as geophone offsets, deviations of the wave propagation path from the assumed straight trajectory, and the traditional soil isotropy assumption), which cannot be resolved even with larger training datasets or more advanced techniques (Thompson et al. 2021). In contrast, epistemic uncertainty is driven by the ML model's limited exposure to certain conditions such as underrepresented soil types and/or V_s ranges. Both types of uncertainty reinforce that perfect predictions are not achievable (Smith et al. 2024), and that some level of error and uncertainty are to be expected.

Collectively, these avenues emphasise a transition from 'black box' ML modelling toward interpretable, uncertainty-aware, and generalisable ML systems suited for geotechnical applications (e.g. Phoon et al. 2022).

Motivation and aim of this study

Most published work on ML-based techniques for in-situ soil parameter characterisation has focused on accuracy. However, to build trust in these techniques and increase their adoption in engineering practice interpretability is essential. To address this, here we investigate the combined application of the XGBoost and Shapely Additive exPlanations

(SHAP) techniques for CPTu-derived prediction and interpretation of V_s.

Methods

Dataset description and preprocessing

Three seismic CPTu datasets were utilised in this study, sourced from separate offshore and onshore field campaigns. These included pre-processed, curated offshore CPTu profiles with push speeds of 2 cm/s from the Dutch sector of the North Sea (Stuyts 2024) and from the German sector of the North Sea (BSH 2024), and onshore CPTu profiles from Austria and Germany (Oberhollenzer et al. 2021). From all three data sources, the curated at source input parameters depth, cone penetration resistance (q_t), sleeve friction (f_s), shoulder pore pressure (u_2), and shear wave velocity (V_s) were extracted. The Stuyts dataset differs from the others in that CPTu inputs were pre-processed and averaged over 1 m intervals at source to align with the sampling rate of V_s measurements, whereas the other datasets used a point-to-point pairing with depth (with a tolerance of ± 4 cm).

The datasets were then filtered to remove any physically invalid measurements, such as negative q_t and f_s measurements, and potentially unrealistic CPTu measurements at depths less than 1 m. Although ISO22476-1:2022 (ISO 2022) recommends discarding the upper 5 m of data, generally this is a conservative threshold. In practice, data from depths below 1–2 m are often considered reliable and here it was decided the amount of training data to be maximized. After this filtering, the datasets were scanned for missing values, and any row with incomplete observations was removed. To enable further geotechnical interpretation and to support a broader feature selection for the XGBoost technique, the normalized CPTu parameters (Q_t, F_r, and B_q) and soil behaviour type index (I_c; e.g. Robertson 2009) were also included (Eqs. 2-5) as input features.

Table 1 Soil types in the training and validation dataset and in the testing dataset. Note that soil types with Vs above 600 m/s are grouped here as "Other" and primarily include non-siliciclastic materials such as chalk from the BSH (2024) dataset (see Ramboll, 2021)

Soil type	Soil behaviour index (I_c)	Frequency	Proportion (%)
Training and validation dataset			
Gravelly to Dense Sand	$I_c < 1.31$	565	7.55
Clean to Silty Sand	$1.31 < I_c < 2.05$	3600	48.10
Silty Sand to Sandy Silt	$2.05 < I_c < 2.60$	1537	20.53
Clayey Silt to Silty Clay	$2.60 < I_c < 2.95$	686	9.16
Silty Clay to Clay	$2.95 < I_c < 3.60$	919	12.28
Organic Soils—Clay	$I_c > 3.60$	62	0.83
Other	Not applicable	116	1.55
Testing dataset			
Gravelly to Dense Sand	$I_c < 1.31$	23	1.51
Clean to Silty Sand	$1.31 < I_c < 2.05$	1062	69.59
Silty Sand to Sandy Silt	$2.05 < I_c < 2.60$	86	18.74
Clayey Silt to Silty Clay	$2.60 < I_c < 2.95$	107	7.01
Silty Clay to Clay	$2.95 < I_c < 3.60$	48	3.15
Organic Soils—Clay	$I_c > 3.60$	0	0.00
Other	Not applicable	0	0.00

$$Q_t = \frac{q_t - \gamma \cdot z}{z(\gamma - \gamma_w)} \tag{2}$$

$$F_r = \frac{f_s}{q_t - \gamma \cdot z} \times 100 \tag{3}$$

$$B_q = \frac{u_2 - \gamma_w \cdot z}{q_t - \gamma \cdot z} \tag{4}$$

where z is the depth in meters, γ is the assumed unit weight of soil (18 kN/m^3), and γ_w is the unit weight of water (9.8 kN/m^3).

$$I_c = \left[(3.47 - \log Q_t)^2 + (\log F_r + 1.22)^2\right]^{0.5} \tag{5}$$

Two additional derived features were included: ratio of q_t to f_s (q_t/f_s) and the product of B_q and F_r ($B_q * F_r$). The engineered feature $B_q * F_r$ was introduced to enhance the model's ability to capture potential non-linear relationships, even though it lacks a direct physical basis, as it combines two parameters that are otherwise unconnected in the existing feature set. The final training and validation dataset consisted of 7485 samples of paired CPTu and V_s parameters (Table 1; dataset provided as an online resource in the supporting information). This dataset was randomly split into two subsets with 95% of the samples used for model training and 5% withheld, i.e. not used during model fitting or feature selection, as a validation set to evaluate the learning process. The random split ensured that the distribution of soil types was broadly preserved in both subsets. Relationships between all the parameters in the dataset is presented in Fig. 1.

To assess model performance, a separate testing set comprising 45 SCPTu profiles from the Dutch sector of the North Sea containing 1526 paired V_s and CPTu measurements (RVO 2023) was used (Table 1; dataset provided as an online resource in the supporting information). Although sourced independently from the processed training dataset, some overlapping CPTu sites with the testing dataset were identified which may have resulted in some training data leakage. As such, while the test data consists of previously unseen measurement intervals and unaveraged CPTu inputs it does not strictly constitute a completely site-independent

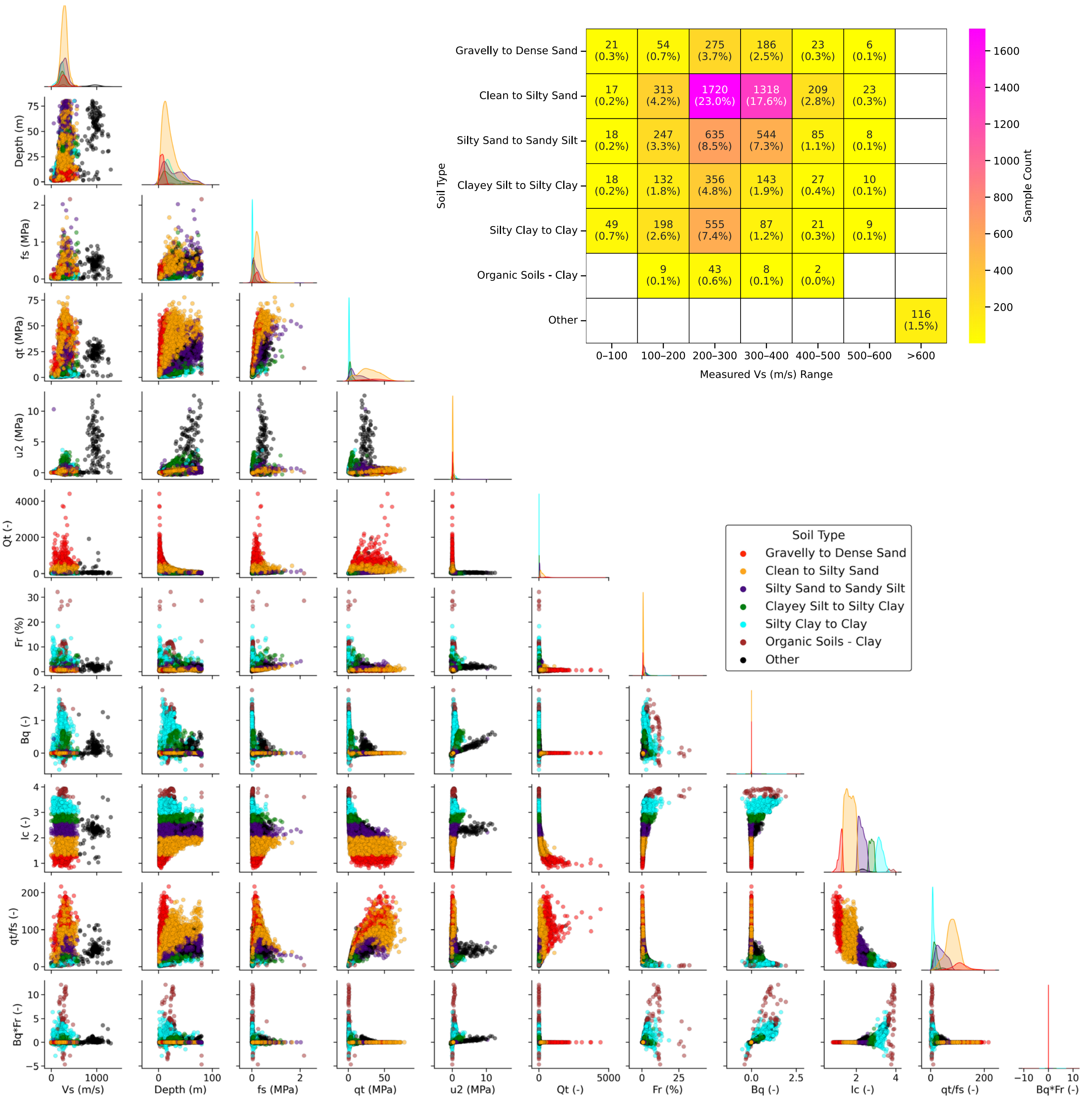

Fig. 1 Pairwise scatter plots and density distribution (main diagonal) of input features in the training and evaluation datasets coloured by soil type. The inset shows a heat map of frequency and corresponding percentages of data points by soil type and V_s range for the testing dataset

evaluation. For each CPTu profile in the testing set, the same input features as those used during training were included. As with the training dataset, the depths of V_s and CPTu measurements in the testing dataset may not exactly match up. Therefore, the depth of each measured V_s value was matched to the predicted V_s from the closest CPTu measurement depth using a nearest-neighbour search implemented with a k-dimensional tree algorithm. Since CPTu measurements are recorded every 0.02 m and seismic measurements every 1 m, a negligible error results from this approximation.

Model development and training

Model framework

The predictive framework that was adopted in this project (Fig. 2) uses XGBoost, an ensemble learning algorithm that

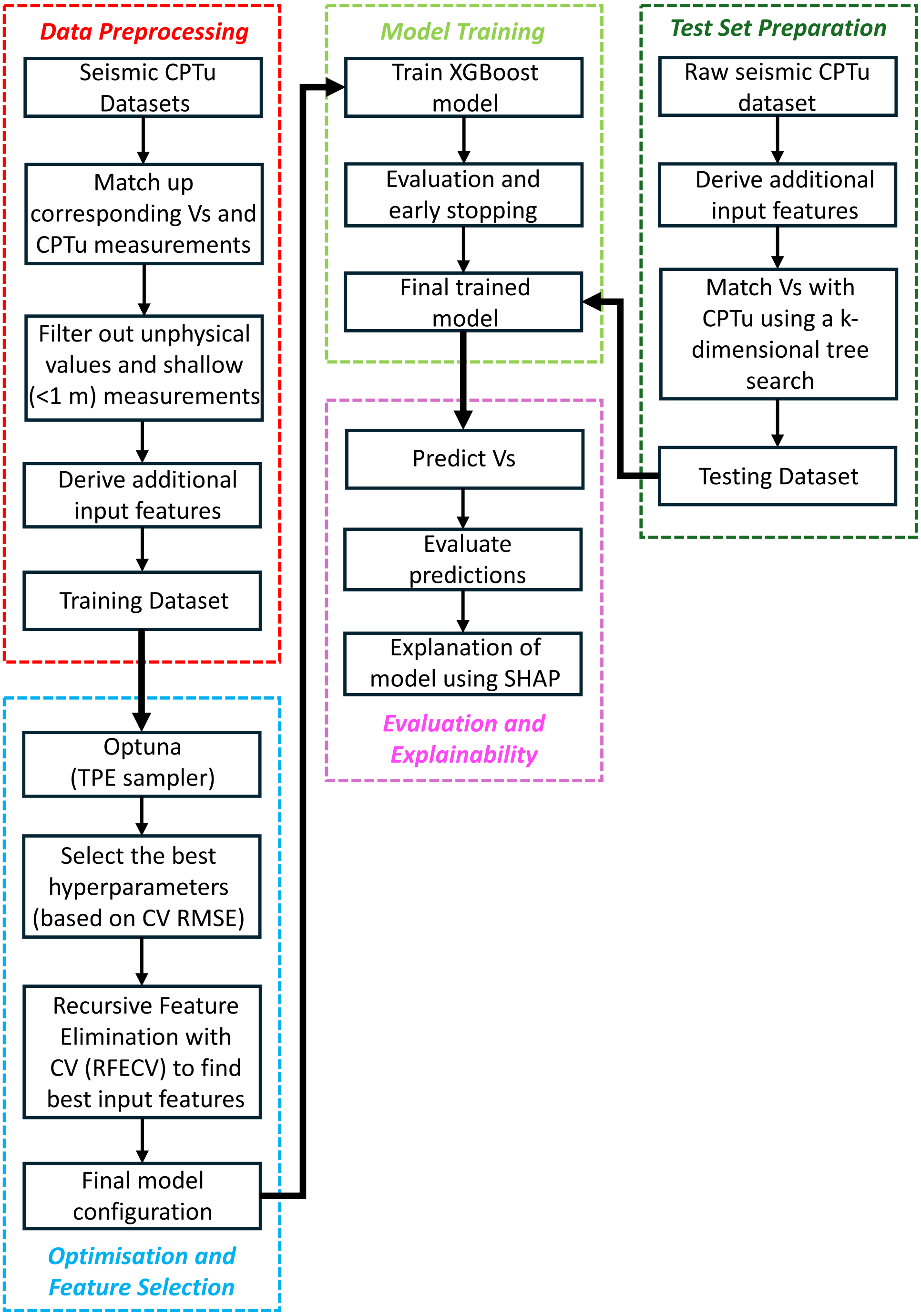
Data Preprocessing
Seismic CPTu Datasets
Match up corresponding Vs and CPTu measurements
Filter out unphysical values and shallow (<1 m) measurements
Derive additional input features
Training Dataset
Optuna (TPE sampler)
Select the best hyperparameters (based on CV RMSE)
Recursive Feature Elimination with CV (RFECV) to find best input features
Final model configuration
Optimisation and Feature Selection
Model Training
Train XGBoost model
Evaluation and early stopping
Final trained model
Predict Vs
Evaluate predictions
Explanation of model using SHAP
Evaluation and Explainability
Test Set Preparation
Raw seismic CPTu dataset
Derive additional input features
Match Vs with CPTu using a k-dimensional tree search
Testing Dataset

Fig. 2 Schematic flow diagram of the methodology behind developing, training and evaluating the performance of the XGBoost model

constructs a sequence of decision trees to perform supervised regression (Chen and Guestrin 2016). The training objective was to minimise the discrepancy between predicted and true V_s values, while also regularising model complexity to prevent overfitting. Formally, XGBoost aims to minimise a regularised objective function:

$$\mathcal{L} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where y_i is the true V_s value for the i-th observation, $\hat{y}_i$ is the model prediction, and $\Omega(f_k)$ is a regularisation term that penalises the complexity of each tree f_k. In this project, the loss function $l(y_i, \hat{y}_i)$ was defined to be the squared error between true and predicted V_s value:

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \tag{7}$$

The prediction for each observation was computed as the sum of outputs from K regression trees:

$$\widehat{y}_i = \sum_{k=1}^{K} f_k\left(X_i\right), \quad f_k \epsilon \mathcal{F} \tag{8}$$

where X_i is the feature vector for the i-th sample and $\mathcal{F}$ is the space of all possible regression trees. The model is trained sequentially, with each tree f_k fitted to the residual errors of the current ensemble, progressively reducing the overall loss. The regularisation term encourages simpler trees and helps prevent overfitting.

Pre-training optimisation and feature selection

Prior to final model training, a sequence of optimisation and feature selection steps were applied to configure and better constrain the model. Hyperparameter optimisation was conducted using the Optuna optimisation framework (Akiba et al. 2019) with a Tree-structured Parzen Estimator (TPE) sampler, a Bayesian optimization algorithm that approximates the distributions of the prior with non-parametric densities (Bergstra et al. 2011), used to efficiently explore a predefined range of hyperparameters over 100 trials (Fig. 3a). The Optuna objective function minimised the root mean squared error (RMSE) between the predicted and true V_s values using five-fold cross-validation (CV) on the training dataset. In each trial, the training data were partitioned into five folds, with the model trained on four folds and validated on the fifth. This process was repeated five times and the mean RMSE across all folds was returned to Optuna to determine the best hyperparameters.

Following hyperparameter optimisation, recursive feature elimination with cross-validation (RFECV) was performed to assess input dimensionality and remove input features that did not contribute to prediction performance. RFECV functions by training a model using the tuned hyperparameters, ranks input features based on importance, and then removes the least important features before retraining and re-evaluating performance. At each iteration, five-fold CV was again used to compute an average validation RMSE, and the process continued until removing further features degraded model performance. The subset of input features that yielded the lowest mean RMSE was then selected, and this reduced feature set was then used in all subsequent training, validation, and testing phases.

Final model training

The final XGBoost model was trained using the optimal hyperparameters and the set of input features identified by RFECV. As mentioned, model fitting was performed using 95% of the dataset, with the remaining 5% withheld as a validation set. The purpose of this validation set was to track generalisation performance during training and support early stopping, which was applied with a patience of 20 rounds. Early stopping terminated training once the validation RMSE failed to improve over 20 consecutive boosting rounds, thereby preventing overfitting and reducing any unnecessary model complexity. The final optimised XGBoost model pipeline is provided as an online resource in the supporting information.

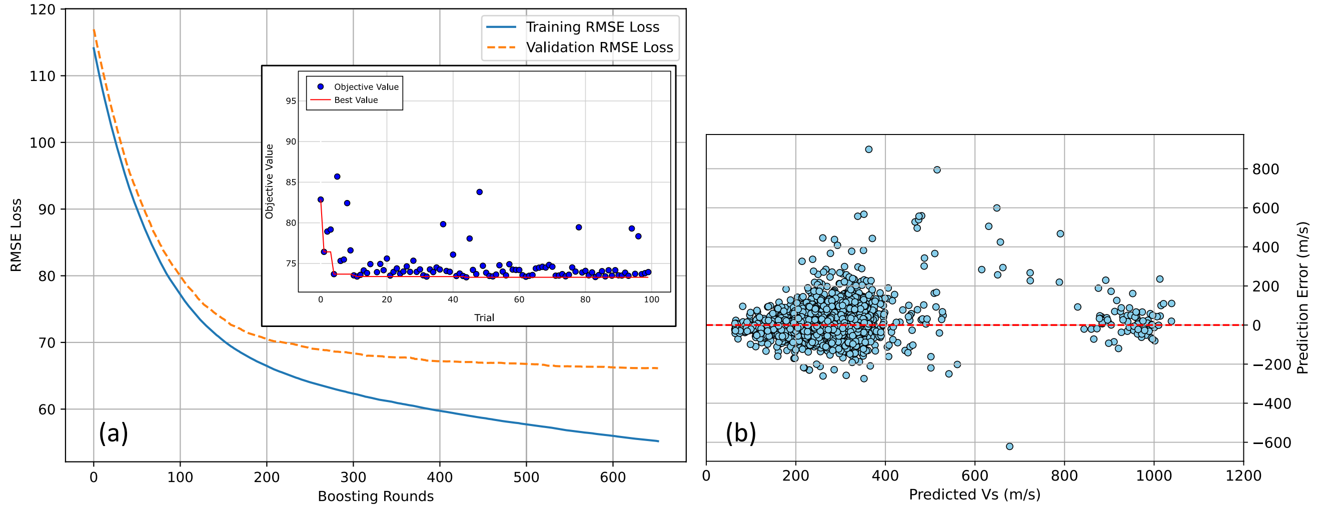

Fig. 3 **a** Training and validation RMSE loss curve over boosting rounds for the optimised XGBoost model. Validation is based on the 5% of data withheld in training. The inset shows the Optuna optimisation history with no significant performance gain after 100 trials. **b** Residual plot for the training dataset

Evaluation metrics

The evaluation focused on both global performance and on the analysis of model behaviour across V_s ranges and soil types. Model accuracy was assessed using three standard statistical metrics, RMSE, mean absolute error (MAE), and the coefficient of determination (R^2) (Eqs. 9-11):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

where $\bar{y}$ is the mean of all true V_s values. RMSE places greater weight on larger errors, while MAE provides a more balanced view of the average prediction errors and is less sensitive to outliers. The R^2 score indicates the proportion of total variance in the target variable that is explained by the model's predictions.

To account for an imbalance in true V_s values across the testing dataset related to an overrepresentation of mid-range values (200–400 m/s) in the training dataset (inset in Fig. 1), the weighted R^2 was also computed. This was achieved by grouping the true V_s values into 50 m/s magnitude bins between 0 and 550 m/s and calculating a weighted average of R^2 across each bin. This weighted R^2 value depends on bin size and was calculated by:

$$R^2{}_{\mathrm{weighted}} = \sum_{j=1}^{m} w_j R_j^2 \tag{12}$$

where $w_j = \frac{n_j}{n}$ is the fraction of test samples in bin j, n_j is the number of samples in that bin, n is the total number of samples, and m is the total number of bins. The weighted R^2 value introduces data density information into the overall R^2 estimate, providing a more balanced assessment of model generalisation across the full V_s range. However, it is noted that increased data density within a bin does not necessarily indicate greater accuracy within that bin.

Prediction errors were defined as the difference between the measured and predicted V_s values ($\epsilon_i = y_i - \hat{y}_i$) such that positive values indicate underprediction and negative values overprediction.

Soil-type-specific evaluation and empirical equation comparison

Prediction performance across different soil types was evaluated using the soil behaviour type index (I_c) value to understand whether the model exhibited consistent performance across different soil types, or whether certain soil types show systematic under- or over- prediction trends and/or larger errors. For performance comparison, an empirical correlation applicable to clays, silts and sands was used. This correlation combines Mayne's (2007) correlation between soil unit weight and normalised shear wave velocity (V_{s1}) with Robertson and Cabal (2010) correlation between soil unit weight and CPTu data (Eqs. 13-17):

$$\gamma = 4.17\ln(V_{s1}) - 4.03 \tag{13}$$

$$\frac{\gamma}{\gamma_w} = 0.27\log_{10}(R_f) + 0.36\log_{10}\left(\frac{q_t}{P_a}\right) + 1.236 \tag{14}$$

where γ is the soil unit weight and γ_w is the unit weight of water (9.8 kN/m^3); R_f is the friction ratio (%), defined as $R_f = 100 \cdot \frac{f_s}{q_t}$; P_a is the atmospheric pressure (0.1 MPa); and V_{s1} is defined as

$$V_{s1} = V_s\left(\frac{P_a}{\sigma'_v}\right)^{0.25} \tag{15}$$

where σ'_v is vertical effective stress under hydrostatic conditions and is defined as

$$\sigma'_v = z(\gamma - \gamma_w) \tag{16}$$

where z is depth. By combining Eqs. 13-16, the following relation between V_s and CPTu data is obtained:

$$V_s = \frac{\exp\left(\frac{\gamma + 4.03}{4.17}\right)}{\left(\frac{P_a}{z(\gamma - \gamma_w)}\right)^{0.25}} \tag{17}$$

SHAP analysis

Interpretability of ML predictions is, arguably, as important as accuracy and uncertainty quantification. To attribute predictions to specific input features and to interpret the trained XGBoost model, the SHapley Additive exPlanations (SHAP) technique was employed. SHAP is a model-agnostic interpretability framework grounded in cooperative game theory, which assigns each feature a Shapley value representing its contribution to a given model output (Lundberg and Lee 2017). The core concept is to treat each prediction as a cooperative game, where features act as players contributing to the final outcome. The SHAP value for a feature represents its average marginal contribution to the model prediction across all possible subsets of features. In practice, exact computation of SHAP values is computationally very expensive for models with many input features. However, for decision tree-based models such as XGBoost, the TreeExplainer algorithm provides an efficient and exact method of computing SHAP values by leveraging the internal structure of the tree ensemble (Lundberg et al. 2020).

Each prediction $\widehat{y}_i$ can then be decomposed into the sum of a model baseline and the total contribution from each individual SHAP value for each feature:

$$\widehat{y}_i = \phi_0 + \sum_{j=1}^{M} \phi_j \quad (18)$$

where ϕ_0 is the expected model output over the training data (i.e. the baseline or global mean value) and ϕ_j is the SHAP value of feature j for a specific prediction. This additive property enables SHAP to provide insights at both global and local levels.

Globally, SHAP values can be aggregated across all samples to assess the average magnitude of each feature's contribution, revealing which input features most strongly influence predictions overall, i.e. their importance. Locally, SHAP values explain individual predictions by showing how each feature increases or decreases the predicted V_s value relative to the baseline. This interpretability framework enables both diagnostic insight into model behaviour and a deeper understanding of the relative influence of CPTu inputs on predicted V_s. However, such influence is not necessarily causal in reality.

Table 2 Optimised hyperparameter values of the XGBoost model (see supporting information, Table S1 for a more detailed description of these parameters)

Hyperparameters	Search span	Optimised values
Number of trees (n_estimators)	200–1200	826
Maximum depth of trees	2–6	6
Learning rate	0.01–0.1	0.010
Subsampling ratio	0.6–0.9	0.610
Column subsampling ratio per tree (colsample_bytree)	0.5–1	0.531
Column subsampling ratio per level (colsample_bylevel)	0.5–1	0.934
L1 Regularisation term on weights (reg_alpha)	0.1–1.5	0.922
L2 Regularisation term on weights (reg_lambda)	1–8	5.035
Minimum loss reduction required to make a further split (gamma)	0.1–2	1.955
Minimum child weight	2–10	2

Results and discussion

Model optimisation and training performance

The predictive performance of the XGBoost model was first optimised using the Optuna framework with a Tree-structured Parzen Estimator (TPE) sampler over 100 trials (Fig. 3a). Rapid convergence was achieved within the first 10 trials with a value of around 73 m/s. The relatively quick convergence and flatlined performance gain in later hyperparameter configurations demonstrates the efficiency of the TPE approach in navigating the hyperparameter space. It also supports that an increased number of trial runs would unlikely improve model performance. The selected hyperparameters chosen by Optuna from the best trial, and consequently used on all the following model training and evaluation tasks, are summarised in Table 2. Notably, the model adopted a relatively small learning rate (0.0102) and high regularisation (reg_lambda = 5.04, reg_alpha = 0.92; Table 2), indicating a conservative training regime to reduce overfitting.

To further safeguard against overfitting, early stopping was employed using the 5% validation set from the training data. The learning curves across boosting rounds for both the training and validation sets display steady declines, with validation loss plateauing after ~ 650 rounds, at which point training was terminated (Fig. 3a). The small divergence between training and validation curves indicates the model's satisfactory generalisation capability on unseen data, as increased divergence can be indicative of excessive overfitting. Figure 3b shows that the prediction errors (i.e. residuals) are mostly scattered around the horizonal—zero error line but with a slight funnel shape between 100 and 400 m/s, indicating increasing prediction error at higher values or a small degree of heteroscedasticity.

Despite the iterative feature removal process, RFECV retained all 10 input features in the final model configuration (depth, f_s, q_t, u_2, Q_t, F_r, B_q, I_c, q_t/f_s, and $B_q * F_r$). This indicates that each feature improved prediction performance during training.

Model testing performance

Figure 4 presents the global predictive performance of the trained model in the testing dataset. Figure 4a shows 48.2% of all V_s predictions within ± 10% of the true value and 79.4% within ± 20%. An increase in scatter and under-prediction is evident at V_s magnitudes above ~ 350 m/s (as also shown in the residual plot presented in the supporting information Figure S1f). This discrepancy aligns with the training data distribution, where only 15% of samples have V_s > 350 m/s, limiting generalisation in the upper V_s range,

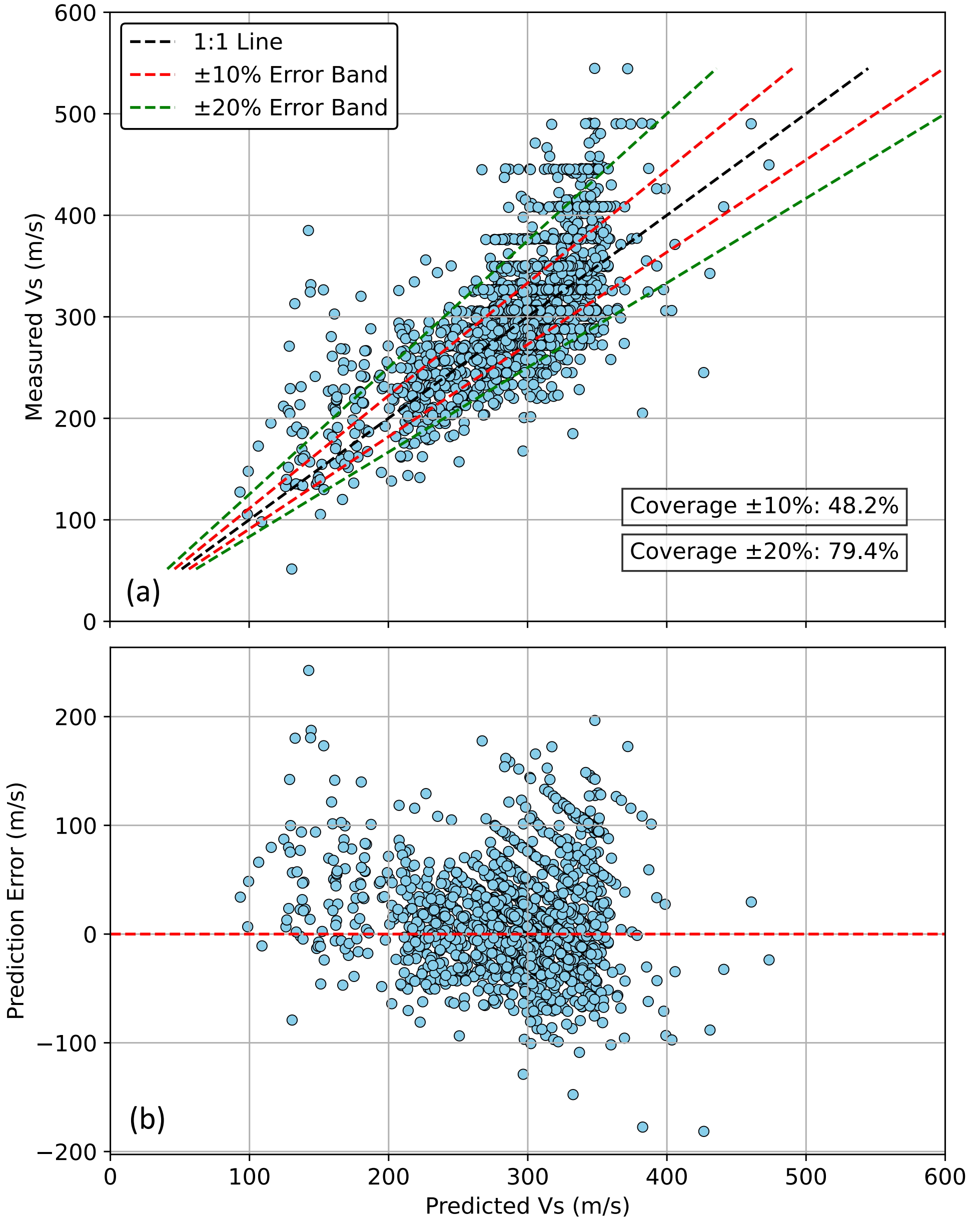

Fig. 4 Global predictive accuracy during testing. In (**a**) the error bands of ± 10% and ± 20% cover 48.2% and 79.4% of predictions, respectively, and are obtained by adding ± 10% and ± 20% to the horizontal axes (prediction axes) values of the 1:1 line

which represents ~ 20% of the testing data. Figure 4b shows that the prediction errors in the testing dataset are randomly scattered around the horizonal—zero error line, there is no clear structure in the shape of the errors, and the error is mostly uniformly distributed across the prediction range (no funnel shape). These observations indicate that the model can capture the underlaying CPTu-V_s relation and generalizes reasonably well. For completeness, the collection of residual plots for the training, validation and testing datasets is presented in the supporting information Figure S1.

The trained XGBoost model demonstrated reasonable predictive capability with a MAE of 38.7 m/s and RMSE of 51.0 m/s when evaluated against the testing dataset of 1526 instances. The overall R^2 was 0.44, reflecting a moderate explanatory power. However, when accounting for uneven distribution in true V_s magnitudes using a weighted R^2 (Eq. 12), the performance improved to 0.68. The MAE and R^2 values for all the CPTu profiles in the testing dataset are presented in the supporting information Figures S2 and S3.

Figure 5 shows a summary of model performance. Additional results comparing the data density distribution for V_s in the training, testing and predicted datasets is provided in the supporting information (Figure S4). The error distribution evaluated via both the Cumulative Distribution Function (CDF) and histogram (Fig. 5a and b) approximates a normal distribution with both mild skewness (0.46) and kurtosis (0.92), indicating low asymmetry and no significant long tail and peakedness. The median error of 9.8 m/s indicates a small underestimation, i.e. small systematic bias, while the histogram-fitted mean error is + 13.9 m/s,

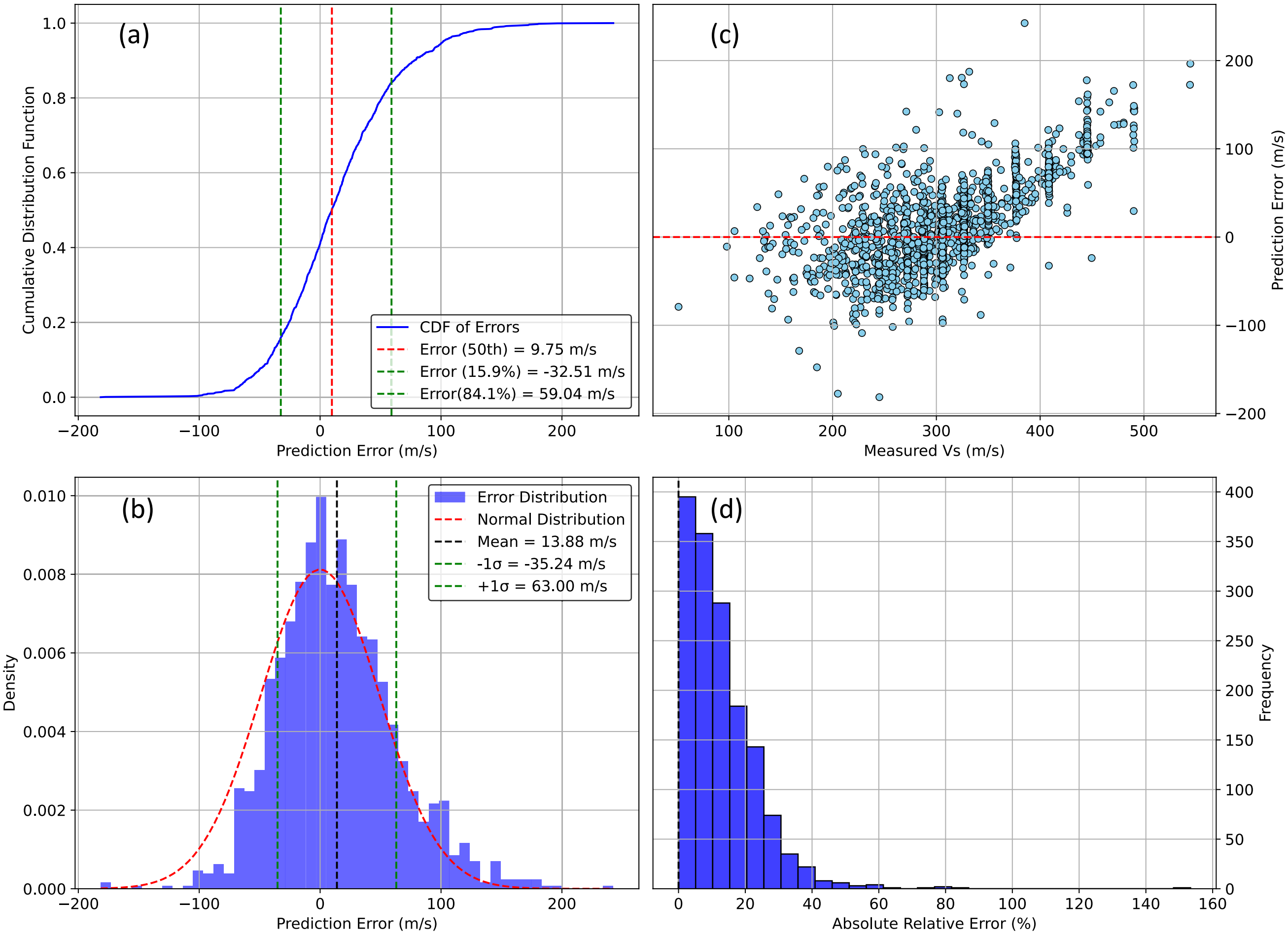

Fig. 5 Summary of global performance for the testing dataset. **a** Cumulative probability distribution of prediction errors. **b** Histogram of prediction errors with fitted normal distribution curve. **c** Prediction errors across the measured V_s range. **d** Frequency of absolute relative errors in 5% bins. In the legend, σ is standard deviation

with a ± 1σ range of − 35.2 to + 63.0 m/s. CDF percentiles (15.9% and 84.1%) span from − 32.5 m/s to + 59.0 m/s, broadly consistent with the ± 1σ range. This near-normal error distribution indicates a statistically predictable pattern of deviations, an important consideration for geotechnical applications. The model exhibits a tendency to overestimate V_s values below 200 m/s, remains approximately unbiased between 200 and 300 m/s, and increasingly underpredicts at $V_s > 300$ m/s (Fig. 5c). Mean absolute relative error analysis shows that 25% of predictions have an error of 5%, and 95% fall within a 30% error (Fig. 5d).

Soil-specific variations are detailed in Figs. 6 and 7. Model error increases with V_s magnitude across all soil types (Fig. 6). For example, in 'clean to silty sand', MAE increases from 22.4 m/s (50–150 m/s) to 145.5 m/s (450–550 m/s), while in 'silty sand to sandy silt', the MAE increases from 17.8 m/s to 131.5 m/s over the same range. Higher errors are observed in fine-grained soils, such as 'silty clay to clay' and 'clayey silt to silty clay', where sample sizes are small (Fig. 6). The largest negative errors (e.g., mean = –21.1 m/s in 'silty clay to clay') indicate a tendency to overpredict V_s in cohesive soils. In contrast, coarse soils such as 'clean to silty sand' exhibit positive errors (mean = + 16.9 m/s), while 'gravelly to dense sand' and 'silty sand to sandy silt' show near-zero median errors (Fig. 7), suggesting more balanced predictions. Skewed error distributions in certain soil classes likely reflect both data imbalance and presence of outliers.

Differences in soil-specific behaviour are further illustrated in the stacked cumulative distribution of absolute relative errors shown in Fig. 8. The model performs best in coarser, sandier soils where 90% of predictions fall within 25–28% relative error for 'clean to silty sand' and 'silty sand and sandy silt', which collectively comprise over 85% of the testing data. In contrast, the 90th percentile error in 'silty clay to clay' is substantially higher at 43%, indicating poorer generalisation in fine-grained soils.

Model performance is influenced by both V_s magnitude and soil type, driven in part by physical variability in CPTu to V_s relationships and imbalanced training data. Fine-grained

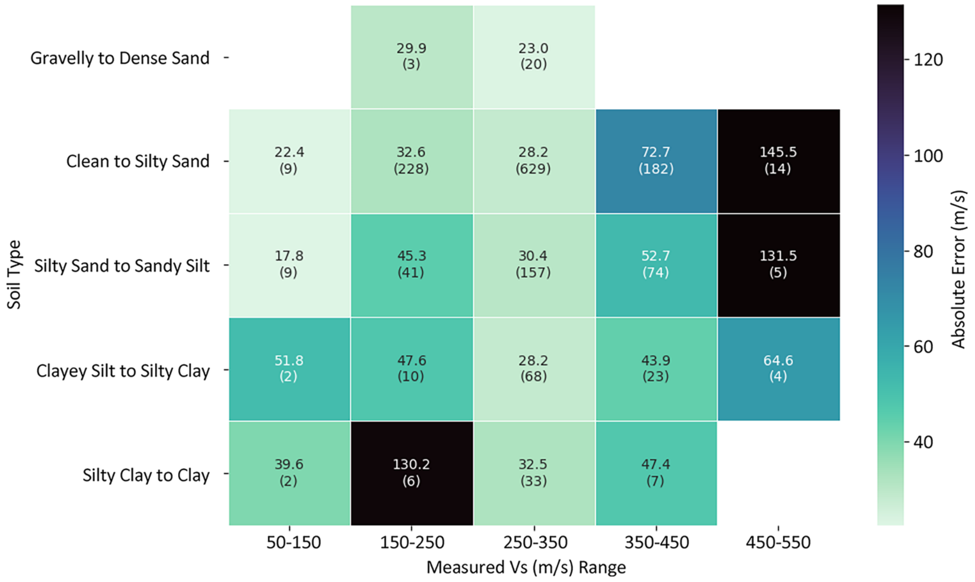

Fig. 6 MAE per soil type across 100 m/s measured ranges for the testing dataset of 1556 samples. The frequency of measured V_s values in each range is shown in brackets

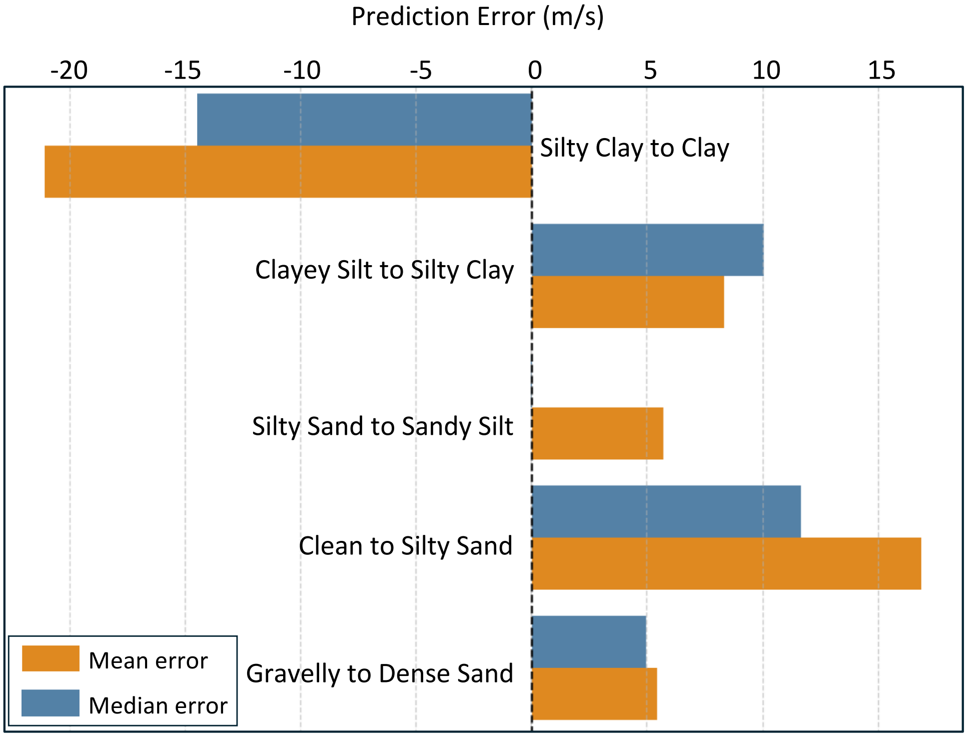

Fig. 7 Median and mean error per soil type. The discrepancy between mean and median error for 'clean to silty sand' and 'silty sand to sandy silt' shows that there are significant outlier predictions in these soils. Except for 'silty clay to clay', all other soils are underpredicted on average, supporting the model's overall prediction error of 9.8 m/s

soils exhibit higher MAEs and a greater proportion of outliers, likely due to their complex microstructures, higher compressibility, and variable drainage conditions during testing (Karlsrud et al. 2005). These conditions complicate the relationship between CPTu inputs and V_s, particularly in partially drained scenarios where the pore pressure response may not reliably reflect stiffness. In contrast, coarse-grained soils, characterised by better drainage and stronger correlations between cone resistance and stiffness, yield lower errors and more predictable responses (Robertson 2009). The model appears to capture these patterns effectively in well-represented soil types.

Another important source of uncertainty arises near soil classification thresholds. When the soil behaviour tracks a soil type boundary such as that between 'silty sand to sandy silt' and 'clayey silt to silty clay', classification ambiguity degrades performance. This uncertainty is not due to data limitations, but rather to the inherent fuzziness in soil type transitions (Robertson 2010), introducing epistemic uncertainty. Addressing this may require additional soil type specific features (e.g. Schorpp et al. 2024).

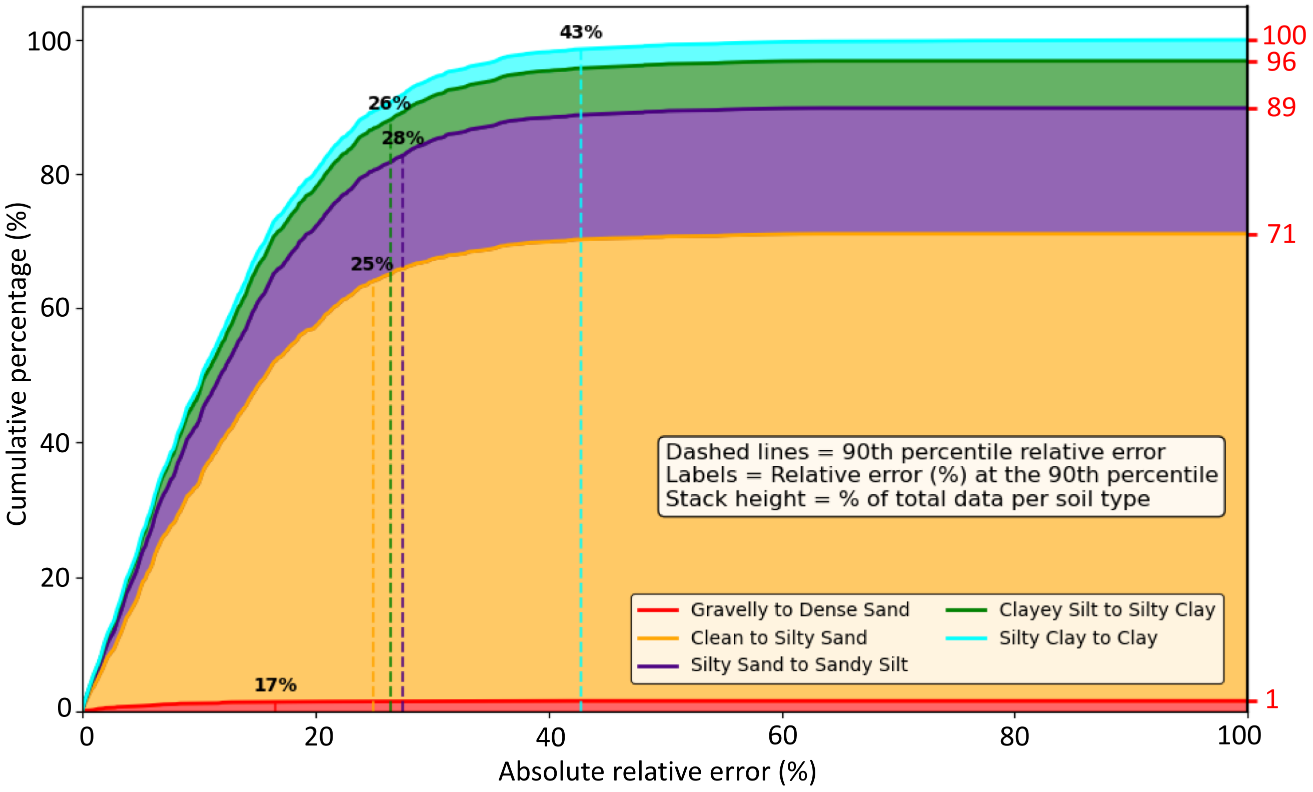

Fig. 8 Stacked cumulative distribution of absolute relative errors by soil type. Each curve represents an individual soil type, with the vertical dashed lines marking the 90th percentile relative error and stack heights indicating the proportion of the testing dataset associated with each soil type

The model demonstrates high fidelity in predicting V_s for most of the dataset, particularly in the mid-range (250–350 m/s), where both data density is large and soil response is more consistent. This is in line with previous studies also using XGBoost and a larger non-open source dataset comprising > 200,000 datapoints (Entezari et al. 2022, 2024). The near-normal, symmetrical distribution of errors with low bias is encouraging for engineering applications, where understanding the range and distribution of potential error is often as important as minimising it (Kotulski and Szczepinski 2009). The reduced performance at lower (< 150 m/s) and higher (> 400 m/s) V_s values highlights the importance of a balanced dataset. Improving prediction in these regions may require targeted data augmentation and more tailored feature engineering.

Overall, these findings emphasize that the XGboost model presented can learn generalisable relationships between CPTu parameters and V_s when sufficient representative training data are available. Still, care must be taken in extending predictions to underrepresented soil types or V_s ranges. Future work should focus on enhancing training coverage and incorporating adaptive modelling strategies for soils with more complex behaviours.

Model performance against an empirical correlation

Benchmarking the XGBoost model performance against Robertson and Cabal (2010) empirical correlation provides insights into the added value of the proposed ML approach (Fig. 9). Across all 1526 test samples, the Robertson and Cabal (2010) based correlation showed an R^2 of -2.4, indicating a performance worse than calculating the mean of the V_s for the entire testing set. This empirical correlation outperformed the XGBoost model in only 20.4% of cases on average, with only 17.8% of total predictions within ± 10% of measured values and 34% within ± 20% (the XGBoost model provides 48.2% and 79.4% coverage, respectively). Hence, the proposed XGBoost model shows superior performance overall, but the performance of both approaches varies across the V_s spectrum (Fig. 9). The percentage of instances where the empirical correlation outperforms the ML model increases consistently at higher V_s magnitudes. This behaviour once again reflects the model's lower performance and ability to generalise at V_s ranges with limited training data. Ultimately, this may support a more nuanced interpretation that empirical correlations and ML models are not mutually exclusive but complementary. The ML model offers higher fidelity in well-characterised regimes, while

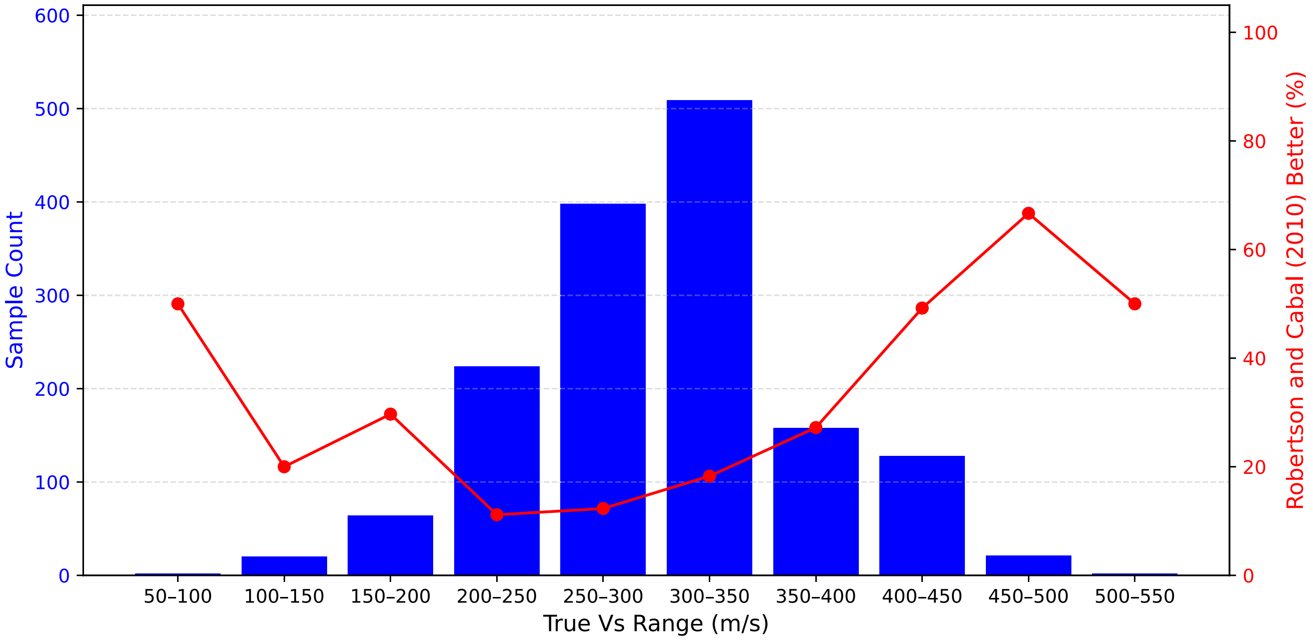

Fig. 9 Relation between testing data grouped in 50 m/s bins and percentage of instances where the Robertson and Cabal (2010) empirical correlation outperforms the XGBoost model

general empirical correlations are most useful in underrepresented zones. Combining both approaches, or exploring hybrid methods, could offer a more robust solution for practical site investigations with varying data quantity, richness and quality (Gao 2024).

Interpretability through SHAP Explanations

Calculated SHAP values provide a unified framework for attributing prediction contributions and behaviours to each input feature, as shown in Fig. 10. The top three contributors for the testing dataset are depth, f_s, and q_t (Fig. 10a), which is consistent with established geotechnical understanding and reinforces the physical relevance of these parameters in controlling V_s behaviour. Input features impact predictions differently depending on their magnitude. For example, high depth, f_s and q_t generally result in large positive V_s contributions with respect to the baseline or global model mean value (i.e. high positive SHAP values), while low to medium u_2 tends to provide negative V_s contributions with respect to the baseline (i.e. negative SHAP values) (Fig. 10b). The variability and skew in SHAP value distributions (Fig. 10c) indicates that the model consistently draws on depth, f_s, and q_t input features to drive predictions, but their influence varies considerably depending on sample. Features like u_2, F_r, q_t/f_s have more compact distributions, suggesting they exert less consistent and/or weaker influence overall. The presence of a few outliers in some features, especially for high values of u_2 providing positive SHAP values (note that most u_2-related SHAP values are negative; Fig. 10b and c), highlights that under specific conditions, certain inputs can exert disproportionately large effects, potentially contributing to mispredictions in those cases.

Figure 11 shows an example of how SHAP operates at the individual point prediction level for the lowest, median, and highest V_s predictions. Each plot decomposes the individual prediction into additive contributions from each input feature, showing how they interact to shift the output from the model's mean prediction of V_s during training (284.9 m/s) towards the final prediction. Depth is the most important feature for the highest (451.8 m/s) and lowest (81.6 m/s) V_s prediction examples and all features contribute with increases (except for u_2) and decreases (except B_q*F_r and F_r), respectively. In contrast, the median prediction (300.2 m/s) example does not show depth as the most important feature and instead shows a more balanced profile with f_s and q_t providing positive contributions and the remaining features negative contributions mostly. The overall result is a slight increase from the model mean, with no single feature substantially dominating the outcome.

These breakdowns highlight that the magnitude of SHAP contributions can vary considerably between samples and helps identify when the model may rely too heavily on a single variable such as depth, i.e. when the difference in V_s between the point sample and the mean in the training dataset is significant, potentially at the expense of mechanical parameters (i.e. q_t, f_s, u_2).

SHAP analysis in this study consistently revealed that depth, f_s, and q_t dominate model predictions, in line with physical understanding of CPTu behaviour and findings

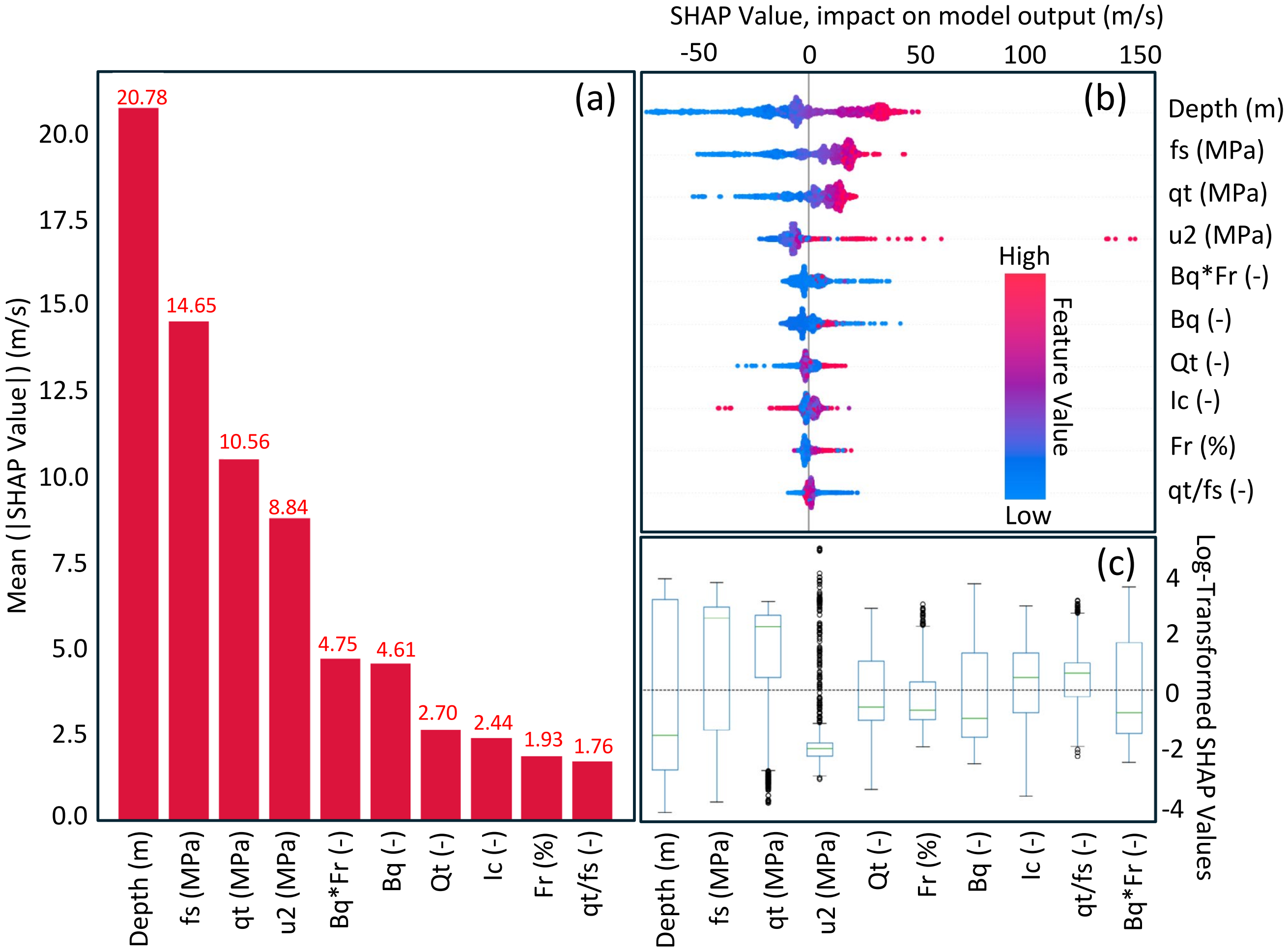

Fig. 10 **a** Mean absolute SHAP value of each input feature and its average effect on predictions. **b** SHAP summary plot showing how each input feature contributes to the final prediction. Features are ordered by importance in the final prediction determined by their mean SHAP absolute value. The horizontal spread within each feature indicates the variability in contribution for different samples, the vertical spread is to avoid overlapping points and so indicates density of data points, while the colour gradient distinguishes the magnitude of each feature. For example, low depth values result in lower V_s predictions. **c** Log-transformed box plot of SHAP values for each feature in the interquartile range, i.e. from the 25th to 75th percentile. Within each blue box, the green horizontal lines indicate the median of the log-transformed SHAP values and the whiskers the range from 5 to 95th percentiles

from another study which also used SHAP (Chala and Ray 2025). This confirms that the model was not driven by spurious patterns but was learning from mechanically meaningful inputs. This analysis also revealed a complicated and heterogeneous feature contribution behaviour. While certain features are generally important (e.g. depth, q_t, f_s), their actual impact varied substantially between samples especially in transitional mixed soils and on soils with V_s velocities that deviate substantially from the mean V_s value of the training dataset.

Overall, the transparency provided by SHAP on how input features control the final V_s prediction allows for a closer physics-based interpretation of the results and for diagnosing when and why relationships between CPTu parameters and V_s may work or break down. This level of transparency is rarely achievable with traditional empirical methods and suggests that interpretable ML has a valuable role to play in geotechnical parameters prediction.

Conclusions

In this study we developed and assessed an interpretable ML framework based on the XGBoost technique, integrated with SHAP, to predict V_s from CPTu data. This approach enables not only accurate prediction but also transparency in the model's internal decision-making logic. SHAP-based interpretability analysis confirms quantitatively that the most influential features driving predictions are the directly measured mechanical parameters q_t and f_s, as well as depth. Notably, the influence of depth increases as predicted V_s values deviate from the mean V_s of the training dataset.

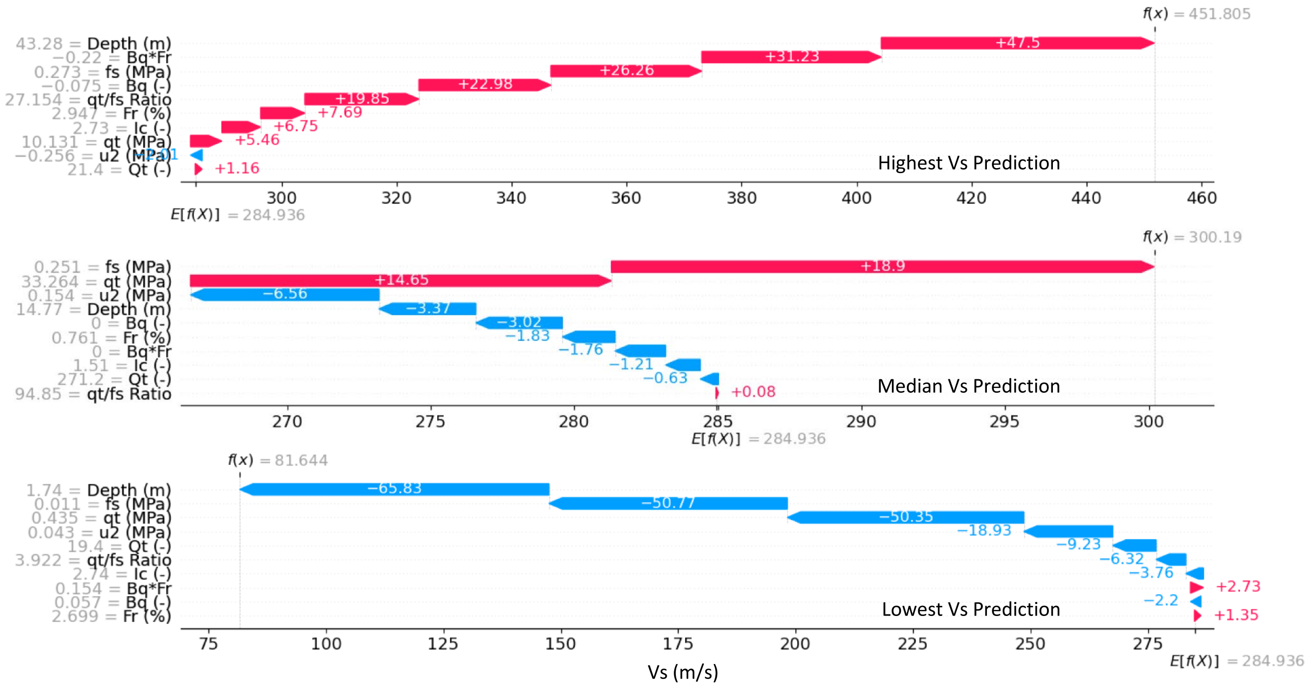

Fig. 11 SHAP waterfall plots for the maximum, median and minimum V_s point prediction (indicated by f(x)). Each input feature provides its own contribution (in m/s) to either increase or decrease the prediction away from the baseline (i.e. global mean value $E[f(x)] = 284.936$ m/s). The y-axis is ordered in terms of feature importance and shows the exact value for each input feature that produced the prediction

The model demonstrates satisfactory predictive performance, achieving a MAE of 39 m/s, a weighted R^2 of 0.68 and with 79.4% of predictions falling within ± 20% of true V_s values. When benchmarked against a widely used empirical correlation, the model exhibits superior performance, particularly within the 250–350 m/s V_s range and in ‘clean to silty sands’. Prediction errors from the ML approach increase at high V_s magnitudes, in fine-grained soils, and in soils with mechanical properties in between soil types; and correspond to features poorly represented in the training data set.

The effectiveness and generalisability of ML-based techniques for CPTu-based V_s prediction are contingent upon the availability of diverse and balanced datasets across soil types and V_s ranges, as well as an improved understanding, and ideally quantification, of epistemic uncertainty associated with sparsely sampled regimes. Ultimately, this study underscores that interpretable ML is not merely a ‘black box’ alternative, but a potentially powerful, transparent tool, capable of supporting offshore geotechnical site investigation and risk assessment. As the availability of publicly accessible geotechnical datasets increases and industry standards continue to evolve, approaches that combine both high predictive accuracy with model interpretability could become central to offshore infrastructure development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11001-025-09602-6.

Acknowledgements We thank two anonymous reviewers for their constructive comments. Susan Gourvenec is supported by the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme. Yuting Zhang is supported through funding provided by the Chair in Emerging Technologies. AI tools were used to edit the manuscript and optimize the code.

Author contributions Héctor Marín-Moreno: Conceptualisation, Methodology, Software, Supervision, Validation, Writing—original draft. James Willis: Conceptualisation, Methodology, Software, Formal analysis, Validation, Writing—original draft. Yuting Zhang: Writing—review & editing. Susan Gourvenec: Writing—review & editing. Héctor Marín-Moreno and James Willis contributed equally to this work.

Data availability The training and validation dataset (python pickle file) and testing dataset (Excel file) are available as an online resource in the supporting information. The trained and optimised XGBoost model pipeline (python joblib file), the environment configuration file used to train and test the XGBoost model (python.yml file), and a template python script to make V_s predictions on individual CPTu(s) from the testing set and other user-provided datasets, are also available as an online resource in the supporting information.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval No applicable.

References

Abdelghani L, Bakhti R (2024) Analyzing the impact of multiple foundation stiffness correlation on the natural frequency of offshore wind turbines. Electron J Struct Eng 24(4):1–7

Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019) Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp 2623–2631)

Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. Artif Intell Rev 54:1937–1967

Bergstra J, Bardenet R, Bengio Y, Kégl B (2011) Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 24:2546–2554

Bundesamt für Seeschifffahrt und Hydrographie (BSH), 2024**.** PINTA database for marine soil investigations. https://pinta.bsh.de/ausschreibungen?lang=en

Chala AT, Ray RP (2023a) Machine learning techniques for soil characterization using cone penetration test data. Appl Sci 13(14):8286

Chala AT, Ray R (2023b) Assessing the performance of machine learning algorithms for soil classification using cone penetration test data. Appl Sci 13(9):5758

Chala AT, Ray RP (2025) Uncertainty quantification in shear wave velocity predictions: integrating explainable machine learning and Bayesian inference. Appl Sci 15(3):1409

Charles J, Gourvenec S, Vardy M (2023) A neural network based approach for recovery of shear stiffness degradation curves. Acta Geotech. https://doi.org/10.1007/s11440-023-01879-4

Chen J, You J, Wei J, Dai Z, Zhang G (2024) Interpreting XGBoost predictions for shear-wave velocity using SHAP: insights into gas hydrate morphology and saturation. Fuel 364:131145

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp 785–794)

Damgaard M (2011) An introduction to operational modal identification of offshore wind turbine structures

Entezari I, Sharp J, Mayne PW (2022) A data-driven approach to predict shear wave velocity from CPTu measurements. Cone penetration testing 2022. CRC Press, pp 374–380

Entezari I, Sharp J, Mayne PW (2024) A data-driven approach to predict shear wave velocity from CPTu measurements: an update. In: Proceedings of the 7th international conference on geotechnical and geophysical site characterization, Barcelona, 18–21 June 2024

Fischer J, Jost O, Richter M, Wiemann J, Fichtner Water & Transportation GmbH (Hamburg) (2019) Presentation and comparison of site investigation methods for offshore wind energy in the European North Seas Countries in the Context of the EU North Seas Energy Cooperation

Gao W (2024) The application of machine learning in geotechnical engineering. Appl Sci 14(11):4712

Gómez D, Stuyts B (2022) Benchmark of small-strain shear modulus on Belgian North Sea soils with bender element testing. In: Proceedings of the 7th international young geotechnical engineers conference, Sydney, Australia, 29 April–1 May 2022, pp.347–352. https://www.issmge.org/publications/online-library

Gourvenec S (2024) Offshore geotechnical challenges of the energy transition. Geomech Energy Environ 39:100584. https://doi.org/10.1016/j.gete.2024.100584

Hamilton RI, Papadopoulos PN (2023) Using SHAP values and machine learning to understand trends in the transient stability limit. IEEE Trans Power Syst 39(1):1384–1397

Hu P, Jeng DS (2025) New challenges in offshore geotechnical engineering developments. J Mar Sci Eng. https://doi.org/10.3390/jmse13030392

Hucker N, Ward I, Manceau S (2019) Measured changes in the natural frequency of offshore wind turbines with monopile foundations. In: Proceedings of the SECED: Conference in Earthquake Risk and Engineering towards a Resilient World

ISO (2022) Geotechnical investigation and testing - Field testing - Part 1: Electrical cone and piezocone penetration test (ISO Standard No. 22476-1:2022). 2nd edition

Jin L, Fuggle A, Roberts H (2025) Effect of database size and composition on machine learning model development to estimate shear wave velocity. In: Geotechnical Frontiers 2025 (pp 95–103)

Karlsrud K, Lunne T, Kort DA, Strandvik S (2005) CPTU correlations for clays. In: Proceedings of the 16th international conference on soil mechanics and geotechnical engineering (pp 693–702). IOS Press

Khawaja L, Asif U, Onyelowe K, Al Asmari AF, Khan D, Javed MF, Alabduljabbar H (2024) Development of machine learning models for forecasting the strength of resilient modulus of subgrade soil: genetic and artificial neural network approaches. Sci Rep 14(1):18244

Koldasbayeva D, Tregubova P, Gasanov M, Zaytsev A, PetroVskaia A, Burnaev E (2024) Challenges in data-driven geospatial modeling for environmental research and practice. Nat Commun 15(1):10700

Kotulski ZA, Szczepinski W (2009) Error analysis with applications in engineering. Springer Science & Business Media

L'Heureux JS, Long M (2017) Relationship between shear-wave velocity and geotechnical parameters for Norwegian clays. J Geotech Geoenviron Eng 143(6):04017013

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2(1):56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst, 30

Marín-Moreno H, Gourvenec S, Charles J (2024) Application of deep neural network combined with dynamic poroelasticity to define seismic velocities and porosity from cone penetrometer data. In: Proceedings of the 7th international conference on geotechnical and geophysical site characterization, Barcelona, 18–21 June 2024

Masters T et al (2024) Evaluation of the challenges present in the obtaining, processing and interpreting useful data from offshore seismic cone penetration testing, In: International conference on geotechnical and geophysical site characterization. https://doi.org/10.23967/isc.2024.004

Mayne PW (2007) In-situ test calibrations for evaluating soil parameters. Charact Eng Proper Nat Soils 3:1601–1652

Oberhollenzer S, Premstaller M, Marte R, Tschuchnigg F, Eharter GH, Marcher T (2021) Cone penetration test dataset Premstaller

Geotechnik. Data Brief 34:106618. https://doi.org/10.1016/j.dib.2020.106618

Phoon KK, Cao ZJ, Ji J, Leung YF, Najjar S, Shuku T, Tang C, Yin ZY, Ikumasa Y, Ching J (2022) Geotechnical uncertainty, modeling, and decision making. Soils Found 62(5):101189

Ramboll (2021) Geotechnical data report of the preliminary investigation of FEP-SITE O-1.3, Bundesamt für Seeschifffahrt und Hydrographie, https://pinta.bsh.de/ausschreibungen?lang=en

Robertson PK (2009) Interpretation of cone penetration tests – a unified approach. Can Geotech J 46(11):1337–1355. https://doi.org/10.1139/T09-065

Robertson PK, Cabal KL (2010) Estimating soil unit weight from CPT. In: 2nd international symposium on cone penetration testing (pp 2–40)

Robertson PK (2010) Soil behaviour type from the CPT: an update. In: 2nd international symposium on cone penetration testing (Vol. 2, No. 56, p. 8). Huntington Beach: Cone Penetration Testing Organizing Committee

RVO (2023) Offshore Wind Energy. http://offshorewind.rvo.nl

Schorpp L, Straubhaar J, Renard P (2024) From lithological descriptions to geological models: an example from the Upper Aare Valley. Front Appl Math Stat 10:1441596

Smith FB, Kossen J, Trollope E, van der Wilk M, Foster A, Rainforth T (2024) Rethinking Aleatoric and Epistemic Uncertainty. arXiv preprint arXiv:2412.20892

Stuyts B, Weijtjens W, Jurado CS, Devriendt C, Kheffache A (2024) A critical review of cone penetration test-based correlations for estimating small-strain shear modulus in North Sea soils. Geotechnics 4(2):604–635

Stuyts B, Suryasentana S (2023) Applications of data science in offshore geotechnical engineering: State of practice and future perspectives. In: 9th international SUT OSIG conference: innovative geotechnologies for energy transition

Thompson A, Jagan K, Sundar A, Khatry R, Donlevy J, Thomas S, Harris P (2021) Uncertainty evaluation for machine learning. NPL Report MS 34. Teddington: National Physical Laboratory. https://doi.org/10.47120/npl.MS34

Trafford A, Ellwood R, Wacquier L, Godfrey A, Minto C, Coughlan M, Donohue S (2022) Distributed acoustic sensing for active offshore shear wave profiling. Sci Rep 12(1):9691

Wair BR et al. (2012) Guidelines for estimation of shear wave velocity Profiles, PEER Report. https://peer.berkeley.edu/sites/default/files/webpeer-2012-08-bernard_r._wair_jason_t._dejong_and_thomas_shantz.pdf

Zhang T, Chai H, Wang H, Guo T, Zhang L, Zhang W (2023) Interpretable machine learning model for shear wave estimation in a carbonate reservoir using LightGBM and SHAP: a case study in the Amu Darya right bank. Front Earth Sci 11:1217384

Zorzi G, Mankar A, Velarde J, Sørensen JD, Arnold P, Kirsch F (2019) Reliability analysis of offshore wind turbine foundations under lateral cyclic loading. Wind Energy Sci Discuss 2019:1–25

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.