IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 August 2024, accepted 21 August 2024, date of publication 27 August 2024, date of current version 5 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450511

RESEARCH ARTICLE

Machine Learning Assisted Tensile Strength Prediction and Optimization of Ti Alloy

JAKA FAJAR FATRIANSYAH[1,2], MUHAMAD RAFI AQILA[1], IPING SUHARIADI[3], ANDREAS FEDERICO[1], DZAKY IMAN AJIPUTRO[1], AGRIN FEBRIAN PRADANA[1], YOSSI ANDREANO[1], MUHAMMAD ALI YAFI RIZKY[1], DONANTA DHANESWARA[1,2], ZAINOVIA LOCKMAN[4], AND SU-MI HUR[5]

[1]Department of Metallurgical and Materials Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Jawa Barat 16424, Indonesia
[2]Advanced Materials Research Center (AMRC), Faculty of Engineering, Universitas Indonesia, Depok, Jawa Barat 16424, Indonesia
[3]Department of Industrial Engineering, Faculty of Engineering, Bina Nusantara University, Jakarta 11480, Indonesia
[4]School of Materials and Mineral Resources Engineering, Engineering Campus, Universiti Sains Malaysia, Penang 14300, Malaysia
[5]School of Polymer Science and Engineering, Chonnam National University, Gwangju 61186, Republic of Korea

Corresponding author: Jaka Fajar Fatriansyah (jakafajar@ui.ac.id)

This work was supported by the Directorate of Research and Development, Universitas Indonesia, through Hibah Publikasi Terindeks Internasional (PUTI) 2023, under Grant NKB-520/UN2.RST/HKP.05.00/2023.

ABSTRACT Understanding the link between material composition and mechanical properties is crucial for material design and optimization. This study utilized four machine learning models—K-Nearest Neighbor (KNN), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN)—to predict tensile strength and examine how alloying composition affects the tensile strength of titanium (Ti) alloys. The first three models are considered simpler machine learning approaches, while the ANN is a deep learning method. The models were trained on publicly available experimental data, using fifteen alloying elements as inputs: Al, C, Cr, Cu, H, Fe, Mo, Ni, Nb, N, O, Si, Sn, V, and Zr. A new framework was proposed for evaluating the best model, which involves running several ML models, assessing metrics such as R^2 value, absolute percentage error distribution, and stability of R^2 through multiple trials, and finally comparing model accuracy using the Diebold-Mariano test. The evaluation showed that all four models achieved good accuracy, with R^2 values above 80%. However, the framework identified KNN as the best model due to its low error rate, the narrowest range in absolute percentage error, and more stable R^2 value. Additionally, the feature importance analysis highlighted how alloying elements impact tensile strength, revealing both linear and non-linear correlations. It was found that increasing Ti content and using alloying elements with an atomic radius smaller than Ti affect the tensile strength. This study illustrates the potential of machine learning in material screening and design for Ti alloys.

INDEX TERMS Machine learning, Ti alloy, mechanical property, tensile strength.

I. INTRODUCTION

Titanium alloys are known to be highly coveted for their outstanding properties, including excellent corrosion resistance, toughness, and strength. These attributes make them indispensable not only for lightweight structural applications but also in a number of industries such as aerospace, high-temperature environments, advanced manufacturing, and biomedicine [1], [2], [3]. Despite their advantages, Ti alloys can be somewhat semi-brittle, exhibiting different mechanical behaviors when compared to metals and ceramics [4]. As a consequence, this necessitates ongoing development of new titanium alloys to ensure greater safety and effectiveness in high-strength applications. A thorough understanding of the relationship between alloy structure and mechanical properties, especially strength, is, therefore, essential for creating these new alloys. However, screening numerous alloy compositions to optimize strength is a time-consuming process [5].

Against this backdrop, machine learning (ML) offers a robust solution for identifying complex patterns between

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

inputs and outputs, proving to be a valuable tool in material science for predicting properties and discovering new materials. It complements experimental work and serves as an alternative to traditional computational methods like density functional theory and molecular dynamics, which are often resource-intensive and time-consuming. With proper model tuning, ML can significantly reduce the time and resources needed for material prediction and generation [6].

Researchers have increasingly applied ML algorithms to predict properties of Ti alloys and to develop new materials. For instance, Li et al. used artificial neural networks to forecast the mechanical properties of forged TC 11 titanium alloys [7]. Tkachenko employed random forests and the Kolmogorov-Gambol polynomial to classify titanium alloys based on various powder characteristics, achieving high classification accuracy [8]. Paul et al. utilized decision tree variants to predict the thermal profiles of additively manufactured titanium alloys with 99% accuracy [9]. Liu et al. applied K-means clustering to link mechanical properties with microstructure [10]. Zhu et al. used the Back Propagation (BP) neural network method to explore the connection between alloy composition and properties in Mo and Cr-added titanium alloys, finding close agreement between ML predictions and experimental results [11]. Zou et al. combined data mining with ML to enhance the strength and flexibility of titanium alloys, utilizing high-throughput first-principles calculations as training data. Though effective, this approach may be less accessible for industrial applications [12]. Zhan et al. applied a random forest model to predict the fatigue life of titanium alloys using both experimental and simulated fatigue life data [13]. Banerjee et al. employed support vector machine (SVM) classifiers to predict the mechanical properties of additively manufactured titanium alloys, incorporating data on laser power, scan speed, hatch spacing, and layer thickness [14]. Zou et al. also applied XGBoost, an enhanced decision tree method, to predict the relative density of Ti-6Al-4V [15].

Although numerous studies have explored the use of machine learning (ML) for modeling titanium (Ti) alloys, there has yet to be research establishing a general correlation between alloy structure and mechanical properties, particularly tensile strength. This paper employs both shallow and deep learning ML techniques to predict and investigate this correlation. The challenge of limited experimental data can affect ML prediction accuracy, as noted by Cui and Gong, who found that the accuracy of regression ML methods declines significantly with smaller sample sizes [16]. However, with the appropriate ML model or algorithm, acceptable prediction performance can still be achieved for materials engineering applications, even with limited data [17]. This study proposes a framework for predicting alloy properties using ML with a small dataset and aims to identify the best ML model based on the average R^2 score for predicting the tensile strength of Ti alloys from composition alone. Additionally, it seeks to explore how various alloying elements influence tensile strength.

II. METHODS

A. DATASET

Experimental data on titanium alloy composition information was obtained from the publicly available MatWeb (https://www.matweb.com/). In this study, titanium alloys containing the main element and 15 (fifteen) elements: Al, C, Cr, Cu, H, Fe, Mo, Ni, Nb, N, O, Si, Sn, V, and Zr, were used. The input and output parameters were elemental composition and tensile strength, respectively, as displayed in Table 1. The "iloc[]" function from the Pandas library was employed to extract input and output variables, which can extract specific columns and rows from the data set. Initially, the total data of titanium alloy obtained from the MatWeb website was 342 data. Following the data cleaning process, the titanium alloy data was then reduced to 213 to be processed in the machine learning model.

TABLE 1. 16 features and 1 output of titanium alloys.

No	Data	Unit	Description
		Input (composition)	
1	Ti	wt. %	Titanium Content in Titanium Alloys
2	Al	wt. %	Aluminium Content in Titanium Alloys
3	C	wt. %	Carbon Content in Titanium Alloys
4	Cr	wt. %	Chromium Content in Titanium Alloys
5	Cu	wt. %	Copper Content in Titanium Alloys
6	H	wt. %	Hydrogen Content in Titanium Alloys
7	Fe	wt. %	Iron Content in Titanium Alloys
8	Mo	wt. %	Molybdenum Content in Titanium Alloys
9	Ni	wt. %	Nickel Content in Titanium Alloys
10	Nb	wt. %	Niobium Content in Titanium Alloys
11	N	wt. %	Nitrogen Content in Titanium Alloys
12	O	wt. %	Oxygen Content in Titanium Alloys
13	Si	wt. %	Silicon Content in Titanium Alloys
14	Sn	wt. %	Tin Content in Titanium Alloys
15	V	wt. %	Vanadium Content in Titanium Alloys
16	Zr	wt. %	Zirconium Content in Titanium Alloys
		Output (mechanical properties)	
1	Tensile Strength	MPa	Tensile Strength of Material

B. MACHINE LEARNING MODELS AND EXPERIMENTAL SETUP

In this study, Python version 3.8 was used to develop property prediction models for the tensile strength of titanium alloys, employing KNN, SVR, XGBoost, and ANN. KNN, SVR, and XGBoost represent simpler machine learning methods, while ANN represents a more advanced deep learning approach. The advantages and disadvantages of these methods are summarized in Table 2. The Scikit-learn Python library was used to build the KNN and SVR models. In the KNN model, the K parameter (Equation (1)) determines the number of nearest data points considered for prediction. The Euclidean distance, used to calculate the distance between the prediction data points and other data points, is described by Equation (2) [18].

$$\hat{y}(x) = \frac{1}{K} \sum x_i \in N_{K(x)} y_i \tag{1}$$

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{2}$$

The radial basis function (RBF) kernel, γ, and C parameters were used to build the SVR model. The SVR model offers various mapping capabilities, and its performance depends significantly on the kernel's ability to project low-dimensional data into higher dimensions and transform linear SVR models into non-linear ones. Equation (3) describes the influence of distance from a data point, affecting the curvature of the line or plane defined by the SVR model. This kernel can be applied to nearly any type of data [19]. Equation (4) demonstrates how parameter C, which is closely related to the slack variable (ξ), impacts the amount of variance the SVR model can tolerate [20]. Parameter C acts as a penalty factor applied to any kernel, balancing model confidence against empirical risk.

$$\gamma = \frac{1}{2\sigma^2} \tag{3}$$

$$T = min : \frac{1}{2}||w||^2 + C \sum_{i-1}^{l} (\xi_i + \xi_i^*) \tag{4}$$

The XGBoost has a unique architecture in the form of incorporated decision trees, an extension of the decision tree, and random forest algorithms. Decision trees in XGBoost are dependent on each other and the training process is serial, which means the training of a decision tree depends on the training results of the previous decision tree [21]. The XGBoost model can overcome the drawback of decision trees and random forest models that are prone to overfitting by applying this principle, thereby improving predictive performance [22]. Prediction produced by the XGBoost model can be determined by referring to Equation (5) [23].

$$\bar{y}_i = y_i^0 + \eta \sum_{K=1}^{n} f_k(U_i) \tag{5}$$

where $\bar{y}_i$ denotes the prediction of the XGBoost model for i-th data, y_i^0 represents the mean of the original parameters in the training data, η is the learning rate, n displays the number corresponding to independent tree structures, and U_i represents the parameter vector. XGBoost offers numerous parameters for fine-tuning, including max_depth (the depth of the tree), min_samples_split (the minimum number of samples required to split an internal node), min_samples_leaf (the minimum number of samples required at a leaf node), max_features (the maximum number of features considered for splitting a node), splitter (the strategy used to select splits at each node), criterion (the metric used to evaluate node splits), min_impurity_decrease (a threshold related to node split quality to limit tree growth), n_estimators (the number of decision trees to build), learning_rate (the shrinkage factor to prevent overfitting), min_child_weight (which controls tree complexity by specifying the minimum number of weights), gamma or min_split_loss (the minimum loss reduction required for further partitioning), subsample (the fraction of observations randomly sampled for each tree), and reg_alpha and reg_lambda (which specify L1 (Lasso) and L2 (Ridge) regularization of the weights) [24].

The ANN model was constructed using the TensorFlow library, with key parameters including learning rate, batch size, number of epochs, number of nodes in each layer, and number of hidden layers. The epoch represents the number of complete cycles through the training dataset. The learning rate determines how quickly the ANN adjusts weights based on gradient loss during each epoch, while batch size indicates the amount of data processed at one time. The model architecture is detailed in Table 2.

TABLE 2. The advantages and disadvantages of KNN, SVR, XGBoost, and ANN.

No	Model	Advantages	Disadvantages
1	KNN	Simple, robust to noisy training data, work well with small number of data set [25]	Computationally expensive, too sensitive to redundant feature [25]
2	SVR	Simple, work well with small number of data set	Difficulty in determining hyperparameter [26]
3	XGBoost	Computationaly efficient, good at handling missing data [27], work well in small number of data set	Not suitable for very complex data.
4	ANN	Adaptive learning machine, good for complex features, robustness on noisy data [28].	Blackbox model [29], not suitable for small number of data set.

The hardware that was used to run the models was a computer desktop with a 12-core 4.6 GHz processor, memory of 32 GB, and storage of 1 TB. The text editor for the Python program was written using Jupyter Notebook. The libraries used were Pandas for data processing, Numpy for

mathematical operations, Scikit Learn for machine learning features, Matplotlib for visualization, Tensorflow, and Keras for machine learning model customization.

C. PREDICTION PERFORMANCE EVALUATION

The R^2 value and Root Mean Square Error (RMSE) calculation were used to assess the quality of the model. The R^2 value was obtained from Equation (6) which is given as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{6}$$

where n, y, $\bar{y}_i$, dan $\hat{y}_i$ represent the sum of the data, the actual value, the average value, and the model's predicted value, respectively. The R^2 value refers to an evaluation metric that shows the amount of the difference between the actual value in the dataset and the predicted outcome of the model. The R^2 value ranges -∞ to 1. If the R^2 score value is near 1, the accuracy level will be higher [30].

The evaluation metric known as root mean square error or RMSE is denoted by the square root of the average difference between the value predicted by the model and the actual value. The RMSE was calculated using the following formula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{7}$$

According to Tatachar, the RMSE value ranges between 0 and $+\infty$ [31]. A value closer to 0 signifies a smaller difference between the predicted and actual values associated with a greater accuracy level.

Mean absolute percentage error (MAPE) is one of the most used evaluation metrics, which represents the mean absolute difference between the actual and the predicted value. The MAPE value was derived from Equation (8) and it is expressed in the form of percentage [32]. A smaller value indicates a smaller absolute difference and vice-versa.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \tag{8}$$

After obtaining the evaluation matric values for each model, the next step is to determine the best model by implementing a framework as shown in Figure 1. In our proposed framework, the best model is chosen not only from the model that has the highest R^2 value but also from its absolute percentage error distribution and fluctuation of the R^2 score based on various trials with the same configuration of data training and testing. The models that have nearly the same R^2 score will be tested using the Diebold-Mariano test to check whether they have a similar degree of accuracy or not.

D. FEATURE IMPORTANCE ANALYSIS

Three feature importance analysis methods were employed to understand the contribution of each alloying element to the tensile strength prediction models: Pearson Correlation Coefficient (PCC), Spearman's rank correlation coefficient (SCC), and Permutation Feature Importance (PFI). PFI is a measure of a feature's significance in relation to the predictive performance of the model. A higher PFI score indicates that the particular feature will exert a greater influence on the predictive model for a given variable. PFI measures how much the model's prediction error increases when the values of the features are switched around. In this paper, PFI was calculated using XGBoost model with higher predictive performance.

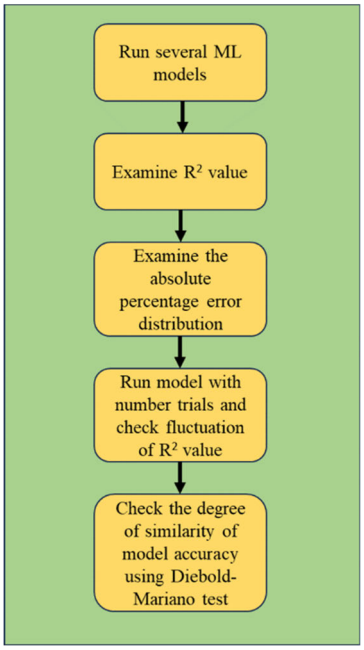

FIGURE 1. The framework for selecting the best ML model.

PCC and SCC analyses were conducted to explore the relationship between alloying elements and tensile strength in more detail, with the results compared to experimental data from the literature. Pearson correlation coefficient (PCC) is a measure to reflect the linear correlation between two variables. The range is between 1 and -1, where positive and negative values give positive and negative associations between two variables. The PCC value was obtained by using the following equation:

$$PCC = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{9}$$

where x and y are the variables to be compared; $\bar{x}$ and $\bar{y}$ are the mean of those two variables. The SCC, on the other hands, measures the monotonic relationship between two variables, where the interval value is similar to that of PCC. SCC value was calculated using the following equation:

$$SPC = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)} \tag{10}$$

where d and n represent difference between two variables and the number of observations, respectively.

Atomic-property-based features were examined using the most effective methods from performance evaluation,

leading to recommendations for selecting alloying elements with atomic radii smaller than titanium to potentially enhance tensile strength. PFI, PCC, and SCC were all integrated into the Atomic-property-based features analysis, providing comprehensive insights into composition-based models for predicting tensile strength in titanium alloys.

E. RESULT AND DISCUSSION

Table 3 and Figure 2 present the statistical details of the dataset and the distribution of tensile strength as the output, respectively. This study used only the composition of titanium alloys as input features, excluding processing parameters. According to predictive regression rules, the number of data points should be at least ten times the number of variables [33]. Given that our dataset comprises only 214 data points, it is relatively small, leading to a larger deviation around the mean.

TABLE 3. Statistical information of the titanium alloys.

Feature	Mean	Min.	Max.	Std. Dev.	
Ti (wt. %)	88.034	44.07	100.00	10.559	Input
Al (wt. %)	3.451	0	13.30	2.741	Input
C (wt. %)	0.061	0	0.20	0.036	Input
Cr (wt. %)	0.377	0	11.00	1.461	Input
Cu (wt. %)	0.031	0	2.50	0.250	Input
H (wt. %)	0.015	0	0.125	0.020	Input
Fe (wt. %)	0.344	0	5	0.512	Input
Mo (wt. %)	1.336	0	15.00	3.065	Input
Ni (wt. %)	1.066	0	55.90	7.521	Input
Nb (wt. %)	0.852	0	45.00	5.177	Input
N (wt. %)	0.033	0	0.20	0.022	Input
O (wt. %)	0.159	0	0.50	0.093	Input
Si (wt. %)	0.042	0	0.60	0.114	Input
Sn (wt. %)	0.815	0	11.00	1.686	Input
V (wt. %)	2.340	0	16.00	3.860	Input
Zr (wt. %)	0.978	0	11.00	1.957	Input
Tensile Strength (MPa)	854.395	220	1450.00	295.820	Output

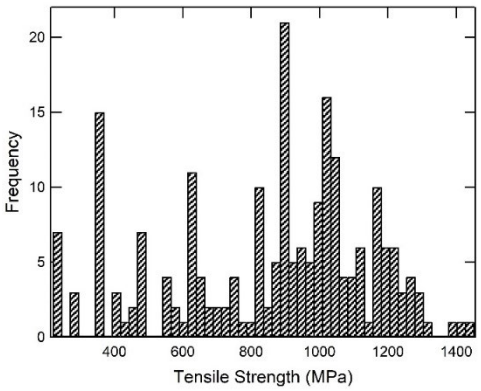

FIGURE 2. Tensile strength distribution of the used dataset.

Table 4 shows the optimized hyperparameters for KNN, SVR, XGBoost, and ANN as determined by Tuner in TensorFlow. Figure 3 illustrates the accuracy of these models,

TABLE 4. The optimized hyperparameter for KNN, SVR, ANN, and XGBoost.

KNN	SVR	ANN	XGBoost
Hyperparameter			
K = 11 Random state = 42	C = 10 Gamma = 0.1 Random state = 0	Layer size = 32-16-32 Optimizer = SGD Learning rate = 0.010 Epoch = 500 Batch size = 16 Random state = 0	n estimators = 97 Max depth = 5 Learning rate = 0.028 Subsample = 1 Criterion = friedman_mse Min samples split = 2 Min samples leaf = 1 Min impurity decrease = 0 Max features = None Random state = 42
MAXIMUM ACCURACY			
86.22%	84.02%	88.69%	83.82%

represented by R^2 scores. For a clearer view, Figure 4 depicts the data distribution between actual and predicted tensile strengths for a selected set of 20 test data points. The x-axis shows the order of 20 test data points. Chicco et al. noted that the R^2 score provides a more informative and accurate measure than other metrics such as symmetric mean absolute percentage error (SMAPE) and mean absolute percentage error (MAPE) [34]. The results indicate that all models achieved good accuracy, with R^2 scores and MAPE values exceeding 80% and falling below 15%, respectively. The discrepancies between the models and experimental results may stem from the exclusion of processing parameters in the input data. Previous research has shown that processing parameters, such as heat treatment, can significantly influence the microstructure and mechanical properties of titanium alloys [11]. Despite this, the models are deemed sufficiently effective for material screening purposes.

Among the four machine learning models used, KNN demonstrated performance comparable to SVR, ANN, and XGBoost, despite its simplicity. The ANN model achieved the highest R^2 value of 88.69%, indicating the best accuracy. However, it is important to note that R^2 scores for ANN can vary with each run due to its inherent stochastic nature [35].

Figure 5 presents a boxplot that measures performance based on the distribution of absolute percentage errors. The plot reveals that while ANN has the lowest mean error, it has a broader distribution range compared to KNN. In contrast, KNN shows the second lowest error with the narrowest distribution range, indicating that it tends to produce results clustered around lower error values. XGBoost, on the other hand, exhibits the highest error and widest distribution range, as evidenced by its lowest R^2 score. Thus, KNN is preferred over XGBoost, despite both models achieving high R^2 values.

The next step in our framework is to assess model reliability by conducting multiple trials with similarly split training and testing data at an 80:20 ratio, as shown in Figure 6. After 10 trials, the R^2 score for the ANN model varied significantly, while the R^2 values for KNN, SVR, and XGBoost remained stable. The ANN mean of the cross-validation value is 82.2 ± 4.2. This suggests that simpler models like KNN, SVR, and XGBoost offer higher reliability due to their

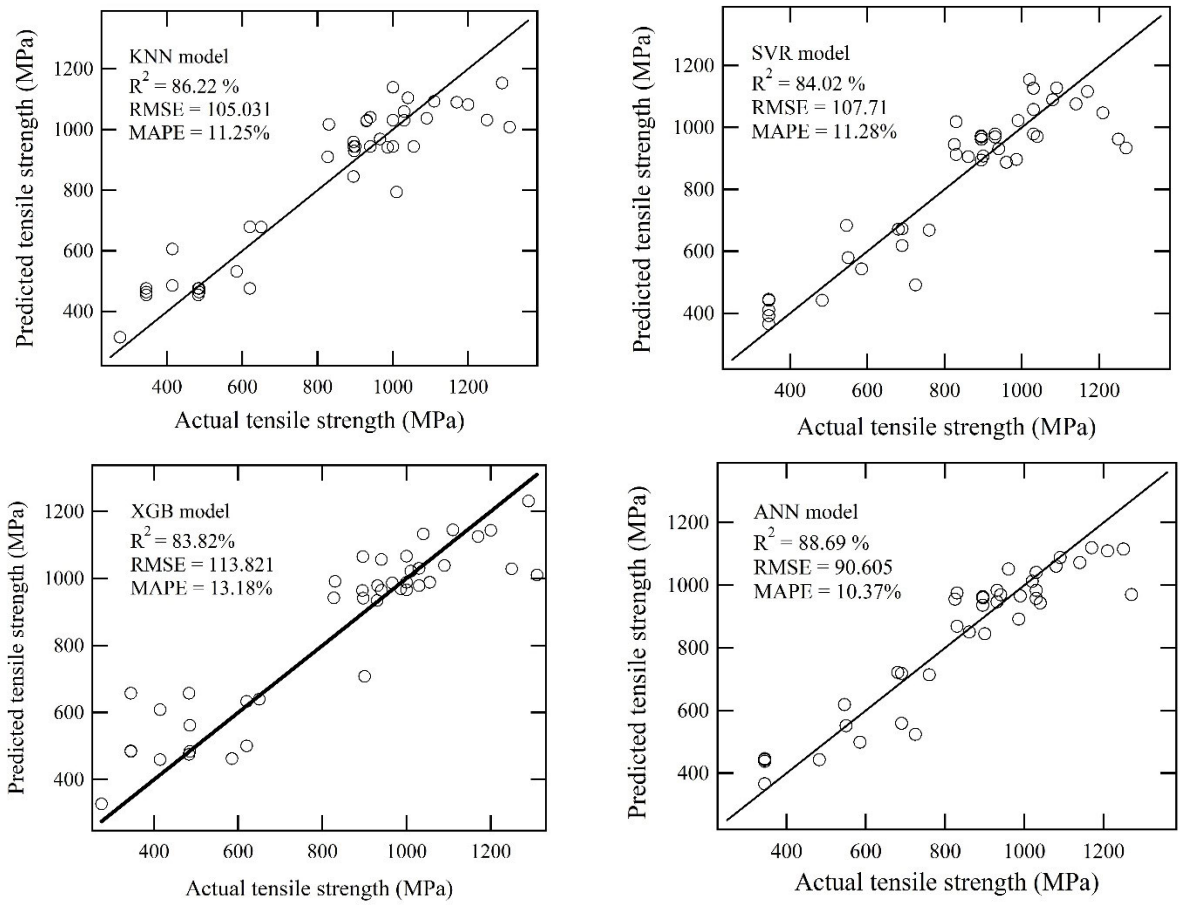

FIGURE 3. R^2, RMSE and MAPE Score of KNN, SVR, XGBoost, and ANN models.

ability to handle smaller datasets more effectively compared to ANN, which requires a more complex relationship between input and output features. Kordos et al. found that KNN with well-tuned parameters performed best in their experiments [36]. Alwosheel et al. recommended that the dataset size should be at least fifty times the number of weights for neural network models [37]. In our case, with 78 weights, this would require at least 3900 datasets, a number challenging to obtain. Therefore, simpler models like KNN, SVR, and XGBoost, which achieved R^2 scores above 80, are preferable as they indicate that the relationship between composition and tensile strength is not overly complex.

In materials science, the limited number of data points is common due to the high cost and complexity of experimental setups. Although generating large datasets through computational methods is feasible, it was not possible in this study. Agrawal and Choudhary [38] highlighted the need to evaluate multiple ML models with limited data and cautioned against relying solely on high R^2 scores due to error distribution issues. Zhang and Ling proposed a targeted property code crude estimation strategy for small datasets [39]. However, this approach may be more challenging for users to implement compared to our strategy, which focuses on selecting an ML model with a high R^2 score.

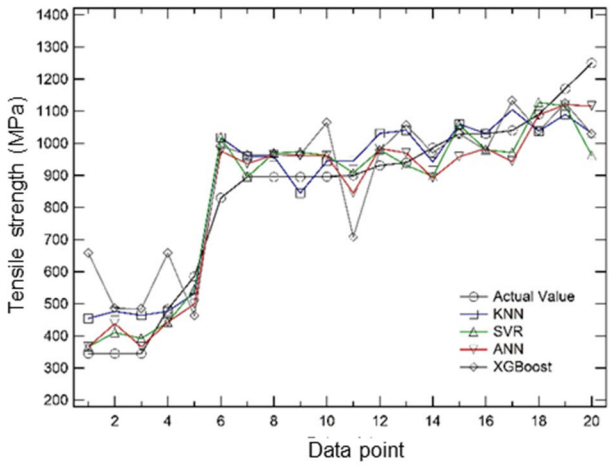

FIGURE 4. Comparisons between actual and predicted tensile strength.

In the next stages of our framework, we analyzed the performance of each model and then compared it with each other by using the Diebold-Mariano test as presented in Table 5. It is worth noting that the test result from the comparison of all models yields statistic values of larger than

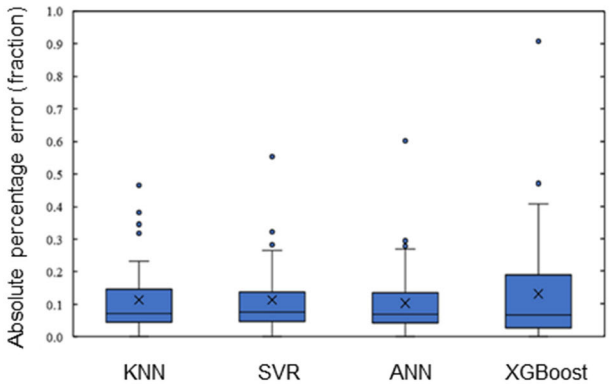

FIGURE 5. Boxplot of prediction absolute percentage error distribution.

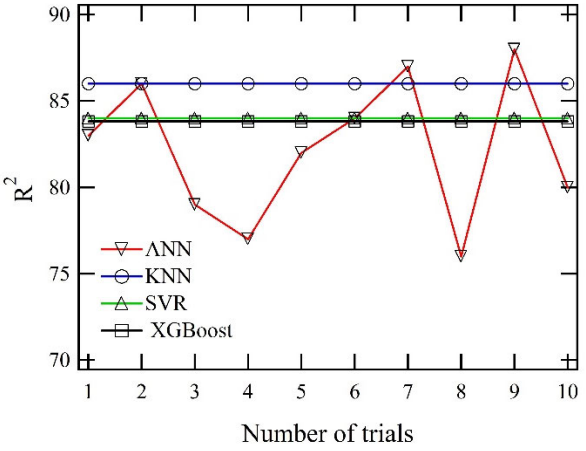

FIGURE 6. R^2 values as a function of the number of trials.

zero indicating that each ML model has a different accuracy from the others. Nonetheless, the p-value reveals that only ANN-XGBoost and SVR-XGBoost model pairs exhibit significant performance differences associated with the values below the generally chosen significant value of 0.05.

Based on the performance evaluation, KNN is identified as the most effective model due to its simplicity, accuracy, and stable performance. Although the p-value indicates no significant performance difference between KNN and other models, the Diebold-Mariano test confirms that KNN's accuracy is distinct from the others.

To investigate the relationship between alloy composition and tensile strength, we calculated the Pearson Correlation Coefficient (PCC), as shown in Table 6. The results indicate that Al has the strongest positive effect on tensile strength, followed by Zr, Sn, V, and Mo. Other alloying elements have a less significant impact on tensile strength compared to these five. Spearman's rank correlation analysis also supports these findings, confirming that Al, Zr, Sn, V, and Mo significantly affect tensile strength.

To validate the reliability of the PCC values from our model, we compared them with existing experimental data. For Al and Sn, adding these elements to titanium alloys in the

TABLE 5. Diebold-mariano test result for each ML model.

Model Comparison	Diebold-Mariano Test Statistic Value	p-value
KNN – SVR	0.290	0.612
SVR – ANN	1.726	0.954
ANN – XGBoost	2.497	0.011
KNN – ANN	1.831	0.958
KNN – XGBoost	0.648	0.260
SVR – XGBoost	1.765	0.047

TABLE 6. PCC and SCC values of titanium alloy.

Alloying Element	PCC Value	SCC value
Ti	-0.579	-0.734
O	-0.369	-0.358
C	-0.339	-0.398
Nb	-0.043	0.013
Cu	-0.032	0.045
N	-0.026	0.048
H	0.085	-0.177
Ni	0.092	-0.103
Fe	0.136	-0.071
Si	0.184	0.202
Cr	0.191	0.213
Mo	0.313	0.467
V	0.358	0.306
Sn	0.365	0.450
Zr	0.376	0.474
Al	0.613	0.531

range of 2 to 6% acts as alpha stabilizers. Huang et al. demonstrated that aluminum increases tensile strength by altering the atomic bond configuration in the alpha-Ti lattice, thereby strengthening it [40]. This change also enhances resistance to dislocation movement. Similarly, Xie et al. found that increasing Sn content boosts yield strength by preventing the transformation from beta to alpha phase. The beta phase's body-centered cubic structure has higher slip resistance compared to the alpha phase's hexagonal close-packed structure, thereby maintaining strength performance [41]. Our findings on Sn align with these experimental results.

In the case of Zr, Si et al. observed that the increase of the Zr element in titanium alloys results in higher strength due to the solid solution strengthening of the α and β phases [42]. Thus, the addition of the Zr element increases tensile strength, which is in line with our PCC calculation. For the V alloying element, An et al. observed that the increase of the V alloying element resulted in an increase in tensile strength despite reductions in fracture strain and impact toughness, and the increase in tensile strength is due to the weaker texture of the α phase in the {0001} plane [43]. For Mo alloying elements, Kotov et al. demonstrated that increasing Mo content to 5% in titanium alloys increased yield and tensile strength due to the promotion of the β-phase fraction above 20% [44].

On the other hand, our PCC calculation showed that O and C alloying elements have a large negative effect on tensile strength. However, observations conducted by Ogden and Jaffee and Morita et al. showed that the increase of O to 5% and C elements resulted in higher tensile strength [45], [46],

TABLE 7. Atomic-property-based features of titanium alloy.

No	Feature	Equation	Mean	Min.	Max.	SD	PCC Value	SCC value	Permutation Feature Importance
1	Ti atomic percentage (a_{Ti})	a_1	0.87	0.49	1.00	0.09	-0.66	-0.703	0.3917
2	Total atomic radius (r_{total})	$\sum_{i=1}^{16} a_i\, r_i$	170.70	162.08	182.52	3.16	-0.597	-0.548	0.5382
3	Atomic radius difference Ti-based (r_{Ti-rel})	$\sqrt{\sum_{i=1}^{16} a_i - (1-\frac{r_i}{r_{Ti}})^2}$	0.11	0	0.21	0.03	0.544	0.539	0.1852
4	Total valence electron (V_{total})	$\sum_{i=1}^{16} a_i\, v_i$	4.04	3.80	7.05	0.42	0.058	-0.233	0.9956
5	Valence electron difference Ti-based (V_{Ti-rel})	$\sqrt{\sum_{i=1}^{16} a_i - (1-\frac{v_i}{v_{Ti}})^2}$	0.04	0	1.14	0.15	0.123	0.484	2.4549
6	Total Pauling electro-negativity (χ_{total})	$\sum_{i=1}^{16} a_i\, \chi_i$	1.25	1.24	1.32	0.01	0.308	0.473	1.3856
7	Electro-negativity difference (χ_{Ti-rel})	$\sqrt{\sum_{i=1}^{16} a_i - (1-\frac{\chi_i}{\chi_{Ti}})^2}$	0.01	0	0.03	0.01	0.021	-0.006	0.2026
8	Total entropy (S_{total})	$\sum_{i=1}^{16} a_i\, S_i$	31.79	30.11	35.71	0.96	-0.235	-0.443	1.0272
9	Entropy difference Ti-based (S_{Ti-rel})	$\sqrt{\sum_{i=1}^{16} a_i - (1-\frac{S_i}{S_{Ti}})^2}$	0.16	0	0.53	0.09	-0.111	-0.212	0.6271

which contradicts our calculation. It is important to note that the PCC only captures linear relationships between input and output features. Our study revealed a non-monotonic relationship between carbon (C) content and the tensile strength of titanium alloys. Ogden and Jaffee observed a similar non-linear relationship between C% content and tensile strength in Ti-7Mn alloys up to 0.3% C content, which supports our findings [45].

To evaluate the importance of each feature on model performance, we used PFI with the SVM model, which showed high accuracy, as illustrated in Figure 7. A higher PFI value indicates a greater contribution of a feature to the model's predictive performance. The results revealed that Al and Ti have the highest PFI values, meaning the model's ability to predict tensile strength significantly depends on the presence of Al and Ti composition data. This underscores the relative importance of Al in comparison to other alloying elements in titanium alloys.

Table 7 presents the atomic-property-based features used to predict tensile strength in titanium alloys. We propose this approach based on the idea that atomic properties are directly related to composition rather than processing conditions. Therefore, we hypothesize that if a machine learning model utilizing atomic properties as inputs can predict tensile strength accurately, then our previous model, which relied solely on atomic species compositions, should also

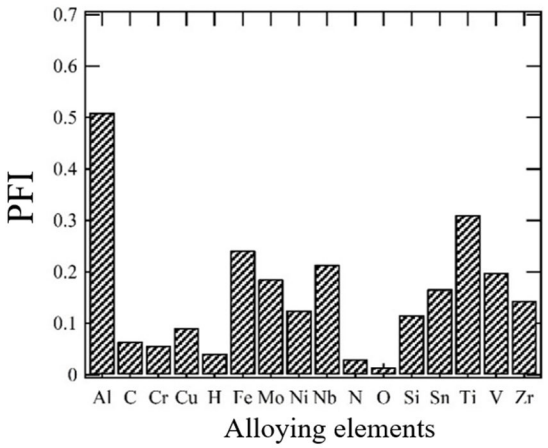

FIGURE 7. Normalized permutation feature importance of 16 features for tensile strength of titanium alloys.

be effective in predicting tensile strength. We employed KNN (with K of 7) and composition data as models and inputs, respectively. The R^2 score, shown in Figure 8, reached 95.17%, demonstrating high model accuracy. This result indicates that atomic-property-based features are capable of predicting tensile strength, validating the use of composition-based features in our initial model. While this

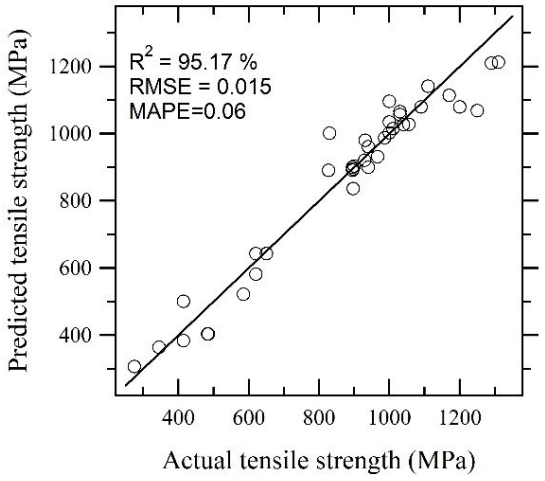

FIGURE 8. Actual versus predicted value of tensile strength of titanium alloy.

does not diminish the importance of processing parameters—which are crucial for further optimizing tensile strength—our composition-based model is sufficient for materials screening applications that do not demand near-perfect accuracy.

For measuring the atomic-property-based correlation to tensile strength, PCC and SCC were conducted as shown in Table 7. In this context, titanium serves as the matrix of the alloy, with the alloying elements acting as solutes within the titanium solvent. The PCC calculations reveal that the difference in atomic radius between the alloying elements and titanium, as well as the atomic percentage of titanium, are the parameters most strongly correlated with tensile strength. Specifically, the atomic radius difference shows a significant positive correlation, while the titanium atomic percentage shows a notable negative correlation. These findings are corroborated by the SCC analysis. The results suggest that tensile strength increases with the addition of alloying elements that have a much smaller atomic radius compared to titanium. Conversely, a high concentration of titanium in the matrix can impair tensile strength.

III. CONCLUSION

Four machine learning models—KNN, SVR, XGBoost, and ANN—were effectively developed to predict the tensile strength of titanium alloys based on their compositions. Our framework identified KNN as the most predictive and stable model compared to SVR, XGBoost, and ANN. The KNN model, utilizing atomic property features, achieved an accuracy of 95.17%, demonstrating that composition-based models are suitable, as these features are only related to composition and not to processing parameters. This result underscores that for pre-screening purposes where accuracy ranges from 85% to 95% is acceptable, composition-based prediction alone is adequate.

Feature importance analysis revealed that the alloying elements Al, Zr, Sn, V, and Mo have the most significant impact on tensile strength compared to other elements. Our modeling showed that the tensile strength of Ti alloys increases with the addition of alloying elements whose atomic radius is smaller than that of titanium. Future research will focus on exploring other mechanical properties, such as fatigue life, hardness, and corrosion resistance. For these properties, however, composition-based features alone may not be sufficient, and incorporating various processing parameters will be necessary.

REFERENCES

[1] G. Chen, Y. Peng, G. Zheng, Z. Qi, M. Wang, H. Yu, C. Dong, and C. T. Liu, "Polysynthetic twinned TiAl single crystals for high-temperature applications," *Nature Mater.*, vol. 15, no. 8, pp. 876–881, Aug. 2016, doi: 10.1038/nmat4677.

[2] N. Stepanov, M. Tikhonovsky, N. Yurchenko, D. Zyabkin, M. Klimova, S. Zherebtsov, A. Efimov, and G. Salishchev, "Effect of cryo-deformation on structure and properties of CoCrFeNiMn high-entropy alloy," *Intermetallics*, vol. 59, pp. 8–17, Apr. 2015, doi: 10.1016/j.intermet.2014.12.004.

[3] K. Lu, "The future of metals," *Science*, vol. 328, no. 5976, pp. 319–320, Apr. 2010, doi: 10.1126/science.1185866.

[4] H. Xiang, Y. Chen, Z. Qi, G. Zheng, F. Chen, Y. Cao, X. Liu, B. Zhou, and G. Chen, "Mechanical behavior of TiAl alloys," *Sci. China Technol. Sci.*, vol. 66, no. 9, pp. 2457–2480, Sep. 2023, doi: 10.1007/s11431-022-2186-9.

[5] J. F. Fatriansyah, I. Suhariadi, H. A. Fauziyyah, I. R. Syukran, F. Hartoyo, D. Dhaneswara, Z. Lockman, A. Fauzi, and M. S. Rohman, "Prediction and optimization of mechanical properties of Ni based and Fe–Ni based super alloys via neural network approach with alloying composition parameter," *J. Mater. Res. Technol.*, vol. 24, pp. 4168–4176, May 2023, doi: 10.1016/j.jmrt.2023.04.065.

[6] J. F. Fatriansyah, E. Kustiyah, S. N. Surip, A. Federico, A. F. Pradana, A. S. Handayani, M. Jaafar, and D. Dhaneswara, "Fine-tuning optimization of poly lactic acid impact strength with variation of plasticizer using simple supervised machine learning methods," *Exp. Polym. Lett.*, vol. 17, no. 9, pp. 964–973, 2023, doi: 10.3144/expresspolymlett.2023.71.

[7] M. Li, X. Liu, and A. Xiong, "Prediction of the mechanical properties of forged TC11 titanium alloy by ANN," *J. Mater. Process. Technol.*, vol. 121, no. 1, pp. 1–4, Feb. 2002.

[8] R. Tkachenko, Z. Duriagina, I. Lemishka, I. Izonin, and A. Trostianchyn, "Development of machine learning method of titanium alloy properties identification in additive technologies," *Eastern-Eur. J. Enterprise Technol.*, vol. 3, no. 12, pp. 23–31, Jun. 2018, doi: 10.15587/1729-4061.2018.134319.

[9] A. Paul, M. Mozaffar, Z. Yang, W.-k. Liao, A. Choudhary, J. Cao, and A. Agrawal, "A real-time iterative machine learning approach for temperature profile prediction in additive manufacturing processes," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2019, pp. 541–550, doi: 10.1109/DSAA.2019.00069.

[10] Z. Liu, J. Zhang, B. He, and Y. Zou, "High-speed nanoindentation mapping of a near-alpha titanium alloy made by additive manufacturing," *J. Mater. Res.*, vol. 36, no. 11, pp. 2223–2234, Jun. 2021, doi: 10.1557/s43578-021-00204-7.

[11] C. Zhu, C. Li, D. Wu, W. Ye, S. Shi, H. Ming, X. Zhang, and K. Zhou, "A titanium alloys design method based on high-throughput experiments and machine learning," *J. Mater. Res. Technol.*, vol. 11, pp. 2336–2353, Mar. 2021, doi: 10.1016/j.jmrt.2021.02.055.

[12] C. Zou, J. Li, W. Y. Wang, Y. Zhang, D. Lin, R. Yuan, X. Wang, B. Tang, J. Wang, X. Gao, H. Kou, X. Hui, X. Zeng, M. Qian, H. Song, Z.-K. Liu, and D. Xu, "Integrating data mining and machine learning to discover high-strength ductile titanium alloys," *Acta Mater.*, vol. 202, pp. 211–221, Jan. 2021, doi: 10.1016/j.actamat.2020.10.056.

[13] Z. Zhan, W. Hu, and Q. Meng, "Data-driven fatigue life prediction in additive manufactured titanium alloy: A damage mechanics based machine learning framework," *Eng. Fract. Mech.*, vol. 252, Jul. 2021, Art. no. 107850, doi: 10.1016/j.engfracmech.2021.107850.

[14] M. Banerjee, A. Banerjee, D. Mukherjee, A. K. Singla, and J. Singh, "Machine learning module for predicting tensile response of SLMed Ti-6Al-4V," in *Advances in Additive Manufacturing and Metal Joining*. Singapore: Springer, 2023, pp. 211–222, doi: 10.1007/978-981-19-7612-4_18.

[15] M. Zou, W. G. Jiang, Q. H. Qin, Y. C. Liu, and M. L. Li, "Optimized XGBoost model with small dataset for predicting relative density of Ti-6Al-4V parts manufactured by selective laser melting," *Materials*, vol. 15, no. 15, p. 5298, 2022.

[16] Z. Cui and G. Gong, "The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features," *Neuroimage*, vol. 178, pp. 622–637, Sep. 2018.

[17] P. Xu, X. Ji, M. Li, and W. Lu, "Small data machine learning in materials science," *NPJ Comput. Mater.*, vol. 9, no. 1, p. 42, Mar. 2023.

[18] R. Goyal, P. Chandra, and Y. Singh, "Suitability of KNN regression in the development of interaction based software fault prediction models," *Proc. IERI*, vol. 6, pp. 15–21, 2014, doi: 10.1016/j.ieri.2014.03.004.

[19] M. F. Rohmah, I. K. G. D. Putra, R. S. Hartati, and L. Ardiantoro, "Comparison four kernels of SVR to predict consumer price index," *J. Phys., Conf. Ser.*, vol. 1737, no. 1, Jan. 2021, Art. no. 012018, doi: 10.1088/1742-6596/1737/1/012018.

[20] P. Rivas-Perea and J. Cota-Ruiz, "An algorithm for training a large scale support vector machine for regression based on linear programming and decomposition methods," *Pattern Recognit. Lett.*, vol. 34, no. 4, pp. 439–451, Mar. 2013, doi: 10.1016/j.patrec.2012.10.026.

[21] C. C. Wang, P. H. Kuo, and G. Y. Chen, "Machine learning prediction of turning precision using optimized XGBoost model," *Appl. Sci.*, vol. 12, no. 15, p. 7739, Aug. 2022.

[22] D. Suenaga, Y. Takase, T. Abe, G. Orita, and S. Ando, "Prediction accuracy of random forest, XGBoost, LightGBM, and artificial neural network for shear resistance of post-installed anchors," *Structures*, vol. 50, pp. 1252–1263, Apr. 2023.

[23] N. M. Shahani, X. Zheng, C. Liu, F. U. Hassan, and P. Li, "Developing an XGBoost regression model for predicting Young's modulus of intact sedimentary rocks for the stability of surface and subsurface structures," *Frontiers Earth Sci.*, vol. 9, Oct. 2021, Art. no. 761990.

[24] A. Pathy, S. Meher, and P. Balasubramanian, "Predicting algal biochar yield using eXtreme gradient boosting (XGB) algorithm of machine learning methods," *Algal Res.*, vol. 50, Sep. 2020, Art. no. 102006.

[25] S. B. Imandoust and M. Bolandraftar, "Application of K-nearest neighbor (KNN) approach for predicting economic events: Theoretical background," *Int. J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 605–610, 2013.

[26] K. Y. Chen and C. H. Wang, "Support vector regression with genetic algorithms in forecasting tourism demand," *Tourism Manag.*, vol. 28, no. 1, pp. 215–226, 2007.

[27] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[28] H. Chaoui, P. Sicard, and W. Gueaieb, "ANN-based adaptive control of robotic manipulators with friction and joint elasticity," *IEEE Trans. Ind. Electron.*, vol. 56, no. 8, pp. 3174–3187, Aug. 2009.

[29] S. Akkoç, "An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system (ANFIS) model for credit scoring analysis: The case of Turkish credit card data," *Eur. J. Oper. Res.*, vol. 222, no. 1, pp. 168–178, Oct. 2012.

[30] M. Steurer, R. J. Hill, and N. Pfeifer, "Metrics for evaluating the performance of machine learning based automated valuation models," *J. Property Res.*, vol. 38, no. 2, pp. 99–129, Apr. 2021, doi: 10.1080/09599916.2020.1858937.

[31] A. V Tatachar, "Comparative assessment of regression models based on model evaluation metrics," *Int. Res. J. Eng. Technol.*, vol. 8, no. 9, pp. 853–860, Sep. 2021.

[32] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecasting*, vol. 32, no. 3, pp. 669–679, Jul. 2016.

[33] F. E. Harrell, K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati, "Regression modelling strategies for improved prognostic prediction," *Statist. Med.*, vol. 3, no. 2, pp. 143–152, Apr. 1984.

[34] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[35] L. Bottou, "Stochastic gradient learning in neural networks," *Proc. Neuro-Nımes*, vol. 91, no. 8, p. 12, Nov. 1991.

[36] M. Kordos, M. Blachnik, and D. Strzempa, "Do we need whatever more than k-NN?" in *Artificial Intelligence and Soft Computing*. Berlin, Germany: Springer, 2010, pp. 414–421, doi: 10.1007/978-3-642-13208-7_52.

[37] A. Alwosheel, S. van Cranenburgh, and C. G. Chorus, "Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis," *J. Choice Model.*, vol. 28, pp. 167–182, Sep. 2018, doi: 10.1016/j.jocm.2018.07.002.

[38] T. Sutojo, S. Rustad, M. Akrom, A. Syukur, G. F. Shidik, and H. K. Dipojono, "A machine learning approach for corrosion small datasets," *NPJ Mater. Degradation*, vol. 7, no. 1, Mar. 2023, Art. no. 18.

[39] Y. Zhang and C. Ling, "A strategy to apply machine learning to small datasets in materials science," *NPJ Comput. Mater.*, vol. 4, no. 1, p. 25, 2018.

[40] S. Huang, Q. Zhao, Z. Yang, C. Lin, Y. Zhao, and J. Yu, "Strengthening effects of Al element on strength and impact toughness in titanium alloy," *J. Mater. Res. Technol.*, vol. 26, pp. 504–516, Sep. 2023, doi: 10.1016/j.jmrt.2023.07.206.

[41] F. Xie, H. Yang, J. Huang, J. Yu, and X. He, "Sn content effects on microstructure, mechanical properties and tribological behavior of biomedical Ti-Nb-Sn alloys fabricated by powder metallurgy," *Metals*, vol. 12, no. 2, p. 255, Jan. 2022, doi: 10.3390/met12020255.

[42] X. Wang, Z. Yin, Y. Yang, S. Wu, G. Li, K. Zhang, H. Wang, and A. Huang, "Influence of oxygen and zirconium contents on the mechanical properties of Ti-23Nb-0.7Ta-Zr-O alloys," *Metals*, vol. 12, no. 6, 2022, doi: 10.3390/met12061018.

[43] F. An, B. Zhang, Y. Yan, and L. Wang, "Effect of vanadium contents on microstructure and mechanical properties of Ti-6Al-xV components produced by wire + arc additive manufacturing," *Mater. Trans.*, vol. 62, no. 8, pp. 1071–1078, 2021, doi: 10.2320/matertrans.MT-M2021031.

[44] A. D. Kotov, M. N. Postnikova, A. O. Mosleh, and A. V. Mikhaylovskaya, "Effect of Mo content on the microstructure, superplastic behavior, and mechanical properties of Ni and fe-modified titanium alloys," *Mater. Sci. Eng., A*, vol. 877, Jun. 2023, Art. no. 145166, doi: 10.1016/j.msea.2023.145166.

[45] H. R. Ogden and R. I. Jaffee, "The effects of carbon, oxygen and nitrogen on the mechanical properties of titanium and titanium alloys," *Titanium Metall. Lab.*, vol. 33, no. 2, 2012, doi: 10.2172/4370612.

[46] M. Morita, S. Suzuki, Y. Kato, W. Li, and O. Umezawa, "Tensile deformation of texture-controlled titanium with high oxygen content at room temperature," *Mater. Sci. Eng., A*, vol. 793, Aug. 2020, Art. no. 139660, doi: 10.1016/j.msea.2020.139660.

JAKA FAJAR FATRIANSYAH received the B.S. degree in theoretical physics and the M.S. degree in materials physics from Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2007 and 2008, respectively, and the Ph.D. degree in applied physics engineering from Hokkaido University, Japan, in 2015.

From 2015 to 2016, he was a Postdoctoral Researcher with the Soft Matter Laboratory, Faculty of Engineering, Hokkaido University. From 2017 to 2019, he was a Lecturer in metallurgical and materials engineering from Universitas Indonesia. From 2019 to 2022, he was an Assistant Professor of metallurgical and materials engineering from Universitas Indonesia, where he has been an Associate Professor of metallurgical and materials engineering, since 2019. His research interests include molecular dynamics applications in materials engineering and drug discovery, finite element methods in risk management, soft matter physics, and machine learning applications in engineering.

Dr. Fatriansyah was the Head of Materials Engineering Chapter of the Indonesian Engineer Association, from 2020 to 2024. He was a recipient of Best Young Lecturer Faculty of Engineering at Universitas Indonesia, in 2023; and the Best Innovator at Universitas Indonesia, in 2023. Since 2022, he has been the Editor-in-Chief of the *Journal of Materials Exploration and Findings*.

MUHAMAD RAFI AQILA received the B.S. degree in metallurgical and materials engineering from Universitas Indonesia, Depok, Indonesia, in 2023. His research interest includes the application of machine learning on materials selection process and alloys.

IPING SUHARIADI received the B.S. degree in physics from Universitas Gadjah Mada (UGM), Jogjakarta, Indonesia, in 2007, the M.S. degree in materials engineering from University Sains Malaysia (USM), Penang, Malaysia, in 2011, and the Ph.D. degree in electronic engineering from Kyushu University, Fukuoka, Japan, in 2014. Since 2015, he has been an Assistant Professor with the Industrial Engineering Department, Bina Nusantara (BINUS) University. His research interest includes the fabrication of micro- or nanostructured surfaces and materials design.

ANDREAS FEDERICO received the B.S. degree in metallurgical and materials engineering from Universitas Indonesia, Depok, Indonesia, in 2023, where he is currently pursuing the M.S. degree with the Materials Engineering Department. His research interests include application machine learning on polymer and synthesis of photocatalyses and various mesoporous materials.

DZAKY IMAN AJIPUTRO received the B.S. degree in metallurgical and materials from Universitas Indonesia, Depok, Indonesia, in 2023. His research interest includes application machine learning on materials selection process and alloys.

AGRIN FEBRIAN PRADANA received the B.Sc. and M.Sc. degrees in chemistry from IPB University, Bogor, Indonesia, in 2015 and 2016, respectively. Since 2022, he has been a Research Assistant with the Computational Laboratory, Department of Metallurgical and Materials Engineering, Faculty of Engineering, Universitas Indonesia. His research interests include computational material, molecular docking, and dynamics simulation.

YOSSI ANDREANO received the B.S. degree in metallurgical and materials engineering from Universitas Indonesia, Depok, Indonesia, in 2024. His research interest includes the application of machine learning on materials selection process and alloys.

MUHAMMAD ALI YAFI RIZKY received the B.S. degree in metallurgical and materials engineering from Universitas Indonesia, Depok, Indonesia, in 2024. His research interest includes the application of machine learning on materials selection process and alloys.

DONANTA DHANESWARA received the B.S. degree in metallurgical engineering and the M.S. and Ph.D. degrees in materials science from Universitas Indonesia, Depok, Indonesia, in 1989, 1995, and 2006, respectively.

In 2024, he received a full professorship in metallurgical and materials engineering from Universitas Indonesia. His research interests include functional engineering of ceramic materials, including mesoporous silica extraction from rice husk ash, corncob waste, and other biomass materials using the sol-gel method, which functions as an adsorbent for liquid waste and heavy metals. He was a recipient of the Best Innovator at Universitas Indonesia, in 2023.

ZAINOVIA LOCKMAN received the Bachelor of Engineering degree in materials science and engineering and the Ph.D. degree in electronic materials (superconductor) from Imperial College London, U.K., in 1999 and 2003, respectively.

She is currently a Professor with University Sains Malaysia (USM), Penang, Malaysia. Her research interests include semiconductor materials, nanomaterials, and synthesis of electronic oxides with nanotubes and nanowires structure for photocatalysis application.

SU-MI HUR received the B.S. degree in chemical engineering from Seoul National University (SNU), South Korea, and the Ph.D. degree from the University of California at Santa Barbara (UCSB), USA. She continued her research career as a Postdoctoral Associate with the Pritzker School of Molecular Engineering, University of Chicago, and Argonne National Laboratory. In 2015, she joined the School of Polymer Science and Engineering, Chonnam National University (CNU), South Korea, and was promoted to a Full Professor, in 2024. Her research is concerned with applying statistical mechanical theory and field-/particle-based coarse-grained simulations to investigate structural, thermodynamic, and dynamic phenomena in polymer-based soft materials. Her interests also lie in designing macromolecular systems and processes assisted with computational techniques including machine learning.