Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Investigation of Support Vector Machines with Different Kernel Functions for Prediction of Compressive Strength of Concrete

Souvik Pal[1], Le Huyen Trang[2,*], Vu Trong Hieu[2], Duc Dam Nguyen[2], Dung Quang Vu[2], Indra Prakash[3]

[1]Department of Computer Science & Engineering, Sister Nivedita University, New Town, Kolkata, India; email: souvikpal22@gmail.com

[2]University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi, Vietnam; email: lehuyentrang0500@gmail.com; damnd@utt.edu.vn, dungvq@utt.edu.vn

[3]DDG (R) Geological Survey of India, Gandhinagar 382010, India; email: indra52prakash@gmail.com

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2024.en.4.2.55-68

***Corresponding author:**
Email address:
lehuyentrang0500@gmail.com

Received: 03/06/2024
Revised: 24/06/2024
Accepted: 24/06/2024

Abstract: In this study, our primary aim is to assess and compare the efficacy of Support Vector Machines (SVM) employing various kernel functions: linear (LIN), polynomial (POL), Radial Basis Function (RBF), and sigmoid (SIG) in predicting the compressive strength of concrete. We generated and validated different models, namely SVM-LIN, SVM-POL, SVM-RBF, and SVM-SIG. Utilizing a dataset comprising 236 samples from the Red River surface water treatment plant project in Hanoi, Vietnam, we partitioned the data into training (70%) and testing (30%) sets for model training and validation. Our analysis employed various validation metrics, including coefficient of correlation (R), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), to assess and compare model performance. Results indicate that SVM-RBF (R = 0.847) outperforms the other models on testing data, followed by SVM-POL (R = 0.7182), SVM-LIN (R = 0.6679), and SVM-SIG (R = 0.0198), respectively. Consequently, our findings suggest that the RBF kernel function is most suitable for training SVM models to predict concrete compressive strength. Therefore, SVM-RBF emerges as a promising tool for the rapid and accurate estimation of concrete compressive strength. This study contributes novel insights by systematically evaluating these models using a comprehensive set of validation metrics, enhancing the robustness and applicability of predictive models in concrete technology.

Keywords: Support Vector Machines, Kernel Functions, Compressive Strength, Concrete Technology, Predictive Modeling.

1. Introduction

Concrete is the most common material used in engineering projects such as dams, building houses, bridges, roads, electricity poles, water tanks, etc. [1] The compressive strength of concrete is a fundamental property that has a significant influence on the performance and durability of structures [2]. Thus, the accurate determination of compressive strength is crucial in the field of civil engineering, particularly for ensuring the safety and reliability of concrete structures [3]. Concrete is a heterogeneous material, and its compressive strength is influenced by a variety of factors including the mix proportions, quality of materials, and curing conditions [4]. Traditional methods of determining compressive strength involve destructive testing, which is time-consuming and costly. Consequently, there is a growing interest in developing predictive models that can estimate the compressive strength of concrete based on its composition and curing parameters. In this context, machine learning (ML) techniques have shown promise due to their robustness and ability to handle non-linear relationships.

Numerous ML techniques and models have been developed and applied effectively for prediction of the compressive strength of various types of concrete. For instance, Paudel, Pudasaini, Shrestha and Kharel [5] used and compared different ML models namely Support Vector Regressor (SVR), Bagging, Random Forest (RF), Multiple Linear Regressor (MLR), AdaBoost Regressor, XGBoost for prediction of the compressive strength of concrete containing fly ash. Ben Seghier, Golafshani, Jafari-Asl and Arashpour [6] applied four ML models including adaptive neuro-fuzzy inference system (ANFIS) , least-square support vector regression (LSSVR), SVR, and multilayer perceptron neural network (MLP-NN) for prediction of the compressive strength of concrete consisting waste glass. El-Mir, El-Zahab, Sbartaï, Homsi, Saliba and El-Hassan [7] compared four ML models including gaussian process regression (GPR), support vector machine (SVM), multiple regression (MR), and regression tree (RT) for prediction of the compressive strength of concrete. Overall, these studies demonstrate the potential of ML models for accurate and effective prediction of the compressive strength of concrete.

Among popular ML models, SVMs are a class of supervised learning models that can be used for both classification and regression tasks [8]. One of the key strengths of SVM is its use of kernel functions, which enable it to operate in a high-dimensional space and capture complex relationships within the data. Previous research has demonstrated the potential of SVM in predicting material properties [9], but there remains a gap in comprehensive studies that compare the performance of different kernel functions specifically for concrete compressive strength prediction. Most existing studies have focused on a single type of kernel, limiting the understanding of how different kernels can influence prediction accuracy [10].

In this study, by systematically evaluating multiple kernel functions, the main objective is to provide deeper insights into the most suitable approaches for applying SVM for prediction of the compressive strength of concrete. This aims to enhance the predictive capabilities of SVM models in the context of concrete technology. For the modeling, a dataset was used, comprising a wide range of mix proportions and curing conditions, ensuring comprehensive coverage of typical concrete compositions and their corresponding compressive strengths. This dataset was obtained from the red river surface water treatment plant project located in Hanoi, Vietnam. The models were evaluated based on their prediction accuracy using metrics such as RMSE, MAE, and R.

2. Materials and Methods

2.1. Data used

In this study, we utilized data collected from red river surface water treatment plant project located in Hanoi, Vietnam. Data of 236 samples were collected, which included the testing results of compressive strength of concrete and other 10 properties including age of concrete, cement type, coarse aggregate 5x20mm, natural sand content, crushed sand content, water content, superplasticizer admixture, slump, water to cement ratio, aggregate to cement ratio. In predictive modeling, the compressive strength of concrete is considered as a dependent variable (output), while other 10 mentioned properties were considered as independent variables (inputs). Selection of 10 properties for prediction of the compressive strength of concrete is based on the literature review of the relevant published works [5,11,12]. Table 1 shows the initial analysis of the variables collected and used for the modeling.

2.2. Methods used

2.2.1. Support Vector Machines (SVMs)

SVMs are a category of supervised machine learning algorithms utilized for both classification and regression tasks. They excel in processing multi-dimensional data and determining the optimal decision boundary, making them highly effective in various applications. [13]. In the context of construction materials, SVM provides flexible tools for material property analysis, performance prediction, and design optimization [14]

The key concept of SVMs lies in finding the optimal hyperplane that separates data points of different classes while maximizing the margin between them. Here are the key components and concepts associated with SVMs:

In the realm of machine learning, a hyperplane stands as a pivotal concept, serving as a critical decision boundary in high-dimensional spaces. In a two-dimensional space, a hyperplane is a straight line, while in higher dimensions, it becomes a plane or hyperplane. Defined as a flat affine subspace, a hyperplane separates data points of different classes, guiding classifiers like Support Vector Machines (SVMs) in making accurate predictions. [15]

Support vectors are the data points that lie closest to the hyperplane and have the most influence on its position and orientation. These points determine the margin of the hyperplane and are crucial for defining the decision boundary.[16]

The margin is the distance between the hyperplane and the nearest data points (support vectors) of each class. SVMs aim to maximize this margin, as a larger margin generally leads to better generalization on unseen data and reduces the risk of overfitting.[17]

In the context of SVMs, a kernel is a function that computes the inner product between pairs of data points in the input space. Kernels play a crucial role in SVMs, especially when dealing with nonlinearly separable data. They enable SVMs to find optimal decision boundaries by implicitly mapping the input features into a higher-dimensional space where the data becomes linearly separable [18].

The linear (LIN) kernel, a foundational component of SVMs, simplifies classification tasks by computing the inner product of input feature vectors in their original space [19]. By defining a linear decision boundary, such as a straight line in two dimensions or a hyperplane in higher dimensions, the linear kernel effectively separates data points of different classes when the relationship between features and classes is predominantly linear [20]. Its simplicity facilitates faster computation and easier interpretability compared to more complex kernel functions, making it a preferred choice for linearly separable datasets [21]. However, its efficacy diminishes when dealing with nonlinear data structures [22]. Mathematically, the linear kernel is defined as $K(x_i, x_j) = x_i . x_j$, where x_i and x_j are the input feature vectors [23].

The polynomial (POL) kernel is a versatile and powerful component of Support Vector Machines (SVMs), extending the capability of SVMs to handle nonlinearly separable data by mapping input features into a higher-dimensional space using polynomial functions [24]. This kernel is defined by its degree, which determines the complexity of the polynomial transformation, allowing it to capture more intricate patterns and relationships within the data [25]. It is defined as $K(x_i, x_j) = (\gamma x_i . x_j + r)^d$, where γ is a scaling parameter r is an offset, and d is the degree of the polynomial.

The RBF kernel is a powerful and widely used kernel function in SVMs that excels in handling complex and nonlinearly separable data

[26]. This kernel computes the similarity between pairs of data points based on their distance, with closer points having higher similarity. The RBF kernel's ability to model complex relationships makes it particularly effective in various applications, including image classification, bioinformatics, and anomaly detection. [27] Its flexibility and robustness contribute to the RBF kernel's popularity as a go-to choice for many real-world machine learning tasks. It is defined as $K(x_{i,}x_j) = \exp(-\gamma \| x_i - x_j \|^2)$, where γ is a scaling parameter that determines the influence of each data point.

The sigmoid (SIG) kernel, also known as the hyperbolic tangent kernel, is used in SVMs to address nonlinear classification and regression challenges. It transforms input features into a higher-dimensional space using the sigmoid function, which is inspired by neural network activation functions. [28] This enables the SVM to model complex and nonlinear relationships within the data. The sigmoid kernel is particularly useful for datasets where the decision boundaries are intricate and not easily defined by linear or polynomial functions, making it a versatile tool in the SVM toolkit for capturing multifaceted data patterns. [29] It is defined as $K(x_i, x_j) = \tanh(\gamma x_i . x_j + r)$, where γ is a scaling parameter and r is an offset.

2.2.2. Validation metrics

To enhance the validation process of machine learning models, three key criteria are often employed: the coefficient of correlation (R), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics collectively provide a comprehensive evaluation of the model's performance, ensuring accuracy, reliability, and robustness in its predictions. Description of these criteria is given below:

R is a statistical measure used in regression analysis to assess the goodness of fit of a model. It represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s). R values range from -1 to 1, where 0 indicates that the model does not explain any of the variance and 1 or -1 indicates that the model explains all the variance in the dependent variable [30]. The formula for R is:

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

where:

y_i represents the actual values of the dependent variable.

$\hat{y}_i$ represents the predicted values of the dependent variable by the regression model.

$\overline{y_i}$ is the mean of the actual values of the dependent variable.

MAE is a widely used metric in regression analysis to measure the accuracy of a model in predicting continuous outcomes. [31] MAE quantifies the average magnitude of the errors between predicted values and actual values, providing a straightforward interpretation of prediction accuracy. [32] Unlike other error metrics, MAE gives equal weight to all individual differences between predicted and actual values, making it a robust measure of model performance. The formula for MAE is as below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (2)$$

where n is the number of data points, y_i represents the actual value of the dependent variable for the i-th observation, $| y_i - \hat{y}_i |$ represents the predicted value of the dependent variable for the i-th observation.

RMSE is a commonly used metric for evaluating the accuracy of a regression model. It measures the average magnitude of the errors between the predicted values and the actual values, providing insight into the model's predictive power. [33] RMSE is particularly valuable because it gives more weight to larger errors, making it sensitive to outliers.[33] The formula for RMSE is as below:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

where n is the number of observations, y_i represents the actual values of the dependent variable, $\hat{y}_i$ represents the predicted values of the dependent variable by the regression model.

3. Methodological flowchart

Methodological flowchart of this study is presented in Fig 1. It includes three main steps. In the first step, the data was randomly split into two parts such as training data (70%) and testing data (30%) [34]. Out of these, training data was used for training the models while testing data was used for validating the models. Selection of training/testing ratio is based on the literature review of the relevant published works [34-36]. In the second step, the SVM was trained using training data with different kernel functions including LIN, POL, RBF and SIG. With different kernel functions used in training SVM, different respective models were built and generated, namely, SVM-LIN, SVM-POL, SVM-RBF, and SVM-SIG. In the final step, different models were validated and compared using testing data with various validation criteria such as R, RMSE, and MAE.

4. Results and discussion

Multiple Support Vector Machine (SVM) models underwent rigorous training and validation processes utilizing distinct sets of training and testing data. The training phase involved the meticulous selection and application of hyperparameters, as delineated in Table 2, to ensure optimal model performance and generalization capability.

Results of training and validating models are shown in Fig 2, Fig 3, Fig 4, Table 3 and Table 4. Fig 2 shows the predicted versus measured compressive strength over the applied models. It can be seen that the predicted and measured compressive strength obtained from SVM-RBF models are the closest compared with those obtained from other models. Fig 3 shows error analysis of the applied models. Fig 4 shows the correlation analysis of the predicted and measured values of the compressive strength of concrete obtained from different applied models. Values of validation metrics used for validation of the models are shown in Table 3 and Table 4. Table 3 shows the values of validation metrics of the models using training data. It can be observed that SVM-POL has the highest value of R (0.9871) and lowest values of RMSE (2.5333) and MAE (1.7194) compared with other models, and SVM-SIG has the lowest value of R (-0.0921) and highest value of RMSE (46.3873) and MAE (35.3873). However, with testing data (Table 4), SVM-RBF has the highest value of R (0.847) and lowest values of RMSE (4.7095) and MAE (4.136), followed by SVM-POL, SVM-LIN, and SVM-SIG, respectively.

From the validation of the models, it can be seen that SVM can be used for accurate prediction of the compressive strength of concrete. It is reasonable as SVM is considered as one of the most effective popular ML models for prediction. It has several main advantages [8,37] such as (1) SVR is robust to outliers due to its use of a margin of tolerance (epsilon-insensitive loss function), which means that small errors within a certain threshold do not affect the model, making it less sensitive to outliers compared to traditional regression methods, (2) The use of kernel functions allows SVR to handle complex non-linear relationships between variables efficiently, and (3) The built-in regularization mechanism helps in managing the trade-off between fitting the training data well and maintaining model simplicity to enhance generalization on new data.

In addition, it can be observed that when training SVM, the choice of kernel function plays a critical role in determining the model's performance. In this work, SVM worked the best with the RBF kernel function compared with other functions (LIN, POL, and SIG) [38,39]. It is because

RBF kernel is often considered better than other kernel functions by its flexibility and robustness in modeling complex, non-linear relationships. It can effectively capture intricate patterns in the data by mapping input features into an infinite-dimensional space, allowing it to handle a wide variety of data distributions and complexities. This capability makes it a versatile choice that performs well across different types of datasets, offering a good balance between accuracy and generalization without requiring extensive feature engineering. In contrast, SIG/POL kernel function involves multiple parameters that need tuning, which can make the optimization process more complex and sensitive. Improper selection of these parameters can lead to either underfitting or overfitting.

Table 1. Descriptive statistics of input and output variables in the study

No	Variables	Unit	Min	Max	Average	Standard deviation
1	Age of Concrete	(day)	1	28	15.644	11.278
2	Cement type	(kg)	220	537	428.699	78.882
3	Coarse aggregate 5x20mm	(kg)	997	1232	1082.301	57.695
4	Natural sand content	(kg)	0	905	368.832	332.833
5	Crushed sand content	(kg)	0	873	415.553	346.732
6	Water content	(l)	135	226	162.676	24.35
7	Superplasticizer Admixture	(l)	0	6.4	4.247	1.523
8	Slump	(cm)	4	20	14.609	2.955
9	Water to cement ratio	-	0.28	0.76	0.412	0.135
10	Aggregate to cement ratio	-	3.09	8.55	4.572	1.22
11	Compressive strength	(MPa)	10.18	67.06	43.466	14.381

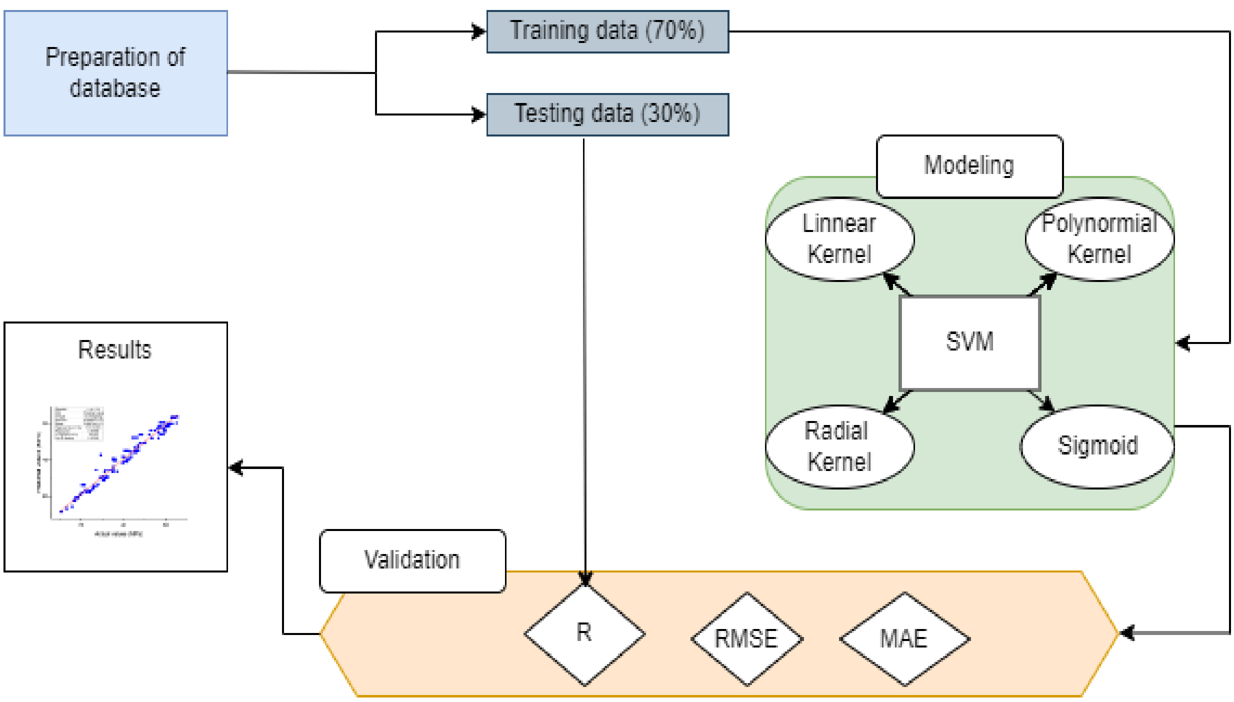
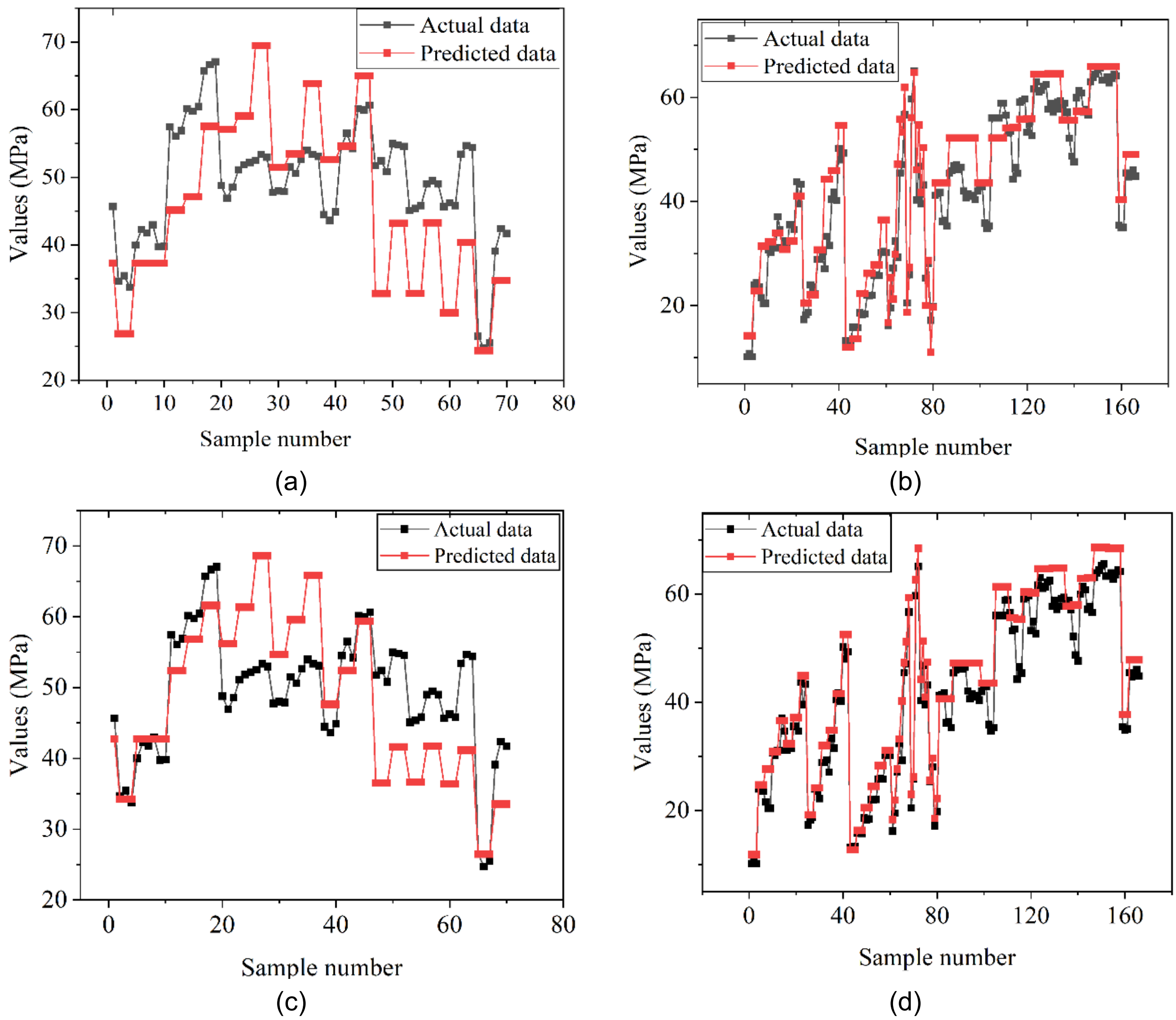
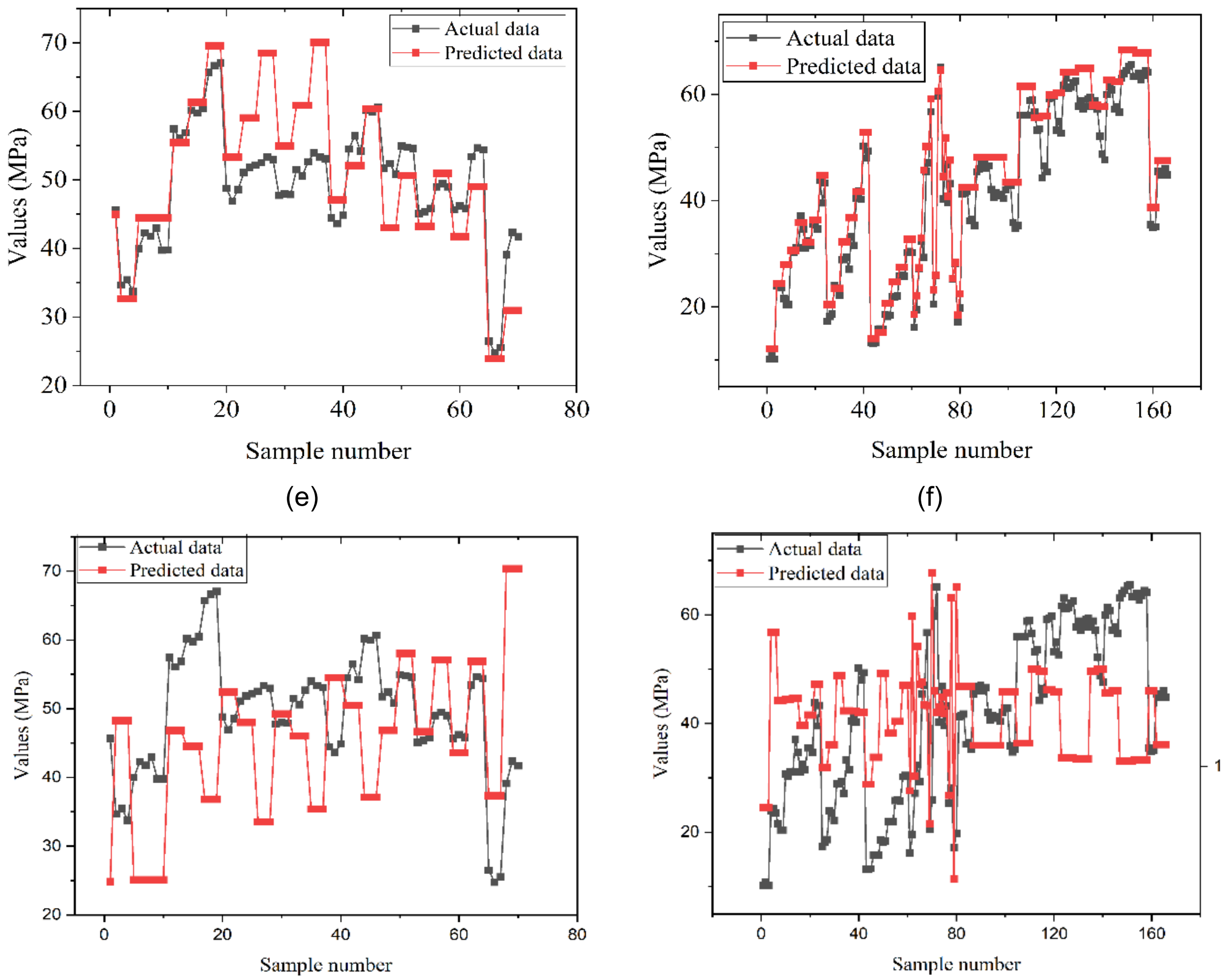


Fig 1. Methodological framework employed in this study

Table 2. Hyper-parameters used for the model development

No	Hyper-parameters	SVM
1	SVM Type	Nu-SVR (regression)
2	Batch type	100
3	Cache size	40.0
4	Coefficient	0.2
5	Cost	1.0
6	Debug	False
7	Degree	3
8	Epsilon	0.002
9	Gamma	0.6
10	Kernel type	LIN, POL, RBF, and SIG
11	Loss	0.1
12	Normalize	True
13	nu	0.5
14	seed	1
15	skrinking	True

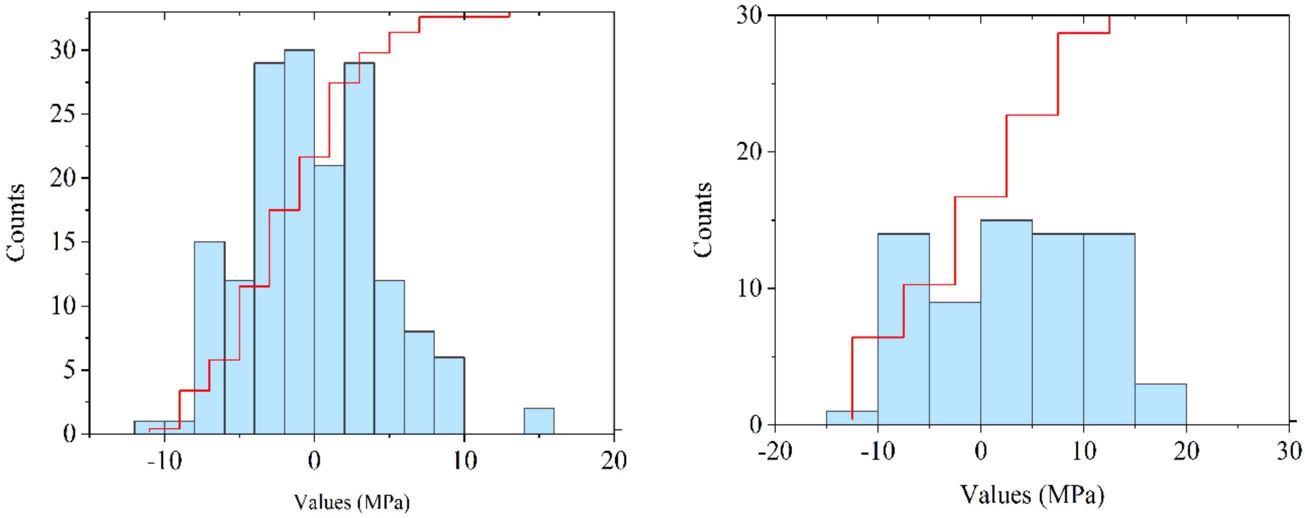




(e) (f)

(g) (h)

Fig 2. Predicted versus measured CSC over the applied models: (a) testing SVM-LIN, (b) training SVM-LIN, (c) testing SVM-POL, (d) training SVM-POL, (e) testing SVM-RAD, (f) training SVM-RAD, (g) testing SVM-SIG, and (h) training SVM-SIG



(a) (b)

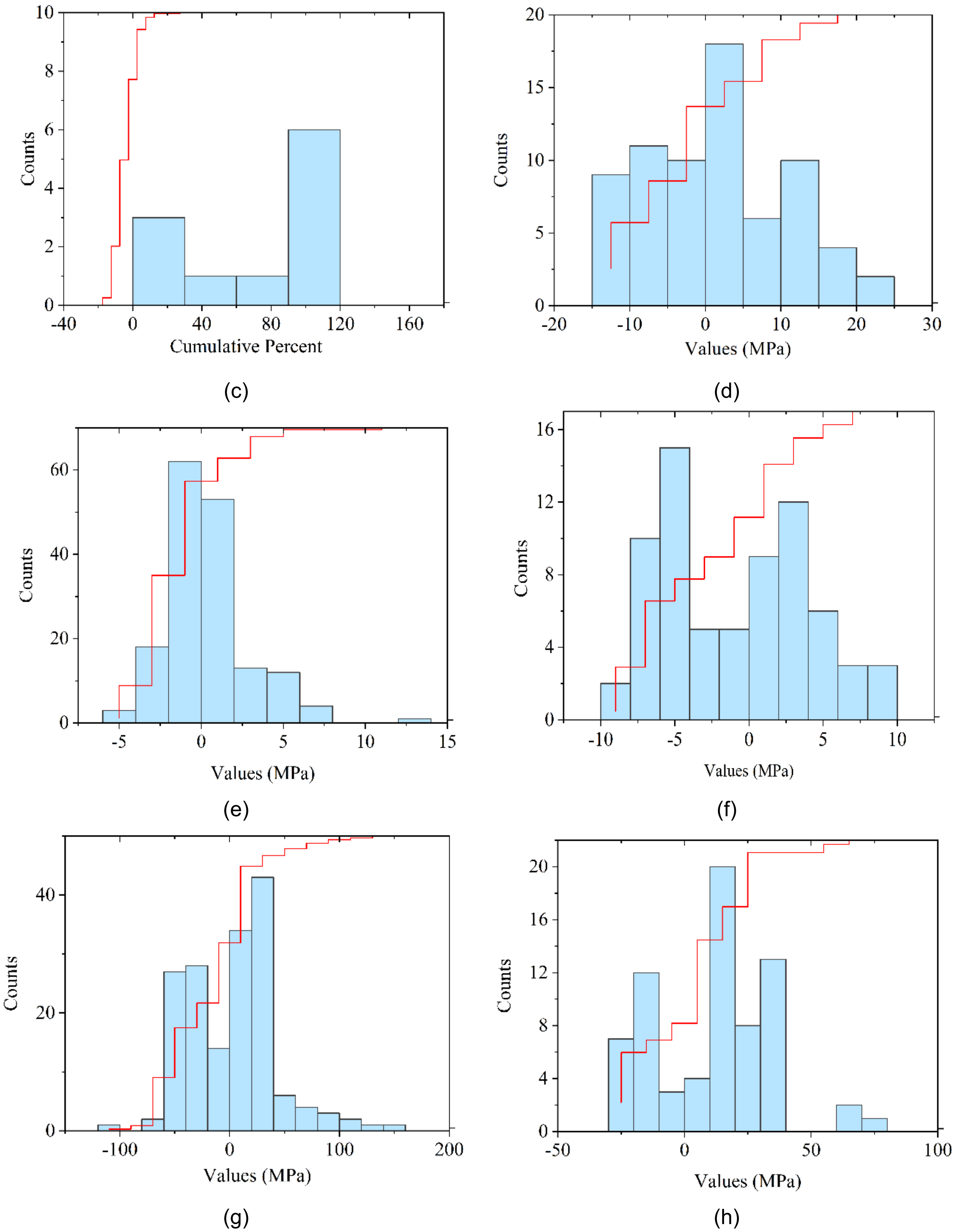


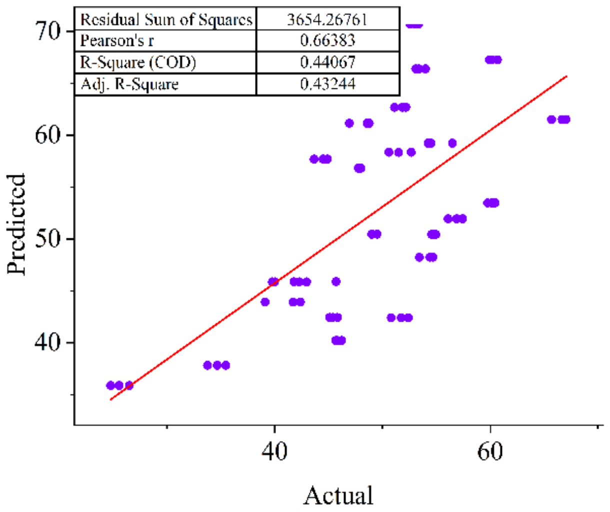
Fig 3. Error analysis of the applied models: (a) training SVM-LIN, (b) testing SVM-LIN, (c) training SVM-POL, (d) testing SVM-POL, (e) training SVM-RBF, (f) testing SVM-RBF, (g) training SVM-SIG, and (h) testing SVM-SIG

Table 4. Training performance of the models

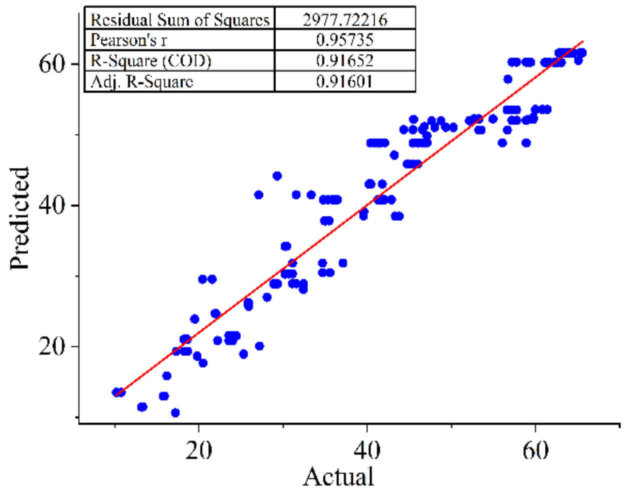
Models \ Metrics	R	MAE (MPa)	RMSE (MPa)
SVM-LIN	0.9585	3.4735	4.4323
SVM-POL	0.9871	1.7194	2.5333
SVM-RBF	0.9846	1.8925	2.7568
SVM-SIG	-0.0921	35.3873	46.3873

Table 5. Validation performance of the models

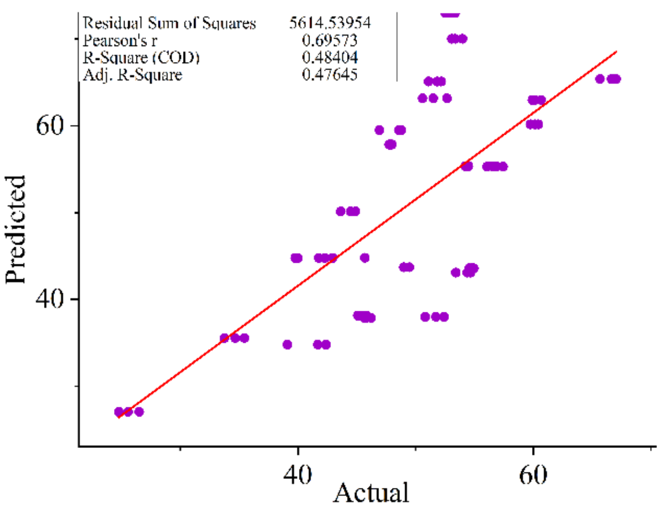
Models \ Metric	R	MAE (MPa)	RMSE (MPa)
SVM-LIN	0.6679	7.1014	8.3108
SVM-POL	0.7182	6.4099	8.0108
SVM-RBF	0.847	4.136	4.7095
SVM-SIG	0.0198	18.8598	26.0528



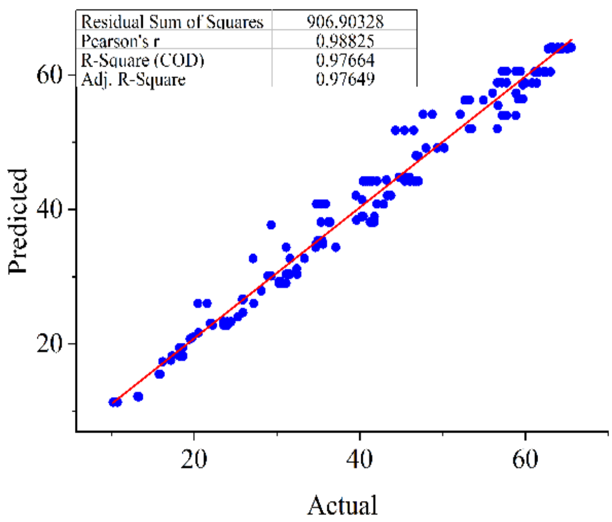
(a)



(b)



(c)



(d)

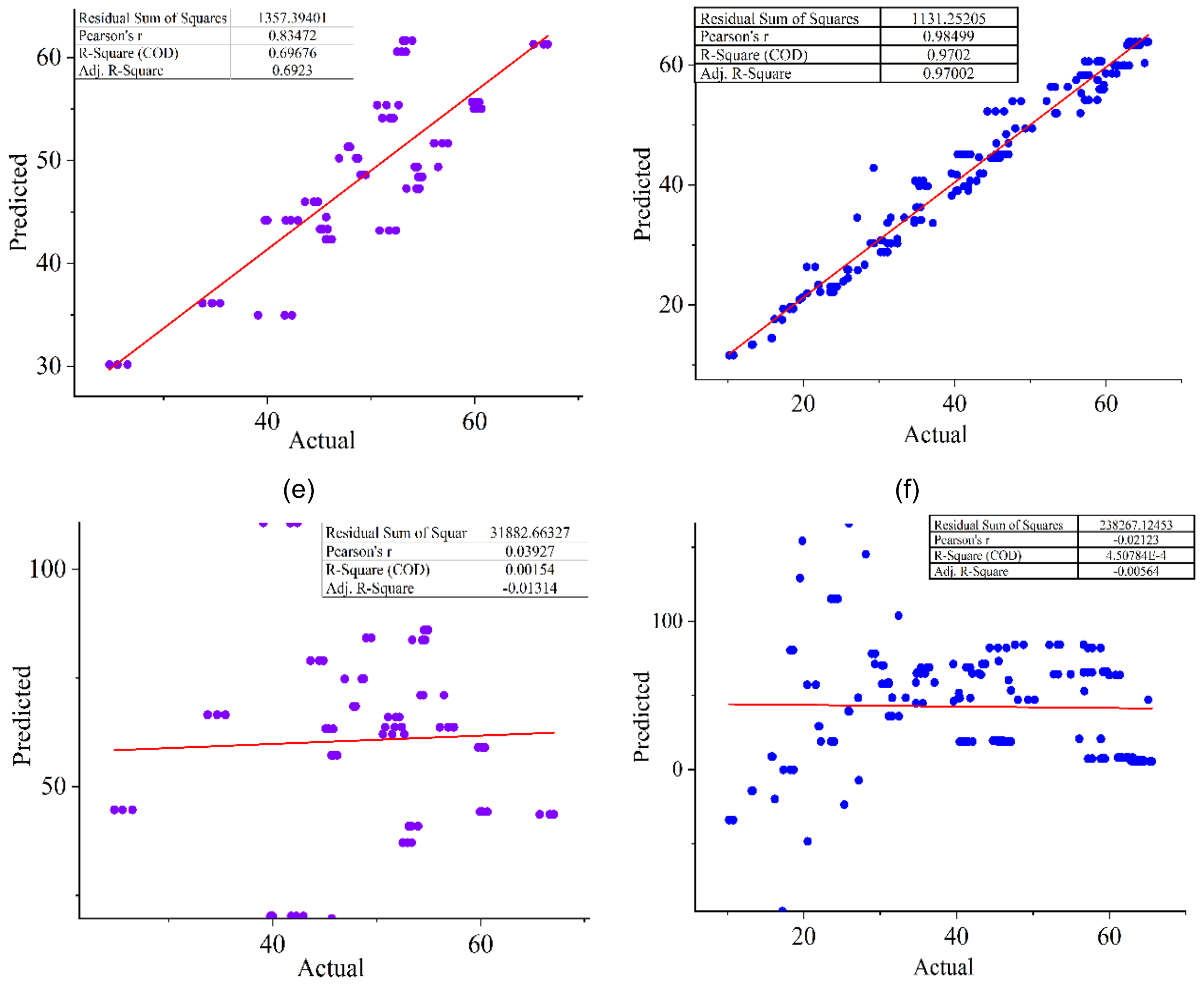


Fig 4. R analysis of the applied models: (a) testing SVM-LIN, (b) training SVM-LIN, (c) testing SVM-POL, (d) training SVM-POL, (e) testing SVM-RAD, (f) training SVM-RAD, (g) testing SVM-SIG, and (h) training SVM-SIG

5. Conclusions

In this study, various SVM models were trained and validated using different kernel functions (LIN, POL, RBF, and SIG) to predict the compressive strength of concrete. The training and testing data were derived from 236 concrete samples collected from the Red River surface water treatment plant project in Hanoi, Vietnam. Validation metrics such as R, RMSE, and MAE were employed to evaluate and compare the models.

Results indicated that SVM-RBF exhibited the highest R value (0.847) and the lowest RMSE (4.7095) and MAE (4.136) among the tested models when evaluated with testing data. Conversely, SVM-POL demonstrated the highest R value (0.9871) and the lowest RMSE (2.5333) and MAE (1.7194) with training data. SVM-SIG, on the other hand, displayed the lowest R value (-0.0921) and the highest RMSE (46.3873) and MAE (35.3873) among all models during training.

The findings suggest that SVM-RBF is the most suitable model for predicting the compressive strength of concrete among the tested kernel functions. Therefore, SVM-RBF holds promise as a reliable and efficient tool for accurately estimating concrete compressive strength. While this study provides valuable insights, certain limitations should be acknowledged. The dataset used was specific to the Red River surface water treatment plant project, potentially limiting generalizability. Additionally, the study solely focused on SVM models without exploring alternative machine learning algorithms, which could offer complementary insights. Future research could address these limitations by incorporating larger and more diverse datasets and comparing SVM models with other methodologies.

Acknowledgments

We thank the University of Transport Technology for supporting this research work.

References

[1]. H. Van Damme. (2018). Concrete material science: Past, present, and future innovations. *Cement and Concrete Research*, 112, 5-24.

[2]. M.S. Shetty. (1982). Concrete technology. *Ram Nagar, S. Chand & Co. Ltd.*

[3]. A.D. T.Q, A.R. Masoodi, A.H. Gandomi. (2023). Unveiling the potential of an evolutionary approach for accurate compressive strength prediction of engineered cementitious composites. *Case Studies in Construction Materials,* 19, e02172.

[4]. Z. Liu, Y.-G. Zhao, L. Ma, S. Lin. (2024). Review on high-strength recycled aggregate concrete: Mix design, properties, models and structural behaviour. *Structures, Elsevier*, 64, 106598.

[5]. S. Paudel, A. Pudasaini, R.K. Shrestha, E. Kharel. (2023). Compressive strength of concrete material using machine learning techniques. *Cleaner Engineering and Technology,* 15, 100661.

[6]. M.E.A. Ben Seghier, E.M. Golafshani, J. Jafari-Asl, M. Arashpour. (2023). Metaheuristic-based machine learning modeling of the compressive strength of concrete containing waste glass. *Structural Concrete*, 24(4), 5417-5440.

[7]. A. El-Mir, S. El-Zahab, Z.M. Sbartaï, F. Homsi, J. Saliba, H. El-Hassan. (2023). Machine learning prediction of concrete compressive strength using rebound hammer test. *Journal of Building Engineering*, 64, 105538.

[8]. A. Roy, S. Chakraborty. (2023). Support vector machine in structural reliability analysis: A review. *Reliability Engineering & System Safety*, 233, 109126.

[9]. K. Stergiou, C. Ntakolia, P. Varytis, E. Koumoulos, P. Karlsson, S. Moustakidis. (2023). Enhancing property prediction and process optimization in building materials through machine learning: A review. *Computational Materials Science,* 220, 112031.

[10]. R. Ramachandra, S. Mandal. (2023). Prediction of fly ash concrete type using ANN and SVM models. *Innovative Infrastructure Solutions*, 8(1), 47.

[11]. M. Hadzima-Nyarko, E.K. Nyarko, H. Lu, S. Zhu. (2020). Machine learning approaches for estimation of compressive strength of concrete. *The European Physical Journal Plus*, 135(8), 682.

[12]. D.-C. Feng, Z.-T. Liu, X.-D. Wang, Y. Chen, J.-Q. Chang, D.-F. Wei, Z.-M. Jiang. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 230, 117000.

[13]. S.F. Hussain. (2019). A novel robust kernel for classifying high-dimensional data using Support Vector Machines. *Expert Systems with Applications*, 131, 116-131.

[14]. H. Sun, H.V. Burton, H. Huang. (2021). Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering,* 33, 101816.

[15]. H. Bhavsar, M.H. Panchal. (2012). A review on support vector machine for data classification. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* 1(10), 185-189.

[16]. L.H. Hamel. (2011). Knowledge discovery with

support vector machines. *John Wiley & Sons*.

[17]. Y.-C. Chen, C.-T. Su. (2016). Distance-based margin support vector machine for classification. *Applied Mathematics and Computation,* 283, 141-152.

[18]. C. Pavithra, M. Saradha. (2016). Classification And Analysis Of Clustered Non-Linear Separable Data Set Using Support Vector Machines. *Migration Letters,* 21(S4), 901-913.

[19]. V. Kecman. (2005). Support vector machines–an introduction. Support vector machines: theory and applications. *Springer*, pp 1-47.

[20]. S. Ghosh, A. Dasgupta, A. Swetapadma. (2019). A study on support vector machine based linear and non-linear pattern classification. 2*019 International Conference on Intelligent Sustainable Systems (ICISS), IEEE*, pp 24-28.

[21]. M.A. Arshad, S. Shahriar, K. Anjum. (2023). The Power Of Simplicity: Why Simple Linear Models Outperform Complex Machine Learning Techniques-Case Of Breast Cancer Diagnosis, arXiv preprint arXiv:2306.02449.

[22]. R.M. Musa, A.P.A. Majeed, N.A. Kosni, M.R. Abdullah. (2020). Machine learning in team sports: performance analysis and talent identification in Beach Soccer & Sepak-takraw. *Springer Nature*.

[23]. A. Carrington. (2018). Kernel methods and measures for classification with transparency, interpretability and accuracy in health care. *University of Waterloo*.

[24]. T. Thanakulketsarat, P. Supnithi, L.M.M. Myint, K. Hozumi, M. Nishioka. (2023). Classification of the equatorial plasma bubbles using convolutional neural network and support vector machine techniques. *Earth, Planets and Space,* 75(1), 161.

[25]. B. Haasdonk, H. Burkhardt. (2007). Invariant kernel functions for pattern analysis and machine learning. *Machine Learning,* 68, 35-61.

[26]. M. Ramachandro, R. Bhramaramba. (2019). Classification of Gene Expression Data Set using Support Vectors Machine with RBF Kernel. *International Journal of Recent Technology and Engineering,* 8(2), 2907-2913.

[27]. D.M. Abdullah, A.M. Abdulazeez. (2021). Machine learning applications based on SVM classification a review. *Qubahan Academic Journal,* 1(2), 81-90.

[28]. D.D. Vico. (2022). Deep learning applied to regression, classification and feature transformation problems. *Universidad Autónoma de Madrid*.

[29]. A. Srivastava, K. Gupta, A. Kumar, W. Ahmed. (2024). Performance Analysis of Multifaceted Disease Prognosis Using Machine Learning. *2nd International Conference on Disruptive Technologies (ICDT), IEEE*, pp 955-960.

[30]. S. Menard. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, 54(1), 17-24.

[31]. S. Kim, H. Kim. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting,* 32(3), 669-679.

[32]. F.M. Talaat, A. Aljadani, B. Alharthi, M.A. Farsi, M. Badawy, M. Elhosseini. (2023). A Mathematical Model for Customer Segmentation Leveraging Deep Learning. Explainable AI, and RFM Analysis in Targeted Marketing. *Mathematics,* 11(18), 3930.

[33]. D. Chicco, M.J. Warrens, G. Jurman. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science,* 7, e623.

[34]. Q.H. Nguyen, H.-B. Ly, L.S. Ho, N. Al-Ansari, H.V. Le, V.Q. Tran, I. Prakash, B.T. Pham. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 1-15.

[35]. B. Anandan, M. Manikandan. (2023). Machine learning approach with various regression models for predicting the ultimate tensile strength of the friction stir welded AA 2050-T8 joints by the K-Fold cross-validation method. *Materials Today Communications,* 34, 105286.

[36]. M.K. Uçar, M. Nour, H. Sindi, K. Polat. (2020). The effect of training and testing process on machine learning in biomedical datasets. *Mathematical Problems in Engineering*, 1-17.

[37]. J. Zhuang, A.C. Midgley, Y. Wei, Q. Liu, D. Kong, X. Huang. (2024). Machine-Learning-Assisted Nanozyme Design: Lessons from Materials and Engineered Enzymes. *Advanced Materials*, 36(10), 2210848.

[38]. S.I. Majid, M. Kumar, N. Sahu, P. Kumar, D.K. Tripathi. (2024). Application of ensemble fuzzy weights of evidence-support vector machine (Fuzzy WofE-SVM) for urban flood modeling and coupled risk (CR) index for ward prioritization in NCT Delhi, India. *Environment, Development and Sustainability*, 1-39.

[39]. L. Ou, B. Liu, X. Chen, Q. He, W. Qian, W. Li, L. Zou, Y. Shi, Q. Hou. (2023). Automatic classification of the phenotype textures of three Thunnus species based on the machine learning SVM algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, 80(8), 1221-1236.