

# COMPREHENSIVE ANALYSIS OF MACHINE LEARNING MODELS FOR PREDICTING CONCRETE COMPRESSIVE STRENGTH

SAIDJON KAMOLOV<sup>1\*</sup>

**ABSTRACT.** Predicting the compressive strength of concrete provides several benefits including insights into its durability, potential lifespan, and its suitability for specific construction projects. Recent advances in machine learning algorithms have prompted their application to compressive strength modeling and prediction. In this paper, we carry out a comprehensive comparative analysis of the existing machine learning algorithms. Unlike, other current studies in this domain, we consider the entire spectrum of algorithms ranging from basic linear models to advanced meta-learning methods. The results show that there exists a trade-off between speed and accuracy of the models. We find that while ensemble models such as stacking and gradient boost provide superior accuracy, they are significantly slower than linear models.

## 1. INTRODUCTION

Concrete is one of the most widely used construction materials in the world due to its durability and cost-effectiveness. One of the most important properties of concrete is its compressive strength. Compressive strength refers to the ability of a concrete specimen to resist forces trying to reduce its size. Accurate prediction of the compressive strength of concrete can provide insights into its durability, potential lifespan, and its suitability for specific construction projects [1].

Traditionally, the compressive strength of concrete has been determined through laboratory tests on cylindrical or cubic specimens [2]. However, these tests can be time-consuming, costly, and require the destruction of the sample. Given the increasing scale and pace of construction projects, there's a growing need for methods that can predict the compressive strength efficiently and, if possible, non-destructively. Over the years, various methods have been proposed to estimate the compressive strength of concrete. Empirical methods, based on mix proportions and the properties of the constituents, have been widely used [3]. However, these methods often require domain expertise and can be limited in accuracy when dealing with novel mixtures or under specific environmental conditions.

---

*Date:* Received: Mar 8, 2024; Accepted: Apr 12, 2024.

\* Corresponding author.

*Key words and phrases.* machine learning, deep learning, civil engineering, concrete, compression strength.

With the advent of computational power and data-driven techniques, machine learning has emerged as a promising tool in many applications [4, 5, 6, 7] including for predicting concrete compression strength. Machine learning models are able to learn complex non-linear relationships between the features (like mix proportions, age, and type of additives) and the target variable (compressive strength) from historical data. These models have the potential to provide accurate predictions without the need for explicit formulae or human-crafted rules.

In this analysis, we evaluate the effectiveness of various machine learning models in predicting the compressive strength of concrete. By comparing their performance and computational efficiency, we aim to offer insights into the best approaches for this crucial task in civil engineering.

While there exist several studies that compare the performance of machine learning algorithms for predicting compressive strength, they are limited in scope. For example, in [8], the authors consider only 3 algorithms: decision tree, random forest, and gradient boosting. Similarly, in [9], the authors analyze only 4 algorithms: decision tree, bagging, gradient boosting, and neural networks. Further, in [10], the authors consider only linear models, kernel models, and tree-based models. To fill the gap in the literature, we perform a comprehensive analysis of a wide spectrum of methods ranging from basic linear models to advanced meta-learning methods.

The paper is organized as follows. Section 2 presents a brief overview of the existing literature. Section 3 provides the details of the machine learning algorithms used in our study. Section 4 contains the details of the dataset used in the study. Section 5 describes the methodology. Section 6 presents the results of the computational experiments and the corresponding discussion. Section 7 concludes the paper.

## 2. LITERATURE REVIEW

Recent advances in machine learning have catalyzed its applications in various fields including finance, pedagogy, medicine, network security, fraud detection, and others [11, 12, 13, 14, 15, 16, 17, 18, 19].

The compressive strength of concrete is a key parameter that dictates the design and evaluation of concrete structures. A plethora of studies have emerged over the recent years, elucidating various facets and methodologies pertinent to understanding and enhancing this fundamental property. This section presents notable contributions and findings from current studies on concrete compressive strength.

A systematic review in [20] was aimed at estimating the compressive strength of modern concrete mixtures employing computational intelligence. The study underscores the importance of compressive strength in the design and evaluation of concrete structures, with a particular emphasis on ultra-high performance concrete. The prediction of concrete's compressive strength is vital in ensuring the quality and progress of construction. Thus, a systematic review in [21], stressed the importance of accurate prediction and presented a bibliometric analysis over a decade, indicating a developing interest and advancements in this domain.

Exploring the potential of machine learning, Li et al. (2023) carried out a comparative study to predict the compressive strength of high-performance concrete. Their work revealed the efficacy of machine learning methods in model validation using experimental data, showing a significant stride towards accurate prediction and better construction quality [22].

A recent study delved into the time variation law of concrete's compressive strength from three distinct perspectives: single, multiple, and material internal factors [23]. It presented mathematical models relative to time under various conditions like carbonization, freeze-thaw cycle, and temperature effect, thereby providing a deeper understanding of the factors affecting concrete's compressive strength over time. The realm of high-strength concrete (HSC) has also witnessed significant contributions. A new study probed into the compressive strength prediction of HSC using hybrid techniques, indicating the role of cementitious materials in achieving superior strength [24]. Modeling and sensitivity analysis of high-performance concrete to various factors were explored in another recent work. The emphasis was on the accurate forecasting of compressive strength for both traditional and high-performance concrete constructions, considering the intricate development processes [25].

The multiaxial strength criteria of concrete has attracted significant attention in the literature. The failure of concrete materials under multiaxial stress was discussed in [26], underscoring the substantial attention this problem has received from the scientific community leading to several experimental studies and research. Exploration into the multiaxial compressive strength of hybrid fiber reinforced concrete (HFRC) provided insights into the uniaxial compressive and tensile strength of HFRC, emphasizing the role of steel and polypropylene fibers [27].

An experimental investigation showed the strength and durability properties of polypropylene fiber reinforced concrete [28]. The study analyzed how polypropylene fibers augment the tensile strength of concrete, thus enhancing its overall compressive strength. The recent technological advances in adding nanomaterials as a reinforcement material in concrete to improve its compressive and flexural strength were discussed in [29]. This innovation is aimed at improving the quality of traditional concrete, indicating a promising avenue for future research and applications in concrete technology.

### 3. MODEL DESCRIPTIONS

We evaluate a wide variety of machine learning algorithms to determine the optimal approach in predicting the compressive strength of concrete.

- **Linear Regression:** Linear regression is one of the simplest and most widely used statistical techniques [30]. It models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The goal of linear regression is to minimize the sum of squared differences between the observed responses in the dataset and the responses predicted by the linear approximation.

- **Lasso (L1 Regularization):** Lasso regression is a type of linear regression that incorporates L1 regularization. In general, regularization adds a penalty to the model parameters to reduce the freedom of the model and prevent overfitting [31]. L1 regularization adds a penalty equal to the absolute value of the magnitude of coefficients, leading to some coefficients being exactly zero. This can be seen as a form of feature selection. The objective function in Lasso regression is given by the following equation:

$$\frac{1}{2n} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \alpha \cdot \|\mathbf{w}\|_1$$

- **Ridge (L2 Regularization):** Ridge regression, also known as Tikhonov regularization, is a type of linear regression that includes a regularization term. The key difference is that Ridge employs L2 regularization, which adds a penalty equal to the square of the magnitude of coefficients. This results in smaller coefficients but doesn't necessarily eliminate them entirely, as it happens in Lasso. The objective function in Ridge regression is given by the following equation:

$$\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \alpha \cdot \|\mathbf{w}\|_2^2$$

- **Support Vector Regression (SVR):** SVR is a regression adaptation of support vector machines (SVMs). It works by trying to predict a continuous value, while still using the principle of maximizing the margin, much like in classification tasks with SVM. The idea is to find a hyperplane that best fits the data, with the consideration of a margin that captures the maximum number of data points [32]. The objective function for SVR is given by the following equation:

$$\min_{\mathbf{w}, b, \zeta, \zeta^*} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n (\zeta_i + \zeta_i^*)$$

subject to the conditions

$$\begin{aligned} y_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b &\leq \varepsilon + \zeta_i, \\ \mathbf{w}^T \phi(\mathbf{x}_i) + b - y_i &\leq \varepsilon + \zeta_i^*, \\ \zeta_i, \zeta_i^* &\geq 0, i = 1, \dots, n \end{aligned}$$

- **Random Forest (RF):** Random forest is an ensemble learning method that operates by constructing multiple decision trees during training. For predictions, the mode of the classes (for classification) or mean prediction (for regression) of the individual trees is returned. Random forests help combat the overfitting problem that can plague individual decision trees [33, 34, 35].
- **Gradient Boosting:** Gradient boosting is an ensemble machine learning technique that builds a series of weak learners, typically decision trees, in a sequential manner. Each tree in the sequence aims to correct the errors made by its predecessor. The core idea is to give more weight to the misclassified or mispredicted data points from the previous tree, ensuring that the next tree in the sequence addresses these errors. The

final prediction is obtained by aggregating the predictions of all the trees [36, 37]. Gradient boosting is particularly effective in scenarios where there's a high bias, as it focuses on reducing bias iteratively.

- **Artificial Neural Network (ANN):** An artificial neural network is a computing system inspired by the biological neural networks in the human brain. It is composed of interconnected nodes (neurons) arranged in layers: input, hidden, and output layers. Due to a recent surge in popularity there currently exist many different types of neural networks including multi-layer perceptron, convolutional, recurrent, graph, generative, and other neural network architectures. These networks can adapt to and learn from data, making them suitable for complex tasks like image and speech recognition, as well as regression and classification problems in structured data [38, 39].

For each layer  $l$  in the neural network, the pre-activation  $\mathbf{z}^{[l]}$  is computed as:

$$\mathbf{z}^{[l]} = \mathbf{w}^{[l]} \mathbf{A}^{[l-1]} + \mathbf{b}^{[l]} \quad (3.1)$$

Then the activation  $\mathbf{A}^{[l]}$  for the hidden layers is computed using:

$$\mathbf{A}^{[l]} = \text{ReLU}(\mathbf{z}^{[l]}) \quad (3.2)$$

Finally, the objective function for a regression problem is the mean squared error.

- **Stacking:** Stacking, also known as "stacked generalization," is an ensemble learning technique designed to improve prediction accuracy by combining multiple models' outputs. It consists of several base models together with a meta-model. Once trained on the data, each of the base models generates predictions which are then treated as meta-features. The meta-features are used to train the meta-model that is responsible for producing the final predictions. When predicting a new instance, each base model provides its prediction. These are then used as input features for the meta-model, which subsequently determines the final prediction for that instance. The overarching goal of the stacking model is to capitalize on the strengths of each individual model, aiming to produce predictions that are more accurate than any single model could achieve on its own [40].

#### 4. DATA

The Concrete Compressive Strength dataset is one of the datasets available in the UCI repository and is particularly relevant for civil engineering applications [41]. The summary of the data is provided in Table 1. It shows that there are 8 input variables and 1 output variable. The dataset pertains to the compressive strength of different samples of concrete based on the volumes of the different ingredients that were used to make them. The target variable is the concrete sample's compressive strength, which can be predicted using the volume proportions of the ingredients.

| Feature                       | Mean   | Std    | Min    | Max     |
|-------------------------------|--------|--------|--------|---------|
| Cement                        | 281.17 | 104.51 | 102.00 | 540.00  |
| Blast Furnace Slag            | 73.90  | 86.28  | 0.00   | 359.40  |
| Fly Ash                       | 54.19  | 63.50  | 0.00   | 200.10  |
| Water                         | 181.57 | 21.35  | 121.80 | 247.00  |
| Superplasticizer              | 6.20   | 5.97   | 0.00   | 32.20   |
| Coarse Aggregate              | 972.92 | 77.75  | 801.00 | 1145.00 |
| Fine Aggregate                | 773.58 | 80.18  | 594.00 | 992.60  |
| Age                           | 45.66  | 63.17  | 1.00   | 365.00  |
| Concrete Compressive Strength | 35.82  | 16.71  | 2.33   | 82.60   |

TABLE 1. Summary statistics for the Concrete Compressive Strength dataset.

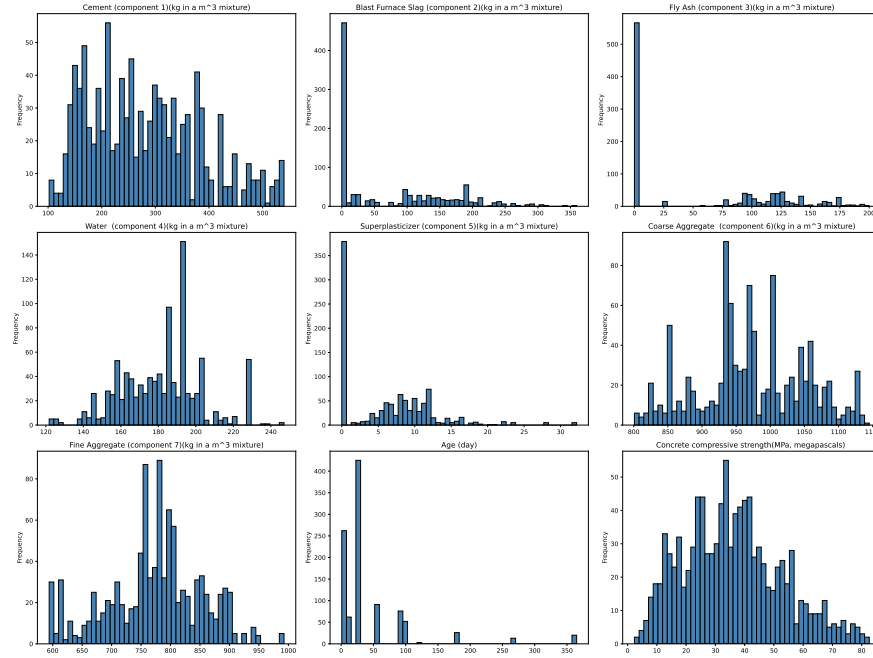


FIGURE 1. Distribution of the feature values.

To better comprehend the data, we analyze the distribution of its feature values as shown in Figure 1. Understanding these distributions is essential for any subsequent analysis or modeling, as they provide insights into the typical values and variations of each feature in the dataset.

The distribution of the **Cement** feature in the dataset is somewhat bimodal, with a notable peak around 150-250 and another significant peak near 500. This suggests that many concrete mixtures either contain a low or a high amount of cement. On the other hand, the **Blast Furnace Slag** feature is heavily skewed to the right, indicating that a large number of mixtures have minimal amounts of blast furnace slag, although there's also a spread of mixtures containing up to

around 250. The distribution of **Fly Ash** is bimodal, with a substantial number of mixtures having no fly ash, and a secondary peak observed around 100-150. This suggests that when fly ash is incorporated, it typically falls within these amounts.

The **Water** feature's distribution is somewhat bell-shaped, albeit with a slight right skew. Most concrete mixtures seem to have a water content ranging from 150-200, but a tail extends to the higher values. For the **Superplasticizer** feature, the data is right-skewed. Many mixtures contain low amounts of superplasticizer, with a noticeable peak around 10. The distribution of **Coarse Aggregate** appears bell-shaped, centering around 950, indicating that the bulk of the mixtures possess a coarse aggregate amount close to this value. Similarly, the **Fine Aggregate** feature's distribution is bell-shaped but with a slight leftward skew, with the majority of mixtures having their fine aggregate amounts in the 600-800 range.

One of the most right-skewed features is **Age**. The majority of the concrete samples are younger, predominantly clustering around 28 days. However, there are a few samples extending up to approximately 365 days or even more. Lastly, when examining the "Concrete compressive strength," the distribution is somewhat bell-shaped with a mild right skew. Most concrete mixtures possess strengths between 20 and 60 MPa, but there's a tail of mixtures that exhibit strengths stretching beyond 80 MPa.

## 5. METHODOLOGY

**5.1. Dataset.** The dataset utilized for this study pertains to the compressive strength of concrete. It comprises multiple features, including the amounts of various ingredients used in concrete mixtures, such as cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate, along with the age of the concrete. The target variable is the concrete's compressive strength. Further details and analysis of the dataset can be found in [Section 4](#).

**5.2. Preprocessing.** Before model training, the dataset underwent preprocessing. In particular, features were scaled to have zero mean and unit variance using the standard scaler. It ensures that all features contribute equally to the model's performance and is particularly crucial for models sensitive to feature scales, such as SVR.

**5.3. Model Selection.** Multiple machine learning algorithms were evaluated:

- Linear Regression
- Lasso
- Ridge
- Support Vector Regression
- Random Forest
- Gradient Boosting
- Artificial Neural Network
- Stacking

We used `sklearn` library to carry out the numerical experiments [42]. The model settings are presented in Table 2.

| Model                     | Settings                                                     |
|---------------------------|--------------------------------------------------------------|
| Linear Regression         | Default settings (no regularization)                         |
| Lasso                     | Alpha (regularization strength) set to 1                     |
| Ridge                     | Alpha (regularization strength) set to 1                     |
| SVR                       | Default settings with kernel='rbf'                           |
| Random Forest             | Default settings                                             |
| Gradient Boosting         | Default settings                                             |
| Artificial Neural Network | Two hidden layers with 64 and 32 neurons, max_iter=1000      |
| Stacking                  | Base models: RF, GB, ANN, SVR; Meta-model: Linear Regression |

TABLE 2. Parameter settings of each model.

**5.4. Evaluation Metrics.** To objectively assess model performance, the following metrics were employed:

- Mean Squared Error (MSE)
- Mean Absolute Error (MAE)
- $R^2$  (Coefficient of Determination)

**5.5. Model Training and Validation.** A 4-fold cross-validation approach was employed for model validation:

- The dataset was partitioned into four subsets. In each iteration, one subset was held out for validation, while the remaining three subsets were used for training.
- The performance metrics were computed for each fold, and the average performance across all folds was used to evaluate each model's overall performance.

## 6. RESULTS AND DISCUSSION

The results of the computational experiments are presented in Table 3. The stacking model outperforms all other models in terms of all the coefficient of determination, demonstrating the strength of ensemble approaches, particularly when base models complement each other. The other ensemble models - random forest and gradient boosting - similarly achieve high predictive performance. Due to their tree-based nature, they are capable of handling non-linear relationships and feature interactions, making them appropriate for intricate datasets like the one in question. On the other hand, while neural networks are potent, they demand careful tuning. The moderate performance observed in our analysis suggests that there is potential for enhancement through more advanced architectures or hyperparameter tuning. In comparison, linear models, including linear regression, ridge regression, and lasso regression, offer quick execution but may not grasp the complex relationships present in the dataset. Specifically, lasso regression's performance implies that either it might not be the ideal model for this

dataset or it requires more detailed tuning. Analysis of the execution times can provide insight on computational efficiency. While linear models are extremely fast and efficient, ensemble models like random forest and gradient boosting take longer. As expected, the stacking model, which combines multiple models, has the longest execution time. It's important to strike a balance between accuracy and execution time. In situations demanding real-time solutions or those with computational restrictions, simpler, quicker models might be favored. Conversely, for non-immediate tasks or when the prediction accuracy is crucial, more intricate models or ensemble techniques could be preferable.

| Model                     | MSE    | MAE  | $R^2$  | Execution Time |
|---------------------------|--------|------|--------|----------------|
| Linear Regression         | 101.53 | 8.05 | 0.6091 | 0.0136         |
| Ridge Regression          | 101.53 | 8.05 | 0.6091 | 0.0020         |
| Lasso Regression          | 130.32 | 9.42 | 0.5086 | 0.0023         |
| Support Vector Regression | 99.60  | 7.99 | 0.6200 | 0.1078         |
| Random Forest             | 32.18  | 3.79 | 0.8642 | 0.6911         |
| Gradient Boosting         | 30.61  | 3.97 | 0.8718 | 0.2281         |
| Neural Network            | 33.31  | 3.80 | 0.8690 | 30.7311        |
| Stacking model            | 31.74  | 3.79 | 0.8760 | 206.33         |

TABLE 3. Performance metrics for different machine learning models using 4-fold cross validation.

## 7. CONCLUSION

Selecting the right model requires a careful consideration of both performance metrics and practical constraints. While ensemble techniques like stacking often provide superior performance, it's crucial to weigh their benefits against increased computational costs. In applications where time is of the essence, simpler models might be more fitting. On the other hand, when the highest predictive accuracy is sought, more sophisticated models or ensemble techniques are worth the computational investment.

## REFERENCES

- [1] Gupta, S., & Sihag, P. (2022). Prediction of the compressive strength of concrete using various predictive modeling techniques. *Neural Computing and Applications*, 34(8), 6535-6545. <https://doi.org/10.1007/s00521-021-06820-y>
- [2] Ahmed, H. U., Mohammed, A. A., Rafiq, S., Mohammed, A. S., Mosavi, A., Sor, N. H., & Qaidi, S. M. (2021). Compressive strength of sustainable geopolymers concrete composites: a state-of-the-art review. *Sustainability*, 13(24), 13502. <https://doi.org/10.3390/su132413502>
- [3] Jin, L., Yu, W., Li, D., & Du, X. (2021). Numerical and theoretical investigation on the size effect of concrete compressive strength considering the maximum aggregate size. *International Journal of Mechanical Sciences*, 192, 106130. <https://doi.org/10.1016/j.ijmecsci.2020.106130>

- [4] Elkhatab, Z., Kamalov, F., Moussa, S., Mnaouer, A. B., Yagoub, M., & Yanikomeroglu, H. (2024). Radio Modulation Classification Optimization Using Combinatorial Deep Learning Technique. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3357628>
- [5] Kamalov, F., Sulieman, H., Moussa, S., Reyes, J. A., & Safaraliev, M. (2023). Nested ensemble selection: An effective hybrid feature selection method. *Heliyon*, 9(9). <https://doi.org/10.1016/j.heliyon.2023.e19686>
- [6] Kamalov, F., Thabtah, F., & Leung, H. H. (2023). Feature selection in imbalanced data. *Annals of Data Science*, 10(6), 1527-1541. <https://doi.org/10.1007/s40745-021-00366-5>
- [7] Kamalov, F., Sulieman, H., Moussa, S., Avante Reyes, J., & Safaraliev, M. (2024). Powering Electricity Forecasting with Transfer Learning. *Energies*, 17(3), 626. <https://doi.org/10.3390/en17030626>
- [8] Ahmad, M., Hu, J. L., Ahmad, F., Tang, X. W., Amjad, M., Iqbal, M. J., ... & Farooq, A. (2021). Supervised learning methods for modeling concrete compressive strength prediction at high temperature. *Materials*, 14(8), 1983. <https://doi.org/10.3390/ma14081983>
- [9] Ahmad, A., Ostrowski, K. A., Maslak, M., Farooq, F., Mehmood, I., & Nafees, A. (2021). Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature. *Materials*, 14(15), 4222. <https://doi.org/10.3390/ma14154222>
- [10] DeRousseau, M. A., Laftchiev, E., Kasprzyk, J. R., Rajagopalan, B., & Srubar III, W. V. (2019). A comparison of machine learning methods for predicting the compressive strength of field-placed concrete. *Construction and Building Materials*, 228, 116661. <https://doi.org/10.1016/j.conbuildmat.2019.08.042>
- [11] Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching*, 6(2). <https://doi.org/10.37074/jalt.2023.6.2.22>
- [12] Das, V., Cherukuri, A. K., Hu, Q., Kamalov, F., & Jonnalagadda, A. (2023). Proactive AI enhanced consensus algorithm with fraud detection in blockchain. In *Blockchain for Cybersecurity in Cyber-Physical Systems* (pp. 259-274). Cham: Springer International Publishing. Chicago [https://doi.org/10.1007/978-3-031-25506-9\\_13](https://doi.org/10.1007/978-3-031-25506-9_13)
- [13] Elreedy, D., Atiya, A. F., & Kamalov, F. (2023). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, 1-21. <https://doi.org/10.1007/s10994-022-06296-4>
- [14] Jonnalagadda, A., Rajvir, M., Singh, S., Chandramouliswaran, S., George, J., & Kamalov, F. (2023). An Ensemble-Based Machine Learning Model for Emotion and Mental Health Detection. *Journal of Information & Knowledge Management*, 22(02), 2250075. <https://doi.org/10.1142/S0219649222500757>
- [15] Kamalov, F., Sulieman, H., & Cherukuri, A. K. (2023, June). Synthetic data for feature selection. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 353-365). Cham: Springer Nature Switzerland. [https://doi.org/10.1007/978-3-031-42508-0\\_32](https://doi.org/10.1007/978-3-031-42508-0_32)
- [16] Kamalov, F., Santandreu, D., Leung, H. H., Johnson, J., & El Khatib, Z. (2023, May). Leveraging computer algebra systems in calculus: A case study with SymPy. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-6). IEEE. <https://doi.org/10.1109/EDUCON54358.2023.10125196>
- [17] Kamalov, F., Elnaffar, S., Sulieman, H., & Cherukuri, A. K. (2023). XyGen: synthetic data generator for feature selection. *Software Impacts*, 15, 100485. <https://doi.org/10.1016/j.simpa.2023.100485>
- [18] Nazir, A., Mitra, R., Sulieman, H., & Kamalov, F. (2023). Suspicious Behavior Detection with Temporal Feature Extraction and Time-Series Classification for Shoplifting Crime Prevention. *Sensors*, 23(13), 5811. <https://doi.org/10.3390/s23135811>
- [19] Zgheib, R., Chahbandarian, G., Kamalov, F., El Messiry, H., & Al-Gindy, A. (2023). Towards an ML-based semantic IoT for pandemic management: A survey of enabling technologies

- for COVID-19. *Neurocomputing*, 528, 160-177. <https://doi.org/10.1016/j.neucom.2023.01.007>
- [20] Nunez, I., Marani, A., Flah, M., & Nehdi, M. L. (2021). Estimating compressive strength of modern concrete mixtures using computational intelligence: A systematic review. *Construction and Building Materials*, 310, 125279. <https://doi.org/10.1016/j.conbuildmat.2021.125279>
- [21] Mylvaganam, N., & Elakneswaran, Y. (2023). A Systematic Review and Assessment of Concrete Strength Prediction Models. *Case Studies in Construction Materials*, e01830. <https://doi.org/10.1016/j.cscm.2023.e01830>
- [22] Li, D., Tang, Z., Kang, Q., Zhang, X., & Li, Y. (2023). Machine Learning-Based Method for Predicting Compressive Strength of Concrete. *Processes*, 11(2), 390. <https://doi.org/10.3390/pr11020390>
- [23] Moccia, F., Yu, Q., Fernández Ruiz, M., & Muttoni, A. (2021). Concrete compressive strength: From material characterization to a structural value. *Structural Concrete*, 22, E634-E654. Chicago <https://doi.org/10.1002/suco.202000211>
- [24] Kashem, A., & Das, P. (2023). Compressive strength prediction of high-strength concrete using hybrid machine learning approaches by incorporating SHAP analysis. *Asian Journal of Civil Engineering*, 1-21. <https://doi.org/10.1007/s42107-023-00707-0>
- [25] Al Yamani, W. H., Ghunimat, D. M., & Bisharah, M. M. (2023). Modeling and predicting the sensitivity of high-performance concrete compressive strength using machine learning methods. *Asian Journal of Civil Engineering*, 1-13. <https://doi.org/10.1007/s42107-023-00658-6>
- [26] Chen, X., Zheng, D., Wu, X., & Li, C. (2023). A Review of Three Common Concrete Multi-axial Strength Criteria from 2010 to 2020. *Archives of Computational Methods in Engineering*, 30(2), 811-829. <https://doi.org/10.1007/s11831-022-09822-8>
- [27] Li, J., Hong, J., Liu, S., Zhou, Y., & Meng, K. (2023). Multiaxial compressive strength of hybrid fiber reinforced concrete: A unified empirical model. *Frontiers in Materials*, 10, 1100868. <https://doi.org/10.3389/fmats.2023.1100868>
- [28] Rashid, M. U. (2020). Experimental investigation on durability characteristics of steel and polypropylene fiber reinforced concrete exposed to natural weathering action. *Construction and Building Materials*, 250, 118910. <https://doi.org/10.1016/j.conbuildmat.2020.118910>
- [29] Ashwini, R. M., Potharaju, M., Srinivas, V., Kanaka Durga, S., Rathnamala, G. V., & Paudel, A. (2023). Compressive and flexural strength of concrete with different nanomaterials: a critical review. *Journal of Nanomaterials*, 2023. Chicago <https://doi.org/10.1155/2023/1004597>
- [30] Gurrib, I., Kamalov, F., & Smail, L. (2021, August). Bitcoin price forecasting: Linear discriminant analysis with sentiment evaluation. In *The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research* (pp. 1-5). <https://doi.org/10.1145/3485557.3485561>
- [31] Senyuk, M., Elnaggar, M. F., Safaraliev, M., Kamalov, F., & Kamel, S. (2023). Statistical method of low frequency oscillations analysis in power systems based on phasor measurements. *Mathematics*, 11(2), 393. <https://doi.org/10.3390/math11020393>
- [32] Kamalov, F., Cherukuri, A. K., Sulieman, H., Thabtah, F., & Hossain, A. (2023). Machine learning applications for COVID-19: a state-of-the-art review. *Data science for genomics*, 277-289. <https://doi.org/10.1016/B978-0-323-98352-5.00010-0>
- [33] Kamalov, F., Moussa, S., Zgheib, R., & Mashaal, O. (2020, December). Feature selection for intrusion detection systems. In *2020 13th international symposium on computational intelligence and design (ISCID)* (pp. 265-269). IEEE. <https://doi.org/10.1109/ISCID51228.2020.00065>
- [34] Senyuk, M., Safaraliev, M., Kamalov, F., & Sulieman, H. (2023). Power system transient stability assessment based on machine learning algorithms and grid topology. *Mathematics*, 11(3), 525. <https://doi.org/10.3390/math11030525>

- [35] Thabtah, F., & Kamalov, F. (2017). Phishing detection: a case analysis on classifiers with rules using machine learning. *Journal of Information & Knowledge Management*, 16(04), 1750034. <https://doi.org/10.1142/S0219649217500344>
- [36] Kamalov, F., Sayari, Z., Gheisari, M., Lee, C. C., & Moussa, S. (2023). Routing algorithms in internet of things complex network with the role of machine learning. *Cybernetics and Physics*, 12(3), 182-193. <https://doi.org/10.35470/2226-4116-2023-12-3-182-193>
- [37] Safaraliev, M., Kiryanova, N., Matrenin, P., Dmitriev, S., Kokin, S., & Kamalov, F. (2022). Medium-term forecasting of power generation by hydropower plants in isolated power systems under climate change. *Energy Reports*, 8, 765-774. <https://doi.org/10.1016/j.egy.2022.09.164>
- [38] Kamalov, F., Smail, L., & Gurrib, I. (2020, December). Forecasting with deep learning: S&P 500 index. In *2020 13th International Symposium on Computational Intelligence and Design (ISCID)* (pp. 422-425). IEEE. <https://doi.org/10.1109/ISCID51228.2020.00102>
- [39] Kamalov, F., Rajab, K., Cherukuri, A. K., Elnagar, A., & Safaraliev, M. (2022). Deep learning for Covid-19 forecasting: State-of-the-art review. *Neurocomputing*, 511, 142-154. <https://doi.org/10.1016/j.neucom.2022.09.005>
- [40] Kamalov, F., Moussa, S., & Avante Reyes, J. (2022). KDE-based ensemble learning for imbalanced data. *Electronics*, 11(17), 2703. <https://doi.org/10.3390/electronics11172703>
- [41] Yeh, I-Cheng. (2007). Concrete Compressive Strength. UCI Machine Learning Repository. <https://doi.org/10.24432/C5PK67>.
- [42] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

<sup>1</sup>FACULTY OF CIVIL ENGINEERING, TAJIK TECHNICAL UNIVERSITY, DUSHANBE, TAJIK-ISTAN.

*Email address:* [said.kamolov@ttu.tj](mailto:said.kamolov@ttu.tj)