frontiers | Frontiers in Materials

TYPE Original Research
PUBLISHED 01 July 2024
DOI 10.3389/fmats.2024.1416918

OPEN ACCESS

EDITED BY
Jeffrey W. Bullard,
Texas A and M University, United States

REVIEWED BY
Salvatore Verre,
University of eCampus, Italy
Ayub Elahi,
University of Engineering and Technology, Taxila, Pakistan

*CORRESPONDENCE
Qianhui Zhang,
20191709@zjgsdx.edu.cn
Hamzeh Ghorbani,
hamzehghorbani68@yahoo.com

RECEIVED 15 April 2024
ACCEPTED 06 June 2024
PUBLISHED 01 July 2024

CITATION
Zhang Q, Jin Y, Wang G, Sun Q and Ghorbani H (2024), Improving stability and safety in concrete structures against high-energy projectiles: a machine learning perspective.
Front. Mater. 11:1416918.
doi: 10.3389/fmats.2024.1416918

Improving stability and safety in concrete structures against high-energy projectiles: a machine learning perspective

Qianhui Zhang[1]*, Yuzhen Jin[1], Guangzhi Wang[2], Qingmei Sun[1] and Hamzeh Ghorbani[3]*

[1]School of Cml Engineering Architecture, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, Zhejiang, China, [2]School of Management Engineering, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, Zhejiang, China, [3]Young Researchers and Elite Club, Islamic Azad University, Ahvaz, Iran

Concrete structures are commonly used as secure settlements and strategic shelters due to their inherent strength, durability, and wide availability. Examining the robustness and integrity of strategic concrete structures in the face of super-energy projectiles is of utmost significance in safeguarding vital infrastructure sectors, ensuring the well-being of individuals, and advancing the course of worldwide sustainable progress. This research focuses on forecasting the penetration depth (BPD) through the application of robust models, such as Multilayer Perceptron (MLP), Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), and K-Nearest Neighbors (KNN) as ML models. The dataset used consists of 1,020 data points sourced from the National Institute of Standards and Technology (NIST), encompassing various parameters such as cement content (Cp), ground granulated blast-furnace slag (GGBFS), fly ash content (FA), water portion (Wp), superplasticizer content (Sp), coarse aggregate content (CA), fine aggregate content (FAA), concrete sample age (t), concrete compressive strength (CCS), gun type (G-type), bullet caliber (B-Cali), bullet weight (Wb), and bullet velocity (Vb). Feature selection techniques revealed that the MLP model, incorporating eight input variables (FA, CA, Sp, GGBFS, Cp, t, FAA, and CCS), provides the most accurate predictions for BPD across the entire dataset. Comparing the four models used in this study, KNN demonstrates distinct superiority over the other methods. KNN, a non-parametric ML model used for classification and regression, possesses several advantages, including simplicity, non-parametric nature, no training requirements, robustness to noisy data, suitability for large datasets, and interpretability. The results reveal that KNN outperforms the other models presented in this paper, exhibiting an R^2 value of 0.9905 and an RMSE value of 0.1811 cm, signifying higher accuracy in its predictions compared to the other models. Finally, based on the error analysis across iterations, it is evident that the final accuracy error of the KNN model surpasses that of the SVM, MLP, and LightGBM models, respectively.

KEYWORDS

strategic structure, concrete structures, bullet penetration depth, machine learning, construction materials

Highlights

- Bullet penetration crucial for protecting personnel and assets.
- 4 ML techniques predicted bullet penetration (14,280 data).
- Feature selection determined accurate MLP model with eight variables.
- KNN predicts ballistic performance in concrete blocks accurately.

1 Introduction

Concrete, a fundamental material used in constructing secure dwellings, is a composite made up of cement, water, and aggregates like sand, gravel, or crushed stone. It serves as a significant reinforcing material in ensuring the construction of robust and durable structures (Brandt, 2008). Its versatility, excellent compressive strength, and ability to withstand various environmental conditions make it an indispensable component in the construction industry (Amaguaña et al., 2023; Srivani and Mohan, 2023). One of the key features that makes concrete stand out is its remarkable compressive strength (Jiao et al., 2023; Kumar et al., 2023). This property makes it an ideal material for supporting heavy loads and resisting deformation under pressure, making it particularly suitable for constructing high-rise buildings, bridges, dams, and other structures where weight-bearing capacity is paramount (Bentz, 2007). Moreover, concrete structures are renowned for their durability and longevity (Rabi et al., 2022). When properly designed and maintained, concrete can withstand the test of time and endure harsh weather conditions, chemical exposure, and other external factors, ensuring the structural integrity of the building over an extended period (Fahimizadeh et al., 2022).

Beyond its strength and durability, concrete offers versatility in forming complex shapes. It can be molded into various forms and sizes, allowing for creative and innovative architectural designs while maintaining structural strength and functionality (Li et al., 2022). Its proven track record over the years has solidified its position as the backbone of modern infrastructure, contributing significantly to the advancement and safety of the built environment (Habibi Rad et al., 2022). Because of the fact that concrete structures possess this inherent ability, the concrete has been widely utilized in the construction of important superstructures and civil infrastructures, such as bridges, dams, hydroelectric power plants, nuclear power plants, and significant strategic and transit hubs (Rich et al., 2022).

There is a necessity for conducting fundamental research on strategic structures. Given the strategic significance of concrete structures, it is essential to assess the structural requirements and security protocols that enable them to withstand potential destructive threats. One such threat is the vulnerability of concrete structures to the forceful impact of super-energetic projectiles (Liu J. et al., 2022). Super-energetic projectiles are projectiles (objects or devices launched or fired through the air) that possess high levels of energy or force. These projectiles are often designed to travel at high speeds and deliver a significant amount of impact or damage upon impact with a target. Many critical infrastructures, such as military bases, power plants, government buildings, and transportation hubs, are potential targets for threats involving super-energetic projectiles (Wang et al., 2023). Understanding how cement and concrete compositions can influence the resistance of these structures to such threats is important for enhancing their protection (Cheng et al., 2023). With the evolving nature of security threats, it's important for governments and corresponding organizations to stay ahead in terms of technology and materials. Research in this area can lead to improvements in building materials and protective measures, ensuring higher levels of national security. By studying the impact of super-energetic projectiles on cement and concrete compositions, civil engineers can better design structures to minimize harm to surrounding communities during such events (Park et al., 2023). The research can also provide valuable insights into developing more robust and resilient construction materials for both strategic and civilian structures, making them safer against various threats, including natural disasters.

It is obvious that the ability of unreinforced concrete to withstand super-energetic projectiles may have its limitations, especially in comparison to reinforced concrete or other specialized materials designed explicitly for such purposes. Today, although structures made with reinforced concrete have taken a very large role in strengthening the strength and protective role of tactical structures, however, for a deep understanding in fundamental research, it is necessary to measure the efficiency of non-reinforced concrete structures in terms of the qualitative role of the structural composition in supporting safety against super-energetic projectiles and its ability to absorb destructive projectile energy. Investigating the role of unreinforced cement and concrete composition on the ability to control and penetrate super-energetic projectiles involves considering various factors that influence the strength and performance of the material against such projectiles.

One key factor is the cement type used in the concrete components. While ordinary Portland cement (OPC) is commonly used, other types such as Portland pozzolana cement (PPC) or slag cement can offer improved properties (Chintalapudi and Pannem, 2022; Derakhshani et al., 2023). Higher-strength cement types generally provide better performance for projectile resistance (Wang et al., 2016). The water-to-cement ratio (w/c) significantly influences both the strength and durability of concrete. A reduced w/c ratio typically results in increased strength and enhanced durability of the concrete mixture (Vu et al., 2009). Too much water in the mix can weaken the concrete (Surahyo et al., 2019) and make it more susceptible to damage from severe local collisions.

The choice of aggregates, such as gravel or crushed stone, and their particle size significantly impact concrete properties (Haach et al., 2011). Well-graded, dense aggregates with suitable particle sizes can enhance the concrete's ability to resist penetration by projectiles (Dancygier et al., 2007).

Admixtures are chemical additives used to modify concrete properties. Certain admixtures like silica fume, fly ash, or superplasticizers can be used to enhance projectile resistance by improving strength, reducing permeability, and increasing the density of the concrete (Wang et al., 2016). The precise proportioning of cement, aggregates, and water in the mix is important for achieving the desired strength and performance of the concrete. Proper mix design is essential for optimizing the material's ability to withstand the impact of intensive external forces (Day, 2006). Adequate curing is essential for concrete to achieve its maximum strength capacity. Proper curing ensures that the cement hydrates fully, forming a strong matrix. Inadequate curing can lead

to lower strength and increased susceptibility to damage (Pawar and Kate, 2020). Lower porosity and permeability are desirable in concrete meant to resist super-energetic projectiles. High porosity allows easier penetration and can lead to spalling when exposed to projectile impacts. The bond between the cementitious matrix and the aggregate particles, known as the aggregate-cement interface, is important for overall concrete strength (Liu Y. et al., 2022). A robust interfacial bond enhances the concrete's capacity to endure dynamic loads and impact stresses.

Concrete gains strength over time due to the ongoing hydration process (Neville and Brooks, 1987). Older concrete is generally stronger and more durable than freshly placed concrete, making the age of the concrete an important factor in its resistance to projectiles.

Data-driven research and the application of machine learning (ML) offer several advantages in various fields, particularly in cement and concrete research. By utilizing data-driven research methods, scientists and engineers can comprehensively analyze vast amounts of data from diverse sources like lab experiments, simulations, and field tests (Brunton and Kutz, 2022). This comprehensive approach leads to more robust and reliable conclusions. Furthermore, ML models have the capability to identify complex patterns and correlations in data that may not be apparent through traditional analytical methods (Uddin et al., 2022). The ability to uncover these intricate relationships helps researchers identify critical factors that influence the performance of concrete. Furthermore, the utilization of ML facilitates the refinement of cement and concrete formulations tailored to particular circumstances and challenges. By utilizing these models, researchers can develop tailored, high-performance materials that excel in particular environments.

Another advantage of ML is that it helps reduce experimental costs (Mohamadian et al., 2021). By using ML models to predict outcomes and behaviors, extensive physical testing becomes less necessary, saving both time and resources. Moreover, ML enables rapid prototyping by generating and evaluating numerous hypothetical scenarios. This ability to quickly assess different possibilities accelerates the research and development process (Beheshtian et al., 2022).

The ML models are also adaptable, as they can be updated and improved as new data becomes available. This ensures that research remains up-to-date and relevant, reflecting the latest information. Lastly, ML allows for multi-factor analysis, enabling the simultaneous examination of multiple influencing factors. This holistic approach provides a thorough understanding of how cement and concrete behave under different conditions.

This study uses a data-driven research model and the application of ML to reveal the hidden, complicated and deep aspects of the complex relationships of the role of concrete individual compositions in creating resistance strength against intense localized collisions with super-energetic objects. This research on the interaction of super-energy projectiles with cement and unreinforced concrete provides a deep understanding based on the insights gained from application of data-driven ML models, which has not been addressed in the field so far. Based on the best knowledge of the authors, this groundbreaking research offers a novel outlook on the utilization of data-driven ML capabilities for the surveillance of quality control and safety aspects in concrete structures. It achieves this by investigating the correlation between the composition of concrete structures and their capacity to withstand the impact of highly energetic projectiles. Firstly, it enhances resistance by providing insights into how different compositions and structural designs respond to high-velocity impacts. This understanding allows engineers to create structures that are more resistant to penetration and damage caused by super-energetic projectiles. Secondly, this research offers cost-effective solutions. By optimizing cement and concrete compositions, it becomes possible to achieve higher levels of protection without significantly increasing construction costs. This allows for the development of structures that provide enhanced security at a reasonable cost. Furthermore, the outcomes of this research can guide the retrofitting of existing strategic structures. By applying the knowledge gained, existing structures can be improved to increase their resilience against emerging threats. Figure 1, show the graphical parameter's effect on the non-reinforced concrete bulk.

2 Research background

Bullet penetration depth (BPD) or projectile impact in concrete blocks represents a important and extensively studied research area that holds significance construction industries. Numerous investigations have been conducted to elucidate the underlying mechanisms of BPD in concrete blocks, as these findings inform building design, protective coatings, and the development of materials resistant to ballistic impact. This review provides a comprehensive overview of existing research pertaining to BPD in concrete blocks.

Lai et al. (2018) present research findings demonstrating the significant advantages of using ultra-high-performance concrete (UHPC) in terms of increased resistance to high-velocity projectiles and enhanced energy absorption capacity compared to conventional concrete. The paper examines various properties of UHPC, including compressive and flexural strengths, energy absorption capability, and penetration depth. Additionally, numerical models are developed to predict the performance of UHPC under various types of multiple bullet impacts (Lai et al., 2018). Zhang et al. (2020) investigate the important parameters influencing the depth of penetration caused by impacts from small caliber non-deformable projectiles in cement-based materials. The authors used two distinct systems to assess the penetration resistance of the material: a drop-weight impact test and a ballistic test. Their observations indicate that concrete exhibits superior penetration resistance compared to mortar, with compressive strength being the most influential parameter affecting penetration depth (Zhang et al., 2020). In another study, Cao et al. (2020) explore the efficiency of ultra-high-performance fiber-reinforced concrete (UHPFRC) as a structural material in high-risk applications with respect to its ballistic tolerance. The study focuses on investigating the behavior of coarse-aggregated layered UHPFRC through numerical simulations (Cao et al., 2020). Additionally, a study conducted by Li et al. in 2020 analyzed the impact of significant design factors of ultra-high-performance fiber-reinforced concrete on its resistance to bullets. The authors conducted experiments to investigate the relationship between impact velocity, concrete thickness, and bullet resistance. The results indicated that increasing the thickness of the concrete

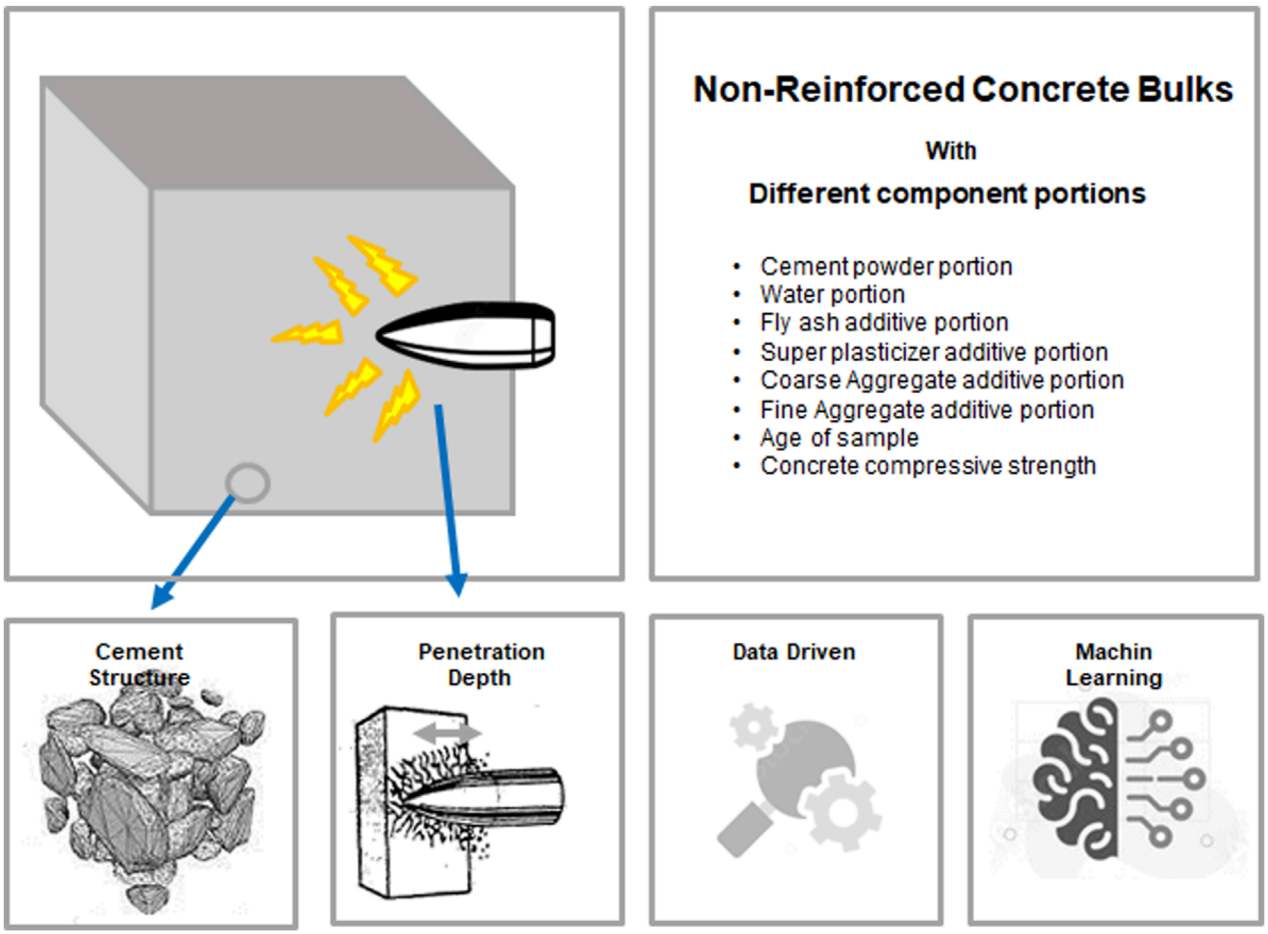

FIGURE 1
Graphical parameter's effect on the non-reinforced concrete bulk.

layer and utilizing high-performance fibers led to improved bullet resistance performance (Li et al., 2020). The research conducted by Mina et al. (2021) undertakes an evaluation of the capability of optimized ultra-high-performance fiber-reinforced concrete (UHPFRC) to withstand high-velocity impacts from projectiles. The findings demonstrate that the new UHPFRC exhibits the capacity to withstand projectile impacts with velocities exceeding 500 m/s, highlighting its potential for use in structures requiring exceptional durability and impact resistance. The study uses a series of impact tests, and the authors conclude that the optimized UHPFRC showcases notable enhancements in impact resistance compared to conventional concrete materials (Mina et al., 2021). In a recent publication in Scientific Reports, Campbell et al. (2022) investigate the impact of bullets on building stones, revealing that such impacts result in conical-shaped craters on the targeted materials. The researchers conducted experiments utilizing sandstone and limestone and observed that the crater shape and size were influenced by the strength and hardness of the respective stones (Campbell et al., 2022). Wu et al. (2022) contribute insights into the deflection of projectiles upon penetrating composite concrete targets. Their article discusses the influence of projectile mass, velocity, and angle of incidence on the deflection response (Wu et al., 2022). A study by Abbas et al. (2023) examines the behavior and effectiveness of reinforced concrete composite blocks when subjected to projectile impact forces. The findings offer valuable insights into understanding the response of such structures when faced with projectile impacts (Abbas et al., 2023).

Prior studies have established that BPD into concrete is influenced by factors such as bullet caliber, bullet velocity, bullet weight, and concrete strength. The incorporation of fiber reinforcement and additives like steel can improve the concrete's resistance. However, previous studies have not utilized ML models to predict the depth of BPD. This research seeks to bridge this gap through the utilization of a substantial dataset and using four robust ML models with diverse computational approaches, thereby offering valuable ML insights for the engineering of defense structures.

3 Materials and methods

3.1 Data gathering

The present research has made use of an extensive and comprehensive dataset consisting of 1,020 datapoints. These

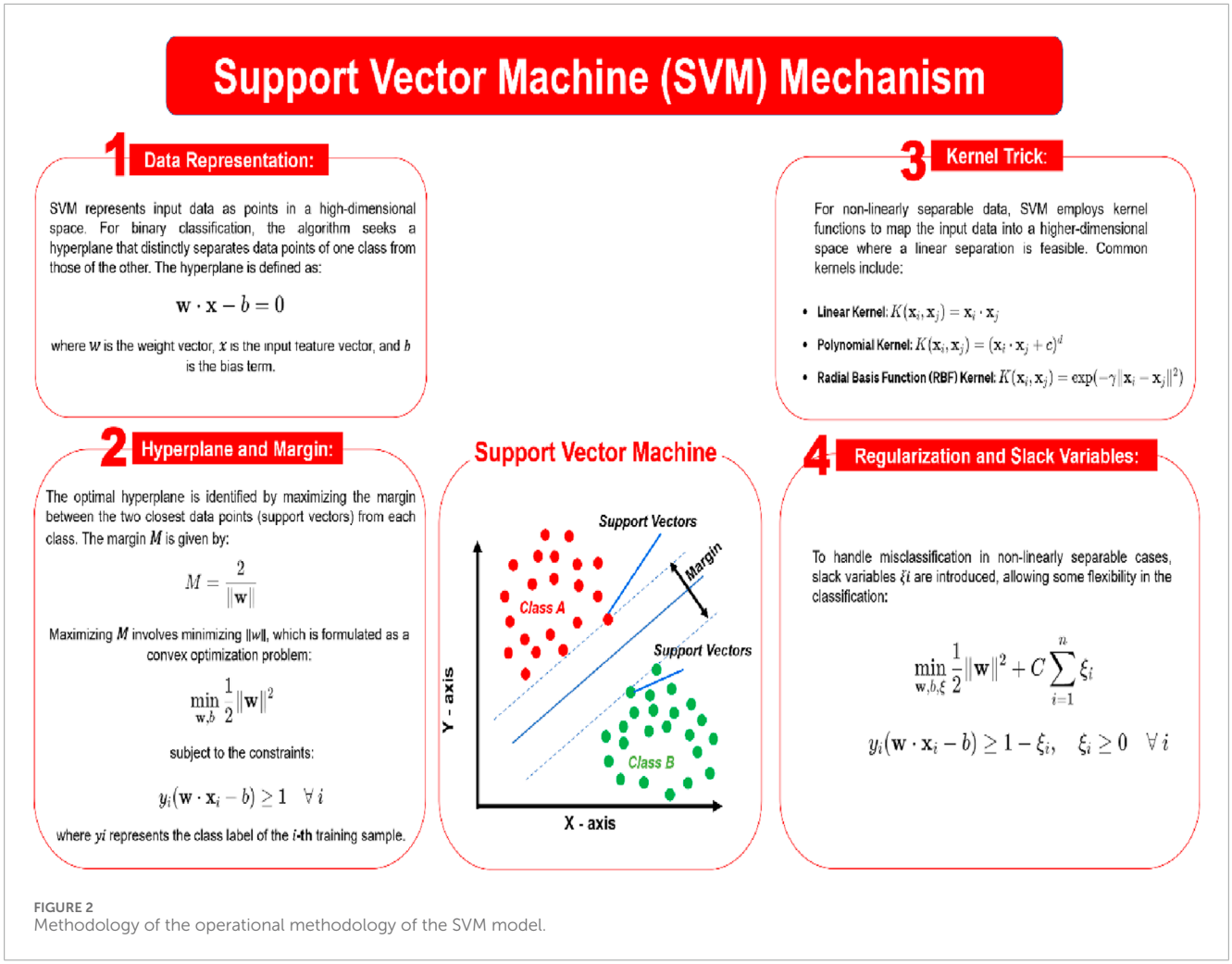

FIGURE 2
Methodology of the operational methodology of the SVM model.

datapoints were obtained from the National Institute of Standards and Technology's (NIST) Open-Source Data (OSD), which encompasses multiple datasets. Additionally, the dataset has been supplemented by the Materials in Extreme Dynamic Environments (MEDE) program of the U.S. Army Research Laboratory (ARL), which provides penetration data and concrete samples. Moreover, the dataset incorporates data from previously published articles. It is noteworthy that the dataset specifically accounts for a constant firing distance and firing angle, with all shots recorded at a safe distance of 10 m and an angle of 90°. This dataset encompasses a wide array of information, including various input features comprising of cement composition (Cp), Ground granulation blast-furnace slag (GGBFS), fly ash content (FA), water portion (Wp), super plasticizing agent (Sp), coarse aggregate content (CA), fine aggregated portion (FA), concrete sample age (t), concrete compressive strength (CCS), gun type (G-type), bullet caliper (B-Cali), bullet weight (Wb), and bullet velocity (Vb). The output variable of interest is the bullet penetration depth (BPD). Each feature contributes to a comprehensive understanding of the factors that influence bullet penetration in cement. The MLP model's ability to predict BPD relies on the interplay of these features, reflecting both the properties of the cement mass and the ballistic parameters of the bullet. By including a diverse set of features, the model can capture the complex physical interactions that occur during bullet impact and penetration, leading to more accurate and reliable predictions.

3.2 Machine learning models

The ML specially ANN models is a powerful tool for predicting key factors across various domains (De Lautour and Omenzetter, 2009; John et al., 2023). Four robust and leading models in their class, including the powerful regressor of Support Vector Machine (SVM), non-parametric technique of K-Nearest Neighbors (KNN), classifier-regressor of Light Gradient Boosting Machine (LightGBM) and the veteran successful artificial neural network of multilayer perceptron (MLP), were proposed for application on the dataset. Normalization is an important preprocessing step in machine learning that involves scaling input variables to a standard range, typically 0 to one or -1 to 1, to ensure that each feature contributes proportionately to the final prediction. In this study some comprehensive measures have taken to address the potential impact of outliers or noise on our feature selection technique. These steps are integral to our methodology, ensuring that proposed model's performance is robust and reliable. The normalization technique used was min-max scaling, which transforms the data

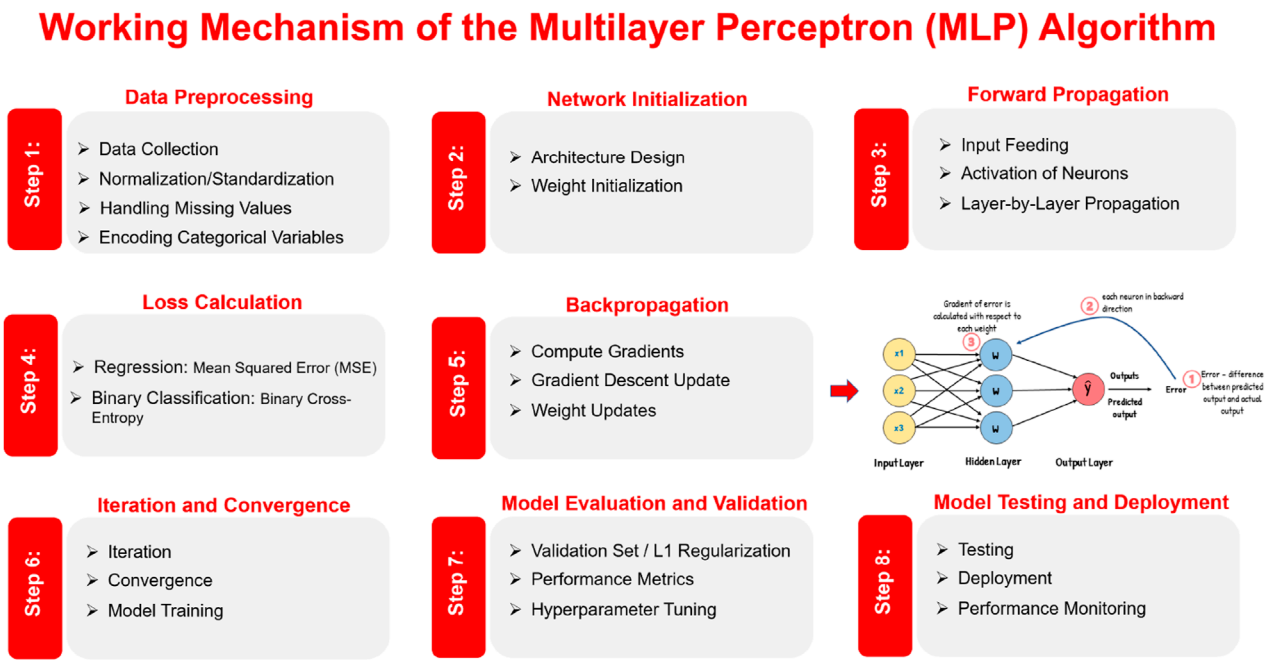

FIGURE 3
Methodology of the working mechanism of the MLP model.

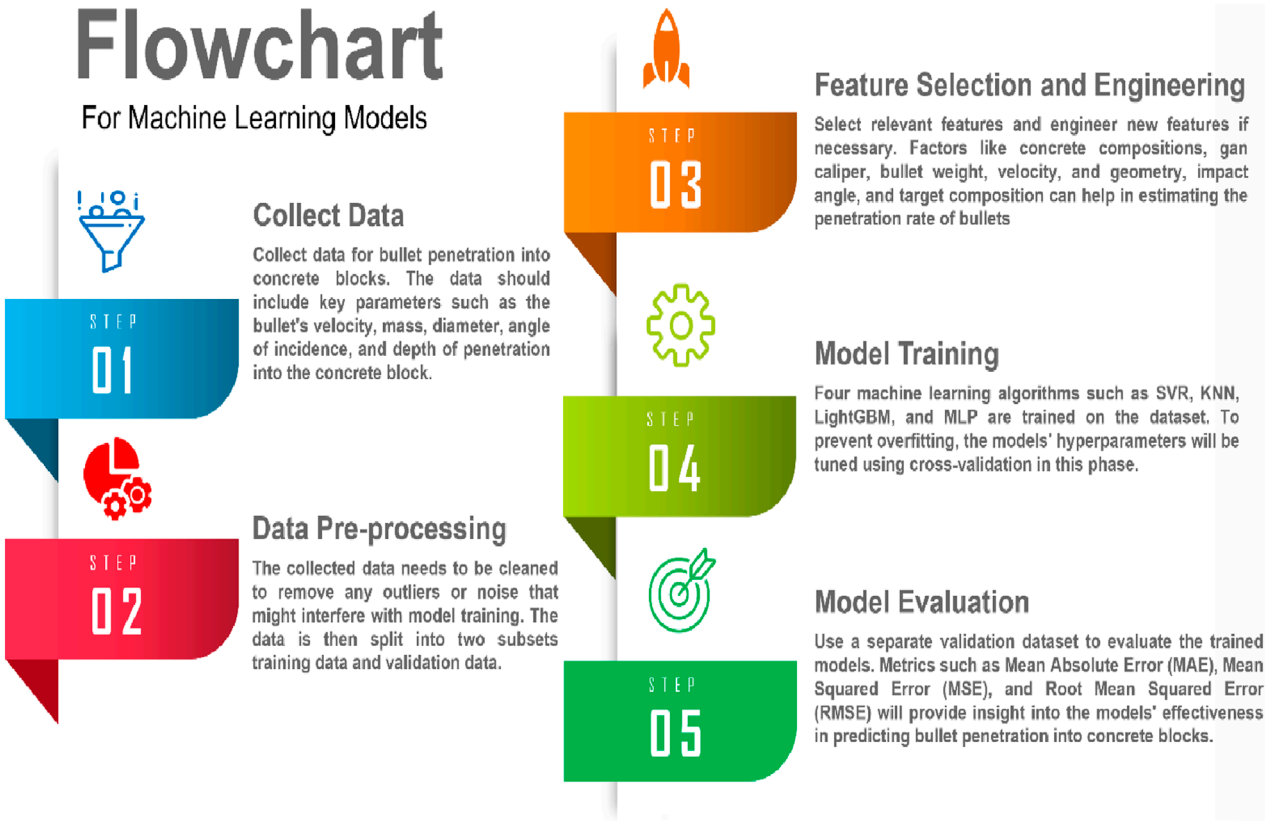

FIGURE 4
Flow chart for implementing four ML models.

TABLE 1 Statistical description of data related to BPD in cement block.

Variable	Cement powder	Ground granulation blast-furnace slag	Fly ash additive	Water portion	Super plasticizer additive	Coarse aggregate additive	Fine aggregate additive	Age of sample	Concrete compressive strength	Gun type	Bullet caliper	Bullet weight	Bullet velocity	Bullet penetration depth
Symbol Unit	Cp (kg/m^3)	GGBFS (kg/m^3)	FA (kg/m^3)	Wp (kg/m^3)	Sp (kg/m^3)	CA (kg/m^3)	FAA (kg/m^3)	t (day)	CCS (MPa)	G-Type (-)	B-Cali (mm)	Wb (grains)	Vb (m/s)	BPD (cm)
Ave.	281.62	73.36	54.14	181.50	6.17	973.67	773.52	45.84	35.82	AK-47	7.62	7.90	715.00	20.17
Std. Dev.	104.72	86.13	64.12	21.42	5.99	77.47	80.47	63.45	16.78	0.00	0.00	0.00	0.00	7.36
Var.	10,955.91	7,411.88	4,107.96	458.20	35.84	5,995.89	6,468.71	4,022.56	281.18	0.00	0.00	0.00	0.00	54.09
Min.	102.00	0.00	0.00	121.75	0.00	801.00	594.00	1.00	2.33	AK-47	7.62	7.90	715.00	5.01
Max.	540.00	359.40	200.10	247.00	32.20	1,145.00	992.60	365.00	82.60	AK-47	7.62	7.90	715.00	30.00
Skew.	0.50	0.81	0.54	0.08	0.92	−0.05	−0.25	3.25	0.42	N/A	N/A	N/A	N/A	−1.11
Kurt.	−0.53	−0.49	−1.33	0.11	1.42	−0.58	−0.12	12.01	−0.33	N/A	N/A	N/A	N/A	−0.11

10.3389/fmats.2024.1416918

TABLE 2 Allocation of unique code for variables in the feature selection step.

Variable	Unit	Code
C_p	(kg/m^3)	ξ1
GGBFS	(kg/m^3)	ξ2
F_A	(kg/m^3)	ξ3
W_p	(kg/m^3)	ξ4
S_p	(kg/m^3)	ξ5
C_A	(kg/m^3)	ξ6
F_{AA}	(kg/m^3)	ξ7
t	(day)	ξ8
CCS	(Mpa)	ξ9
G-Type	-	ξ10
B-Cali	(mm)	ξ11
W_b	(grains)	ξ12
V_b	(m/s)	ξ13

according to Eq. 1:

$$x' = max(x) - min(x)x - min(x) \quad (1)$$

where x represents the initial value, x′ denotes the normalized value, and min(x) and max(x) signify the minimum and maximum values of the feature, respectively.

SVM is a meticulous regression and classification model that operates by identifying the hyperplane that optimally separates the data points into distinct classes. In the case of a numerical dataset, SVM seeks to locate and adjust a hyperplane that maximizes the margin between the different classes. To achieve this, SVM initially transforms the numerical features into a higher-dimensional space using a kernel function, such as the radial basis function (RBF) or polynomial kernel. This enables SVM to identify a nonlinear decision boundary capable of separating the data points effectively. Subsequently, SVM endeavors to identify the hyperplane optimizes the separation margin between distinct classes. The margin refers to the distance between the hyperplane and the closest data points from each class. By maximizing this margin, SVM aims to establish a robust decision boundary that remains resilient to noise and can generalize well to new data instances. In scenarios where the data is not linearly separable, SVM permits a certain degree of misclassifications through the introduction of slack variables that penalize such misclassifications. The degree of penalty is controlled by a regularization parameter, which helps prevent overfitting issues (Pisner and Schnyer, 2020). Figure 2 represents the operational methodology of the SVM model.

The KNN model is a type of non-parametric learning method suitable for both regression and classification. Its methodology entails identifying the K nearest neighbors within the training dataset relative to a specific input data point and leveraging their labels or values to forecast the label or value of the input data point. For numerical datasets, K-Nearest Neighbors (KNN) employs a distance metric, like Euclidean or Manhattan distance, to gauge the resemblance between data points. The algorithm assesses the input data point against each instance in the training set, selecting the K closest neighbors based on their proximity. Notably, one of the key strengths of KNN lies in its simplicity and user-friendliness, as it does not require any assumptions about the data distribution (Abu Alfeilat et al., 2019). The most commonly used distance metric in KNN, calculated as (Eq. 2):

$$d_{(x,y)} = \sqrt[p]{\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)} \quad (2)$$

where, $d_{(x,y)}$ is distribution distance of sample data between test and train data; x_i is the test data sample; y_i is the train data sample.

LightGBM, is a gradient-boosting framework specifically designed to handle vast and complex datasets with numerous dimensions. Its operation involves creating multiple decision trees, with each subsequent tree aiming to correct the deficiencies of its predecessor. When dealing with numerical datasets, LightGBM uses a histogram-based model to bin the numerical features into discrete values. This approach reduces memory usage and accelerates the training process. LightGBM also utilizes a feature parallelism technique, dividing the dataset into smaller subsets to train multiple decision trees in parallel. During training, LightGBM optimizes a loss function, such as mean squared error or log loss, using a gradient-based approach to minimize the discrepancy between predicted and actual values. It uses a leaf-wise growth strategy, adding leaves to the tree based on the largest reduction in the loss function. One noteworthy feature of LightGBM is its ability to effectively handle imbalanced datasets through weighted sampling and gradient-based re-sampling techniques. Additionally, it supports early stopping to mitigate overfitting issues and enhance the generalization performance of the model (Al Daoud, 2019).

The MLP is a widely-used type of Artificial Neural Network (ANN) that consists of multiple layers of artificial neurons, drawing inspiration from the structure of the human brain. The MLP model operates by iteratively adjusting the weights (W) and biases (B) of the neurons in the network to minimize the associated error between the estimated final target and the actual final target. This iterative adjustment process is known as backpropagation. For the application of the MLP model on numerical datasets, the data must first undergo preprocessing to ensure it is in a suitable numerical format. Once the numerical dataset has been prepared, it is then divided into three distinct subsets: training, validation, and test sets.

During the training phase, the MLP model aims to optimize the weights and biases of the neurons in the network using the training set. This is achieved by repeatedly applying backpropagation models to minimize the error between the predicted output and the actual output for each example in the training set. Once the intricate training process is completed, the network is considered sufficiently trained for validation purposes. During the validation phase, the validated dataset is utilized to evaluate the performance of the network and ensure its adherence to the expected output specifications. Unlike the training phase that involved adjusting the weights and biases of the network's neurons,

TABLE 3 Feature selection performance ranking by MLP model.

Number of input variables	Input variables	RMSE (cm)
1	ξ5	0.95956
2	ξ2, ξ5	0.84695
3	ξ5, ξ2, ξ9	0.75423
4	ξ9, ξ5, ξ2, ξ7	0.63155
5	ξ2, ξ7, ξ9, ξ5, ξ6	0.52545
6	ξ9, ξ8, ξ5, ξ2, ξ6, ξ7	0.42166
7	ξ1, ξ7, ξ9, ξ8, ξ5, ξ7, ξ6	0.37214
8	ξ3, ξ6, ξ5, ξ2, ξ1, ξ8, ξ7, ξ9	0.35234
9	ξ1, ξ11, ξ3, ξ2, ξ7, ξ6, ξ9, ξ5, ξ8	0.38563
10	ξ8, ξ11, ξ1, ξ6, ξ3, ξ5, ξ12, ξ9, ξ2, ξ7	0.40633
11	ξ11, ξ2, ξ4, ξ9, ξ7, ξ8, ξ3, ξ12, ξ6, ξ1, ξ5	0.43525
12	ξ6, ξ5, ξ7, ξ4, ξ12, ξ3, ξ8, ξ9, ξ11, ξ1, ξ2, ξ13	0.45426
13	ξ10, ξ3, ξ1, ξ9, ξ11, ξ13, ξ8, ξ8, ξ2, ξ7, ξ12, ξ5, ξ6	0.47452

the validation dataset assesses the network's effectiveness, allowing for necessary adjustments to maintain optimal performance levels (Camacho Olmedo et al., 2018). Figure 3 illustrates the operational methodology of the Multilayer Perceptron (MLP) model.

3.3 Feature selection

Feature selection stands out as a essential technique in the ML data driven researches. Its principal objective is to discern a subset of relevant features that substantially influence the prediction of BPD. By adeptly diminishing the dimensionality of input data, feature selection elevates the efficacy of ML models. Its significance stems from its capability to pinpoint pertinent and informative features within the dataset, while concurrently eliminating extraneous and duplicative ones. This not only alleviates computational burdens but also guards against overfitting challenges (Saeys et al., 2007). In academic research, diverse models have emerged to streamline feature selection, comprising filter, wrapper, and embedded techniques. Filter methods employ statistical or correlation-based metrics to evaluate the relevance of individual features. Conversely, wrapper methods employ a designated learning model to assess subsets of features that yield optimal performance. Embedded methods, however, rely on the learning model to ascertain feature importance. Feature selection profoundly influences both the efficacy and efficiency of ML models. It enhances model accuracy and generalization capacity while concurrently diminishing the computational resources needed for analysis.

The flowchart technique shown in Figure 4 is used to predicted the BPD into a concrete block using ML models.

4 Results and discussion

4.1 Data analysis

In any ML-driven research frameworks, the detailed description of the dataset is of paramount importance and must not be overlooked. This section offers an exhaustive account of the data employed in the study, detailing its scope, size, and quality. A thorough understanding of the dataset is crucial to ensure the validity and reliability of the research findings. Without a detailed data description, replicating the study or verifying its results by other researchers becomes impractical. Moreover, it is essential to ensure that the dataset used in ML research is unbiased and representative of the target population.

The dataset used in this ML-driven research paper consists of a comprehensive collection of information regarding various variables pertinent to the study. Table 1 provides an extensive summary of the input and output variables critical for predicting the BPD in concrete blocks.

4.2 Feature selection and engineering

The process of feature engineering plays an important role in determining the optimal combination of input parameters required for hybrid models. Given the versatility of neural networks in addressing various regression, clustering, classification, and stochastic problems, the MLP model was chosen as the foundational model for implementing the feature selection technique. Through the analysis of various data record subsets for both training and testing phases, the MLP model proficiently diminishes the count

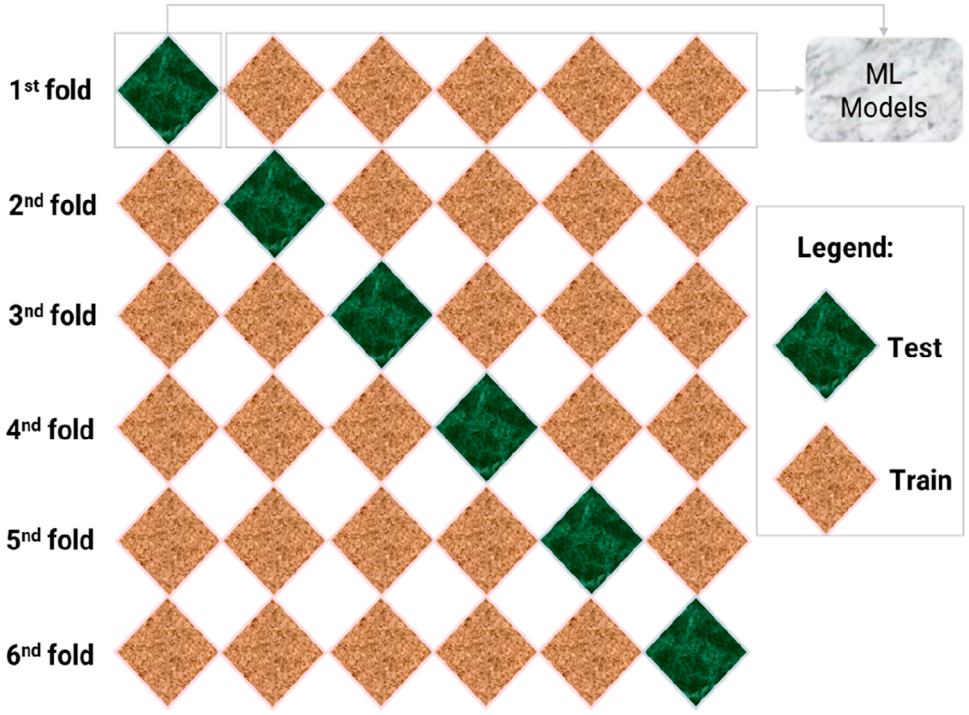

FIGURE 5
K-fold validation based on determination of select the test/train evaluated.

of independent variables. Iterative experimentation revealed that a two-layer MLP, configured with eight neurons in the initial hidden layer and six neurons in the subsequent hidden layer, optimally serves as the feature selection mechanism. This configuration is most effective in minimizing the Root Mean Square Error (RMSE) of the predicted numerical values for BPD. The determination of the specific architecture of the MLP model with two hidden layers and eight and six nodes was based on a thorough and methodical analysis of relevant literature, empirical experimentation, and optimization techniques to ensure the model's effectiveness and efficiency for the given task.

The primary objective of the feature selection technique is to identify the optimal configuration of the 13 input variables listed in Table 2. This entails identifying the most favorable set of features capable of yielding the lowest RMSE values for penetration depth. Failing to identify the ideal feature combination could result in inaccurate or erroneous outcomes, thereby undermining the credibility of the entire implementation process. Hence, selecting the most suitable features becomes imperative in order to maximize the precision and reliability of the analysis.

The findings of the feature selection analysis have been succinctly presented in Table 3 to facilitate comprehension. The table clearly indicates the attainment of an optimal pore pressure root mean square error (RMSE) value of 0.35234 cm, achieved through the utilization of a distinct combination of eight variables. The significance of this discovery cannot be overstated, as it underscores the critical role of feature selection in enhancing the accuracy and precision of the obtained results. It is important to emphasize that any model incorporating more than 13 features is expected to exhibit a higher RMSE value in comparison to the identified combination of eight variables (ξ3, ξ6, ξ5, ξ2, ξ1, ξ8, ξ7, and ξ9), which has been demonstrated as the most effective. Therefore, the identification of the appropriate feature combination becomes imperative in order to achieve optimal outcomes.

The Cp plays an important role in determining strength as it binds the cement grains together and influences the amount of BPD. Similarly, the Sp also has a significant impact on strength. The compressibility of cement, measured by the CCS, affects the BPD due to its influence on the amount of compression. Additionally, the industry by-product known as GGBFS is utilized as a strength-enhancing additive in cement. The CAA is an important parameter for determining compressive strength, representing the pressure applied to the cement for strength evaluation. The t is a fundamental parameter for cement and its durability since cement gradually loses its effectiveness over time, thereby affecting BPD. The FA significantly affects the rheological, mechanical, and strain properties of cement. Furthermore, the FAAA greatly impact cement mortar and consequently influence BPD. On the other hand, the G-Type, B-Cali, Wb, and Vb remain constant and are not highly influential in feature selection for predicting BPD.

Following the determination of the optimal architecture for the ML, several techniques were investigated to create distinct, mutually exclusive training and test data subsets for model evaluation. Initially, a method was employed where

10.3389/fmats.2024.1416918

TABLE 4 Determination of statistical parameter related to BPD prediction using artificial intelligence models for train, test and validation data.

Dataset	Models Units	APD (%)	AAPD (%)	SD (cm)	RMSE (cm)	R^2 -
Train data	KNN	−0.055	0.249	0.176	0.1756	0.9981
	SVM	0.033	0.405	0.361	0.3610	0.8986
	MLP	0.088	0.362	0.385	0.3850	0.8898
	LightGBM	0.035	0.430	0.405	0.4051	0.8758
Test data	KNN	−0.025	0.324	0.181	0.1811	0.9905
	SVM	−0.090	0.531	0.367	0.3677	0.8969
	MLP	−0.012	0.454	0.363	0.3634	0.8883
	LightGBM	0.050	0.476	0.406	0.4063	0.8743
Validation data	KNN	0.046	0.326	0.185	0.1863	0.9914
	SVM	0.034	0.540	0.350	0.3503	0.8973
	MLP	0.068	0.540	0.340	0.3408	0.8890
	LightGBM	−0.117	0.459	0.445	0.4475	0.8815

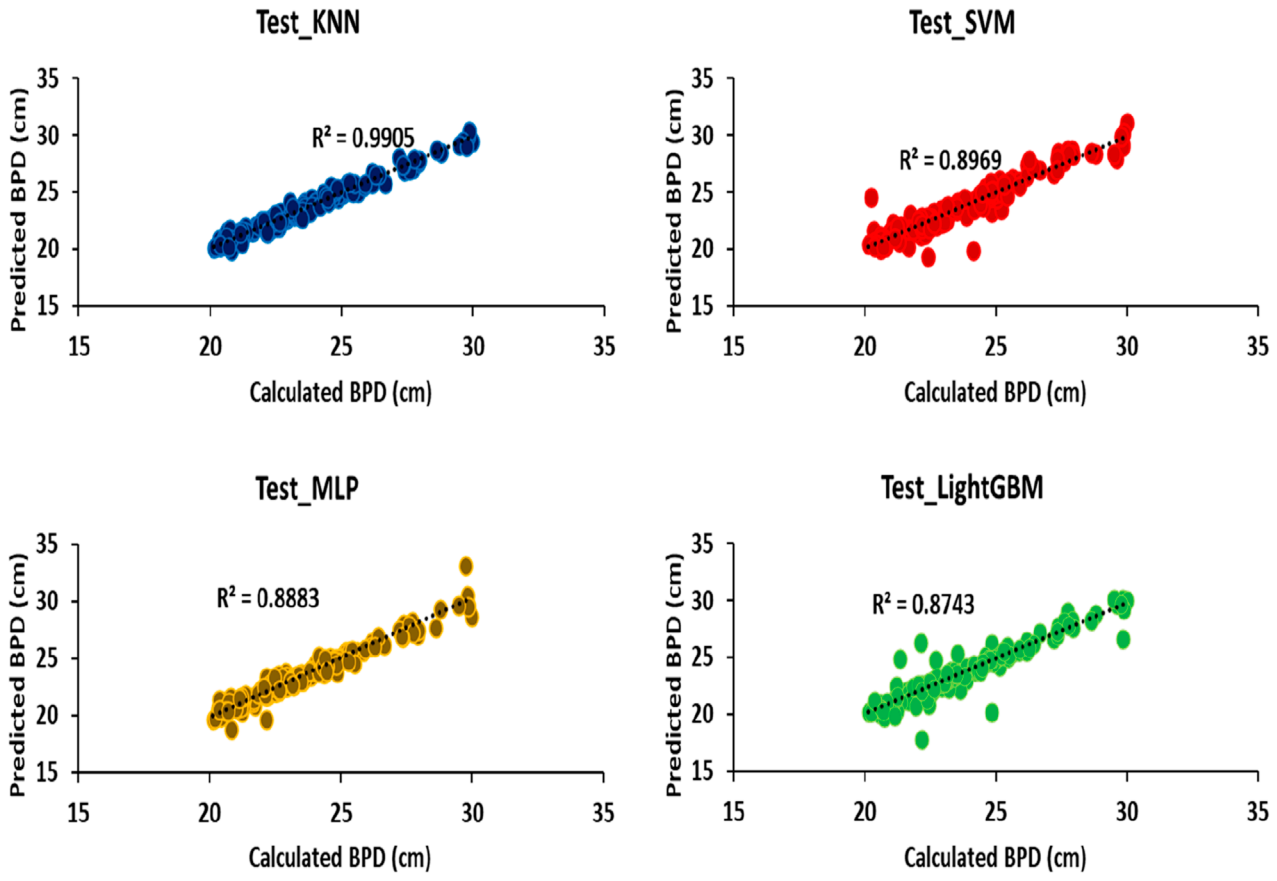

FIGURE 6
Illustrates a cross-plot to prediction of BPD based on the testing dataset utilizing four ML models such as SVM, KNN, LightGBM, and MLP.

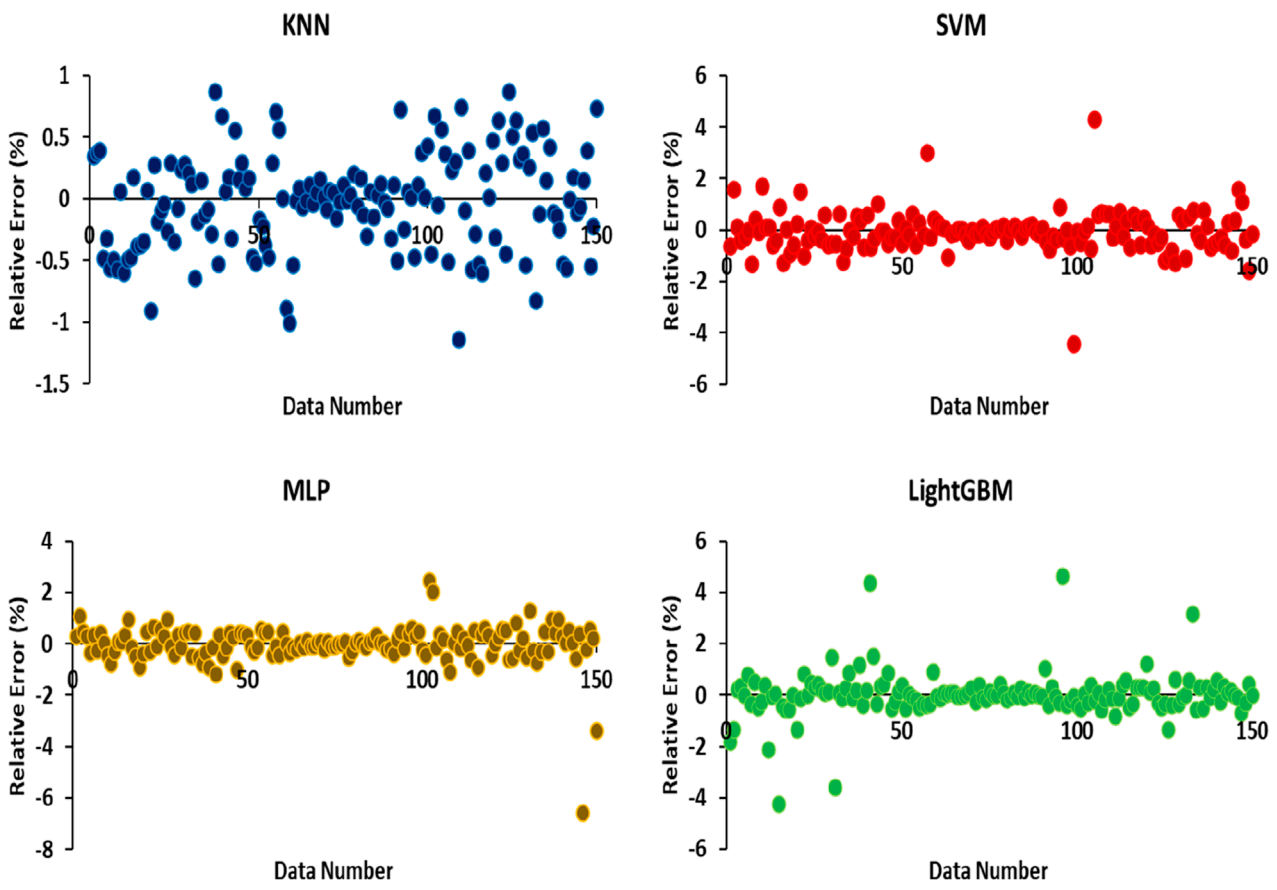

FIGURE 7
Illustrates of the relative error for prediction of BPD based on the test dataset utilizing four ML models such as SVM, KNN, LightGBM, and MLP.

30% of the available data records were randomly selected for the test subset, with the remaining 70% allocated to the training subset. However, this approach fell short in preventing overfitting during feature selection, as certain features disproportionately influenced the predictions. To address this issue, a more effective approach was adopted using the 6-fold cross-validation technique, which effectively mitigated the problem of overfitting. The rationale behind choosing 6-fold cross-validation for evaluating feature selection performance was to balance computational efficiency, statistical robustness, and the trade-off between bias and variance in performance estimation.

The 6-fold cross-validation approach entails partitioning the complete dataset into six mutually exclusive segments. One segment is used as the evaluation subset, while the remaining five serve as the training subset. Each configuration involves assessing the machine learning model 60 times, corresponding to 10 repetitions for each training and test subset combination. The model with the lowest Root Mean Square Error (RMSE) for predicting BPD, as compared to the observed BPD, was selected for each of the ten training/test subset combinations. The mean of the ten optimal RMSE values across six iterations was considered indicative of the predictive accuracy and overall performance of the feature selection process. This 6-fold validation procedure is illustrated in Figure 5.

4.3 Model performance

This research aims to establish and focus on four main ML based strategies: SVM, KNN, LightGBM, and MLP. These models were trained on an extensive dataset comprising experimental data encompassing various bullet types and concrete block configurations. The findings of our study demonstrate that data-driven ML models can be remarkably effective in predicting BPD in concrete blocks, with potential applications in diverse real-world scenarios.

During the evaluation and comparison of these models use various metrics, including Eqs 3–7. The Average Percentage Difference (APD), Absolute Average Percentage Difference (AAPD), Standard Deviation (SD), Root Mean Square Error (RMSE), and the Correlation Coefficient (R^2) are metrics used to evaluate model performance. Specifically, the Correlation Coefficient (R^2) quantifies the proportion of variance in the dependent variable that is attributable to the independent variable.

$$\mathrm{APD} = \frac{\sum_{i=1}^{n}\left(\frac{\mathrm{BPD_{Cal.}}-\mathrm{BPD_{Pred.}}}{\mathrm{BPD_{Cal.}}}\mathrm{x}\,100\right)_i}{n} \tag{3}$$

$$\mathrm{AAPD} = \frac{\sum_{i=1}^{n}\left|\left(\frac{\mathrm{BPD_{Cal.}}-\mathrm{BPD_{Pred.}}}{\mathrm{BPD_{Cal.}}}\mathrm{x}\,100\right)_i\right|}{n} \tag{4}$$

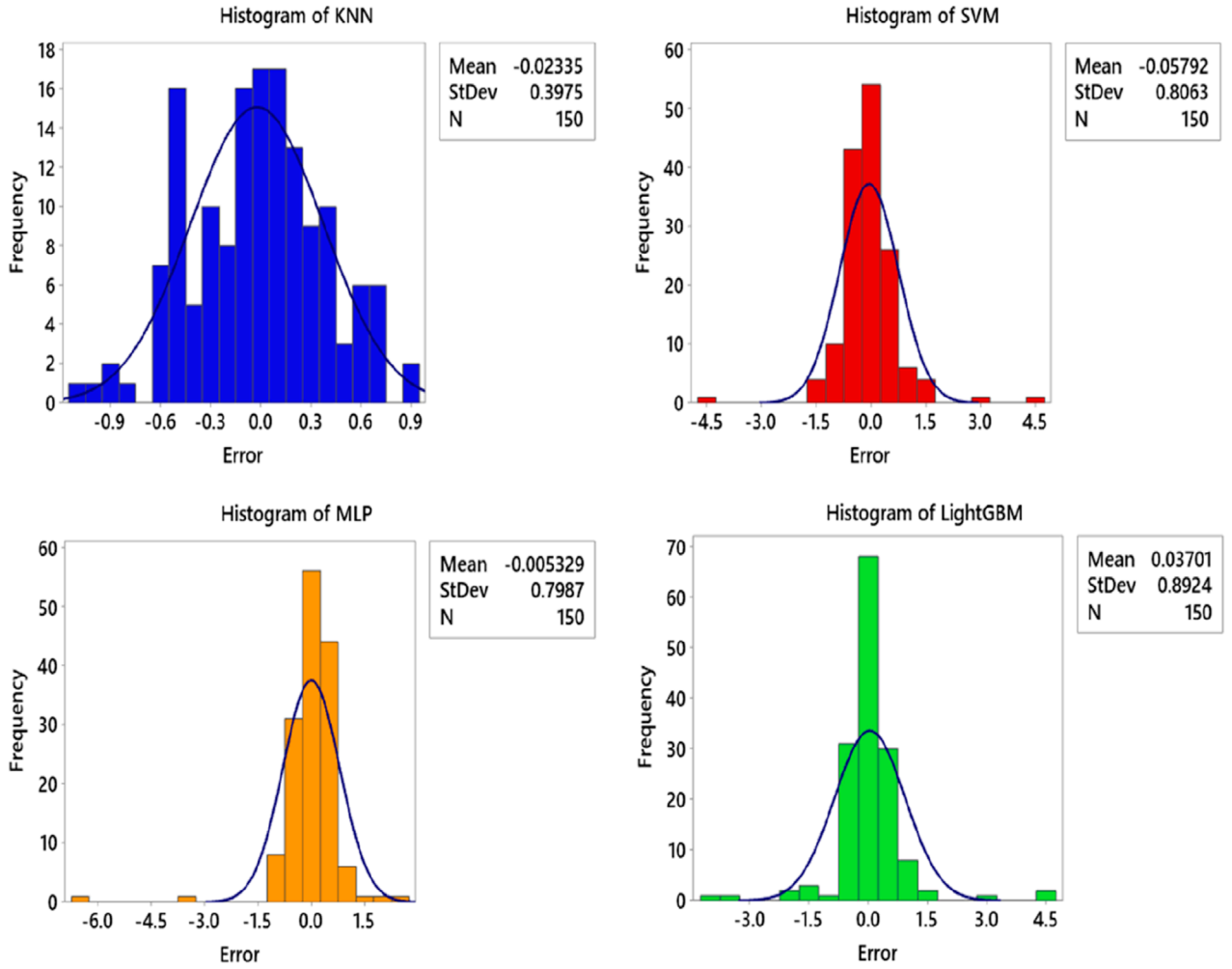

FIGURE 8
Error histogram plot prediction of BPD using four ML models (SVM, KNN, LightGBM, and MLP).

$$SD = \sqrt{\frac{\sum_{i=1}^{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(BPD_{Cal.i} - BPD_{Pred.i}\right)\right)_i - \left(\frac{1}{n}\sum_{i=1}^{n}\left(BPD_{Cal.i} - BPD_{Pred.i}\right)\right)imean\right)^2}{n-1}} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(BPD_{Cal.i} - BPD_{Pred.i}\right)^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(BPD_{Pred.i} - BPD_{Cal.i}\right)^2}{\sum_{i=1}^{N}\left(BPD_{Pred.i} - \frac{\sum_{I=1}^{n} BPD_{Cal.i}}{n}\right)^2} \quad (7)$$

Four robust models, namely, SVM, KNN, LightGBM, and MLP, were used to predict the BPD. Each model underwent separate execution to train and validate the models, followed by independent testing. To determine the accuracy, the dataset was divided into three subsets. Specifically, 70% of the data records were assigned to the training subset, 15% were allocated for independent testing, and the remaining 15% were set aside for model validation.

This article has comprehensively evaluated the performance of multiple models, including SVM, KNN, LightGBM, and MLP, for the prediction of BPD. The outcomes of these models have been meticulously presented in Table 4, incorporating the data values from the training, testing, and validation subsets.

The findings of the validation data set are presented in Table 4, revealing the exceptional performance of the KNN model with respect to the root mean square error (RMSE), average percentage difference (APD), absolute average percentage difference (AAPD), and standard deviation (SD), which stand at 0.1811 cm−0.025, 0.324, and 0.181, respectively. A comparison of the KNN model's precision with the other three models on the validation subset underscores its superiority in accurately predicting BPD. Specifically, the SD value of 0.181 achieved by the KNN model surpasses the respective SD values of 0.406, 0.367, and 0.363 obtained by the evaluated SVM, MLP, and LightGBM models.

The R-square parameter serves as a important statistical measure for evaluating and comparing various parameters. It evaluates the model's fit to the data by measuring the proportion of variance in the dependent variable that is explained by the independent

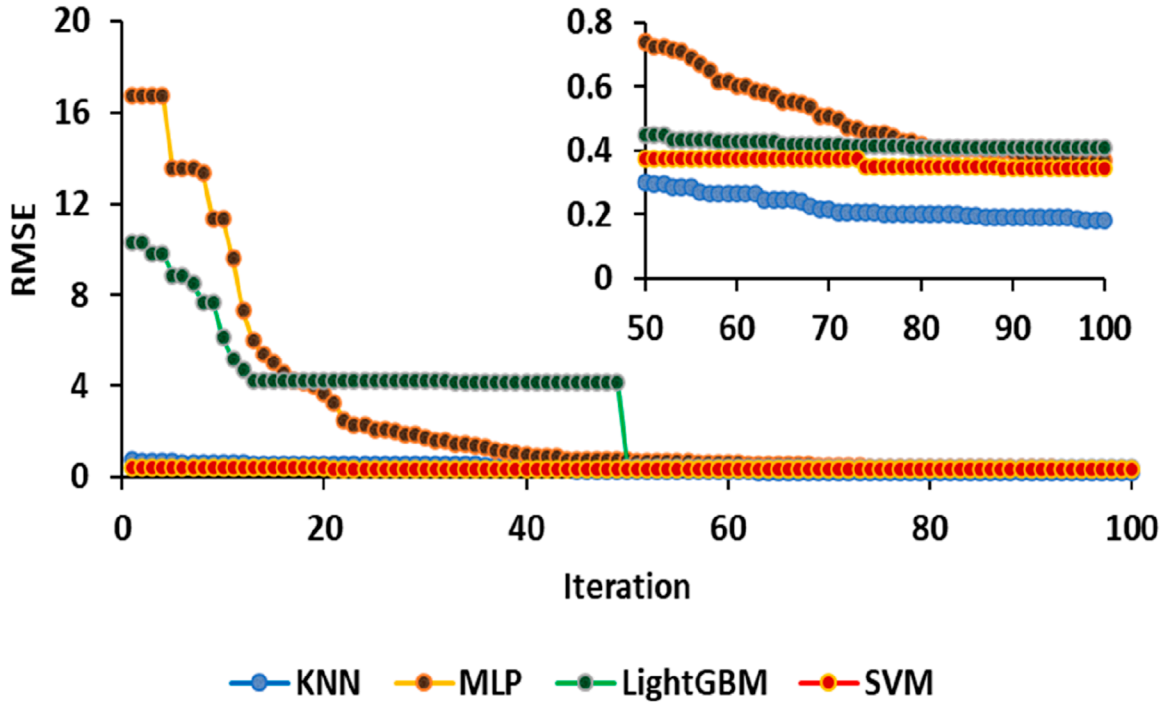

FIGURE 9
Determination of error vs. iteration using four ML models (SVM, KNN, LightGBM, and MLP).

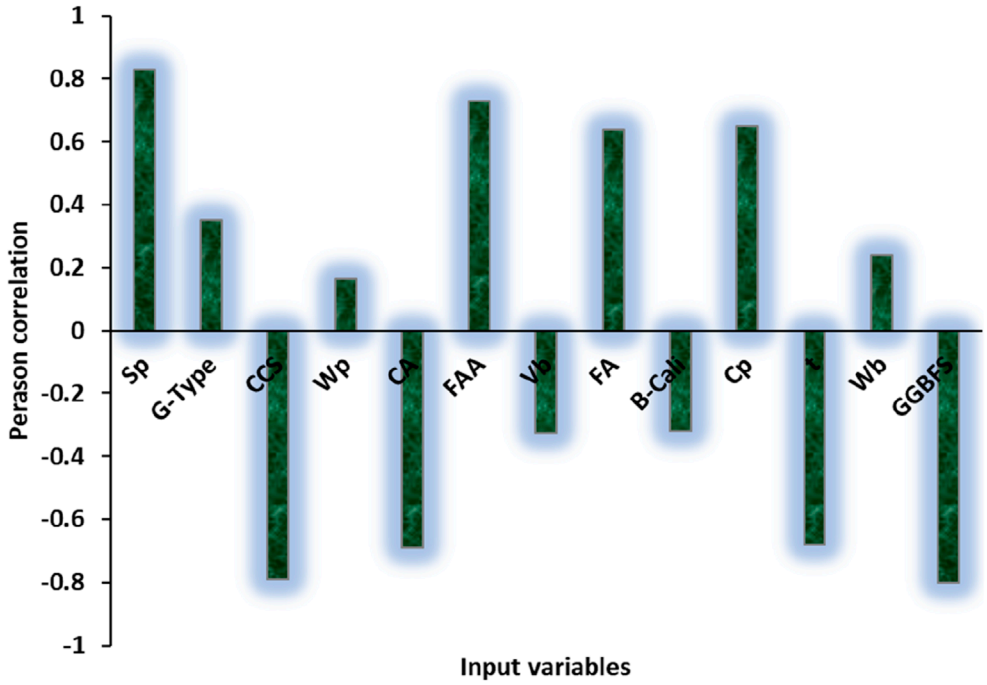

FIGURE 10
Illustration for the Pearson correlation versus input variables for prediction of BPD based on ML models.

variables. Figure 6 presents cross plots illustrating the predicted BPD values for the training, testing, and validation datasets, demonstrating significantly higher prediction accuracy compared to other evaluated models.

To evaluate the accuracy of the KNN model, the results shown in Table 4 and Figure 6 for the training, testing, and validation subsets were analyzed. The analysis indicates that the KNN model achieves exceptionally low errors for BPD, with RMSE values of 0.4389,

0.1811, and 0.1863 cm, and R^2 values of 0.9965, 0.9970, and 0.9969, respectively. Comparing the BPD prediction performance of the four ML models, the KNN model emerged as the best performer, followed by the SVM and MLP models, while the LightGBM model demonstrated the least favorable prediction performance. These findings provide valuable insights into the effectiveness of different ML models for BPD prediction and underscore the importance of parameter optimization in achieving accurate and reliable predictions across various applications.

Figure 7 illustrates the relative error for predicting of BPD based on the test dataset using four ML models: SVM, KNN, LightGBM, and MLP. This figure effectively demonstrates the performance accuracy of the models for RE. The graphical representation in this figure allows for the conclusion that the models' performance accuracy and the computational errors for each test dataset have been effectively determined. According to this figure, it is evident that the computational error for KNN is within the range of $-1 < RE < 1$, while for the other models, it is within the range of $-8 < RE < 4$. This clearly indicates the high accuracy of the KNN model for accurate and precise prediction of BPD.

Figure 8 presents an error histogram plot that showcases the prediction errors for BPD using four robust ML models: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), LightGBM, and MLP. Each of the five histograms represents the prediction error BPD for each model and demonstrates a normal distribution of predict errors centered around zero, with a relatively narrow spread and no discernible positive or negative bias. This plot enables a comprehensive analysis of the models' performance, facilitating the identification of the model that exhibits the best performance with a normal error distribution. Upon thorough examination, it becomes apparent that the KNN model's normal distribution of data surpasses that of the other models, indicating a superior and more accurate standard deviation. Considering the comparison of these models based on the information presented in Table 4 and Figure 8, it can be concluded that the performance accuracy of the models can be ranked as follows: KNN > SVM > MLP > LightGBM.

Figure 9 illustrates the error rates for the SVM, KNN, LightGBM, and MLP models as a function of iteration. The findings of this study indicate that the MLP and LightGBM models initially exhibit higher error values, which gradually decrease over time. However, this trend is not observed in the KNN and SVM models.

Upon careful analysis of the figure, it becomes apparent that the MLP model achieves higher accuracy than the other models at the beginning of the iteration. At the threshold of 17, it surpasses the LightGBM model with a lower error value. In the subsequent phase, at iteration 50, the LightGBM model surpasses the MLP model, demonstrating superior performance accuracy. Conversely, these variations are not observed in the KNN and SVM models, where the performance accuracy consistently decreases from the start to the end of the iteration.

Examining the zoomed-in portion of the figure (iteration 60–100) provides a clearer depiction of the ongoing performance patterns of these models. As established, the KNN model exhibits the highest performance accuracy among the models, with SVM, MLP, and LightGBM following in descending order of accuracy.

Pearson's coefficient (R) is a robust method for determining the relative relationship between input and output variables. It expresses the correlation coefficient on a scale from −1 to +1, where +1 signifies the strongest positive correlation, −1 signifies the strongest negative correlation, and 0 indicates no correlation. The calculation of Pearson's correlation coefficient is given by Eq. 8.

$$R = \frac{\sum_{i=1}^{n}\left(\partial_i - \overline{\partial}\right)\left(\propto_i - \overline{\propto}\right)}{\sqrt{\sum_{i=1}^{n}\left(\partial_i - \overline{\partial}\right)^2 \sum_{i=1}^{n}\left(\propto_i - \overline{\propto}\right)^2}} \tag{8}$$

Figure 10 compares Pearson and Spearman correlation coefficients, offering valuable insights into the relationships between the input variables and BPD. Notably, a negative correlation was observed with CCS, CA, Vb, B-Cali, t, and GGBFS, while a positive correlation was found with Sp, G-Type, Wp, FAA, FA, Cp, and Wb. By using both Pearson and Spearman correlation methods, it is possible to develop models based on the predicted outcomes for BPD. Analysis of this figure indicates that certain parameters, including Sp, GGBFS, CCS, FAA, CA, t, Cp, and FA, have a significant effect on BPD.

5 Conclusion

The evaluation of concrete structures' ability to withstand projectile penetration holds great significance, particularly from a general defense standpoint. The objective of this investigation is to predict the penetration depth (BPD) utilizing robust models, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), and Multilayer Perceptron (MLP) as ML models. The dataset used consists of 1,020 datapoints sourced from the National Institute of Standards and Technology (NIST), encompassing various parameters such as cement content (Cp), ground granulation blast-furnace slag (GGBFS), fly ash content (FA), water portion (Wp), superplasticizer content (Sp), coarse aggregate content (CA), fine aggregate content (FAA), concrete sample age (t), concrete compressive strength (CCS), gun type (G-type), bullet caliper (B-Cali), bullet weight (Wb), and bullet velocity (Vb).

The BPD serves as the output variable, and 4 ML models were developed using the dataset: 70% of the data was allocated for training the models, while 15% was set aside for testing and validation each. Feature selection techniques revealed that the MLP model, incorporating eight input variables (FA, CA, Sp, GGBFS, Cp, t, FAA, and CCS), provides the most accurate predictions for BPD across the entire dataset. Hyperparameter tuning using cross-validation will be used in this phase to mitigate the risk of overfitting. The results indicate that the average of the ten best Root Mean Square Error (RMSE) values obtained from six iterations serves as a representative measure of the prediction accuracy and performance of the feature selection process.

Comparing the four models used in this study, KNN demonstrates a distinct superiority over the other methods. KNN, a non-parametric ML model used for classification and regression, possesses several advantages, including simplicity, non-parametric nature, no training requirements, robustness to noisy data, suitability for large datasets, and interpretability. The results reveal that KNN outperforms the other models presented

in this paper, exhibiting an R^2 value of 0.9905 and an RMSE value of 0.1811 cm, signifying higher accuracy in its predictions compared to the other models. Finally, based on the error analysis across iterations, it is evident that the final accuracy error of the KNN model surpasses that of SVM, MLP, and LightGBM models, respectively.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The corresponding authors will provide data upon reasonable requests for academic purposes. Requests to access these datasets should be directed to HG, hamzehghorbani68@yahoo.com.

Author contributions

QZ: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Visualization, Writing–original draft, Writing–review and editing. YJ: Conceptualization, Data curation, Software, Supervision, Writing–original draft, Writing–review and editing. GW: Conceptualization, Methodology, Resources, Visualization, Writing–original draft, Writing–review and editing. QS: Conceptualization, Formal Analysis, Resources, Software, Supervision, Validation, Writing–original draft, Writing–review and editing. HG: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abbas, H., Al-Dabaan, M., Siddiqui, N., Almusallam, T., and Al-Salloum, Y. (2023). Performance of reinforced concrete composite wall systems under projectile impact. *J. Mater. Res. Technol.* 23, 3062–3090. doi:10.1016/j.jmrt.2023.01.187

Abu Alfeilat, H. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., et al. (2019). Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big data* 7, 221–248. doi:10.1089/big.2018.0175

Al Daoud, E. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *Int. J. Comput. Inf. Eng.* 13, 6–10. doi:10.5281/zenodo.3607805

Amaguaña, M., Guamán, L., Gómez, N. B. Y., Khorami, M., Calvo, M., and Albuja-Sánchez, J. (2023). Test method for studying the shrinkage effect under controlled environmental conditions for concrete reinforced with coconut fibres. *Materials* 16, 3247. doi:10.3390/ma16083247

Beheshtian, S., Rajabi, M., Davoodi, S., Wood, D. A., Ghorbani, H., Mohamadian, N., et al. (2022). Robust computational approach to determine the safe mud weight window using well-log data from a large gas reservoir. *Mar. Petroleum Geol.* 142, 105772. doi:10.1016/j.marpetgeo.2022.105772

Bentz, D. P. (2007). Cement hydration: building bridges and dams at the microstructure level. *Mater. Struct.* 40, 397–404. doi:10.1617/s11527-006-9147-3

Brandt, A. M. (2008). Fibre reinforced cement-based (FRC) composites after over 40 years of development in building and civil engineering. *Compos. Struct.* 86, 3–9. doi:10.1016/j.compstruct.2008.03.006

Brunton, S. L., and Kutz, J. N. (2022). Data-driven science and engineering: machine learning, dynamical systems, and control. Cambridge, United Kingdom: Cambridge University Press.

Camacho Olmedo, M. T., Paegelow, M., Mas, J.-F., and Escobar, F. (2018). *Geomatic approaches for modeling land change scenarios. An introduction*. Springer.

Campbell, O., Blenkinsop, T., Gilbert, O., and Mol, L. (2022). Bullet impacts in building stone excavate approximately conical craters, with dimensions that are controlled by target material. *Sci. Rep.* 12, 17634. doi:10.1038/s41598-022-22624-z

Cao, Y. Y. Y., Yu, Q., Tang, W. H., and Brouwers, H. J. H. (2020). Numerical investigation on ballistic performance of coarse-aggregated layered UHPFRC. *Constr. Build. Mater.* 250, 118867. doi:10.1016/j.conbuildmat.2020.118867

Cheng, D., Pu, W., and Huang, L. (2023). "Study on dynamic characteristics of high speed projectile impacting concrete target," in Mechatronics and Automation Technology: Proceedings of ICMAT 2022, 349.

Chintalapudi, K., and Pannem, R. M. R. (2022). Enhanced chemical resistance to sulphuric acid attack by reinforcing Graphene Oxide in Ordinary and Portland Pozzolana cement mortars. *Case Stud. Constr. Mater.* 17, e01452. doi:10.1016/j.cscm.2022.e01452

Dancygier, A. N., Yankelevsky, D. Z., and Jaegermann, C. (2007). Response of high performance concrete plates to impact of non-deforming projectiles. *Int. J. Impact Eng.* 34, 1768–1779. doi:10.1016/j.ijimpeng.2006.09.094

Day, K. W. (2006). *Concrete mix design, quality control and specification*. CRC Press.

De Lautour, O. R., and Omenzetter, P. (2009). Prediction of seismic-induced structural damage using artificial neural networks. *Eng. Struct.* 31, 600–606. doi:10.1016/j.engstruct.2008.11.010

Derakhshani, A., Ghadi, A., and Vahdat, S. E. (2023). Study of the effect of calcium nitrate, calcium formate, triethanolamine, and triisopropanolamine on compressive strength of Portland-pozzolana cement. *Case Stud. Constr. Mater.* 18, e01799. doi:10.1016/j.cscm.2022.e01799

Fahimizadeh, M., Pasbakhsh, P., Mae, L. S., Tan, J. B. L., and Raman, R. K. S. (2022). Multifunctional, sustainable, and biological non-ureolytic self-healing systems for cement-based materials. *Engineering* 13, 217–237. doi:10.1016/j.eng.2021.11.016

Haach, V. G., Vasconcelos, G., and Lourenço, P. B. (2011). Influence of aggregates grading and water/cement ratio in workability and hardened properties of mortars. *Constr. Build. Mater.* 25, 2980–2987. doi:10.1016/j.conbuildmat.2010.11.011

Habibi Rad, M., Mojtahedi, M., Ostwald, M. J., and Wilkinson, S. (2022). A conceptual framework for implementing lean construction in infrastructure recovery projects. *Buildings* 12, 272. doi:10.3390/buildings12030272

Jiao, H., Wang, Y., Li, L., Arif, K., Farooq, F., and Alaskar, A. (2023). A novel approach in forecasting compressive strength of concrete with carbon nanotubes as nanomaterials. *Mater. Today Commun.* 35, 106335. doi:10.1016/j.mtcomm.2023.106335

John, S. K., Nadir, Y., Cascardi, A., Arif, M. M., and Girija, K. (2023). Effect of addition of nanoclay and SBR latex on fly ash-slag geopolymer mortar. *J. Build. Eng.* 66, 105875. doi:10.1016/j.jobe.2023.105875

Kumar, R., Rai, B., and Samui, P. (2023). A comparative study of prediction of compressive strength of ultra-high performance concrete using soft computing technique. *Struct. Concr.* 24, 5538–5555. doi:10.1002/suco.202200850

Lai, J., Yang, H., Wang, H., Zheng, X., and Wang, Q. (2018). Properties and modeling of ultra-high-performance concrete subjected to multiple bullet impacts. *J. Mater. Civ. Eng.* 30, 04018256. doi:10.1061/(asce)mt.1943-5533.0002462

Li, P. P., Brouwers, H. J. H., and Yu, Q. (2020). Influence of key design parameters of ultra-high performance fibre reinforced concrete on in-service bullet resistance. *Int. J. Impact Eng.* 136, 103434. doi:10.1016/j.ijimpeng.2019.103434

Li, W., Lin, X., Bao, D. W., and Xie, Y. M. (2022). *A review of formwork systems for modern concrete construction*. Amsterdam, Netherlands: Elsevier, 52–63.

Liu, J., Li, J., Fang, J., Su, Y., and Wu, C. (2022a). Ultra-high performance concrete targets against high velocity projectile impact–a-state-of-the-art review. *Int. J. Impact Eng.* 160, 104080. doi:10.1016/j.ijimpeng.2021.104080

Liu, Y., Wang, J., Hu, S., Cao, S., and Wang, F. (2022b). Enhancing the mechanical behaviour of concretes through polymer modification of the aggregate-cement paste interface. *J. Build. Eng.* 54, 104605. doi:10.1016/j.jobe.2022.104605

Mina, A. L., Petrou, M. F., and Trezos, K. G. (2021). Resistance of an optimized ultra-high performance fiber reinforced concrete to projectile impact. *Buildings* 11, 63. doi:10.3390/buildings11020063

Mohamadian, N., Ghorbani, H., Wood, D. A., Mehrad, M., Davoodi, S., Rashidi, S., et al. (2021). A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. *J. Petroleum Sci. Eng.* 196, 107811. doi:10.1016/j.petrol.2020.107811

Neville, A. M., and Brooks, J. J. (1987). *Concrete technology*. Longman Scientific and Technical England.

Park, G.-K., Lee, M., Lee, N., Park, G.-J., and Kim, S.-W. (2023). Dynamic structural responses of high-performance fiber-reinforced cement composites panels subjected to high-velocity projectile impact loadings. *Compos. Struct.* 306, 116581. doi:10.1016/j.compstruct.2022.116581

Pawar, Y., and Kate, S. (2020). Curing of concrete: a review. *Int. Res. J. Eng. Technol.* 7, 1820–1824.

Pisner, D. A., and Schnyer, D. M. (2020). "Support vector machine," in *Machine learning* (Elsevier), 101–121.

Rabi, M., Shamass, R., and Cashell, K. A. (2022). Structural performance of stainless steel reinforced concrete members: a review. *Constr. Build. Mater.* 325, 126673. doi:10.1016/j.conbuildmat.2022.126673

Rich, K. M., Rizzuto, N. M., and Zieger, S. (2022). *The aesthetic life of infrastructure: race, affect, environment*. Northwestern University Press.

Saeys, Y., Inza, I., and Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics* 23, 2507–2517. doi:10.1093/bioinformatics/btm344

Srivani, G., and Mohan, U. V. (2023). "Study on strength properties of concrete by partial replacement of cement with sugarcane bagasse ash and coarse aggregate with coconut shells," in Materials Today: Proceedings.

Surahyo, A., Surahyo and Luby (2019). *Concrete construction*. Springer.

Uddin, S., Ong, S., and Lu, H. (2022). Machine learning in project analytics: a data-driven framework and case study. *Sci. Rep.* 12, 15252. doi:10.1038/s41598-022-19728-x

Vu, X. H., Malecot, Y., Daudeville, L., and Buzaud, E. (2009). Effect of the water/cement ratio on concrete behavior under extreme loading. *Int. J. Numer. Anal. Methods Geomechanics* 33, 1867–1888. doi:10.1002/nag.796

Wang, S., Le, H. T. N., Poh, L. H., Feng, H., and Zhang, M.-H. (2016). Resistance of high-performance fiber-reinforced cement composites against high-velocity projectile impact. *Int. J. Impact Eng.* 95, 89–104. doi:10.1016/j.ijimpeng.2016.04.013

Wang, W., Song, X., Yang, J., Liu, F., and Gao, W. (2023). Experimental and numerical research on the effect of ogive-nose projectile penetrating UR50 ultra-early-strength concrete. *Cem. Concr. Compos.* 136, 104902. doi:10.1016/j.cemconcomp.2022.104902

Wu, Y., Tao, X., Liu, Y., Zhang, Q., and Xue, Y. (2022). Analysis on deflection of projectile penetrating into composite concrete targets. *Materials* 15, 7871. doi:10.3390/ma15227871

Zhang, F., Poh, L. H., and Zhang, M.-H. (2020). Critical parameters for the penetration depth in cement-based materials subjected to small caliber non-deformable projectile impact. *Int. J. Impact Eng.* 137, 103471. doi:10.1016/j.ijimpeng.2019.103471

Glossary

AAPD	Absolute average percentage difference
APD	Average percentage difference
ARL	Army research laboratory
B-Cali	Bullet caliper
BPD	Bullet penetration depth
CA	Coarse aggregate content
CCS	Concrete compressive strength
Cp	Cement content
$\mathbf{d_{(x,y)}}$	Distribution distance of sample data between test and train data
DTIC	Defense technical information center
FA	Fly ash content
FAA	Fine aggregate content
GGBFS	Ground granulation blast-furnace slag
G-type	Gun type
KNN	K- means, k-nearest neighbors
LightGBM	Light gradient boosting machine
max(x)	Maximum values of the feature
MEDE	Materials in extreme dynamic environments
min(x)	Minimum values of the feature
ML	Machine learning
NIST	National institute of standards and technology
OSD	Open-source data
RMSE	Root mean square error
SD	Standard deviation
Sp	Superplasticizer content
SVM	Support vector machine
t	Concrete sample age
Vb	Bullet velocity
Wb	Bullet weight
Wp	Water portion
x	The original value
$\mathbf{x'}$	The normalized value
$\mathbf{x_i}$	Test data sample
$\mathbf{y_i}$	Train data sample