Materials Today Communications 36 (2023) 106545

Contents lists available at ScienceDirect

Materials Today Communications

journal homepage: www.elsevier.com/locate/mtcomm

Predicting adhesion strength of micropatterned surfaces using gradient boosting models and explainable artificial intelligence visualizations

I.U. Ekanayake [a], Sandini Palitha [b], Sajani Gamage [c], D.P.P. Meddage [d], Kasun Wijesooriya [d], Damith Mohotti [d,*]

[a] Department of Computer Engineering, Faculty of Engineering, University of Peradeniya, Sri Lanka
[b] Department of Civil Engineering, Swinburne University of Technology, Hawthorn, Australia
[c] Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Sri Lanka
[d] School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT 2600, Australia

ARTICLE INFO

Keywords:
Machine learning
Bioinspiration
Fibrillar adhesives
Gradient boosting
Explainable AI

ABSTRACT

Fibrillar dry adhesives are widely used due to their effectiveness in air and vacuum conditions. However, their performance depends on various factors. Previous studies have proposed analytical methods to predict adhesion strength on micro-patterned surfaces. However, the method lacks interpretation on which parameters are critical. This research utilizes gradient-boosting machine learning (ML) algorithms to accurately predict adhesion strength. Additionally, explainable machine learning (XML) methods are employed to interpret the underlying reasoning behind the predictions. The analysis demonstrates that gradient boosting models achieve a high correlation coefficient ($R > 0.95$) in accurately predicting pull-off force on micro-patterned surfaces. The use of XML methods provides insights into the importance of features, their interactions, and their contributions to specific predictions. This novel, explainable, and data-driven approach holds potential for real-time applications, aiding in the identification of critical features that govern the performance of fibrillar adhesives. Furthermore, it improves end-users' confidence by offering human-comprehensible explanations and facilitates understanding among non-technical audiences.

1. Introduction

Automated gripping devices have become popular in the context of industrial digitalization in the recent decade [1]. These devices grasp and handle objects with various geometries and sizes. Among such devices, a novel implementation of hairy/fibrillar dry adhesives has been inspired by spiders, beetles, and geckos [2–4]. As highlighted in related work, splitting an adhesive pad into many fibrils provides benefits to improve adhesive contact. The same concept has been used in handling operations [5,6], soft robots [7], climbing and crawling [8], docking mechanisms, etc. Generally, these adhesives comprise arrays (of fibrils) connected to a backing layer and work using Van der Waals interactions in both air and vacuum conditions. Their performance changes with multi-scale contact engagement [9–11]. If adhesion strength in the fibrillar stage is considered, it strongly depends on the size of a fibril, mechanical properties, and geometric design. At the array level, statistical variations and load-sharing efficiency control their performance [12–14].

Given that fibrillar adhesives are advantageous to non-patterned adhesives or other gripping technologies, their working principle relies on Van der Waals interactions. These interactions have a limited (short) range and may be over a few nanometers. Interfacial defects can result in a loss of intimate contact, reducing adhesion [15,16]. Bacca et al. [14] and Booth et al. [16] investigated the effect of misalignment on the pull-off force. They described parameters such as fibril length, array size, and fibril spacing that can affect the sensitivity of adhesion to alignment errors. Furthermore, they provided a regression model to predict pull-off force F_P as a function of misalignment angle θ (Eq. 1).

$$F_p = \begin{cases} Nf_{\max} \left[1 - \dfrac{\pi a^2 E}{2f_{\max}}(n-1)\tan\theta\dfrac{d}{h}\right], \tan\theta \le \dfrac{f_{\max}}{\pi a^2 E} & \dfrac{h}{d(n-1)} \\ \dfrac{Nf_{\max}}{2n} \left[1 + \dfrac{f_{\max}}{\pi a^2 E}\right] \dfrac{1}{\tan\theta}\dfrac{h}{d}, \tan\theta > \dfrac{f_{\max}}{\pi a^2 E} & \dfrac{h}{d(n-1)} \end{cases} \quad (1)$$

$n^2 = N$ is the number of fibrils in a square array, E is Young's modulus, f_{max} represents the pull-off force of a single fibril, d is center-to-center

* Corresponding author.
E-mail address: d.mohotti@unsw.edu.au (D. Mohotti).

https://doi.org/10.1016/j.mtcomm.2023.106545
Received 12 May 2023; Received in revised form 12 June 2023; Accepted 26 June 2023
Available online 27 June 2023

spacing, h = length of the fibril, and a denotes the radius of the fibril. Nevertheless, this model can be used under the limits of a rigid backing layer and full contact. Samri et al. [17] mentioned that such conditions are difficult to meet in real-life applications.

Recent in-situ observations allow us to determine the adhesion strength of individual fibrils. Kim et al. [18] employed a bi-lateral end-notched flexural test and single-leg bend test in order to measure the adhesion strength. However, they stated that engineering experiments like this are often sensitive to global inelastic deformations occurring away from the surface. The adhesion performance of a single fibril depends on interfacial defects, imperfections, and contamination or dust [19]. The size of a defect may randomly vary across the contact, resulting in a distribution of pull-off force over an array. Booth et al. [13] introduced a framework to predict adhesion force based on Weibull statistics and reference elongation at detachment which can be obtained experimentally. Bettscheider et al. [20] developed continuum models to predict the local adhesion strength of micropatterned surfaces. Berardo et al. [21] used both numerical and experimental methods to investigate the adhesive friction of micropatterned surfaces. Their numerical model showcased a reasonable agreement with experimental results. On the other hand, home-built equipment has been used to measure the adhesion strength of these micropatterned surfaces [22,23]. However, building such instruments required a special set of expertise. Estimating the performance of fibrillar adhesives can be solved using numerical methods despite the time it takes to monitor these operations. Such constraints make it difficult to implement these methods, especially in robotic applications. Related work highlights that the experimental methods do not readily highlight the factors that govern adhesion strength and numerical methods need to be improved to handle the sensitivities in model parameters. Booth and Hensel [24] argue that it is important to identify the factors that govern the adhesion strength by using either a statistical method.

To investigate such relationships, the use of ML provides a suitable alternative. It is noted that only one recent study has used ML to predict adhesive strength by using visually observed features [17]. An acceptable accuracy has been obtained for predicting pull-off force F_P. However, their study did not highlight the importance of ML interpretation. The existing methods of predicting adhesion strength lack a proper interpretation of the models and predictions. Also, different model architectures can be used to improve the accuracy of the predictions. The lack of transparency makes traditional ML approaches challenging to be implemented in real applications as it diminishes the end-user's trust.

Explainable/interpretable ML reveals the hidden relationships and the underlying reasoning of conventional ML-based predictions. It assists in identifying the importance of features and elucidates the ML model's inner workings. Interestingly, XML provides the underlying reasoning behind predictions. Thus, explainable methods develop end-users confidence in ML-based predictions. It has become much more prevalent in many fields (e.g., business, data science, and engineering) as a result of human-comprehensible explanations [25,26].

To the best of the authors' knowledge, no related studies have used XML to predict the pull-off force of micro-patterned surfaces using XML. In this study, we used gradient-boosting ML models (histogram gradient boosting (HGB), extreme gradient boosting (XGB), and gradient boosting (GB)). The reason to use these classical tree-based gradient boosting models is that they have been often used in the engineering context because of their ability to identify non-linear relationships efficiently [27,28]. Also, tree-based models generally are easy to optimize compared to complex models such as neural networks [29,30]. In addition, we not only used a single explanation method but also used several explanation methods (Shapley additive explanations (SHAP), Local interpretable model-agnostic explanations (LIME), and Model agnostic language for explorations and explanation (DALEX)). Therefore, the study critically evaluates how each gradient boosting model performs predictions and how each explainable method ranks/weighs its feature importance. This study is therefore novel and important as it achieves the following objectives: (a) using gradient boosting ML models to predict adhesion strength to improve prediction accuracy; (b) revealing the underlying reasoning of adhesion strength predictions on micro-patterned surfaces by using XML; (c) involving different explanation methods to critically evaluate feature importance and reveal the ML model's inner workings; (d) improving end-user's confidence in ML-based adhesion strength predictions by revealing causality of predictions.

From a different viewpoint, using XML cross-validates predictions by using experimental data. The overall study helps to signify the importance of ML explanations and how each ML model's inner working changes with others. Furthermore, the study emphasizes that using XML does not cost accuracy or model complexity but rather supports decision-making criteria by providing human-comprehensible explanations. Section 2 provides brief details on ML and XML. The rest of the manuscript is arranged such that Section 3 describes data used for ML models. In Section 4, the performance of gradient boosting models is evaluated. Section 5 articulates the novel black-box explanations and critically investigates the inner-working of ML models using three different explanations and finally, Section 6 concludes the study.

2. Machine learning (ML) and explainable machine learning (XML)

2.1. ML models

The authors proposed using three gradient boosting algorithms, namely, histogram gradient boosting (HGB), extreme gradient boosting (XGB), and gradient boosting (GB). A brief introduction to these three models has been provided in the latter subsections. Gradient boosting is a technique that ensembles (combining multiple models) simple models that works better than an individual model. All modelling and programs were carried out using Python language. For the model implementation, we used the Scikit library [31].

2.1.1. Histogram gradient boosting regressor (HGB)

HGB combines gradient boosting with histograms to efficiently handle numerical features. It creates an ensemble of decision trees sequentially, with each tree trained to correct the errors made by the previous trees [32,33]. By using histograms, it discretizes the continuous numerical features into bins, which helps in reducing the computational complexity and memory requirements [32]. The algorithm optimizes a loss function through gradient descent, improving the model's prediction accuracy with each iteration. For samples $> 10{,}000$, this method is much efficient than the typical gradient boosting regressor.

2.1.2. Extreme gradient boosting regressor (XGB)

XGB is a widely-used ML algorithm known for its exceptional performance in various domains [30,34–36]. It sequentially builds an ensemble of trees, where each tree learns from the errors of the previous ones [37]. The algorithm employs a combination of regularization techniques, such as tree pruning (removing specific branches (subtrees)) and column subsampling, to control overfitting and enhance generalization. XGB also utilizes a sophisticated optimization algorithm to efficiently find the best split points during tree construction. It has become a popular choice in data science due to its speed, scalability, and robustness [28].

2.1.3. Gradient boosting regressor (GB)

GB is an algorithm that sequentially constructs an ensemble of decision trees to make accurate predictions [38]. The algorithm starts with an initial weak model, usually a shallow decision tree, and then iteratively adds more trees to the ensemble [39,40]. Each subsequent tree is trained to correct the errors of the previous ones by minimizing a loss function using gradient descent. GB combines the predictions of multiple weak models to generate a final prediction.

2.2. XAI

Explainable machine learning (XML) enhances ML users' confidence by providing causality of predictions [41,42]. The use of explainable models can be described in two ways. First, simple models such as decision trees are self-explainable at each depth. When models become complex, a post-hoc explanation is mostly preferred (e.g. LIME [43], RISE [44], and SHAP [45], etc). Recently, Liang et al. [46] did a comprehensive review on ML interpretability methods. The authors used data-driven perturbation methods as post-hoc methods [47–49]. Perturbation is performed by masking a region of the input dataset. A set of perturbations creates disturbances that can result in new predictions, and new predictions are subsequently compared with original predictions (obtained using an undisturbed sample). Hence, the significance of different regions of the input domain can be calculated. Even though these methods are categorized under the perturbation method, unique rules and strategies are implemented within each method.

The authors noticed that SHAP and LIME are widely used in ML-related studies. Their working principle is different from one another [30],[50]. In addition, we used the DALEX explanation to compare the feature importance with the other two explanations.

2.2.1. Local interpretable model-agnostic explanations (LIME)

LIME is an instance-based explanation method. It aims to provide insights into how an ML model arrives at its predictions [43]. It achieves this by performing tests on the model's predictions when alterations are made to the input data. It is done by tweaking the values of individual features and observing the resulting impact on the model's predictions. This process involves creating a new dataset by perturbing the original sample [51,52]. The predictions obtained from these perturbed inputs are then weighed, considering their proximity to the original predictions. In addition, LIME generates explanations automatically by approximating the original model locally with an interpretable one. By employing these techniques, LIME provides interpretable explanations for individual predictions made by complex ML models [53–55].

2.2.2. SHAP (Shapley Additive Explanations)

SHAP is an explanation method introduced by Lundberg and Lee [45]. It offers a comprehensive understanding of ML models, both as a whole and on an instance-by-instance basis [27,56,57]. SHAP is based on concepts from game theory [58], where inputs are treated as players and predictions act as payouts in a cooperative game. The goal of SHAP is to determine the contribution of each player (input feature) to the overall game (prediction). It offers several versions, such as DeepSHAP, Kernel SHAP, LinearSHAP, and TreeSHAP. For this study, we utilized Tree-SHAP to explain the predictions of ML models.

2.2.3. Descriptive machine learning explanation (Dalex)

DALEX is an interpretable framework that focuses on providing comprehensive explanations for complex models. It offers both global and local explanations, allowing us to understand model behavior at different levels of granularity [59,60]. The approach is model-agnostic, meaning they extract essential information from ML models regardless of their internal structure. Srinath and Gururaja [61] stated that DALEX has good computation speed.

3. Data description

In a recent study, the data set consists of ten micro-patterned specimens (S1 to S10) fabricated using polydimethylsiloxane [17]. Specimens had been cured at 95 °C for 1 h or 75 °C for 2 h. The specimens were subjected to adhesion measuring experiments, and a dataset was generated. Descriptive statistics of the data set are provided in Table 1. For adhesion measurement, specimens were brought in contact with the target object. Specimens were compressed 105 μm and subsequently retracted at a rate of 1 mm per minute. The highest tensile force observed was noted as pull-off force, F_P. Alignment angles varied between 0 and 1 in 0.1 steps. More details are provided in the study conducted by Samri et al. [17].

Image analysis had been used for feature extraction in which the number of attached fibrils (N_f), contact area (A), which is the summation of real contact areas of each fibril, and length of misalignment vector were obtained (Fig. 1). Samri et al. [17] gave a reasonable attempt by only considering features obtained from visual observations. However, we included the misalignment angle and Weibull modulus as we intend to reveal underlying relationships using XML. The pairwise correlation plot only indicates the direct correlation between the input and output features. An inferior correlation between a specific input and output does not indicate that the specific feature always has less impact on the model output. By considering such parameters, XML can determine the actual importance and local and global impact on the model output.

Booth et al. [16] made an attempt to predict adhesion strength based on M_a. However, as mentioned earlier, their method encountered challenges when applied in real-world applications. On the other hand, Samri et al. [17] developed an ML model using parameters A, M_v, and N_f. Although they acknowledged the significance of W_m, they did not incorporate it into their model. Despite achieving an R^2 of 0.84 with their best model, the authors hypothesize that crucial parameters were overlooked in each instance. Therefore, these five parameters (M_a, A, M_v, N_f, and W_m) were used to predict F_P. The pair-wise correlation of these parameters is depicted in Fig. 2.

Fig. 2 indicates that all independent parameters showcase a good correlation with F_P except for W_m. Weibull modulus had a poor correlation with pull-off force. Misalignment angle had a good negative correlation with F_P. The feature, W_m did not show any moderate correlation with other input parameters. Between A, N_f, and M_v, we observed nearly perfect correlations where the coefficient of correlation is close to 1. With such correlation, it was reasonable to use a simple model such as a linear model to predict F_P. However, the proposed linear equation by Samri et al. [17] achieved an R^2 of 0.13, whereas their best-performed model, support vector regression, reached an R^2 of 0.84. After observing these validation scores, we expected a possible non-linear relationship to have formed between these inputs and predicting variables. Therefore, we used gradient boosting algorithms to build ML models.

On one hand, parameter A represents the contact area, which is the sum of the real contact areas of individual fibrils, whether in full or partial contact. N_f refers to the number of attached fibrils, and M_v represents the difference between the half-width of the array (W) and the distance between the center of mass in full contact and the center of

Table 1
Descriptive statistics of the data set.

	Description	Mean	Minimum	Maximum	Standard Deviation	Skewness	Kurtosis
M_a	Misalignment angle (°)	0.49	0	1	0.316	0.00104	-1.218
A	Contact area (mm^2)	44.5	2.23	115.56	35.1	0.742	-0.952
$M_v = W\text{-}\lvert\vec{v}\rvert$	Misalignment length (mm)	5.18	0.009	10.9	3.57	0.529	-1.23
N_f	Number of fibrills	107.1	7	242	81.1	0.655	-1.15
F_P	Pull-off Force (N)	2.37	0.045	14.25	2.37	2.143	6.01
W_m	Weibull modulus	6.48	2	13.9	3.971	0.863	-0.72
R_e	Reference elongation (mm)	0.39	0.28	0.79	0.146	1.883	2.56

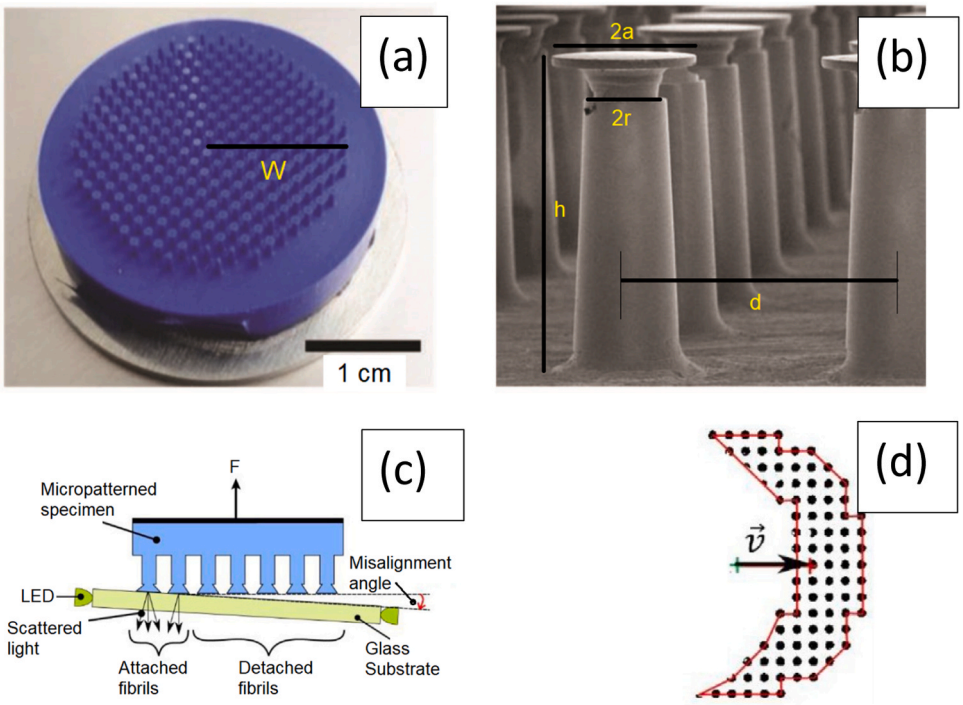


Fig. 1. (a) Optical image of a micro-patterned specimen; (b) mushroom-shaped typical fibrils; (c) Setup used to measure adhesion; and (d) misalignment vector shown in black arrow (black dots denote attached fibrils).

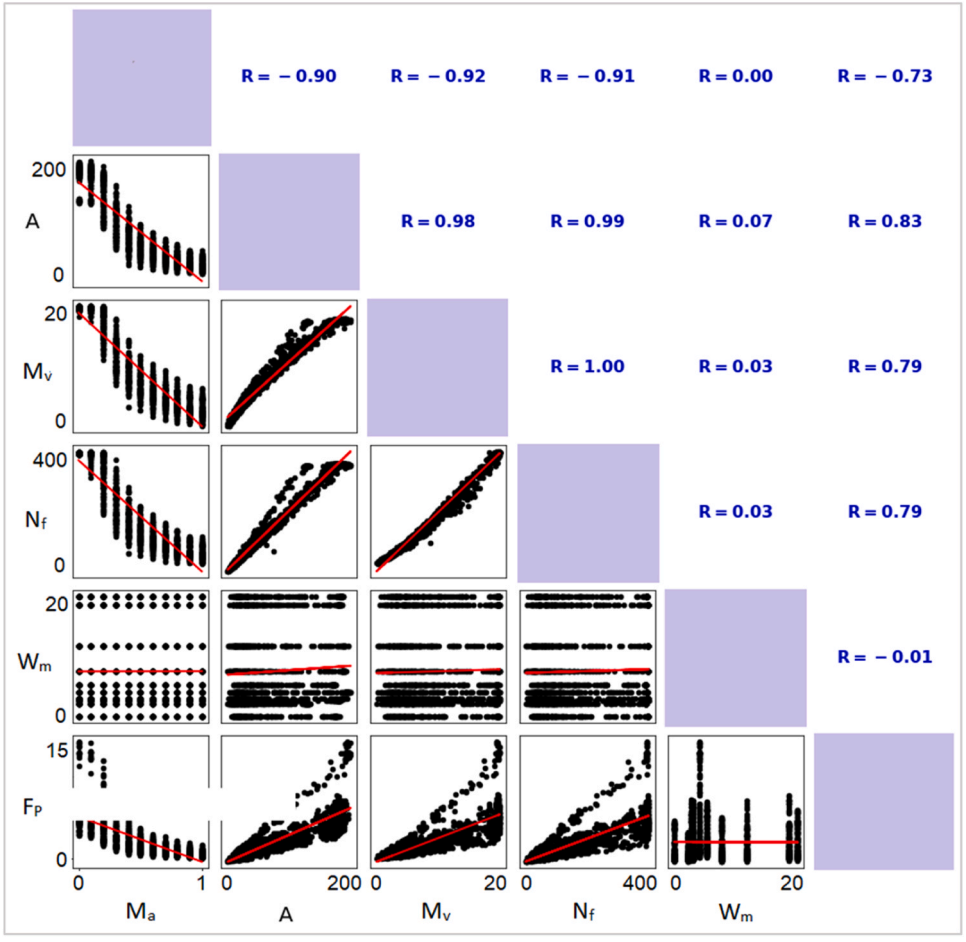


Fig. 2. Pair-wise correlation plot.

mass of the specimen in partial contact ($|\vec{v}|$). When feature A decreases from the full contact stage, the remaining two parameters also decrease in a linear manner. However, in the work by Samri et al. [17], these three parameters were used in their model without considering their dependencies, as indicated by the pairwise correlation plot shown in present study. It prudent to assume that the correlation values alone do not indicate which parameter among these three holds dominance in determining F_P. Consequently, all three parameters were included in the model, as their individual contributions could be determined using XML methods.

4. Performance evaluation of ML models

Hyperparameters of an ML model are useful to optimize the model performance. The method, “grid search" was used for the optimization. This method considers various combinations of hyperparameters and generates models. Subsequently, these models are evaluated to obtain the optimum values for hyperparameters. The optimized hyper-parameter values are presented in **ANNEX A1**.

Foremost, the training size was systematically adjusted from 90% to 10% in 5% intervals, with the corresponding test (validation) set

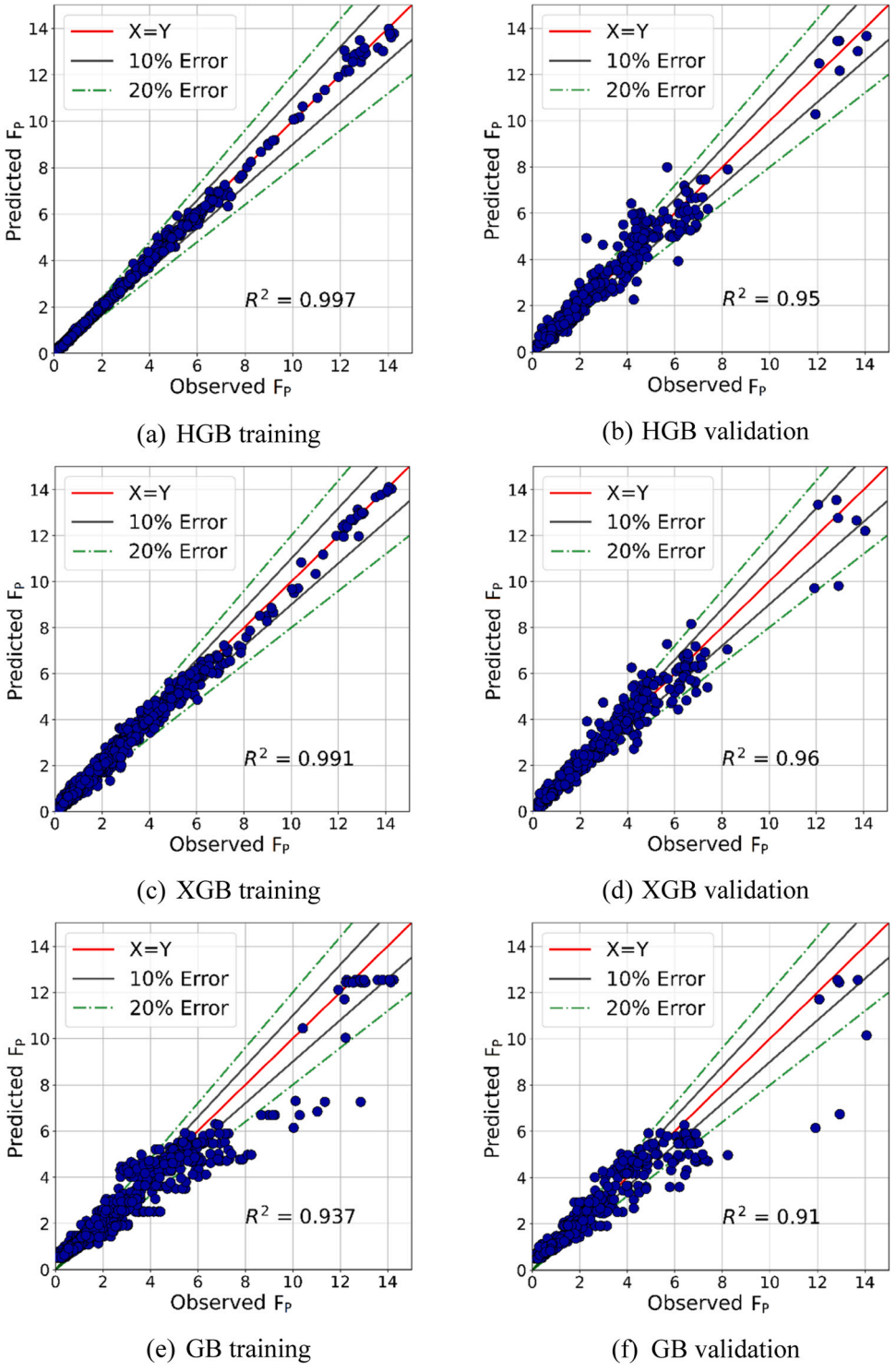


Fig. 3. Training and validation (testing) predictions.

changing from 10% to 90% accordingly. For each different split, three models were trained and tested to evaluate the accuracy of model training and validation, measured by the coefficient of determination (R^2). Upon analysis, it was observed that the gap between the training and testing scores was minimized when the training-to-testing percentages were set at 70:30 (as outlined in **ANNEX A2**). Notably, larger gaps between the training and testing scores may indicate potential overfitting issues. For model training, 70% (1146 out of 1637 instances) of the dataset was available. The rest of the dataset (30%) was kept for validation (testing).

Fig. 3 depicts the training and validation scores of each model. All three models were accurate in predicting F_P. For example, training scores (R^2) of all three regression models were higher than 0.93. HGB regressor reached $R^2 = 0.997$, the highest training score, whereas the XGB regressor reached $R^2 = 0.991$. However, the validation score of XGB is slightly higher than the corresponding score of the HGB regressor. The scatter plot of validation predictions showcased that HGB accurately predicts higher F_P values than XGB regressor, whereas the XGB regressor has fewer deviations for lower F_P values. The lowest validation accuracy was obtained for the conventional gradient boosting regressor, which achieved an R^2 of 0.91.

With minor differences between training and validation, it is evidence that models do not overfit training data. It is noteworthy that GB regressor underestimates F_P values resulting in a lower accuracy compared to the remaining models. Considering both training and validation sets, Fig. 4 shows the variation of F_P vs. M_a for specimens 01, 05, and 10. Accordingly, variation is reasonably well captured in all three models showcasing the superiority of gradient boosting algorithms.

Table 2 summarizes the performance indices of ML models. All performance indices were calculated using equations stated in **ANNEX A3**. For this purpose, we have used the coefficient of correlation (R^2) for model training, coefficient of determination (R), Mean absolute error (MAE), root mean square error (RMSE), and fractional bias (FB) for model validation.

Accordingly, model HGB has reached R= 0.99 for training and R= 0.98 for validation predictions. Corresponding RMSE and MAE values of validation are 0.48 and 0.3. Fractional bias indicates that HGB model predictions (validation) are overestimations. However, FB is closer to zero, meaning predictions are very close to experimental values. The same trend occurs for the XGB model as well. Both XGB and HGB models achieved the best accuracies across the three models. Even though the training RMSE value is 0.23 for XGB and 0.11 for HGB, both models achieved RMSE= 0.48 for validation. Despite training error indices, the XGB model is the best-performed model in terms of validation data. A slightly positive FB value describes that predictions are overestimations. The Pearson correlation of the GB regressor has reached 0.96 for the training set and 0.95 for the validation test. However, RMSE and MAE values are moderately higher than obtained values for XGB and HGB models. Overall, these indices showcase that models are accurate and do not consist of overfits or under-fits. Therefore, all models can be taken for the explanation process.

Table 2
Comparison of performance indices of ML models.

Model	Performance Index	Training	Validation
HGB	R	0.99	0.98
	R^2	0.997	0.95
	RMSE	0.11	0.48
	MAE	0.06	0.30
	FB	0	0.00995
XGB	R	0.99	0.98
	R^2	0.991	0.96
	RMSE	0.23	0.48
	MAE	0.16	0.29
	FB	0	0.00626
GB	R	0.96	0.95
	R^2	0.937	0.91
	RMSE	0.65	0.73
	MAE	0.41	0.44
	FB	0	0.01248

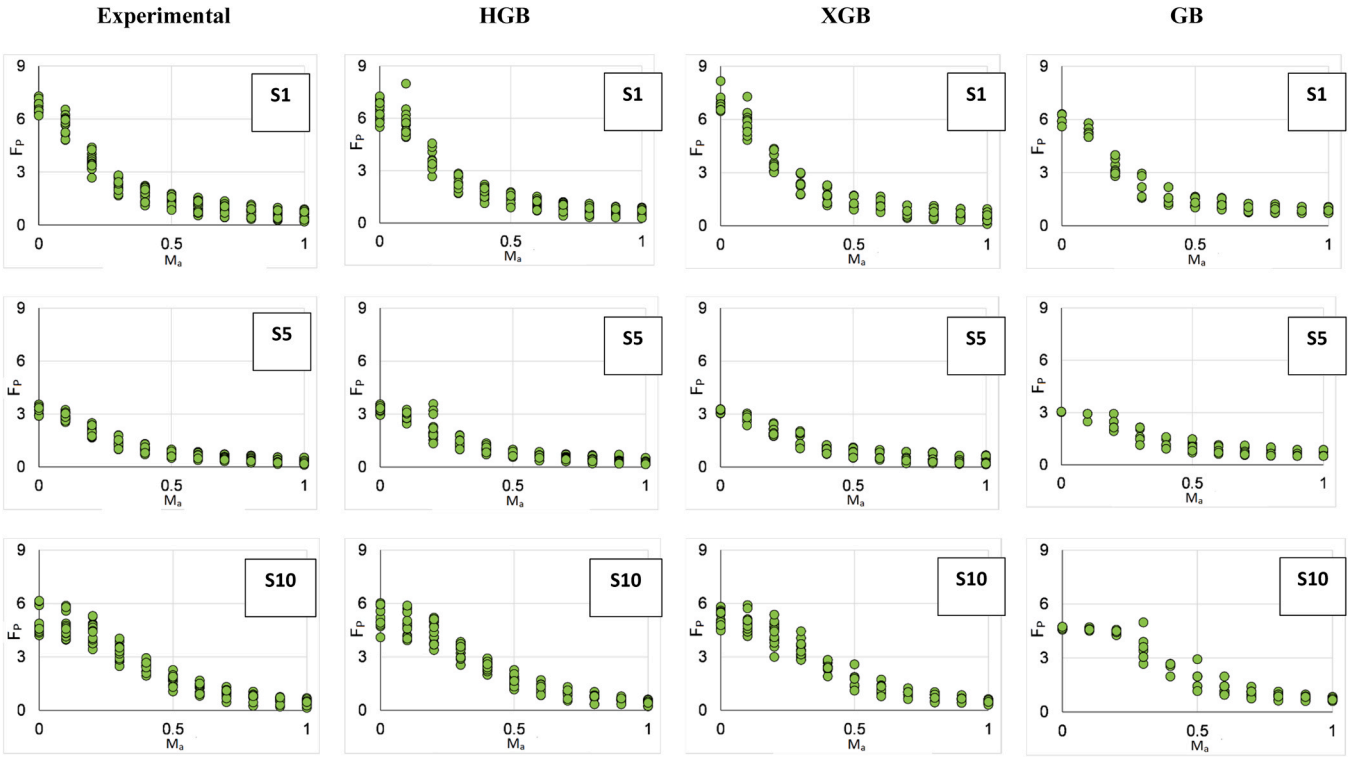


Fig. 4. Pull-off force (F_P) vs. Misalignment angle (M_a) for specimens 01 (S1), 05 (S5) and 10 (S10).

5. Machine learning model interpretation

So far, gradient boosting ML methods were used to predict the adhesion strength of micro-patterned surfaces. Samri et al. [17] did a similar investigation by incorporating distinct ML models. However, their study did not explicitly reveal the underlying reasoning behind model predictions. Despite the accuracy of ML models, predictions consist of hidden relationships where the end-user is unaware of the underlying logic or interaction of input features.

5.1. Global explanations

Three XML models (SHAP, LIME, DALEX) were used to interpret model predictions. Among these models, SHAP and DALEX provide global and local (instance) explanations, whereas LIME only provides instance-based explanations. They use unique methods to calculate feature importance. Fig. 5 shows the global feature explanation obtained from SHAP for three gradient boosting models.

SHAP concludes that the importance of each input feature is the same for all three models. The contact area is the dominant input feature, whereas lower values of A reduce F_P, and higher values of A increase F_P. However, the range of SHAP values is not the same for all models. For example, the feature importance of A for the HGB model extends beyond − 2 with respect to XGB and GB models. SHAP reveals moderately lower W_m feature importance than XGB and HGB models. It is noted that the GB regressor does not reflect a considerable effect from either N_f or M_a whereas the remaining models showcase a moderate SHAP value variation.

Interestingly, the impact of N_f is different between XGB and HGB. SHAP explanation depicts that lower values of N_f have little effect on F_P for the HGB model and higher values of N_f have a low impact on F_P for the XGB model. Even though the importance of the first three features is comparable across each ML model, there can be inconsistencies in the explanations of lower-ranked features. Since this dataset is not significantly large, the authors believe that the explanations tend to be moderately algorithm-dependent rather than data-dependent. This highlights how different ML models learn, even though they achieve the same accuracy. However, the authors recommend investigating this point in a subsequent study.

Based on the correlation observed in Fig. 2, it can be inferred that out of the features A, M_v, and N_f, the feature A holds dominance in predicting F_P. Therefore, it is reasonable to assume that M_v and N_f are dependencies of A. This finding could be valuable in future studies aiming to enhance empirical equations for predicting F_P.

The previous detailed global explanation is a unique feature of the SHAP explanation, and even DALEX does not provide such an interpretation. Nevertheless, both SHAP and DALEX generate absolute global feature importance as shown in Fig. 6. To compare feature values, we used feature contribution as a percentage by dividing each feature's importance by the sum of absolute values of all features. The order of features is not the same even for the same gradient boosting model. In the HGB model, SHAP determines that M_v is the third most dominant variable, whereas DALEX recognizes N_f for the corresponding position. Despite their unique feature importance value, the percentage of contribution is comparable. Both explanation methods reveal that the overall least contribution is from M_a. It is interesting to notice that W_m has become the second most dominant feature even though it has not been included in a previous study (examined in the partial dependency section).

When the XGB model is considered, the SHAP value of feature A is comparatively higher than the SHAP value observed in the HGB model. A similar observation is found for M_v and the remaining features show a comparable variation with the HGB model. According to SHAP interpretation, the number of fibrils has the lowest contribution, which contradicts all other explanations in Fig. 6. SHAP and DALEX explanations for the XGB model look comparable up to the third feature from the top. SHAP gives 64% importance to feature A, while DALEX gives 59% importance to the same feature. Both explanations give 19% importance to the Weibull modulus. However, SHAP underestimates the contribution of M_v as 8% compared to DALEX, which is 14%.

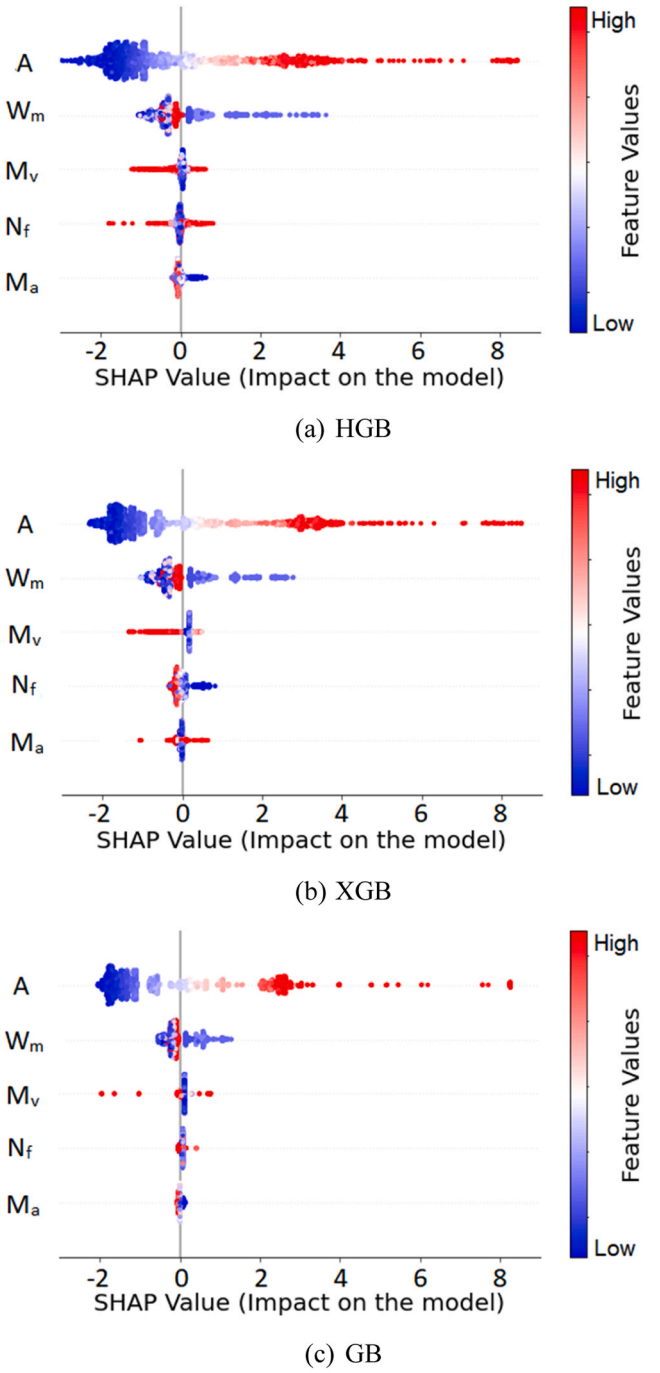


Fig. 5. SHAP global plots for ML model.

The explanation of the GB model is markedly different from that observed from XGB and HGB. Particularly, the DALEX explanation recognized an 82% contribution from feature A which is larger than the value obtained from the SHAP explanation (58%). Also, DALEX highlights almost 0% contribution from M_a and M_v. However, both explanations showcase a similar order of feature importance for the GB model.

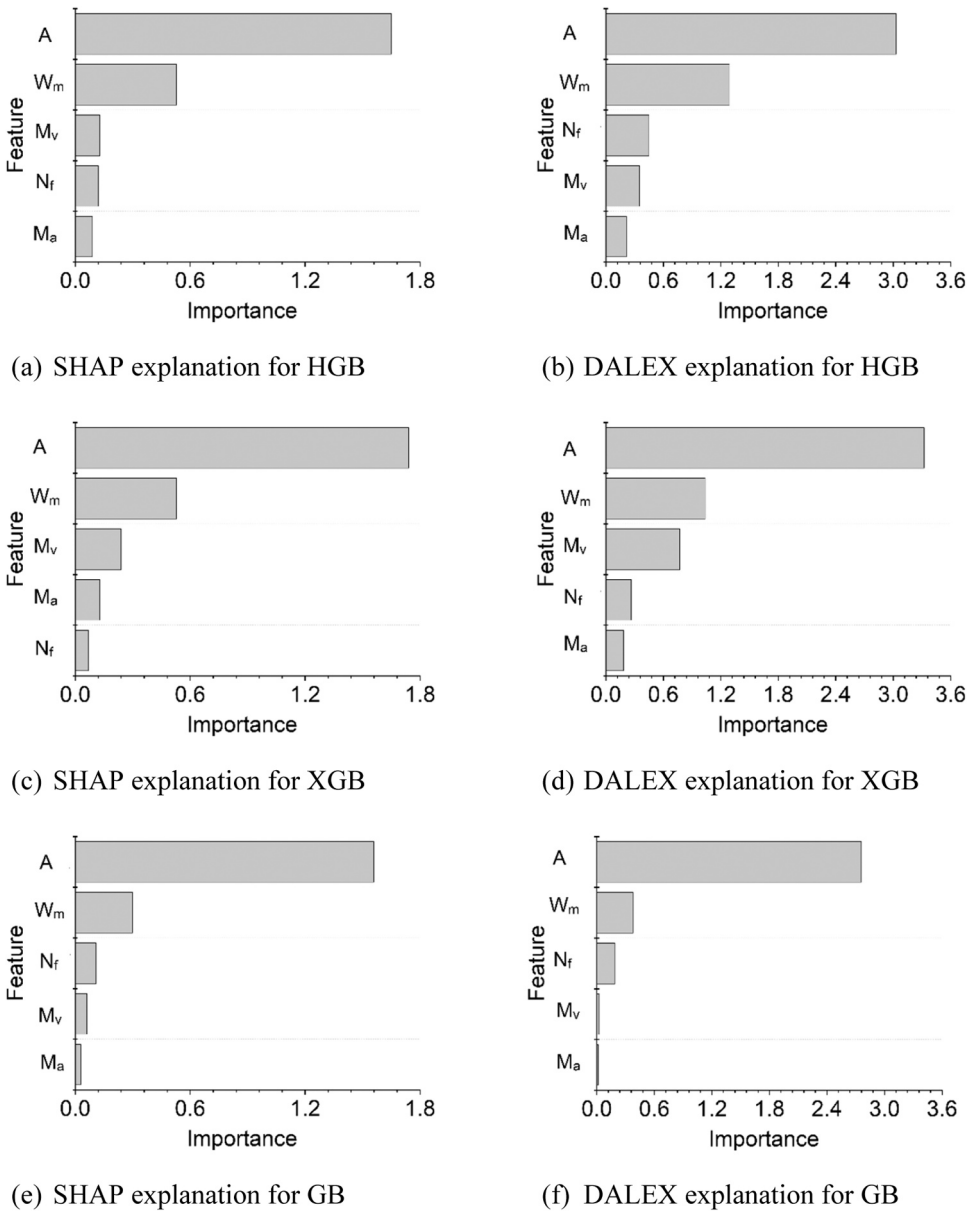


Fig. 6. Absolute global feature importance.

5.2. *Instance-based (local) explanations*

Global explanation describes an overview of model predictions and the importance of features. However, particular instances may contribute differently from features not highlighted in the global explanation. In addition, specific instances may require an explanation of "how that value was predicted". For such occasions, the local explanation becomes effective. Figs. 7 and 8 showcase two distinct instances which were explained using SHAP, LIME, and DALEX for each ML model.

For the particular instance (Fig. 7), the contact area of the specimen holds significance. Also, it positively impacts model output, indicating that an increase in contact area causes an increase in F_P. Secondly, W_m has a positive impact on model output. For HGB and XGB models, the effect of W_m is comparable. However, the remaining three features hold relatively lower significance. Moreover, their order of importance changes with the model used and explainable method. SHAP explanation indicates that N_f holds the third-highest positive impact on the model. The same feature has been identified as a negative contributor when the LIME explanation is used. All three explainable models showcase similarities when identifying the least important feature of the XGB model. However, all three explainable methods display that magnitude of negative contributing features is lower than the XGB model.

Fig. 8 indicates that these instances have different feature importance than global explanations. According to all three explainable methods, HGB decides that the effect of W_m is most important for this particular instance. The prediction, F_P increases as $W_m = 3.8$. SHAP displays that the effect of M_v holds the second most importance whereas LIME chooses M_a. DALEX's explanation interprets that the effect of N_f holds second position and the effect of N_f, M_v, and M_a are all positive. SHAP and DALEX agree that negative contribution occurs from A when it is equal to 37.4. However, LIME assigns a lower negative contribution to feature A, considering that $N_f = 91$ has a greater negative effect on this instance. If the contribution percentage is considered, A= 37.4 has a 50% negative contribution and $W_m = 3.8$ has a 28% positive contribution according to the SHAP explanation (Fig. 8). Nevertheless, LIME showcases that contribution of W_m= 3.8 is 84% and limits the

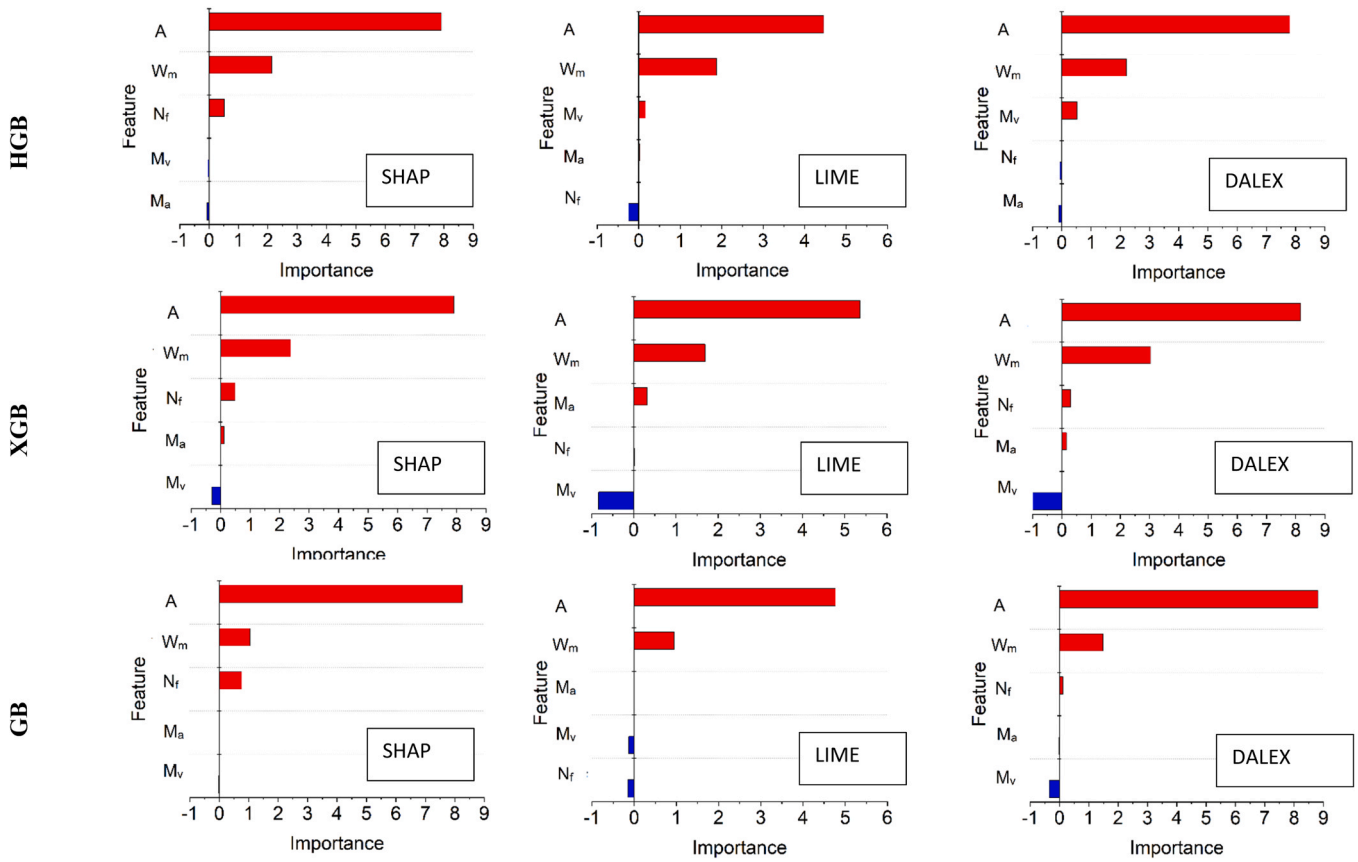


Fig. 7. Instance based explanation for; A = 112.87, W_m = 4.4, M_v = 10.86, N_f = 238, and M_a = 0.1, F_P (experimental) = 12.94 (red color for positive and blue color for negative).

contribution of A= 37.4–6%. DALEX assigns different percentages to each feature with respect to SHAP and LIME. For instance, the feature contribution of W_m = 3.8 is 45% and the corresponding negative contribution of A= 37.4 is 18%. SHAP and DALEX assign positive feature contributions to M_v = 4.99 while LIME assigns negative feature contributions.

The SHAP explanation of the XGB regressor is comparable with the HGB regressor. However, a negative contribution from N_f and M_a has been reduced with respect to HGB regressor. LIME explanation differs from the corresponding explanation for HGB as it assigns A= 37.4 as the highest negative contribution. On the other hand, zero contribution from M_v = 4.99 obtained through LIME for the HGB regressor has increased in the presence of the XGB regressor. A markedly different variation is observed for DALEX interpretation, where the effect of M_v becomes the highest positive. DALEX confirms that the effects of M_a and N_f are insignificant. SHAP assigns 61% negative contribution from A= 37.4% and 34% positive contribution from W_m= 3.8. Corresponding values become 14.6% for A= 37.4% and 67% for W_m= 3.8 for LIME explanation. DALEX decides a 30% positive contribution from M_v = 4.99% and 50% negative contribution from A= 37.4.

Except for feature contribution from W_m and A, the explanation of the GB regressor is different from XGB and HGB models. DALEX assigns almost zero contributions from N_f, M_v, and M_a. Figs. 7 and 8 witness the unique working methodology of each ML model and the distinct working principle of each explanation method. Overall these explanation methods provide realistic interpretations which adhere to experimental observations. For example, the sample instance (Fig. 7) has a lower F_P value due to a lower contact area. The advantage of black-box interpretation is that it is consistent even if many input features are involved. Complex instances can be explained without intervention from a domain expert. Explanations are concise and human-comprehensible, so it is convenient for a non-technical community to understand.

5.3. Partial dependence plots

Fig. 9 shows partial dependence plots obtained using SHAP. The dependence plot describes the feature impact and most interacting variable for a particular feature. If we consider the contact area of the specimen, the most interacted variable is N_f. Higher features of N_f always interact with higher features of A. Moreover, higher A and N_f feature values result in a positive SHAP value (feature importance on model output). All three gradient boosting models showcase a comparable variation.

For feature M_v, the most interacted variable is W_m. SHAP plot showcases a mixed variation where both lower and higher feature values interact with corresponding features of M_v. Despite deviations observed in some instances, overall variation is stalled. SHAP explanations for XGB and HGB describe that fluctuations can be anticipated when higher feature values of M_v are considered. A similar variation is observed for the next feature, N_f. SHAP thinks that N_f mostly interacted with W_m for all regressors. Furthermore, the HGB model differs from stalled variation observed for XGB and GB models.

The feature M_a displays a gradually decreasing variation with its feature values. M_a primarily interacts with N_f for HGB regressor, W_m for XGB and GB regressors. Both XGB and HGB regression models reveal that higher feature values of interacting variables are mostly associated with lower features of M_a. When W_m is considered, ML models identify distinct features. For example, SHAP explains that W_m interacts with M_v for the HGB model and W_m interacts mostly with N_f for XGB and GB models. It is noteworthy that none of the features often interact with M_a. As expected in the feature selection, the feature W_m had a significant impact on the model despite the less correlation observed with F_P. The

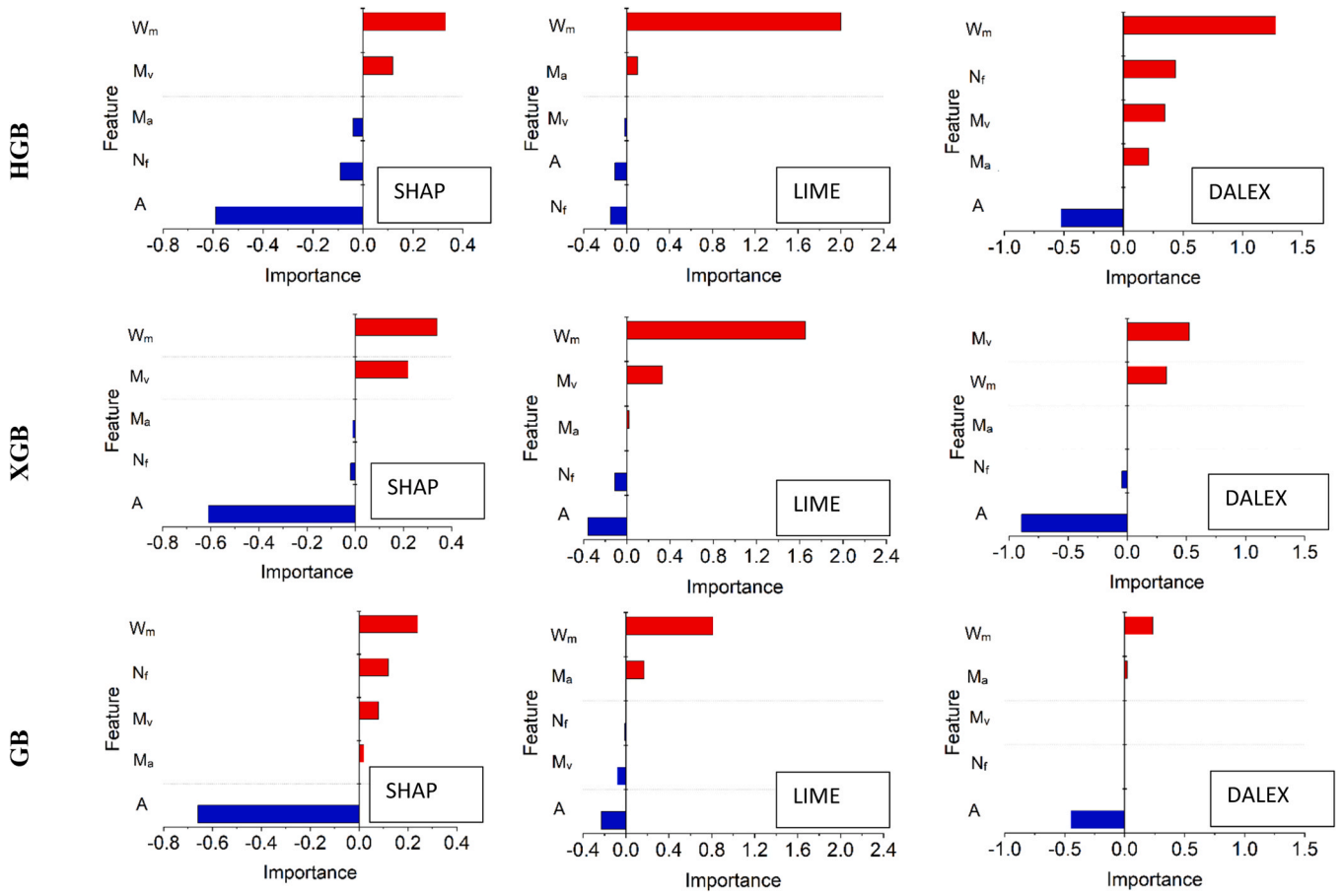


Fig. 8. Instance based explanation for; A = 37.4, W_m = 3.8, M_v = 4.99, N_f = 91, and M_a = 0.4, F_P (experimental) = 2.2 (red color for positive and blue color for negative).

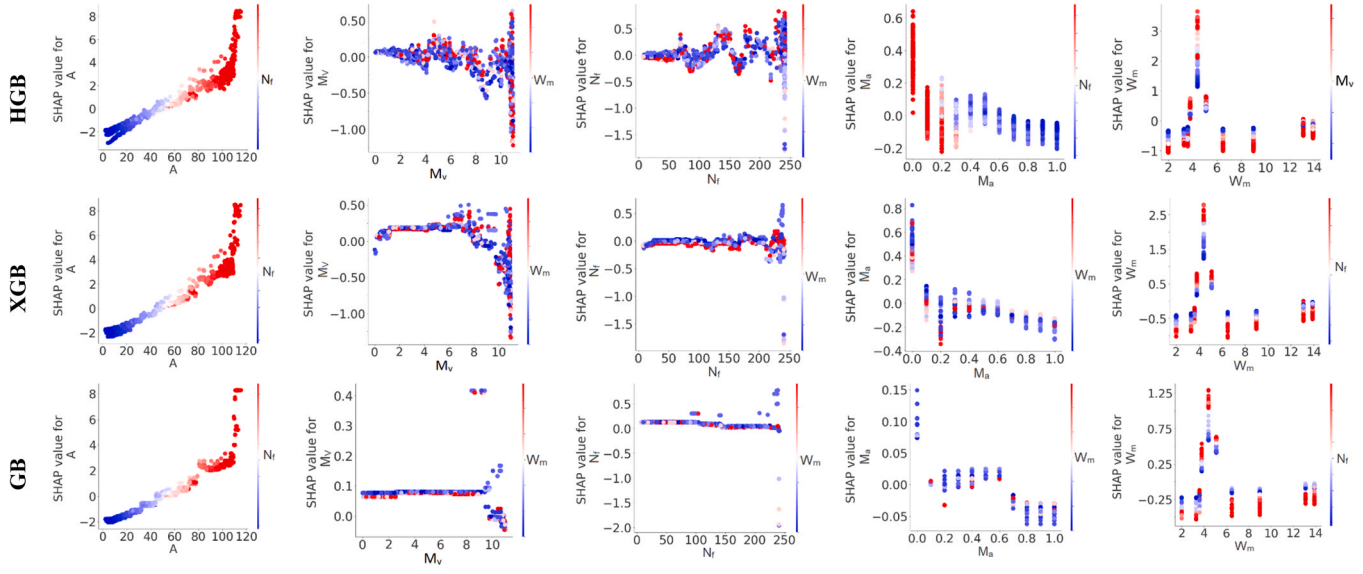


Fig. 9. Partial dependence plots obtained from SHAP.

SHAP curve exhibit that the W_m near 5.0 can drastically increase F_P.

Likewise, Fig. 10 showcases dependence plots obtained using DALEX for each ML model. An increase in feature A always increases the pull-off force, F_P. Conversely, higher features of M_v decrease prediction F_P. Overall, features N_f and W_m have a moderate effect on F_P prediction. However, the local impact observed for W_m at 5, is also observed in the DALEX explanation. This local variation can be the reason that SHAP global importance ranked W_m as the second most dominant variable.

6. Conclusions

This study used gradient boosting ML methods to predict the

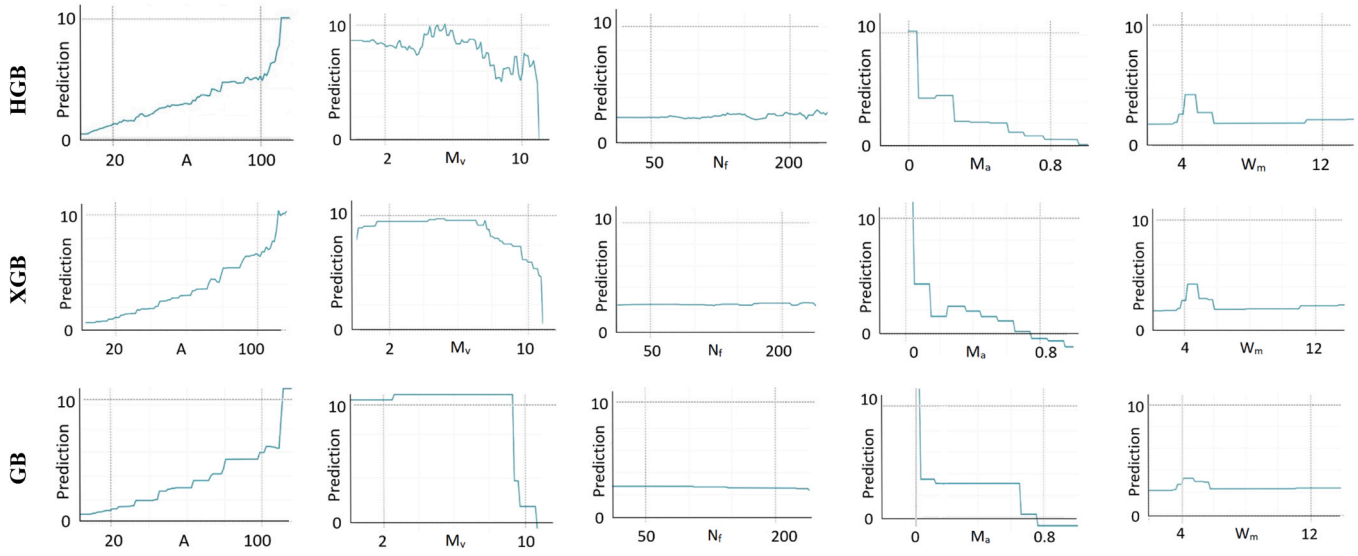


Fig. 10. Partial dependence plots obtained from DALEX.

adhesive strength of micro-patterned surfaces. Models were subsequently interpreted using explainable methods. The following conclusions were made from the study.

- Gradient boosting (GB), extreme gradient boosting (XGB), and histogram-based gradient boosting (HGB) accurately predicted adhesion strength (F_P) on micro-patterned surfaces. XGB and HGB models achieved $R^2 = 0.99$ for training and $R^2 > 0.95$ for validation. Compared to related studies, gradient boosting models are more efficient and enhance prediction accuracy.
- SHAP, LIME, and DALEX explained the underlying reasons behind the predictions by ranking features based on their contributions. Instance-based explanations from SHAP, LIME, and DALEX offer insights into how a specific F_P is obtained, transforming black-box models into glass-box models. These explanations instill confidence in end-users regarding machine learning (ML)-based approaches in engineering applications.
- Weibull modulus significantly impacts the F_P despite showing a weaker correlation with F_P. The misalignment angle exhibits the lowest global feature importance across all models. These explanations assist both technical and non-technical communities in identifying important factors for adhesion force on micro-patterned surfaces without requiring expert intervention.
- In this study, the percentage contribution was considered to compare feature importance values. By employing multiple explanation methods for different ML models, the community gains a deeper understanding of the inner workings of each ML model and the ranking systems employed by different explanation methods. These explainable machine learning (XML) and ML-based frameworks facilitate decision-making regarding adhesive characteristics of micro-patterned surfaces.

CRediT authorship contribution statement

I.U. Ekanayake: Investigation, Data preprocessing, Modelling & Programming, **Sandini Palitha:** Investigation, Methodology, Modelling & Programming, **Sajani Gamage:** Writing – original draft, Formal analysis and Validation, **D. P. P. Meddage:** Writing original draft, Formal analysis, **Kasun Wijesooriya:** Supervision, Writing – review & editing, Conceptualization, **Damith Mohotti**: Supervision, Writing – review & editing, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgements

We sincerely thank Ms. Manar Samri and Prof. Eduard Arzt for providing data for the study. We thank the Department of Civil Engineering, the University of Moratuwa for facilitating research work.

ANNEX A1

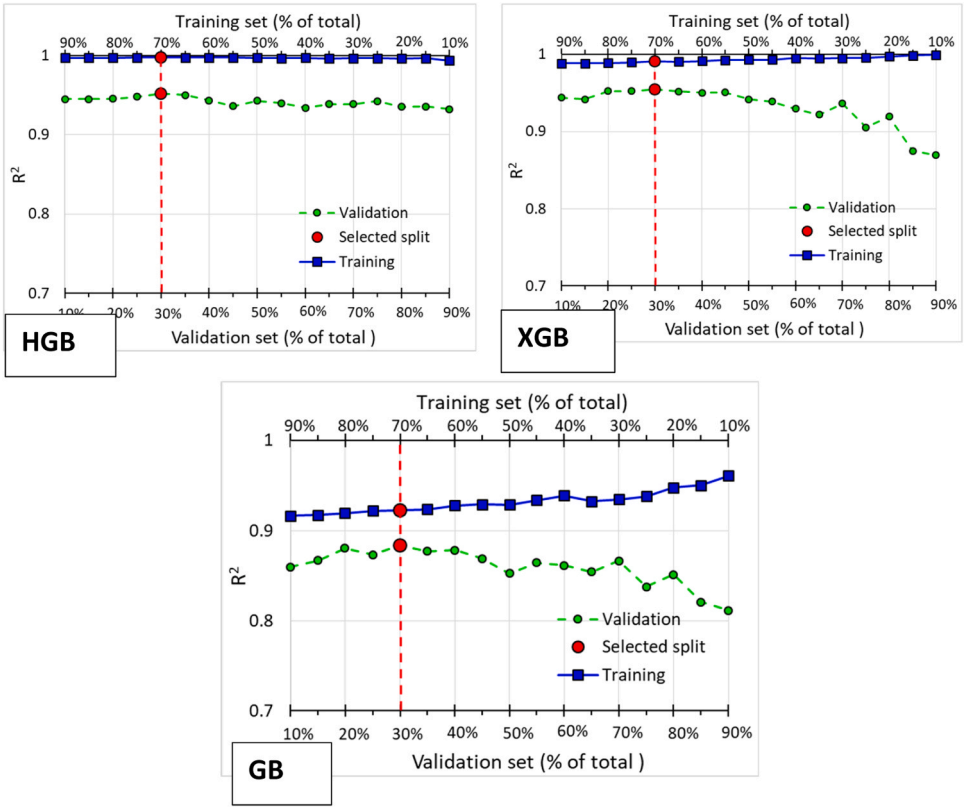
HGB		XGB		GB	
Hyperparameter	Value	Hyperparameter	Value	Hyperparameter	Value
Maximum depth	3	Maximum depth	3	criterion	Friedman_mse
Learning rate	0.2	Gamma	0.0001	Maximum depth	3
Maximum iteration	500	Learning rate	0.3	Learning rate	0.1
Regularization	0	Number of Estimators	50	Minimum sample split	2
Verboose	0	Random state	154	minimum sample leaf	2
Maximum bins	255	Reg_Alpha	0.0001	Bootstrap	FALSE
min_samples_leaf	20	Base score	0.5	Minimum impurity decrease	0

(*continued on next page*)

(*continued*)

HGB		XGB		GB	
Hyperparameter	Value	Hyperparameter	Value	Hyperparameter	Value
				Number of Estimators	25
				Alpha	0.9

ANNEX A2



.

ANNEX A3

$$R^2 = \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(P_i - \overline{O_i})^2} \quad \text{(a)}$$

$$R = \frac{N\sum_{i=1}^{N}(P_i.O_i) - (\sum_{i=1}^{N}P_i.\sum_{i=1}^{N}O_i)}{\sqrt{(N\sum_{i=1}^{N}O_i{}^2 - (\sum_{i=1}^{N}O_i)^2)}.\sqrt{N\sum_{i=1}^{N}P_i{}^2 - (\sum_{i=1}^{N}P_i)^2)}} \quad \text{(b)}$$

$$MAE = \frac{\sum_{i=1}^{N}|O_i - P_i|}{N} \quad \text{(c)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \quad \text{(d)}$$

$$\text{Fractional Bias} = \frac{2(\overline{P_i} - \overline{O_i})}{(\overline{P_i} + \overline{O_i})} \quad \text{(e)}$$

P_i and O_i denote prediction and experimental values, respectively. $\overline{O_i}$ and $\overline{P_i}$ refer to the mean value of the experimental and predicted set.

References

[1] G. Fantoni, et al., Grasping devices and methods in automated production processes, CIRP Ann. vol. 63 (2) (2014) 679–701, https://doi.org/10.1016/j.cirp.2014.05.006.

[2] K. Autumn, et al., Adhesive force of a single gecko foot-hair, Art. no. 6787, Nature vol. 405 (6787) (2000), https://doi.org/10.1038/35015073.

[3] W. Federle, Why are so many adhesive pads hairy? J. Exp. Biol. vol. 209 (14) (2006) 2611–2621, https://doi.org/10.1242/jeb.02323.

[4] S. Gorb, R. Beutel, Evolution of locomotory attachment pads of hexapods, Naturwissenschaften vol. 88 (12) (2001) 530–534, https://doi.org/10.1007/s00114-001-0274-y.

[5] R. Hensel, K. Moh, E. Arzt, Engineering micropatterned dry adhesives: from contact theory to handling applications, Adv. Funct. Mater. vol. 28 (28) (2018) 1800865, https://doi.org/10.1002/adfm.201800865.

[6] S. Song, D.-M. Drotlef, C. Majidi, M. Sitti, Controllable load sharing for soft adhesive interfaces on three-dimensional surfaces, Proc. Natl. Acad. Sci. USA vol. 114 (22) (2017) E4344–E4353, https://doi.org/10.1073/pnas.1620344114.

[7] S.N. Gorb, M. Sinha, A. Peressadko, K.A. Daltorio, R.D. Quinn, Insects did it first: a micropatterned adhesive tape for robotic applications, Bioinspiration Ampmathsemicolon Biomim. vol. 2 (4) (2007) S117–S125, https://doi.org/10.1088/1748-3182/2/4/S01.

[8] E.W. Hawkes, E.V. Eason, A.T. Asbeck, M.R. Cutkosky, The gecko's toe: scaling directional adhesives for climbing applications, IEEEASME Trans. Mechatron. vol. 18 (2) (2013) 518–526, https://doi.org/10.1109/TMECH.2012.2209672.

[9] E. Arzt, H. Quan, R.M. McMeeking, R. Hensel, Functional surface microstructures inspired by nature – From adhesion and wetting principles to sustainable new devices, Prog. Mater. Sci. vol. 120 (2021), 100823, https://doi.org/10.1016/j.pmatsci.2021.100823.

[10] M. Kamperman, E. Kroner, A. del Campo, R.M. McMeeking, E. Arzt, Functional adhesive surfaces with 'gecko' effect: the concept of contact splitting, Adv. Eng. Mater. vol. 12 (5) (2010) 335–348, https://doi.org/10.1002/adem.201000104.

[11] J.A. Booth, R. Hensel, Perspective on statistical effects in the adhesion of micropatterned surfaces, Appl. Phys. Lett. vol. 119 (23) (2021), 230502, https://doi.org/10.1063/5.0073181.

[12] R. Hensel, J. Thiemecke, J.A. Booth, Preventing catastrophic failure of microfibrillar adhesives in compliant systems based on statistical analysis of adhesive strength, ACS Appl. Mater. Interfaces vol. 13 (16) (2021) 19422–19429, https://doi.org/10.1021/acsami.1c00978.

[13] J.A. Booth, V. Tinnemann, R. Hensel, E. Arzt, R.M. McMeeking, K.L. Foster, Statistical properties of defect-dependent detachment strength in bioinspired dry adhesives, p. 20190239, J. R. Soc. Interface vol. 16 (156) (2019), https://doi.org/10.1098/rsif.2019.0239.

[14] M. Bacca, J.A. Booth, K.L. Turner, R.M. McMeeking, Load sharing in bioinspired fibrillar adhesives with backing layer interactions and interfacial misalignment, J. Mech. Phys. Solids vol. 96 (2016) 428–444, https://doi.org/10.1016/j.jmps.2016.04.008.

[15] V. Barreau, R. Hensel, N.K. Guimard, A. Ghatak, R.M. McMeeking, E. Arzt, Fibrillar elastomeric micropatterns create tunable adhesion even to rough surfaces, Adv. Funct. Mater. vol. 26 (26) (2016) 4687–4694, https://doi.org/10.1002/adfm.201600652.

[16] J.A. Booth, M. Bacca, R.M. McMeeking, K.L. Foster, Benefit of backing-layer compliance in fibrillar adhesive patches—resistance to peel propagation in the presence of interfacial misalignment, Adv. Mater. Interfaces vol. 5 (15) (2018) 1800272, https://doi.org/10.1002/admi.201800272.

[17] M. Samri, J. Thiemecke, E. Prinz, T. Dahmen, R. Hensel, E. Arzt, Predicting the adhesion strength of micropatterned surfaces using supervised machine learning, Mater. Today vol. 53 (2022) 41–50, https://doi.org/10.1016/j.mattod.2022.01.018.

[18] W.-S. Kim, I.-H. Yun, J.-J. Lee, H.-T. Jung, Evaluation of mechanical interlock effect on adhesion strength of polymer–metal interfaces using micro-patterned surface topography, Int. J. Adhes. Adhes. vol. 30 (6) (2010) 408–417, https://doi.org/10.1016/j.ijadhadh.2010.05.004.

[19] R.M. McMeeking, E. Arzt, A.G. Evans, Defect dependent adhesion of fibrillar surfaces, J. Adhes. vol. 84 (7) (2008) 675–681, https://doi.org/10.1080/00218460802255558.

[20] S. Bettscheider, et al., Breakdown of continuum models for spherical probe adhesion tests on micropatterned surfaces, J. Mech. Phys. Solids vol. 150 (2021), 104365, https://doi.org/10.1016/j.jmps.2021.104365.

[21] A. Berardo, G. Costagliola, S. Ghio, M. Boscardin, F. Bosia, N.M. Pugno, An experimental-numerical study of the adhesive static and dynamic friction of micro-patterned soft polymer surfaces, Mater. Des. vol. 181 (2019), 107930, https://doi.org/10.1016/j.matdes.2019.107930.

[22] E. Kroner, R. Maboudian, E. Arzt, Adhesion characteristics of PDMS surfaces during repeated pull-off force measurements, Adv. Eng. Mater. vol. 12 (5) (2010) 398–404, https://doi.org/10.1002/adem.201000090.

[23] N. Cañas, M. Kamperman, B. Völker, E. Kroner, R.M. McMeeking, E. Arzt, Effect of nano- and micro-roughness on adhesion of bioinspired micropatterned surfaces, Acta Biomater. vol. 8 (1) (2012) 282–288, https://doi.org/10.1016/j.actbio.2011.08.028.

[24] J.A. Booth, R. Hensel, Perspective on statistical effects in the adhesion of micropatterned surfaces, Appl. Phys. Lett. vol. 119 (23) (2021), 230502, https://doi.org/10.1063/5.0073181.

[25] V. Belle, I. Papantonis, Principles and practice of explainable machine learning, Front. Big Data vol. 4 (2021), 688969, https://doi.org/10.3389/fdata.2021.688969.

[26] R. Roscher, B. Bohn, M.F. Duarte, J. Garcke, Explainable machine learning for scientific insights and discoveries, IEEE Access vol. 8 (2020) 42200–42216, https://doi.org/10.1109/ACCESS.2020.2976199.

[27] P. Meddage, I. Ekanayake, U.S. Perera, H.M. Azamathulla, M.A. Md Said, U. Rathnayake, Interpretation of machine-learning-based (Black-box) wind pressure predictions for low-rise gable-roofed buildings using shapley additive explanations (SHAP, Art. no. 6, Buildings vol. 12 (6) (2022), https://doi.org/10.3390/buildings12060734.

[28] D.P.P. Meddage, I.U. Ekanayake, S. Herath, R. Gobirahavan, N. Muttil, U. Rathnayake, Predicting bulk average velocity with rigid vegetation in open channels using tree-based machine learning: a novel approach using explainable artificial intelligence, Art. no. 12, Sensors vol. 22 (12) (2022), https://doi.org/10.3390/s22124398.

[29] Geoffrey K.F. Tso, K.W.Yau Kelvin, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, Energy 32.9 (2007) 1761–1768.

[30] I.U. Ekanayake, D.P.P. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP), Case Stud. Constr. Mater. vol. 16 (2022), e01059, https://doi.org/10.1016/j.cscm.2022.e01059.

[31] F. Pedregosa, et al., Scikit-learn: machine learning in python, J. Mach. Learn. Res. vol. 12 (85) (2011) 2825–2830.

[32] H. Nhat-Duc, T. Van-Duc, Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification, Autom. Constr. vol. 148 (2023), 104767, https://doi.org/10.1016/j.autcon.2023.104767.

[33] A. Guryanov, Histogram-based algorithm for building gradient boosting ensembles of piecewise linear decision trees (in Lecture Notes in Computer Science), in: W.M. P. van der Aalst, V. Batagelj, D.I. Ignatov, M. Khachay, V. Kuskova, A. Kutuzov, S. O. Kuznetsov, I.A. Lomazova, N. Loukachevitch, A. Napoli, P.M. Pardalos, M. Pelillo, A.V. Savchenko, E. Tutubalina (Eds.), Analysis of Images, Social Networks and Texts, Springer International Publishing, Cham, 2019, pp. 39–50, https://doi.org/10.1007/978-3-030-37334-4_4 (in Lecture Notes in Computer Science).

[34] A. Ogunleye, Q.-G. Wang, XGBoost model for chronic kidney disease diagnosis, IEEE/ACM Trans. Comput. Biol. Bioinform. vol. 17 (6) (2020) 2131–2140, https://doi.org/10.1109/TCBB.2019.2911071.

[35] Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, C. Li, Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration, Eng. Comput. vol. 38 (5) (2022) 4145–4162, https://doi.org/10.1007/s00366-021-01393-9.

[36] J. Nobre, R.F. Neves, Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets, Expert Syst. Appl. vol. 125 (2019) 181–194, https://doi.org/10.1016/j.eswa.2019.01.083.

[37] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16, Association for Computing Machinery, New York, NY, USA, . 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[38] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. vol. 29 (5) (2001) 1189–1232.

[39] S. Demir, E.K. Sahin, An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost, Neural Comput. Appl. vol. 35 (4) (2023) 3173–3190, https://doi.org/10.1007/s00521-022-07856-4.

[40] M.H.L. Louk, B.A. Tama, Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system, Expert Syst. Appl. vol. 213 (2023), 119030, https://doi.org/10.1016/j.eswa.2022.119030.

[41] M.A. Ahmad, C. Eckert, A. Teredesai, Interpretable machine learning in healthcare. *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, in BCB '18, Association for Computing Machinery,, New York, NY, USA, 2018, pp. 559–560, https://doi.org/10.1145/3233547.3233667.

[42] O. Sagi, L. Rokach, Explainable decision forest: transforming a decision forest into an interpretable tree, Inf. Fusion vol. 61 (2020) 124–138, https://doi.org/10.1016/j.inffus.2020.03.013.

[43] M.T. Ribeiro, S. Singh, C. Guestrin, 'Why Should I Trust You?': Explaining the Predictions of Any Classifier HLT-NAACL Demos 2016 doi: 10.1145/2939672.2939778.

[44] V. Petsiuk A. Das K. Saenko RISE: Randomized Input Sampling for Explanation of Black-box Models ArXiv180607421 Cs Jun. 2018.Accessed: Apr. 11, 2021. [Online]. Available⟨http://arxiv.org/abs/1806.07421⟩.

[45] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, Curran Associates Inc., NY, USA, . 2017, pp. 4768–4777.

[46] Y. Liang, S. Li, C. Yan, M. Li, C. Jiang, Explaining the black-box model: a survey of local interpretation methods for deep neural networks, Neurocomputing vol. 419 (2021) 168–182, https://doi.org/10.1016/j.neucom.2020.08.011.

[47] Y. Aydin, B. Dizdaroğlu, Blotch detection in archive films based on visual saliency map, Complexity (2020), https://doi.org/10.1155/2020/5965387.

[48] R. Fong, A. Vedaldi, Interpretable Explanations of Black Boxes by Meaningful Perturbation 2017 IEEE Int. Conf. Comput. Vis. ICCV Oct. 2017 3449 3457 doi: 10.1109/ICCV.2017.371.

[49] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks. Computer Vision – ECCV 2014, Springer, Cham, 2014, pp. 818–833, https://doi.org/10.1007/978-3-319-10590-1_53.

[50] M. Moradi, M. Samwald, Post-hoc explanation of black-box classifiers using confident itemsets (doi: DOI), Expert Syst. Appl. (2021), https://doi.org/10.1016/j.eswa.2020.113941.
[51] D. Garreau, U. Luxburg, Explaining the explainer: a first theoretical analysis of LIME. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, PMLR, . 2020, pp. 1287–1296. Accessed: Jun. 10, 2023. [Online]. Available: ⟨https://proceedings.mlr.press/v108/garreau20a.html⟩.
[52] D. Garreau, U. von Luxburg, Looking Deeper into Tabular LIME." arXiv, Jul. 18 2022 doi: 10.48550/arXiv.2008.11092.
[53] K.R. Chowdhury, A. Sil, S.R. Shukla, Explaining a black-box sentiment analysis model with local interpretable model diagnostics explanation (LIME) (in Communications in Computer and Information Science), in: M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, T. Ören, V.R. Sonawane (Eds.), Advances in Computing and Data Sciences, Springer International Publishing, Cham, 2021, pp. 90–101, https://doi.org/10.1007/978-3-030-81462-5_9 (in Communications in Computer and Information Science).
[54] T. Grimes, E. Church, W. Pitts, L. Wood, Explanation of unintended radiated emission classification via LIME arXiv, Sep 08 2020 doi: 10.48550/arXiv.2009.02418.
[55] H. Hakkoum, A. Idri, I. Abnane, Artificial neural networks interpretation using LIME for breast cancer diagnosis (Advances in Intelligent Systems and Computing), in: Á. Rocha, H. Adeli, L.P. Reis, S. Costanzo, I. Orovic, F. Moreira (Eds.), Trends and Innovations in Information Systems and Technologies, Springer International Publishing, Cham, 2020, pp. 15–24, https://doi.org/10.1007/978-3-030-45697-9_2 (Advances in Intelligent Systems and Computing).
[56] I. Ullah, K. Liu, T. Yamamoto, M. Zahid, A. Jamal, Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction, Travel Behav. Soc. vol. 31 (2023) 78–92, https://doi.org/10.1016/j.tbs.2022.11.006.
[57] D.-C. Feng, W.-J. Wang, S. Mangalathu, E. Taciroglu, Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls, J. Struct. Eng. vol. 147 (11) (2021) 04021173, https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115.
[58] G. Owen, Game Theory, Emerald Group Publishing, 2013.
[59] H. Baniecki, W. Kretowicz, P. Piatyszek, J. Wisniewski ,P. Biecek dalex: Responsible Machine Learning with Interactive Explainability and Fairness in Python." arXiv Oct. 11, 2021 doi: 10.48550/arXiv.2012.14406.
[60] P. Biecek, DALEX: Explainers for Complex Predictive Models in R, *J Mach Learn Res*, 2018.
[61] T. Srinath, G. H.s Explainable machine learning in identifying credit card defaulters Glob. Transit. Proc. vol. 3 1 2022 tanuj doi: 10.1016/j.gltp.2022.04.025.