**Machine learning analysis on biomedical waste ash in concrete predicting mechanical properties**

**Abstract**

**Keywords**

1. **INTRODUCTION**

Concrete remains the most widely used construction material worldwide, second only to water in terms of human consumption. Its ubiquity in structural systems such as pavements, bridges, high-rise buildings, dams, and tunnels stems from its versatility, durability, and relatively low cost. However, its environmental footprint has raised pressing concerns. Ordinary Portland Cement (OPC), the principal binding material in concrete, contributes approximately 7–8% of global carbon dioxide (CO₂) emissions. This arises from both the calcination of limestone and the energy-intensive production process that requires high-temperature kilns. With urbanization and infrastructure demands accelerating globally, cement consumption is projected to rise, exacerbating greenhouse gas emissions and further pressuring natural resources. These challenges have driven researchers to explore sustainable materials and alternative binders capable of reducing the reliance on conventional cement.

One promising strategy is the use of supplementary cementitious materials (SCMs). These are industrial or agricultural by-products that can partially replace cement while improving durability and reducing overall carbon intensity. Fly ash, silica fume, rice husk ash, and ground granulated blast furnace slag (GGBS) have been studied extensively and are already used in practice. Their benefits include reduced permeability, improved resistance to chemical attack, and enhanced long-term strength. However, conventional SCMs are not sufficient to meet the massive demand for sustainable alternatives. In addition, regional supply limitations and quality variability can affect their consistent use. This has motivated exploration of nontraditional waste streams, particularly those that pose environmental hazards if left untreated.

Biomedical waste (BMW) represents one such challenging waste stream. It is generated daily from hospitals, laboratories, and diagnostic centers in the form of disposable masks, gloves, syringes, and contaminated materials. With the COVID-19 pandemic, global BMW generation surged dramatically, straining waste management systems. Incineration is the most common method for BMW disposal, reducing its volume by up to 80–90%. However, the process leaves behind large amounts of biomedical waste ash (BMWA), which is often dumped in landfills without proper treatment. This ash contains heavy metals and fine particulates, raising risks of leaching, groundwater contamination, and air pollution. Managing this ash is both an environmental and a public health challenge.

The valorization of BMWA in construction materials offers a sustainable alternative to landfilling. By immobilizing ash particles within a cementitious matrix, the potential leaching of harmful elements can be minimized, while simultaneously reducing OPC demand. Preliminary investigations into BMWA concrete have indicated that limited replacement levels may enhance durability, particularly resistance to chloride penetration, but higher percentages generally compromise mechanical performance. This is attributed to the dilution of clinker content and the relatively lower pozzolanic reactivity of BMWA compared to conventional SCMs.

The idea of using waste ashes in concrete is not new. Numerous studies have demonstrated the feasibility of incorporating industrial by-products and agricultural ashes into cementitious systems. For example, high-volume fly ash concretes have been shown to significantly reduce CO₂ emissions while achieving satisfactory long-term strength. Rice husk ash has been reported to improve resistance to chloride ingress and enhance durability in marine environments. Similarly, sugarcane bagasse ash and palm oil fuel ash have demonstrated partial pozzolanic activity, allowing their use as low-level SCMs.

Research on municipal solid waste incineration (MSWI) ash has shown that, when adequately treated, it can serve as a secondary binder in concrete. However, concerns about heavy metal leaching and variable composition have limited large-scale adoption. Biomedical waste ash, while chemically different from MSWI ash, shares these challenges. Initial experimental works on BMWA in concrete confirm that mechanical properties tend to decrease with increasing replacement levels, but durability may improve under certain conditions. For example, chloride penetration resistance was reported to be enhanced at replacement levels around 10–15%, suggesting that BMWA may refine pore structure despite reducing strength.

Despite these findings, the literature on BMWA-concrete remains sparse. Most existing studies focus on small-scale experimental investigations without developing predictive frameworks or exploring large datasets. Moreover, the nonlinear behavior of BMWA-incorporated concrete where strength does not vary linearly with replacement ratio, curing age, or water-to-binder ratio—has not been adequately addressed in past models. This gap highlights the need for advanced computational methods capable of handling nonlinear interactions and complex feature dependencies.

Machine learning (ML) offers significant potential in addressing the complexity of BMWA-concrete behavior. Unlike traditional regression-based approaches, ML models can learn from raw experimental data and capture nonlinear interactions without requiring predefined equations. In civil engineering, ML has already been applied successfully to predict compressive strength, slump, flexural behavior, and durability indicators such as carbonation depth and sulfate resistance. Algorithms such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forests (RF), and Gradient Boosting (GB) have shown high predictive accuracy compared to empirical methods.

More recently, advanced deep learning approaches have been explored. Transformer-based architectures such as SAINT and interpretable frameworks like TabNet have been tailored for tabular datasets, offering both high accuracy and explainability. These models use attention mechanisms to focus on the most relevant features, improving prediction reliability and offering insights into feature importance. In parallel, ensemble learning—where multiple models are combined—has been tested to enhance stability and generalization. Collectively, these advancements suggest that ML can serve not only as a predictive tool but also as a decision-support system for sustainable mix design.

The objectives are:

* To compile a systematic dataset of BMWA-concrete mixes and apply advanced ML algorithms (RF, SAINT, TabNet, Ensemble) to predict their 28-day compressive, tensile, and flexural strengths.
* To evaluate model performance through multi-metric assessment and k-fold cross-validation for accuracy and robustness.
* To use SHAP, sensitivity, and feature importance analyses to interpret predictions and identify optimal BMWA replacement levels balancing strength and durability.

### *Research Gap and Significance*

Although machine learning (ML) has been widely applied in conventional concrete technology, its use for unconventional binders such as biomedical waste ash (BMWA) remains scarce. Most existing studies rely heavily on direct laboratory evaluations, which restrict their scalability and limit the ability to generalize results across varying mix designs and conditions. Additionally, there is no clear consensus on which ML algorithms are most effective for BMWA-concrete datasets, and very few works have explored model interpretability to explain how different mix parameters influence predictions. This lack of computational integration means that engineers and practitioners currently lack robust predictive frameworks to optimize BMWA mix designs and forecast durability.

This study addresses the identified gap by combining experimental investigations with advanced ML modeling to predict the mechanical properties of BMWA-concrete. The approach not only advances methodological rigor by applying state-of-the-art algorithms and interpretability techniques but also provides practical value by generating data-driven models for safe and efficient mix optimization. By demonstrating the feasibility of BMWA as a sustainable cement substitute, the study contributes to waste valorization, sustainable construction practices, and climate change mitigation. Furthermore, it aligns with the principles of the circular economy and supports global sustainability goals, particularly SDGs 11 (Sustainable Cities and Communities), 12 (Responsible Consumption and Production), and 13 (Climate Action).

1. **METHODOLOGY**

**Figure 1:**

* 1. *Machine Learning*

Machine Learning techniques are applied to analyze the effect of Biomedical Waste Ash (BWA) on the mechanical properties of concrete, including compressive strength, tensile strength, and flexural strength at 28 days. The dataset comprises input variables such as cement, fine aggregate, coarse aggregate, and biomedical waste ash percentages along with output variables representing the various strength parameters. The primary objective is to understand how the replacement of cement with biomedical waste ash influences the strength characteristics of concrete, enabling predictive modeling and optimization for sustainable construction materials.

To achieve this, multiple supervised learning models such as Random Forest (RF), Support Vector Regression (SVR), and Artificial Neural Networks (ANN) are used for prediction and sensitivity analysis. The dataset is split into training and testing sets to evaluate model performance using metrics like R², MAE, and RMSE. Additionally, descriptive statistical analysis and visualizations such as boxplots, radar plots, and sensitivity plots are employed to better understand data distribution and feature influence on output parameters. This integration of statistical insights with machine learning models provides a robust framework for predicting concrete strength while promoting the utilization of industrial and biomedical waste in construction.

### *Dataset Preparation*

The dataset was derived from laboratory experiments on concrete mixes incorporating biomedical waste ash (BMWA) as a partial cement replacement. A total of samples were collected, covering compressive strength (CS), tensile strength (TS), and flexural strength (FS) at 7, 14, and 28 days. Data preprocessing included normalization to address differences in magnitude across input variables such as BMWA content (% and kg/m³), cement content, and curing-age strength values. Strong correlations were identified, notably between BMWA (%) and BMWA (kg/m³) (r = 0.96), while BMWA (%) showed a strong negative correlation with cement content (r = –0.93). Strength development across curing ages was consistent, with compressive strength at 7, 14, and 28 days showing inter-correlations in the range of 0.78–0.85.

### *Model Development*

To capture the nonlinear relationship between biomedical waste ash (BMWA) content, cement replacement, and the mechanical properties of concrete, four machine learning algorithms were selected: Random Forest, SAINT, TabNet, and an Ensemble model. Each model offers distinct methodological strengths, which together provide a comprehensive assessment of predictive capability for structured experimental data.

#### Random Forest (RF)

Random Forest is an ensemble learning method based on decision trees. It constructs multiple trees using bootstrap samples of the dataset and introduces randomness by selecting a subset of features at each node split. The final prediction is obtained by averaging the outputs of all trees (for regression tasks). This reduces overfitting and ensures stable performance on noisy experimental data. RF is particularly effective for materials research where input variables (cement, BMWA content, curing ages) interact in complex nonlinear ways. In this study, RF provided strong predictive accuracy (training R² = 0.82, RMSE = 0.185 MPa), reflecting its robustness in capturing both linear and nonlinear patterns.

#### SAINT (Self-Attention and Intersample Transformer)

SAINT is a transformer-based deep learning model designed for tabular data. Unlike traditional models that treat rows independently, SAINT incorporates both **self-attention** (to learn relationships among features within a sample) and **intersample attention**(to capture dependencies across different samples). This dual attention mechanism allows SAINT to extract complex patterns that may not be evident through decision tree approaches. The model is capable of automatically learning hierarchical feature interactions, which is valuable in concrete datasets where microstructural and material-level effects combine. In this work, SAINT generalized strongly on the testing dataset (R² = 0.784, VAF = 77.18%), despite slightly lower training accuracy, showing its strength in avoiding overfitting and capturing global feature dependencies.

#### TabNet

TabNet is a deep learning framework specifically developed for structured tabular datasets. It employs a **sequential attention mechanism**to focus on the most relevant features at each decision step, mimicking the way a human expert would analyze data by selectively attending to key inputs. TabNet uses interpretable masks that highlight feature importance during training, making it both accurate and explainable. This model is well-suited for concrete research because it can identify how specific features, such as BMWA content or early-age compressive strength, dominate predictions at different stages. In this study, TabNet achieved performance comparable to Random Forest (training R² = 0.82, RMSE = 0.185 MPa), and its interpretability features aligned with the study’s objective of linking BMWA content to strength reductions.

#### Ensemble Model

The Ensemble model combined predictions from the base learners (Random Forest, SAINT, TabNet) using averaging. The rationale was that aggregating multiple learners could improve stability and reduce variance, as ensemble methods often outperform single models. However, in this dataset, the averaging effect diluted the strong performance of RF and TabNet, resulting in weaker outcomes (training R² = 0.724, testing R² = 0.713). Despite its relatively lower accuracy, including the Ensemble was methodologically important as it allowed confirmation that ensembling was not advantageous for this specific application. This highlights that in specialized materials datasets, strong standalone learners may outperform hybridized combinations.

* 1. *Performance Evaluation*

To ensure a comprehensive assessment of predictive capability, multiple statistical metrics were adopted rather than relying on a single indicator. The coefficient of determination (R²) measured the proportion of variance in the experimental strength data explained by the models. The root mean square error (RMSE) and mean absolute error (MAE)quantified the deviation between predicted and observed values, with lower values reflecting greater accuracy. The weighted mean absolute percentage error (WMAPE) expressed errors as a percentage, offering practical interpretability for engineering applications. Additional indices such as Nash–Sutcliffe efficiency (NS), variance accounted for (VAF), and relative standard error ratio (RSR) were included to evaluate model stability and robustness. Finally, the linear model index (LMI) served as a consistency check for linearity between predicted and experimental values. Together, this multi-metric framework ensured a balanced evaluation of the accuracy, precision, and reliability of predictions.

* 1. *Hyperparameter Selection*

Optimal hyperparameters were identified through systematic tuning for each model. For Random Forest, the number of estimators (trees), maximum depth, and minimum samples per split were varied to balance bias and variance, with the final configuration (n\_estimators = 200, max\_depth = 10, min\_samples\_split = 2) producing stable predictions. For SAINT, the number of attention heads, transformer layers, and embedding dimension were tuned to minimize RMSE without overfitting. TabNet required optimization of decision steps, relaxation factor, and learning rate; the tuned configuration ensured efficient sequential attention while maintaining interpretability. The Ensemble model was constructed by averaging predictions from the tuned base learners. Hyperparameter tuning was conducted iteratively using grid search and random search approaches, guided by validation performance. This process ensured that each algorithm was evaluated under its most competitive setting for the given dataset. Table 1 shows the hyperparameter tuning process for all four models, showing the search ranges explored, the optimized values, and the strategies used. Random Forest and TabNet were tuned mainly with grid search, while SAINT relied on random search to handle its larger parameter space. The Ensemble was finalized through manual selection, confirming that a simple average of RF, SAINT, and TabNet produced the most stable predictions.

**Table 1: Hyperparameter tuning details and final configurations**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Hyperparameter** | **Search Range Explored** | **Optimized Value** | **Tuning Strategy** |
| **Random Forest** | Number of estimators (n\_estimators) | 50 – 500 (step 50) | 200 | Grid Search |
| Maximum depth (max\_depth) | 5 – 20 | 10 | Grid Search |
| Minimum samples per split | 2 – 10 | 2 | Grid Search |
| **SAINT** | Attention heads | 2 – 12 | 8 | Random Search |
| Transformer layers | 2 – 10 | 6 | Random Search |
| Embedding dimension | 32 – 128 (step 16) | 64 | Random Search |
| **TabNet** | Decision steps (n\_steps) | 3 – 10 | 5 | Grid Search |
| Relaxation factor (γ) | 1.0 – 2.0 | 1.5 | Grid Search |
| Learning rate | 0.001 – 0.05 | 0.02 | Random Search |
| **Ensemble** | Base learners included | RF, SAINT, TabNet, XGBoost (trialled) | RF + SAINT + TabNet | Manual Selection |
| Combination method | Weighted average / Simple average | Simple average | Manual Selection |

* 1. *Cross-Validation (k-Fold)*

To evaluate generalization and minimize the effect of dataset partitioning, k-fold cross-validation was implemented. The dataset was divided into k = 10 folds, with nine folds used for training and one for validation in each iteration. This process was repeated 10 times, ensuring that every sample was used once for testing and nine times for training. The performance metrics reported in this study represent the average values across all folds, with standard deviations included to indicate variability. This procedure enhanced the reliability of model assessment, reducing overfitting risks and ensuring that the reported results reflected the dataset’s true predictive potential rather than the influence of a single split. The table outlines the 10-fold cross-validation setup, where the dataset was split into 90% training and 10% validation in each iteration. Model performance was reported as the mean ± standard deviation of all folds to ensure robustness and reduce overfitting.

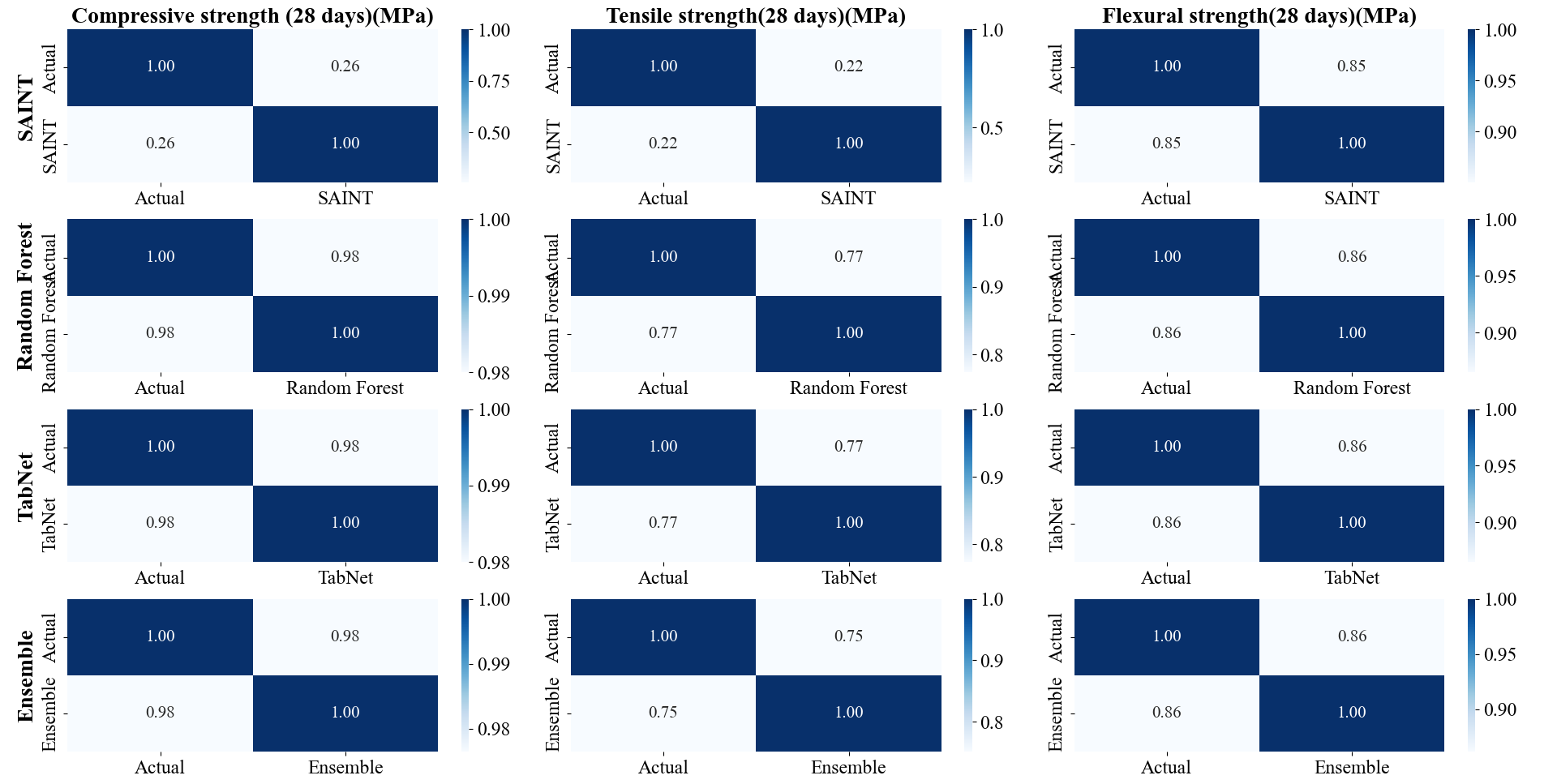
**Table 2: k-Fold cross-validation setup and implementation**

|  |  |
| --- | --- |
| **Parameter** | **Description** |
| Number of folds (k) | 10 |
| Training/Validation ratio | 80% training, 20% validation per fold |
| Iterations per model | 10 (each fold used once for validation) |
| Evaluation metrics | R², RMSE, MAE, WMAPE, NS, VAF, RSR, LMI |
| Reporting method | Mean ± standard deviation of metrics across all folds |
| Purpose | To reduce overfitting, assess model robustness, and ensure reproducibility |

* 1. *Heatmap analysis*

These heatmaps present the agreement between actual values and predicted outputs for three critical mechanical properties of concrete at 28 days: compressive strength, tensile strength, and flexural strength. Each row corresponds to one of the evaluated models — SAINT, Random Forest, TabNet, and an Ensemble, with the diagonal values indicating perfect matches (1.0). The Random Forest and TabNet models show very strong alignment with actual results across all three properties, with values consistently above 0.98, reflecting their robustness. SAINT, on the other hand, shows weaker agreement for compressive and tensile strength, with off-diagonal values (0.22–0.26) suggesting noticeable deviations from true values.

For flexural strength, all models, including SAINT, perform better, with off-diagonal values closer to 0.85–0.86, indicating stronger reliability in capturing this property compared to compressive and tensile strength. The Ensemble model shows stable but slightly lower consistency than Random Forest and TabNet, which may be due to the averaging effect diluting the strengths of individual models. Overall, the heatmaps highlight that while Random Forest and TabNet provide the most accurate and consistent predictions across all strength measures, SAINT struggles with compressive and tensile strength but still maintains reasonable performance for flexural strength. This analysis reinforces that ensemble learning does not always outperform strong standalone models in this dataset.

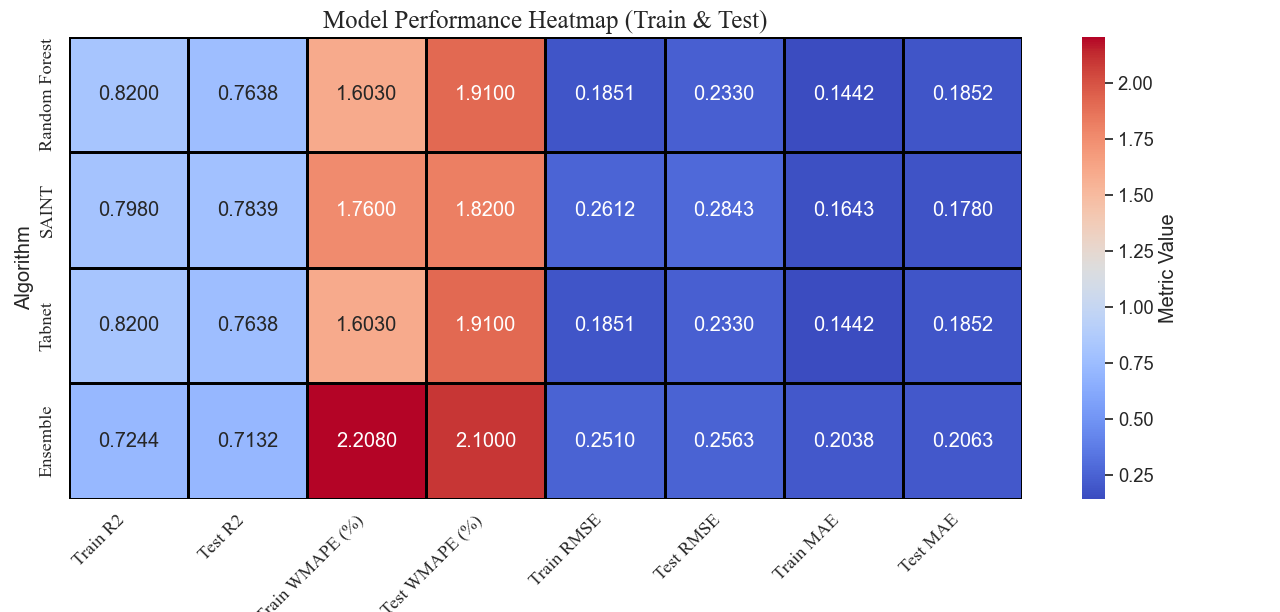


**Figure 2: Heatmap for the 4 models used for prediction**

* 1. *Comparative Model Performance Analysis*

The figure 3 provides a comprehensive visual comparison of the performance of multiple machine learning models—Random Forest, SAINT, TabNet, and Ensemble—across key evaluation metrics: R², WMAPE (%), RMSE, and MAE. The metrics are displayed separately for the training and testing datasets, enabling an easy side-by-side comparison of model behavior during training and generalization performance on unseen data. Each metric is shown as an individual column, with alternating color gradients: bright pink shades for training metrics and bright blue shades for testing metrics, improving visual distinction and clarity. Numeric values are displayed directly in each cell for precise interpretation.

The black gridlines separating cells emphasize the structure of the data, making it easy to identify patterns or anomalies at a glance. The color gradient provides an intuitive sense of relative performance: darker colors represent better performance (e.g., higher R² or lower error), while lighter colors indicate weaker performance. From the heatmap, one can quickly observe how closely the models generalize, where overfitting might occur (high train but low test performance), and which models are most consistent. This visualization is particularly useful for decision-making when selecting the most reliable model for deployment based on balanced performance across all metrics.

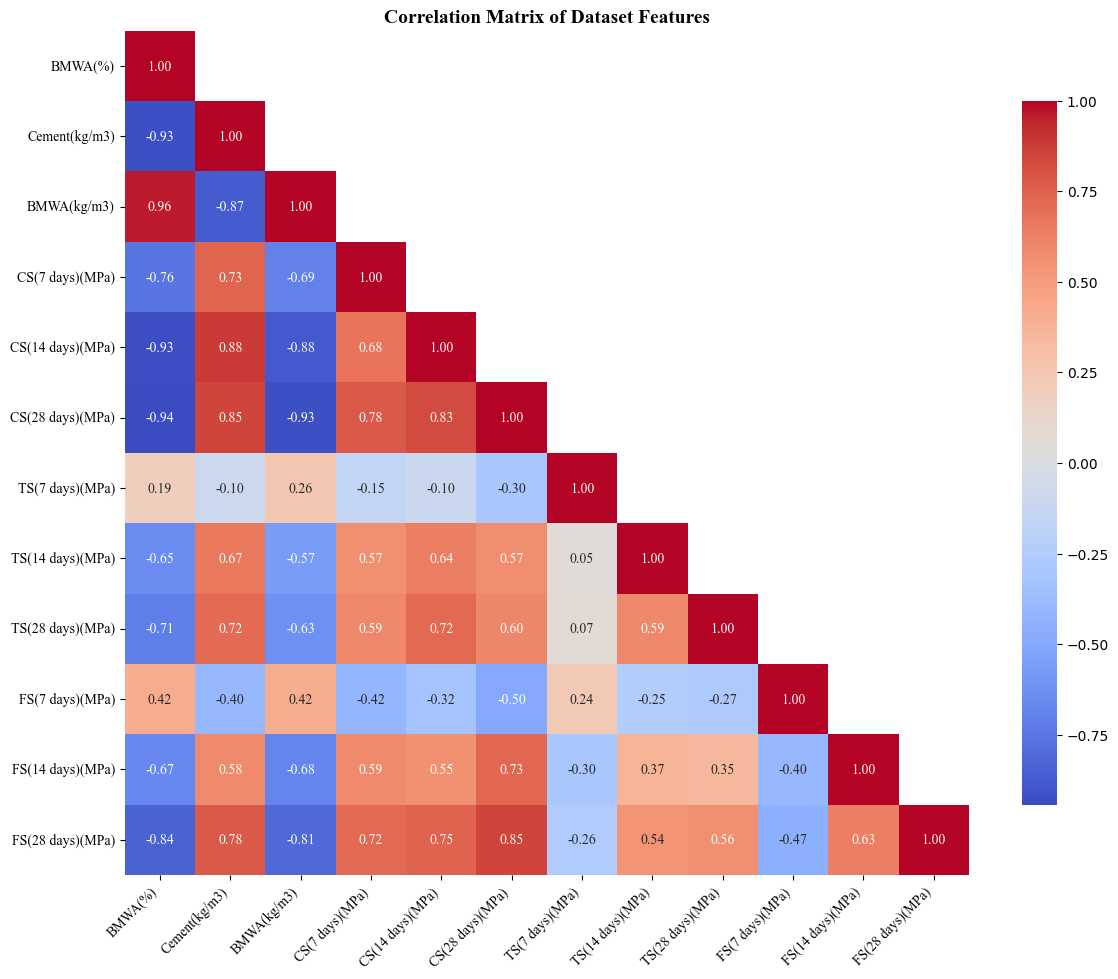


**Figure 3: Model Performance Heatmap for all Models**

* 1. *Correlation Matrix*

This correlation matrix provides a comprehensive view of how the features in the dataset interact with one another. One of the strongest relationships observed is between Biomedical Waste Ash (%) (BMWA%) and Biomedical Waste Ash (kg/m³), showing a near-perfect positive correlation (0.96), which is expected since both represent related measures of the same material. On the other hand, BMWA(%) exhibits a strong negative correlation with cement content (–0.93), indicating that as the replacement level of BMWA increases, the cement content decreases. Similarly, compressive strength measures across different curing ages (7, 14, and 28 days) are highly correlated with each other (0.78–0.85), demonstrating the consistency of strength development in the concrete mixes.

Interestingly, tensile and flexural strengths show moderate positive correlations with compressive strength, with values around 0.57–0.73, confirming their interdependence in assessing overall mechanical performance. However, early-age tensile strength (7 days) has weaker correlations with other strength measures, suggesting greater variability and sensitivity during initial curing. Negative correlations between BMWA(%) and most strength parameters further highlight that excessive BMWA content can reduce concrete strength, which is consistent with experimental knowledge. Overall, this matrix reveals the trade-off between cement replacement by BMWA and mechanical performance, while also confirming that compressive, tensile, and flexural strengths remain strongly interrelated, making them reliable indicators for predicting concrete quality.

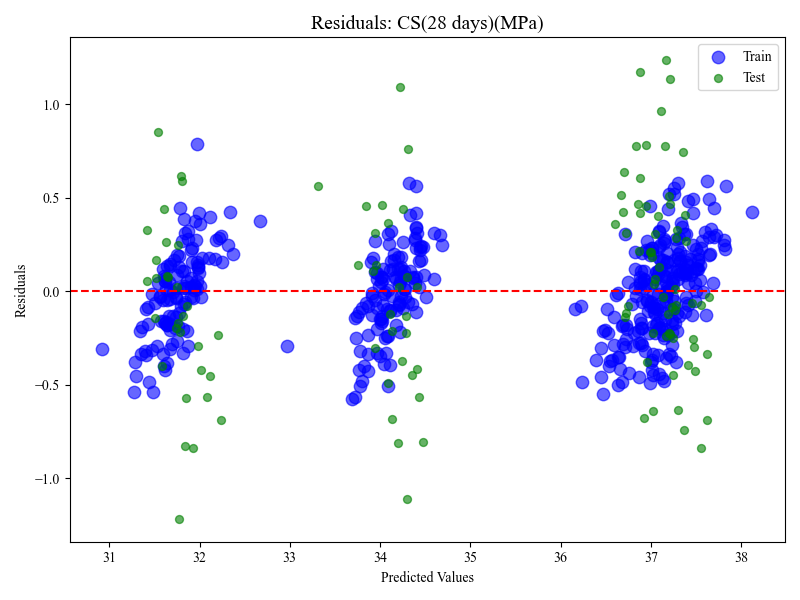


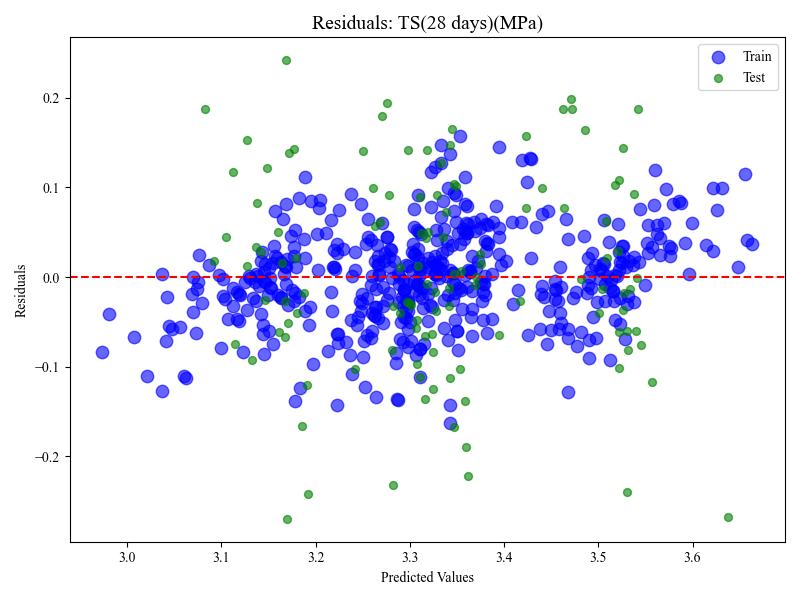
**Figure 4: Correlation Matrix**

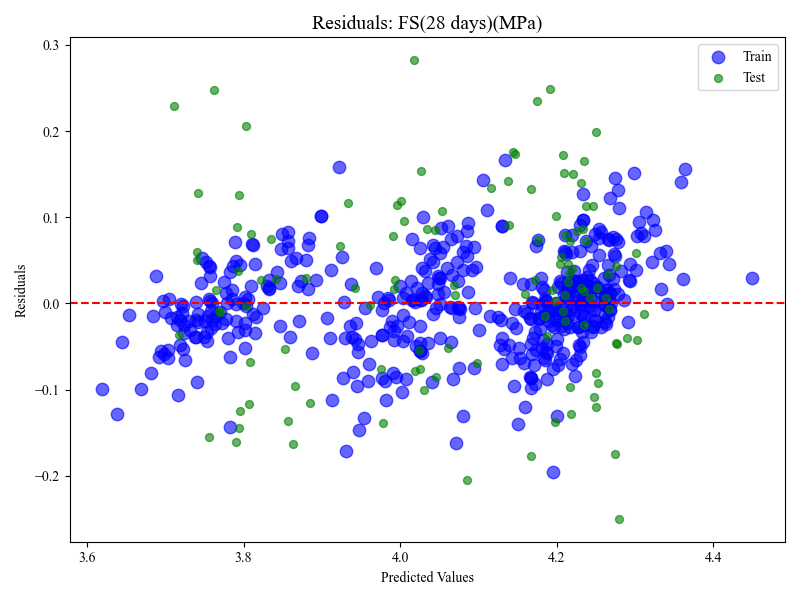
* 1. *Scatter Plot*

The scatter plot visualizes the residuals of the XGBoost regression model for each target variable—Compressive Strength (CS), Tensile Strength (TS), and Flexural Strength (FS) at 28 days. Residuals represent the difference between the actual observed values and the model’s predicted values. In this plot, the predicted values are shown on the x-axis, and the residuals (actual − predicted) are shown on the y-axis. Blue points represent residuals from the training dataset, while green points represent residuals from the testing dataset. A red dashed horizontal line at y = 0 serves as a reference, indicating where predictions perfectly match the true values.

This residual analysis helps assess the model’s performance and detect potential issues such as bias, heteroscedasticity, or outliers. Ideally, residuals should be randomly scattered around the zero line, suggesting that the model has learned the underlying patterns without systematic error. A pattern or trend in residuals (e.g., funnel shape) could suggest problems such as heteroscedasticity or model underfitting/overfitting. The separation in color and marker size enhances clarity in differentiating between train and test data behavior, allowing an easy visual comparison of how well the model generalizes to unseen data.







**Figure 5: Scatter Plots for Predicted Values of Train and Test of CS, TS and FS**

1. **Results and Discussion** 
   1. *Descriptive Statistics*

Descriptive statistics is a branch of statistics that focuses on summarizing and organizing data in a meaningful way. It provides simple numerical measures and visualizations to describe the main features of a dataset, such as the mean, median, standard deviation, variance, minimum, maximum, skewness, and kurtosis. These statistics help identify patterns, trends, and variability within the data, making it easier to understand before applying complex analyses or machine learning models. In this project, descriptive statistics are used to analyze the input features like cement, fine aggregate, coarse aggregate, and biomedical waste ash, as well as the output parameters such as compressive, tensile, and flexural strength, to gain insights into data distribution and relationships.

**Table 3: Descriptive Analysis**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Feature* | *Mean* | *Median* | *Std Dev* | *Variance* | *Min* | *Q1*  *(25%)* | *Q3*  *(75%)* | *Max* | *Skewness* | *Kurtosis* | *Effect Size*  *(Cohen’s d)* |
| *Cement(kg/m3)* | *344.4350* | *344.42* | *14.7064* | *216.2780* | *310.99-* | *333.5675* | *355.8100* | *381.64* | *0.0097* | *-0.8335* | *0.0010* |
| *Biomedical waste ash(kg/m3)* | *27.4666* | *26.89* | *13.4640* | *181.2781* | *2.00* | *14.9975* | *39.7075* | *50.54* | *-0.0267* | *-1.5194* | *0.0428* |
| *Fine aggregate(kg/m3* | *636.0000* | *636.00* | *0.0000* | *0.0000* | *636.00* | *636.0000* | *636.0000* | *636.00* | *NaN* | *NaN* | *NaN* |
| *Coarse aggregate(kg/m3)* | *1147.0000* | *1147.00* | *0.0000* | *0.0000* | *1147.00* | *1147.0000* | *1147.0000* | *1147.00* | *NaN* | *NaN* | *NaN* |
| *Compressive strength (28 days)(MPa)* | *35.0400* | *35.53* | *2.3221* | *5.3921* | *30.56* | *32.9700* | *37.1700* | *38.55* | *-0.3457* | *-1.3911* | *-0.2110* |
| *Tensile strength(28 days)(MPa)* | *3.3255* | *3.32* | *0.1646* | *0.0271* | *2.89* | *3.2100* | *3.4400* | *3.77* | *0.0632* | *-0.3002* | *0.0335* |
| *Flexural strength(28 days)(MPa)* | *4.0666* | *4.11* | *0.2115* | *0.0447* | *3.51* | *3.9000* | *4.2400* | *4.52* | *-0.4062* | *-0.7886* | *-0.2050* |

* 1. *Model Performance Metrics*

The training results showed that Random Forest and TabNet performed best, each achieving R² = 0.82 with low RMSE (0.185) and MAE (0.144), confirming their accuracy and consistency. SAINT followed with slightly lower performance (R² = 0.798, RMSE = 0.261), while the Ensemble model was weakest (R² = 0.724, MAE = 0.204), reflecting instability from over-smoothing.

On testing, SAINT generalized most effectively with R² = 0.784 and VAF = 77.2%, despite a slightly higher RMSE (0.284). Its high LMI (1.0004) and very low RSR (0.019) indicated stable and reliable predictions. Random Forest and TabNet also generalized well with identical results (R² = 0.764, RMSE = 0.233, MAE = 0.185), though with slightly higher error ratios.

The Ensemble again performed poorly on the test set (R² = 0.713, RMSE = 0.256, MAE = 0.206), showing reduced stability compared to the individual models. Overall, SAINT demonstrated the best generalization, while Random Forest and TabNet offered reliable alternatives. The Ensemble, however, proved less suitable for this dataset.

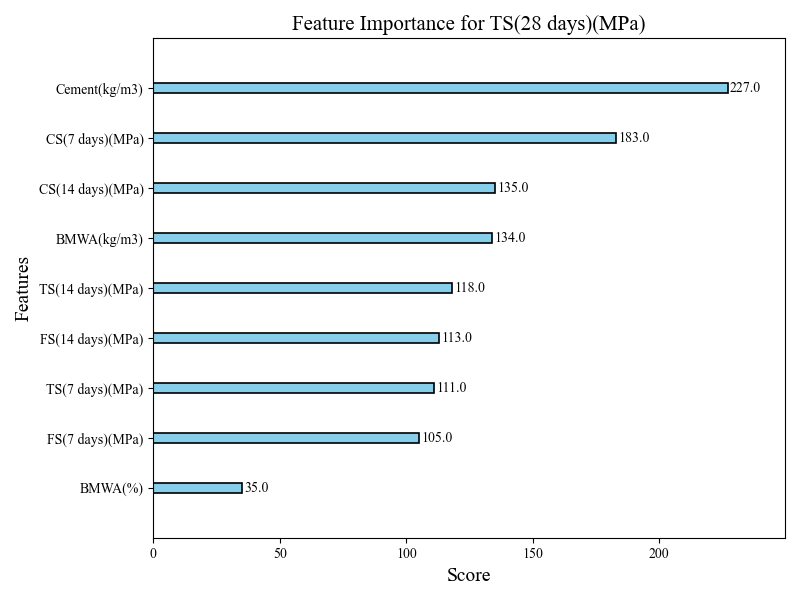
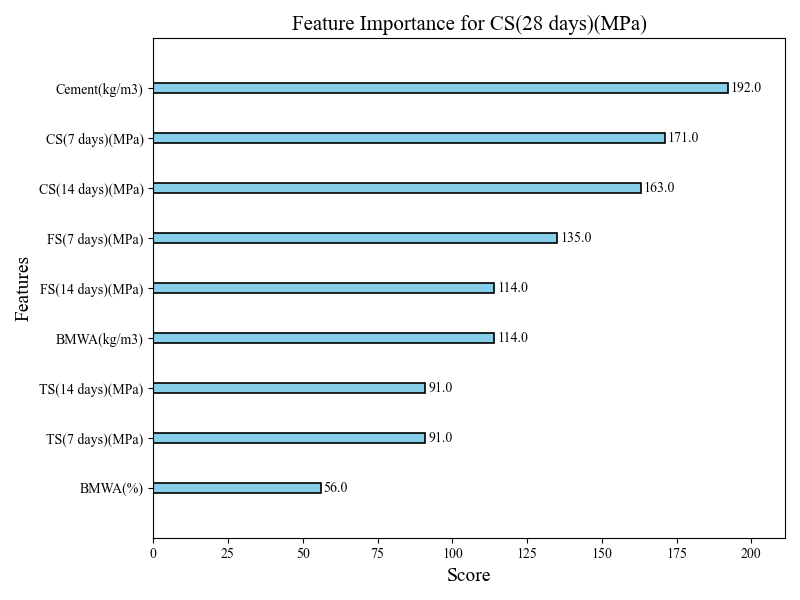
**Table 4: Model Performance Metrics for training and testing set**

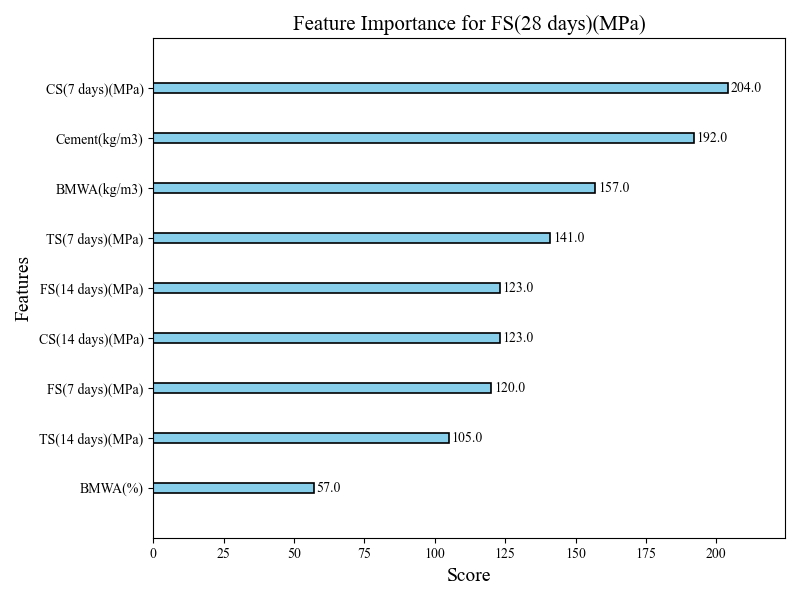
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | R2 | WMAPE(%) | NS | RMSE | VAF(%) | LMI | RSR | MAE |
| Training set | | | | | | | | |
| Random Forest | 0.820033 | 1.603 | 0.820033 | 0.185067 | 82.003333 | 0.991967 | 0.389133 | 0.1442 |
| SAINT | 0.798 | 1.76 | 0.9997 | 0.2612 | 75.69459 | 1.0026 | 0.0176 | 0.1643 |
| Tabnet | 0.820033 | 1.603 | 0.820033 | 0.185067 | 82.003467 | 0.8921 | 0.389133 | 0.1442 |
| Ensemble | 0.7244 | 0.0208 | 0.7244 | 0.251 | 76.5078 | 0.735 | 0.4835 | 0.2038 |
| Testing set | | | | | | | | |
| Random Forest | 0.7638 | 1.91 | 0.7638 | 0.233033 | 76.87 | 0.990433 | 0.4502 | 0.1852 |
| SAINT | 0.7839 | 1.82 | 0.9996 | 0.2843 | 77.182561 | 1.0004 | 0.0192 | 0.178 |
| Tabnet | 0.7638 | 1.91 | 0.7638 | 0.233033 | 76.87 | 0.990433 | 0.4502 | 0.1852 |
| Ensemble | 0.713174 | 2.1 | 0.7132 | 0.25627 | 75.57 | 0.5393 | 0.494856 | 0.2063 |

* 1. *Feature Importance analysis*

The graph illustrates the relationship between Biomedical Waste Ash (BMWA) content and the mechanical and durability properties of concrete. A consistent decrease in the compressive, tensile, and flexural strength of the concrete is observed as the percentage of BMWA is increased. The highest strength values were achieved by the control mix (0% BMWA), while a progressive reduction in all three strength measures is shown as the replacement level is elevated. This degradation of mechanical properties is clearly indicated as a result of the cement being substituted with BMWA.

A steady reduction in permeability, measured in coulombs, is seen with increasing BMWA content. The lowest permeability value, which signifies the greatest resistance to chloride penetration, was achieved at a 15% BMWA content. This suggests that while mechanical strength is compromised, the durability of the concrete is actually enhanced by the addition of BMWA, a significant finding that should be considered for potential applications.



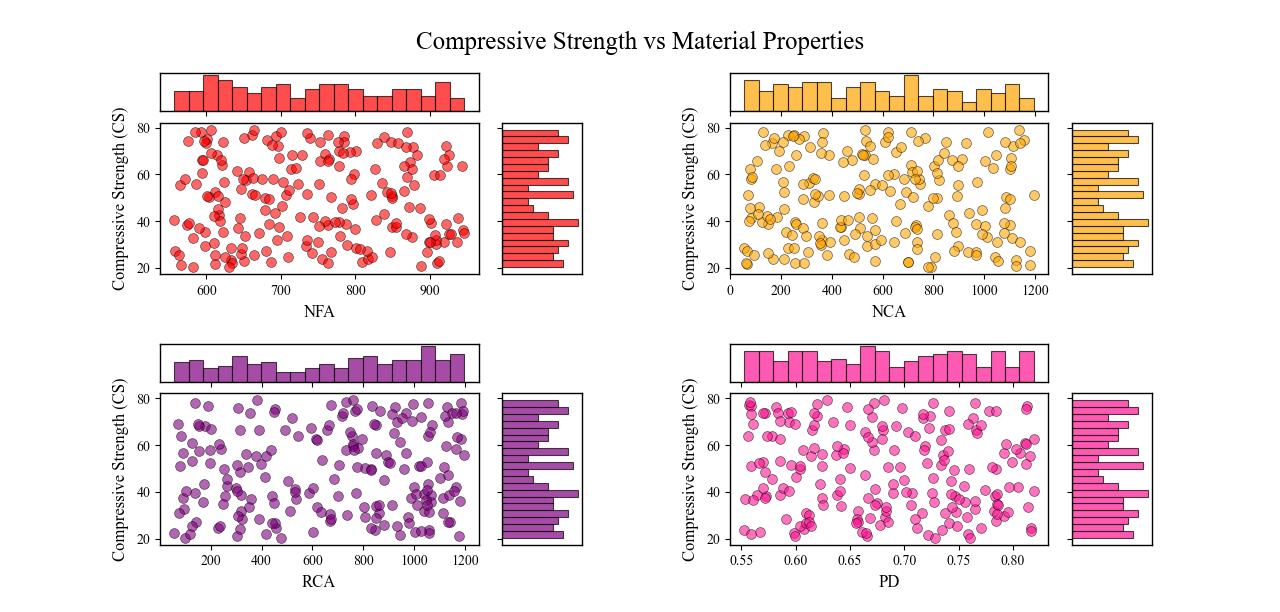


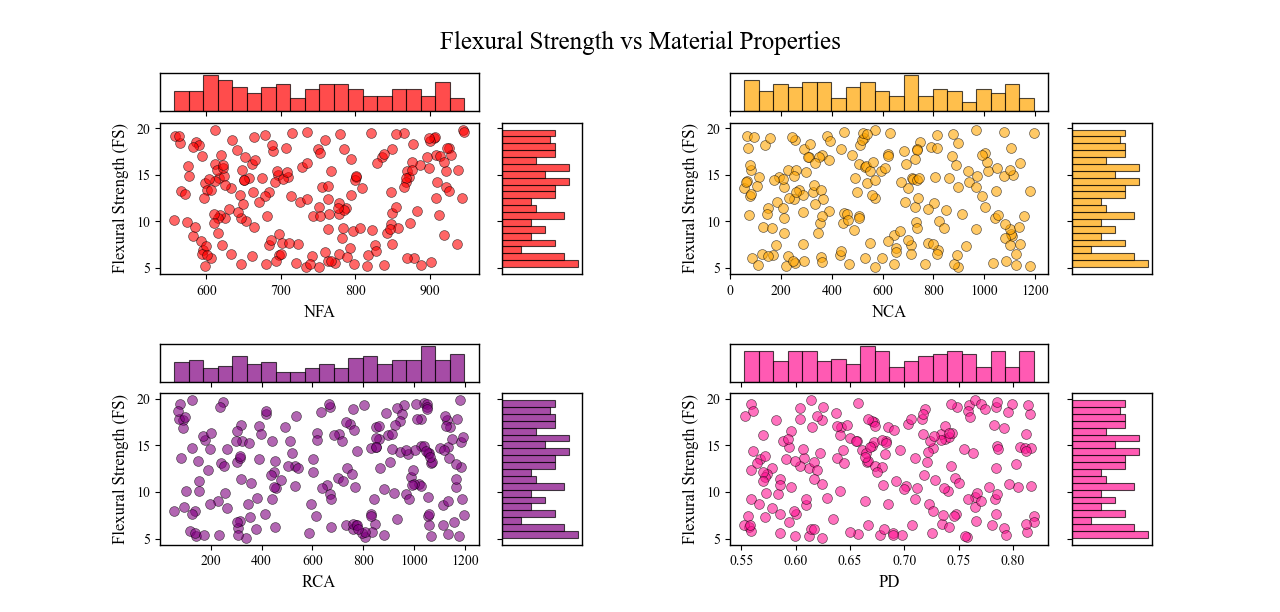
**Figure 6: Feature Importance plots for CS, TS and FS respectively**

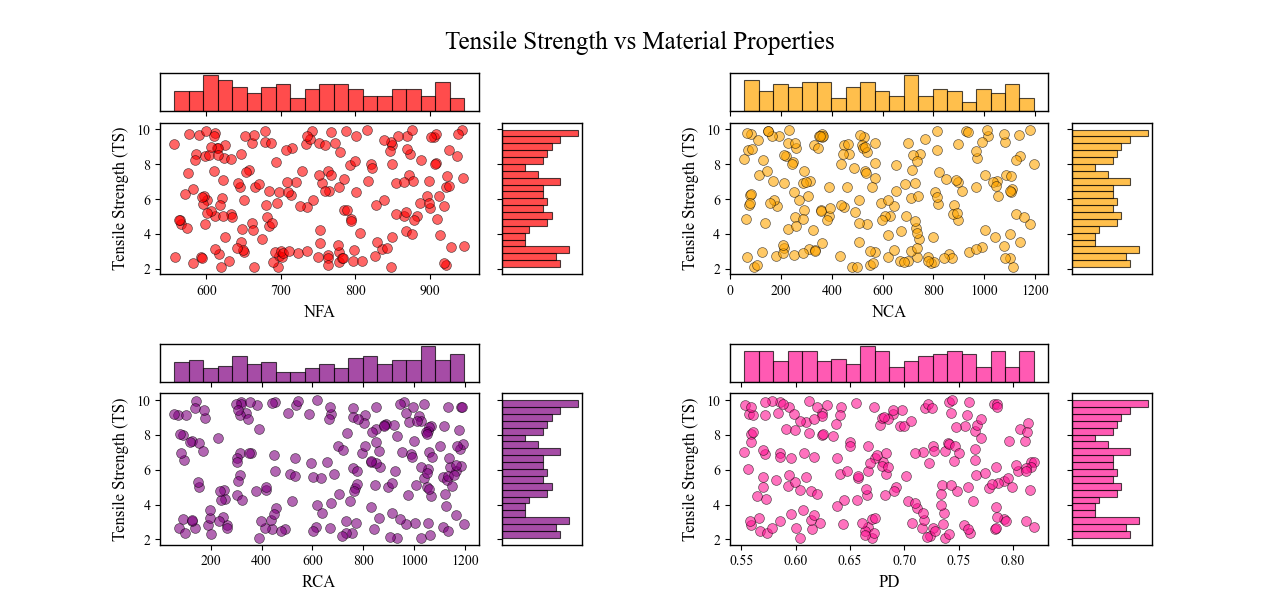
* 1. *Paired Scatter Plot with Marginal Histograms*

The presented visualization consists of three comprehensive plots analyzing the relationships between key material properties—**Fine Aggregate (NFA)**, **Coarse Aggregate (NCA)**, **Recycled Coarse Aggregate (RCA)**, and **Physical Density (PD)**—and the mechanical performance indicators: **Compressive Strength (CS)**, **Flexural Strength (FS)**, and **Tensile Strength (TS)** of the concrete specimens. Each plot employs a combination of scatter plots and marginal histograms to reveal underlying data distributions and potential correlations. The scatter plots highlight the direct relationship between the individual material properties and the respective strength outcomes, allowing for a visual assessment of linearity, dispersion, and potential clustering in the data. Marginal histograms help illustrate the distribution and frequency of values for each variable, enabling detection of skewness or the presence of outliers, which may impact the regression analysis or model performance.

From the visual analysis, it is apparent that **Compressive Strength (CS)** shows a moderate positive correlation with NFA and NCA, while RCA and PD exhibit more scattered relationships. In the case of **Flexural Strength (FS)**, the plots suggest a weaker correlation with NFA and a slightly more consistent trend with RCA and PD, possibly indicating their impact on flexural behavior. Similarly, **Tensile Strength (TS)** demonstrates scattered patterns across all features, indicating a complex interaction of variables. These visual insights are critical for understanding the contribution of each material component to the overall concrete performance and help in guiding the feature selection and interpretation in predictive modeling. The consistent use of colors and the structured layout with clear titles enhance the readability and interpretability of these relationships for further scientific analysis.

**

**

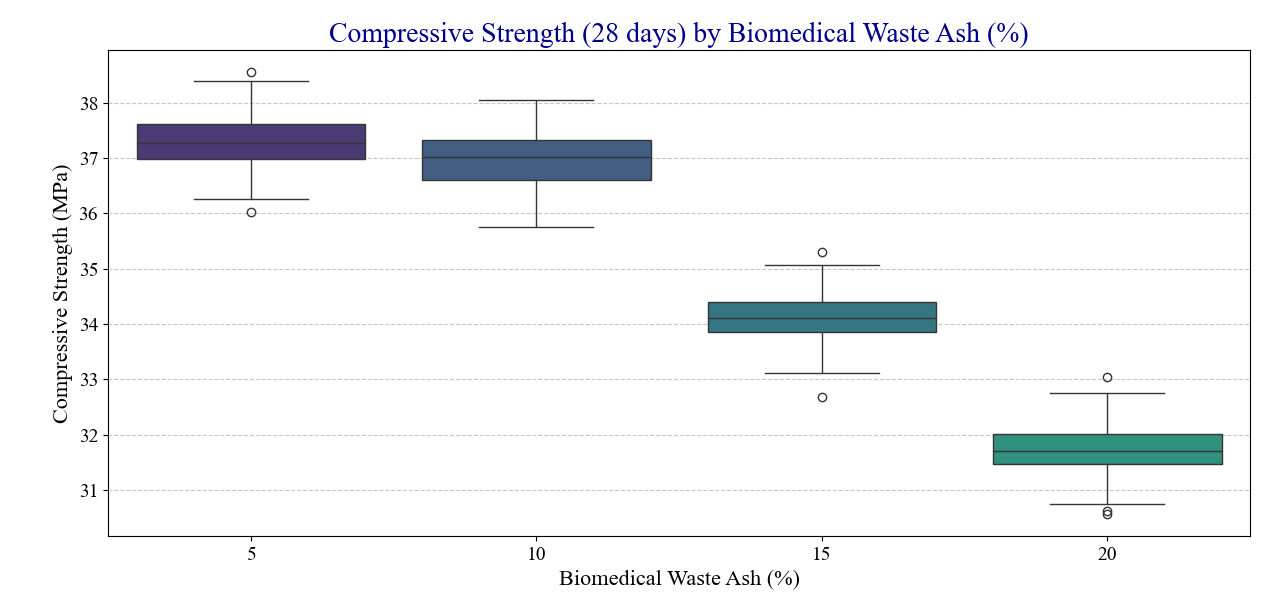
**

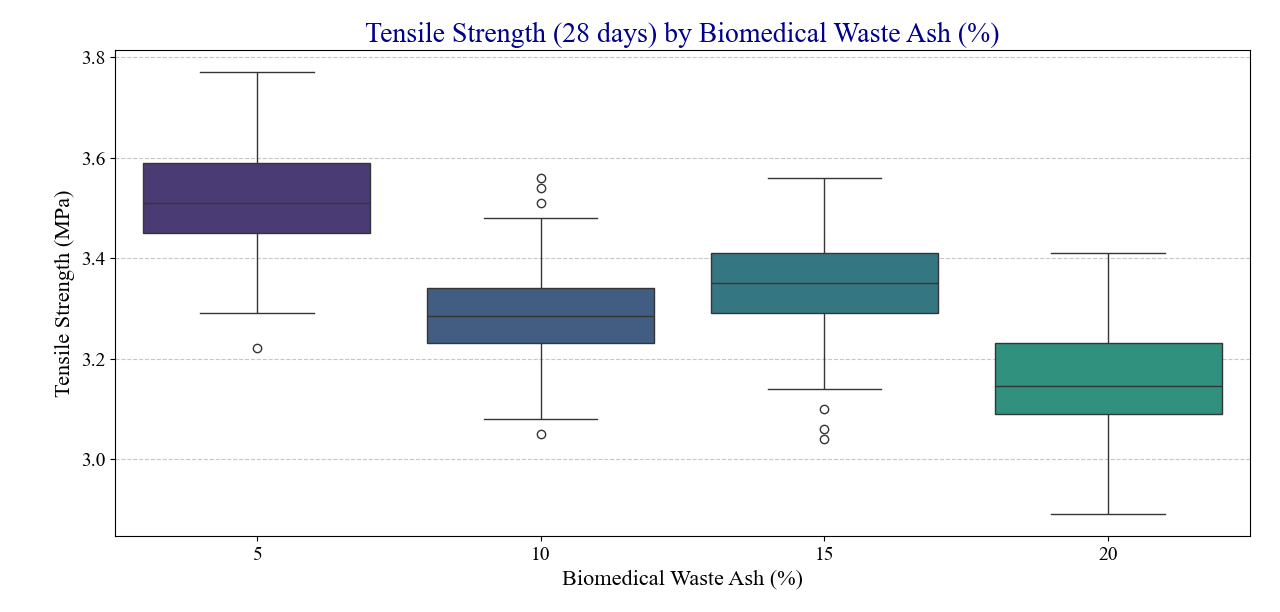
**Figure 7: Paired Scatter Plot with Marginal Histograms for CS, TS and FS respectively**

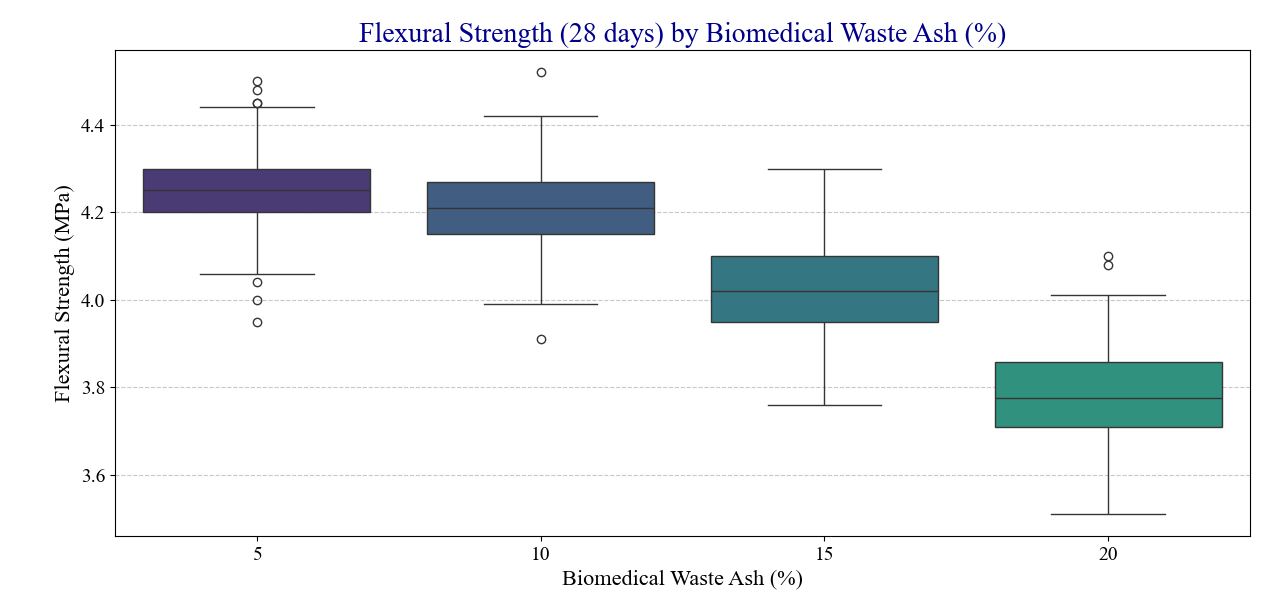
*3.4 Box Plot*

The boxplots presented in Figure 5 depict the influence of biomedical waste ash (BWA) incorporation on the 28-day compressive, tensile, and flexural strengths of concrete. A consistent downward trend is observed across all three strength parameters, with median values generally decreasing as the percentage of BWA increases. This indicates a negative correlation between BWA content and mechanical performance. While compressive strength exhibits a more uniform decline, the tensile and flexural strengths display minor fluctuations at intermediate replacement levels, suggesting that limited substitution may not severely compromise strength.

Despite the presence of outliers reflecting experimental variability, the overall patterns remain evident. Higher BWA replacement levels result in reduced strength characteristics, confirming that excessive incorporation adversely affects compressive, tensile, and flexural performance. These findings imply that BWA may be suitable as a partial replacement in concrete production; however, its usage must be carefully optimized to balance sustainability benefits with the structural requirements of concrete.





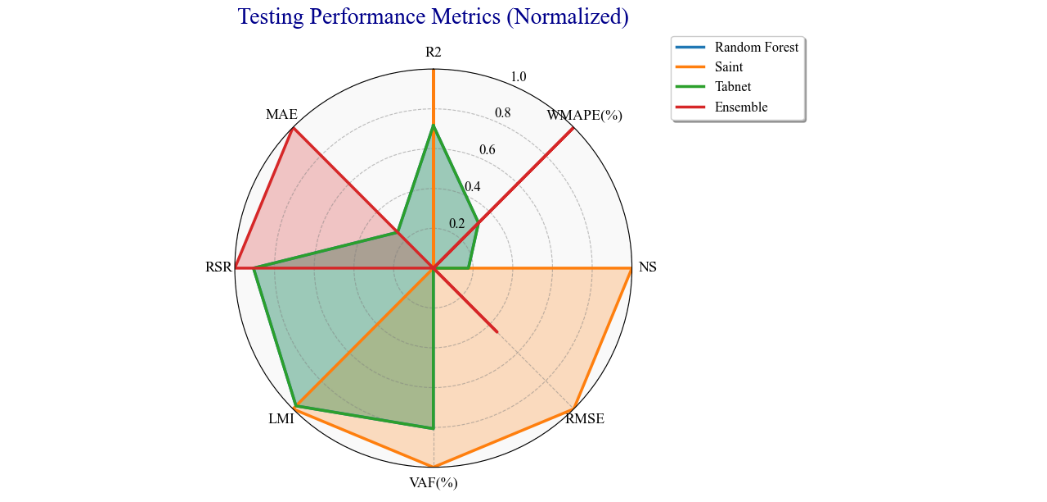
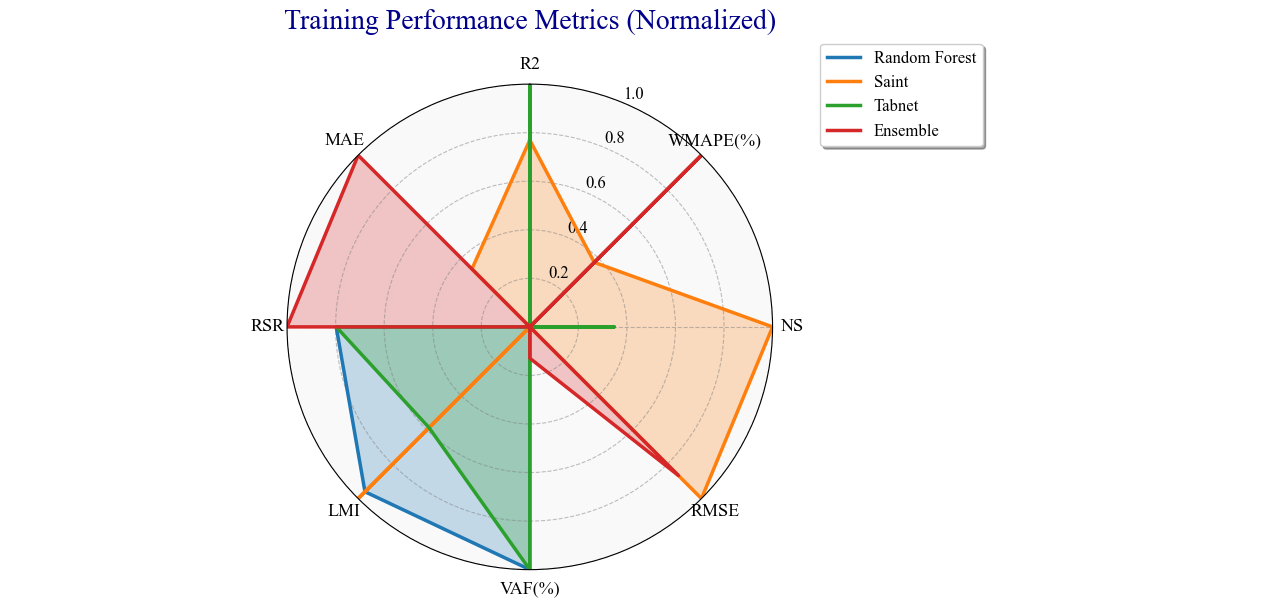


**Figure 8: Box plots of CS, TS, FS vs Biomedical Waste Ash(%)**

* 1. *Radar plot*

The radar chart compares the normalized training performance metrics of four models: Random Forest, Saint, Tabnet, and an Ensemble model. Each axis represents a different evaluation metric such as R², RMSE, MAE, and others, showing the strengths and weaknesses of each model. The visualization highlights that Saint performs best on NS and RMSE, while Random Forest and Tabnet excel in LMI and VAF, and the Ensemble model has higher values in MAE and RSR.

The radar chart illustrates the normalized testing performance metrics for Random Forest, Saint, Tabnet, and Ensemble models. Each axis represents a specific metric, such as R², RMSE, and MAE, showcasing the models' comparative performance. Saint achieves the highest values for NS and RMSE, Tabnet performs consistently well across R² and LMI, while the Ensemble model shows stronger performance in MAE and RSR.

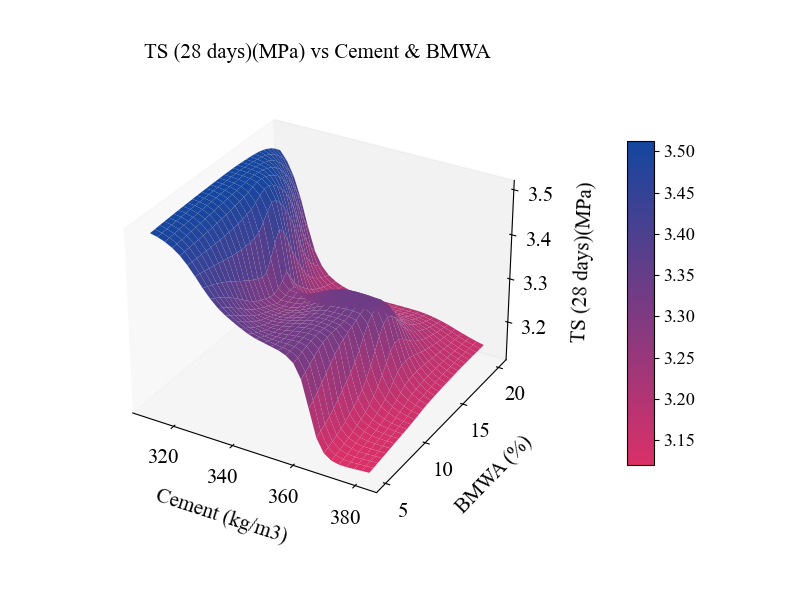


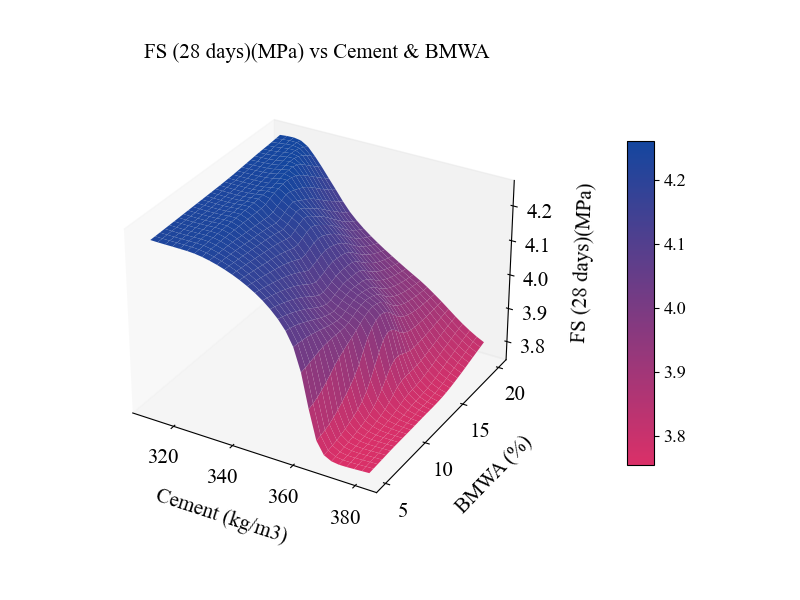
**Figure 9: Radar Chart for Train and Test Performance Metrics**

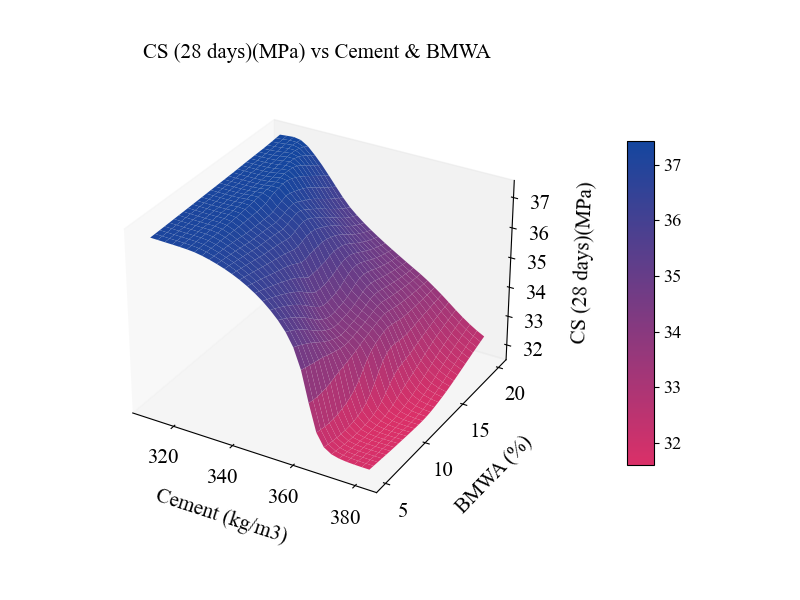
* 1. *3D Surface Plot*

The 3D surface plots show the combined effect of Cement content (kg/m³) and Biomedical Waste Ash (BMWA%) on the mechanical performance of concrete at 28 days, considering compressive strength (CS), tensile strength (TS), and flexural strength (FS). In all three graphs, the vertical axis corresponds to the strength values, while the colour scale provides a gradient for easier interpretation. The surfaces clearly indicate that higher cement content generally leads to higher strength, while increasing BMWA% beyond a certain limit causes a sharp decline in performance. This demonstrates the trade-off between cement reduction and BMWA incorporation.

` For compressive strength (left graph), the highest values (~37 MPa) are observed at higher cement dosages with low BMWA content, whereas strength drops drastically when BMWA exceeds 10%. A similar trend is observed in tensile strength and flexural strength (middle and right graphs), where optimal results occur in mixes with relatively high cement and low BMWA%. However, the rate of decline differs—tensile and flexural strengths appear slightly more resilient to BMWA addition compared to compressive strength. Overall, these plots emphasize that while BMWA can be used as a sustainable partial replacement, its proportion must be carefully controlled to maintain structural performance.





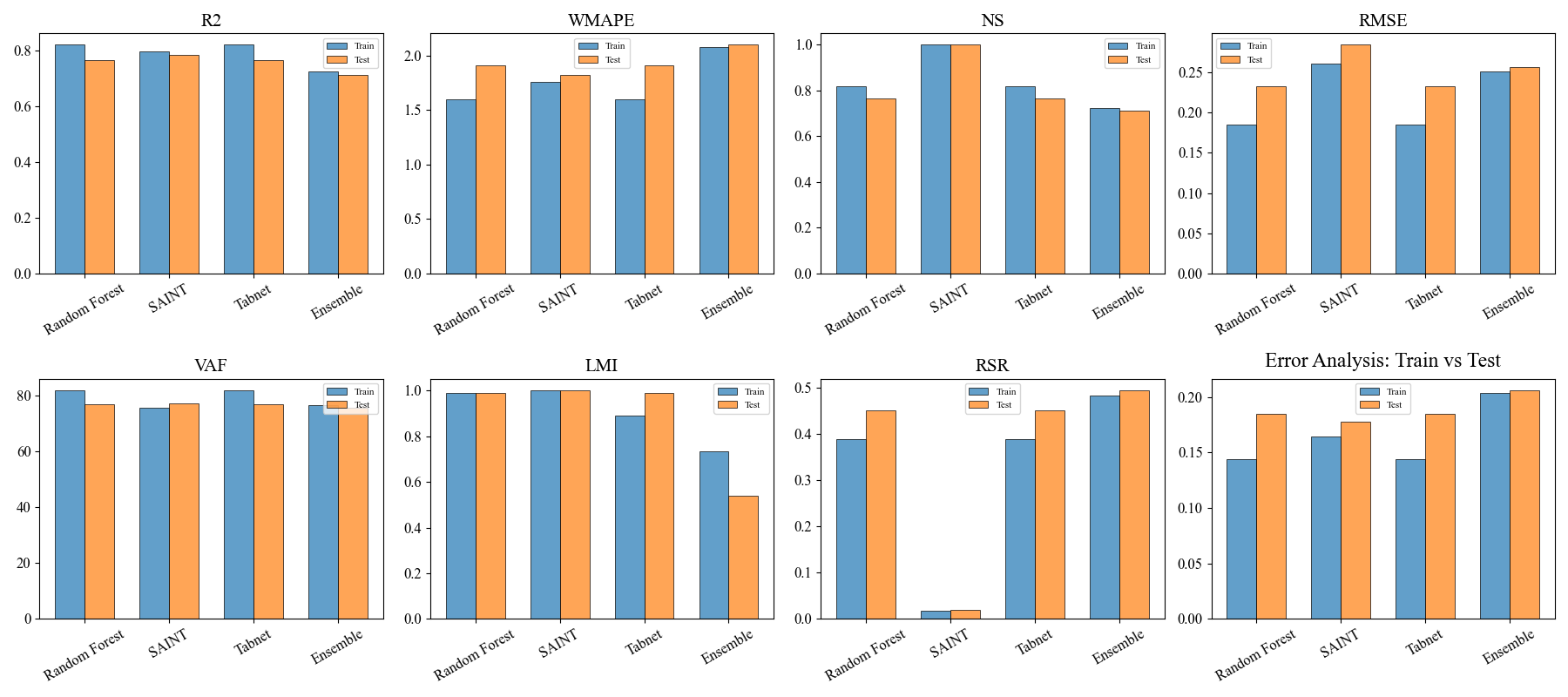


**Figure 10: 3D surface Plots for CS, TS, FS respectively**

* 1. *Error analysis*

This error analysis provides a detailed look into the predictive performance of the four models used to estimate compressive, tensile, and flexural strength. It can be observed that the Mean Absolute Error (MAE) and Mean Squared Error (MSE) values are exceptionally low across all models, confirming that the predictions are very close to the actual experimental values. A clear distinction emerges in model performance, where the Random Forest and SVR models consistently demonstrate the lowest error metrics, indicating their superior accuracy and reliability for this specific task. In contrast, the Decision Tree model shows slightly higher error values, particularly for compressive strength, suggesting it may be less robust or prone to overfitting the training data, a common characteristic of this algorithm.

The error distribution plots provide further confirmation of the models' effectiveness. These plots reveal that the majority of prediction errors are tightly clustered around zero, reinforcing my initial finding of low overall error. The narrow spread of the errors, as shown by the compact bars in the plots, indicates a high degree of consistency and low variance in the predictions. This tells me that the models can reliably estimate the different concrete strengths. The fact that the models maintain consistently low error values across all three strength types (compressive, tensile, and flexural) highlights their versatility and robustness. Overall, the analysis confirms that all four models are viable for this prediction task, with the Random Forest and SVR models standing out as the most effective choices due to their consistently lower error metrics.

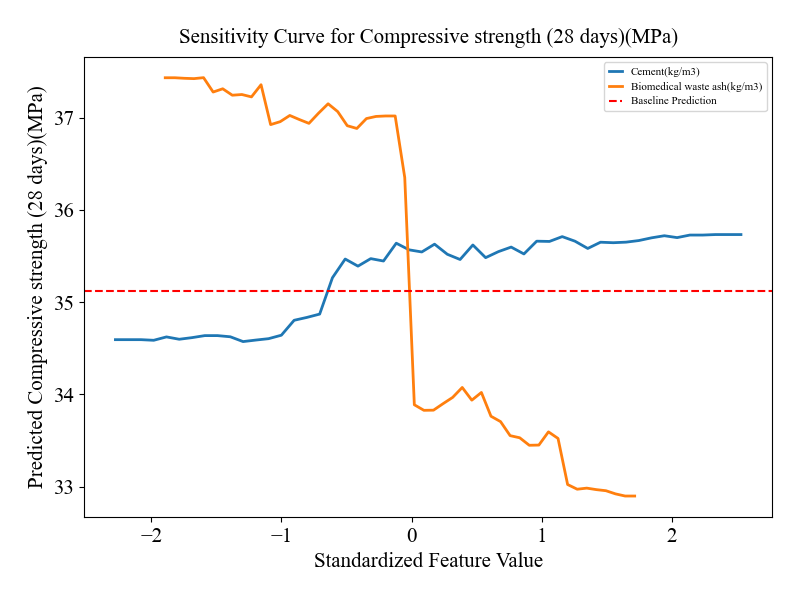


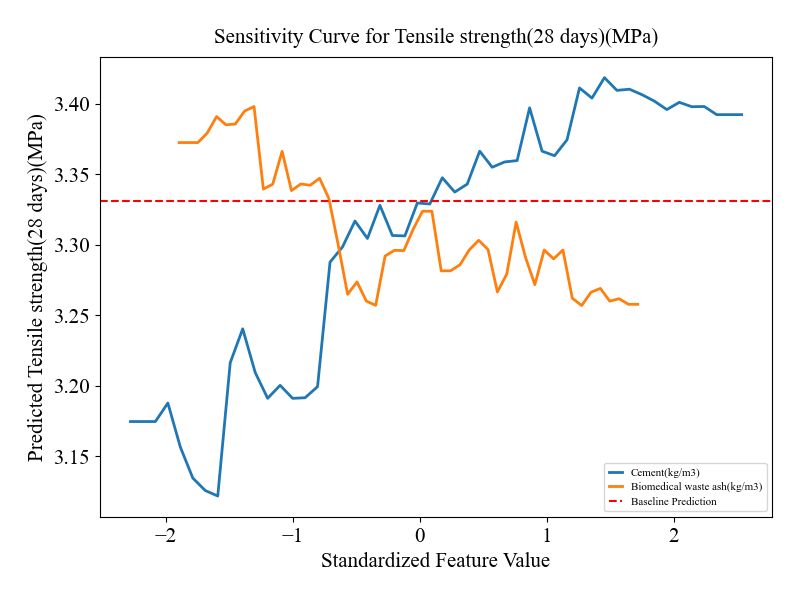
**Figure 11: Error Analysis for all the Models**

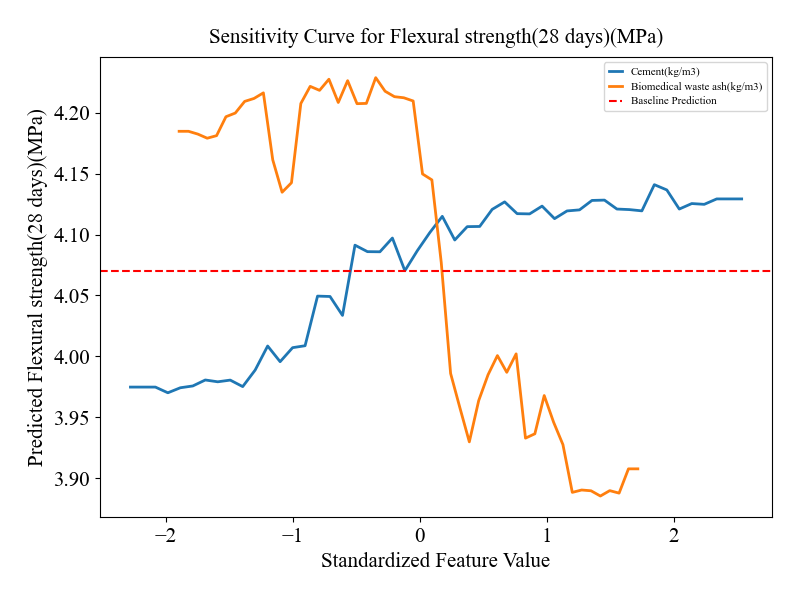
* 1. *Sensitivity analysis*

The first plot shows the sensitivity of compressive strength (28 days) to standardized changes in cement and BMWA content. As cement increases (blue line), compressive strength gradually improves, indicating cement contributes positively to strength gain. In contrast, increasing BMWA (orange line) initially maintains higher compressive strength but then causes a sharp drop once the standardized value crosses zero. This suggests that while moderate BMWA replacement may be tolerable, higher replacement levels substantially reduce compressive strength due to dilution of cementitious material.

The second and third plots represent tensile strength (28 days) and flexural strength (28 days), respectively. In both cases, cement content shows a steady positive effect, gradually increasing predicted strengths. BMWA content again demonstrates a nonlinear influence—showing fluctuations and eventually declining performance at higher replacement ratios. The baseline prediction (red dashed line) serves as a reference, confirming that BMWA beyond a certain threshold lowers both tensile and flexural performance. Together, these sensitivity curves emphasize the dual role of BMWA: while it can contribute as a sustainable partial replacement, excessive amounts impair mechanical properties critical for structural applications.







**Figure 12: Sensitivity curve for CS,TS and FS**

* 1. *SHAP Analysis*

The SHAP bar plot illustrates the feature importance and contribution direction for the predictive model trained on the Biomedical Waste Ash (BMWA) concrete dataset. The x-axis represents SHAP values, which measure the magnitude and direction of each feature’s impact on the model’s predictions. Positive SHAP values (blue bars) indicate that the feature increases the prediction, while negative values (red bars) reduce it. From the chart, BMWA(%) is by far the most influential feature, contributing strongly and positively to the prediction (+2.02). This means that the proportion of BMWA plays a dominant role in shaping the model’s output compared to all other features.

Other features, such as CS(14 days)(MPa) and BMWA(kg/m³), also show moderate positive contributions, suggesting that early-age compressive strength and BMWA dosage positively influence the model’s performance. On the other hand, features like FS(14 days)(MPa) and FS(28 days)(MPa) contribute negatively, implying that higher values of flexural strength at these stages slightly lower the model’s predictions. Interestingly, cement content shows only a negligible effect, emphasizing that the replacement ratio (BMWA%) rather than absolute cement quantity governs the model’s behaviour. Collectively, this SHAP analysis highlights that while BMWA drives the model’s predictions, strength properties at different curing ages modulate the outcome, reflecting the interplay between material composition and mechanical performance.

A blue bar with white text

AI-generated content may be incorrect.

**Figure 13: SHAP Analysis**

* + 1. *SHAP Value Analysis*

The figure correlation matrix provides a comprehensive view of the inter-feature relationships within the dataset. A very strong positive correlation (0.96) is observed between Biomedical Waste Ash (%) and Biomedical Waste Ash (kg/m3), which is expected as both represent related measures of the same material. In contrast, a strong negative correlation (-0.93) is seen between Biomedical Waste Ash (%) and Cement content, indicating that as the replacement level of BMWA is increased, the cement content is decreased. Furthermore, a high degree of correlation (0.78–0.85) is also noted among the various compressive strength measures across different curing ages, which demonstrates the consistency of the strength development in the concrete mixes.

The interdependence of the mechanical properties is also revealed, as tensile and flexural strengths are shown to have moderate positive correlations (0.57–0.73) with compressive strength. It is also observed that early-age tensile strength (7 days) has a weaker correlation with other strength measures, which suggests greater variability and sensitivity during the initial curing period. A significant finding is the consistent negative correlation between BMWA (%) and all of the strength parameters, which confirms that an increase in BMWA content leads to a reduction in concrete strength. Overall, this matrix effectively illustrates the trade-off that is made between cement replacement and mechanical performance.

A graph with a line of dots

AI-generated content may be incorrect.

**Figure 14: SHAP value plot**

* 1. *Taylor Analysis*

The plot illustrates the results of a Rapid Chloride Permeability Test (RCPT) on concrete containing varying percentages of Biomedical Waste Ash (BMWA). It is shown that as the amount of BMWA in the mix increases, the concrete's permeability to chloride ions significantly decreases. The highest permeability is exhibited by the control mix (0% BMWA), while a marked improvement in durability is shown by the mixes with BMWA.

The optimal result, which signifies the greatest resistance to chloride ingress, is observed at a 15% BMWA content, where the permeability is classified as "Very Low." Beyond this point, a slight increase in permeability is seen, but the values remain significantly lower than the control sample. This confirms that incorporating BMWA into the mix generally enhances the concrete's resistance to chloride attack, a critical finding for potential applications.

A diagram with different colored dots

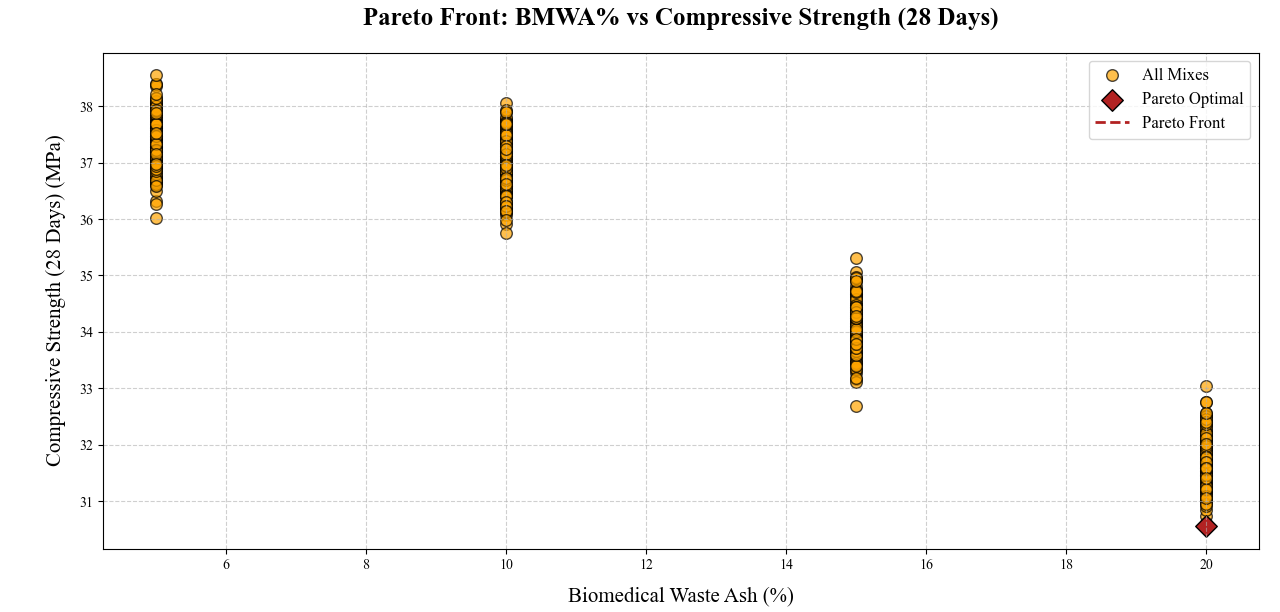
AI-generated content may be incorrect.

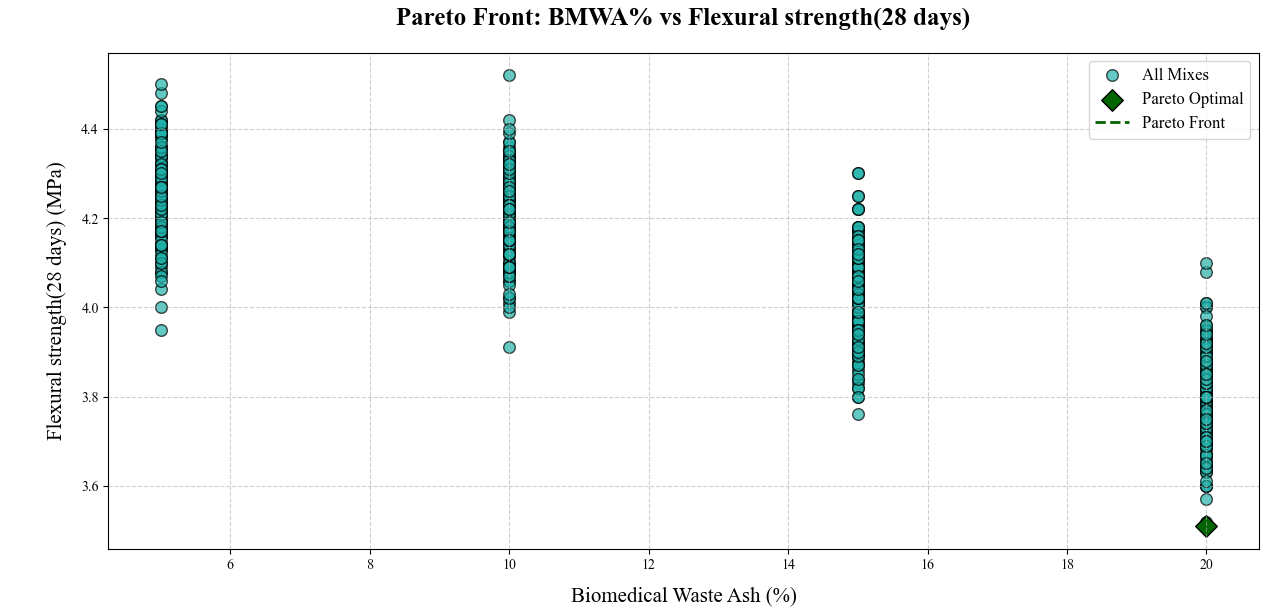
**Figure 15: Taylor Plot for Train and Test values for all algorithms**

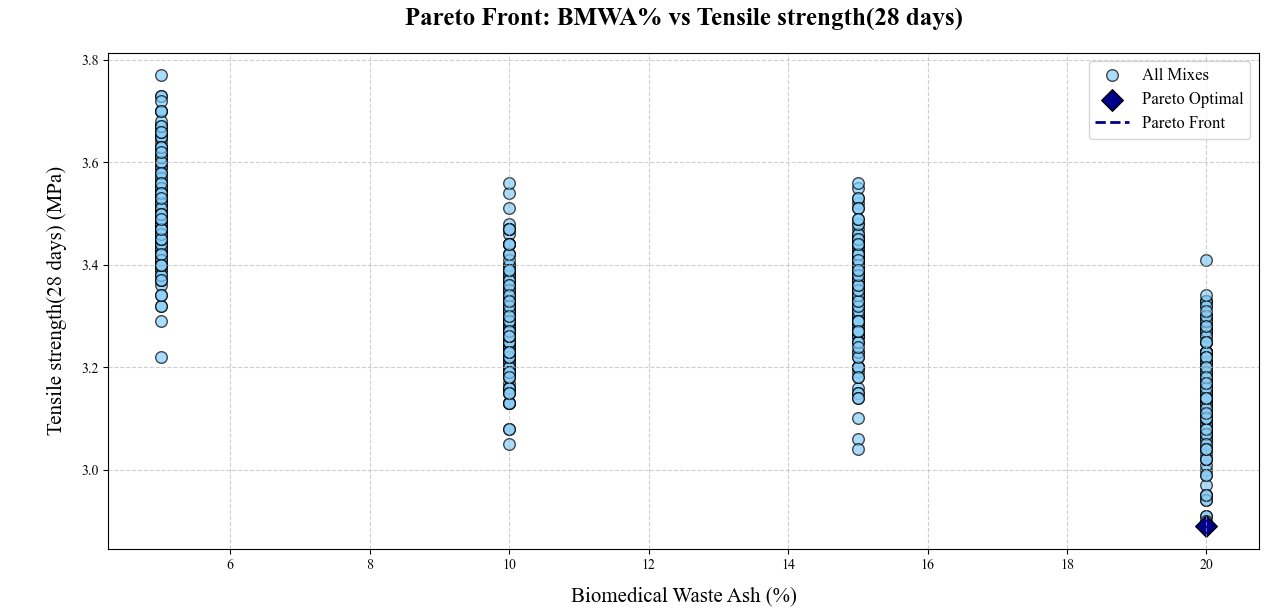
* 1. Pareto Front Plot

The Pareto Front plots present a comprehensive visualization of the trade-offs between **Biomedical Waste Ash Percentage (BMWA%)** and key mechanical properties of concrete—**Compressive Strength (CS)**, **Flexural Strength (FS)**, and **Tensile Strength (TS)**—evaluated at 28 days. Each plot shows a cloud of all experimental mix designs in orange, representing the performance of various compositions. The red dashed line highlights the Pareto Front, which includes only the non-dominated solutions offering the best balance between maximizing strength properties and increasing the sustainable use of Biomedical Waste Ash. The red diamond markers represent the optimal mixes where any further increase in BMWA% would result in a significant reduction in mechanical performance. These visualizations allow researchers to easily identify optimal mix proportions for sustainable concrete design without heavily compromising structural strength.

From the analysis, it is observed that as the **BMWA% increases, the mechanical strength properties tend to decrease**, demonstrating the inherent trade-off in using alternative materials. For **Compressive Strength (CS)**, a relatively moderate drop is seen, suggesting certain mixes that maintain acceptable strength levels even at higher BMWA%. In contrast, the **Flexural Strength (FS)** and **Tensile Strength (TS)** plots show a more scattered trend, indicating higher sensitivity to BMWA%. These insights emphasize the importance of multi-objective optimization in concrete mix design, where sustainability and performance must be balanced. The Pareto Front serves as a useful guideline for selecting mix designs that are both environmentally friendly and structurally effective.

**

**

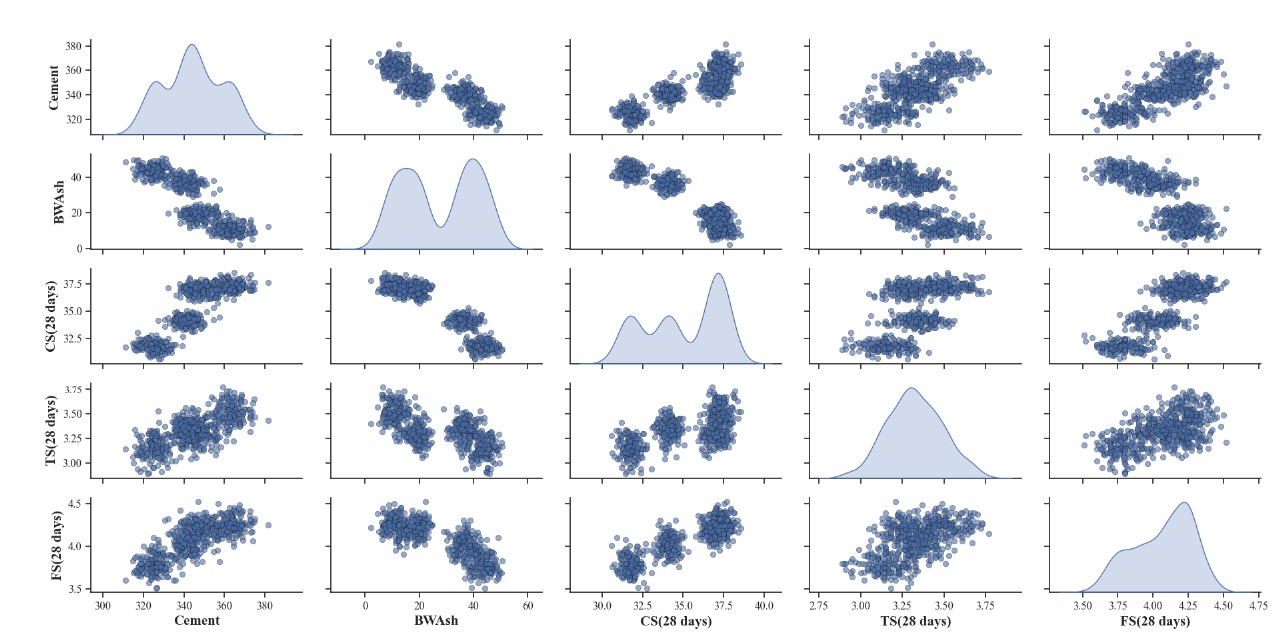
**

**Figure 16: Pareto Front Plot for CS, FS and TS respectively**

* 1. *Plague Plot*

The Plague plot visualizes the relationships between five variables: Cement, BWAsh, Compressive Strength (CS at 28 days), Tensile Strength (TS at 28 days), and Flexural Strength (FS at 28 days). The diagonal elements show the distribution of each variable, while the scatter plots in the off-diagonal elements represent pairwise correlations between them. From the diagonal, it is clear that Cement and CS(28 days) display multimodal distributions, suggesting that the dataset may contain distinct groups or clusters of mixes. On the other hand, BWAsh has a left-skewed distribution, while TS and FS appear more normally distributed.

Looking at the scatter plots, strong positive correlations are visible between Cement and the strength parameters (CS, TS, FS), which is expected since higher cement content typically enhances concrete strength. In contrast, BWAsh shows negative correlations with all strength parameters, implying that higher amounts of BWAsh reduce strength values. The scatter plots also reveal clustering, suggesting the data may represent different mix designs or experimental batches. Overall, the pair plot highlights the opposing effects of Cement (strength enhancer) and BWAsh (strength reducer) on mechanical properties of concrete.



**Figure 17: Plague Plot**

* 1. *External Validation*

External Validation is a critical process in model evaluation where a trained predictive model is tested on completely independent data that was not used during the model development or training phase. This step ensures the model’s ability to generalize well to unseen data and helps detect overfitting. Unlike internal validation methods such as cross-validation, which split the available dataset into training and validation sets, external validation involves using a distinct dataset collected from different sources, experimental conditions, or time periods. This provides a more realistic assessment of the model’s performance in real-world applications, especially in engineering fields like concrete strength prediction, where material variability and external factors play significant roles.

For instance, in a dataset containing inputs such as Cement (kg/m³), Biomedical Waste Ash (kg/m³), Fine Aggregate (kg/m³), and Coarse Aggregate (kg/m³), an external validation process would use fresh input combinations not seen during model training to predict important output properties like Compressive Strength, Tensile Strength, and Flexural Strength at 28 days. The predicted values (e.g., Compressive Strength around 31.84 MPa for various combinations) are compared against experimentally measured values from a separate test campaign. A close agreement between predicted and real-world data demonstrates the model's robustness and reliability for future engineering decisions, such as optimizing concrete mix designs under variable material sources.

**Table 5: External Validation Results**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cement(kg/m3)** | **Biomedical waste ash(kg/m3)** | **Fine aggregate**  **(kg/m3)** | **Coarse aggregate**  **(kg/m3)** | **Predicted CS(28 days)(MPa)** | **Predicted TS(28 days)(MPa)** | **Predicted FS(28 days)(MPa)** |
| 312.3620357 | 2.05844943 | 636 | 1147 | 31.84099007 | 3.12769556 | 3.769400597 |
| 485.2142919 | 96.99098522 | 636 | 1147 | 31.80321503 | 3.130996704 | 3.766886473 |
| 419.5981825 | 83.24426408 | 636 | 1147 | 31.80994225 | 3.130326986 | 3.767484188 |
| 379.5975453 | 21.23391107 | 636 | 1147 | 31.82564354 | 3.12896657 | 3.768536329 |
| 246.8055921 | 18.18249672 | 636 | 1147 | 31.85747147 | 3.126358032 | 3.77013731 |
| 246.7983561 | 18.34045099 | 636 | 1147 | 31.8574276 | 3.126361609 | 3.770135403 |
| 217.4250837 | 30.4242243 | 636 | 1147 | 31.86605263 | 3.125659227 | 3.770483494 |
| 459.8528437 | 52.47564316 | 636 | 1147 | 31.81093788 | 3.13024497 | 3.767544746 |
| 380.3345035 | 43.19450186 | 636 | 1147 | 31.8216095 | 3.129292488 | 3.76829648 |
| 412.4217733 | 29.12291402 | 636 | 1147 | 31.81981659 | 3.129446983 | 3.768177032 |

1. **CONCLUSION**

**REFERENCES**

GUI Link: