**Analysis and Prediction of 28-Day Tensile Strength Using Artificial Neural Networks1.**

**1. Introduction**  
Tensile strength is a fundamental property in material science, indicating a material’s resistance to failure under tension (Ashby & Jones, 2018). Accurate prediction of tensile strength is crucial in evaluating structural performance in construction and civil engineering. This study analyzes tensile strength development over time—specifically at 7, 14, 21, and 28 days—to predict the 28-day tensile strength using an ANN. The project supports ongoing research in material behavior and offers a baseline machine learning workflow for similar datasets.

**2. Dataset Overview**

**2.1 Dataset Description**  
The dataset comprises 180 samples, it includes:

* Sample Type (categorical)
* Tensile Load and Tensile Strength at 7, 14, 21, and 28 days

There are no missing values. The data distribution is consistent across samples and time points.

**2.2** **Statistical Summary of Tensile Strength (N/mm²)**

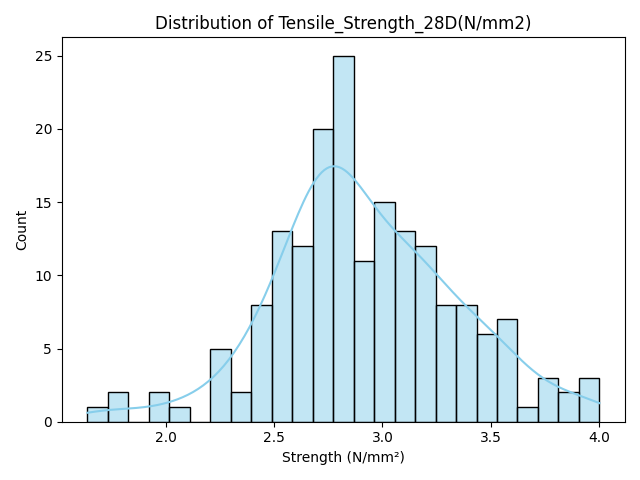
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **7 Days** | **14 Days** | **21 Days** | **28 Days** |
| Count | 180 | 180 | 180 | 180 |
| Mean | 2.76 | 2.82 | 3.11 | 2.92 |
| Std. Dev. | 0.25 | 0.33 | 0.44 | 0.42 |
| Min | 2.08 | 1.76 | 2.03 | 1.64 |
| 25th Percentile | 2.60 | 2.66 | 2.76 | 2.66 |
| Median | 2.78 | 2.81 | 3.12 | 2.85 |
| 75th Percentile | 2.93 | 3.02 | 3.45 | 3.20 |
| Max | 3.44 | 3.62 | 4.01 | 4.00 |

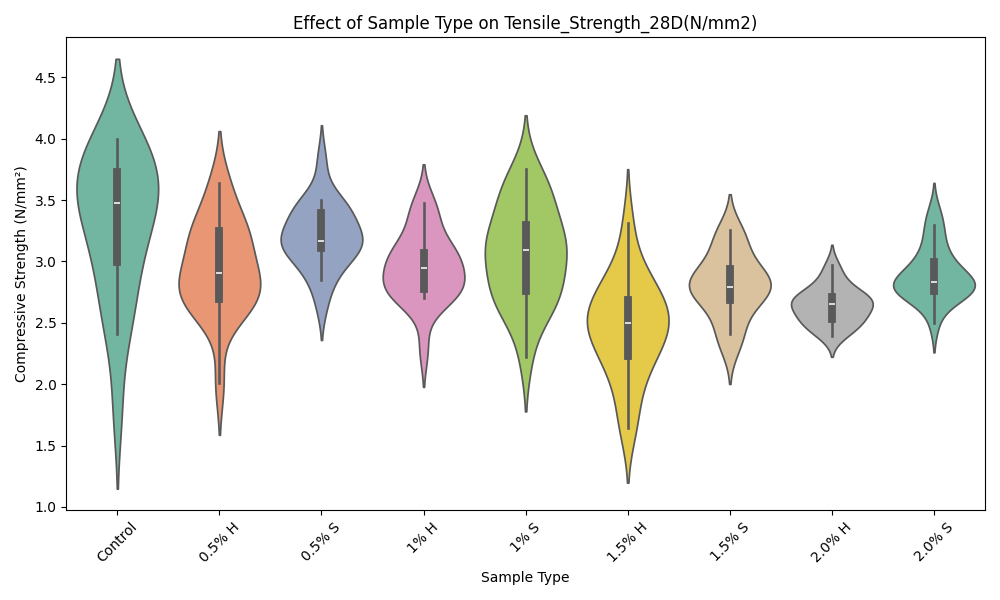
**3. Exploratory Data Analysis (EDA)**  
EDA was instrumental in understanding the distribution and variability of tensile strength values:

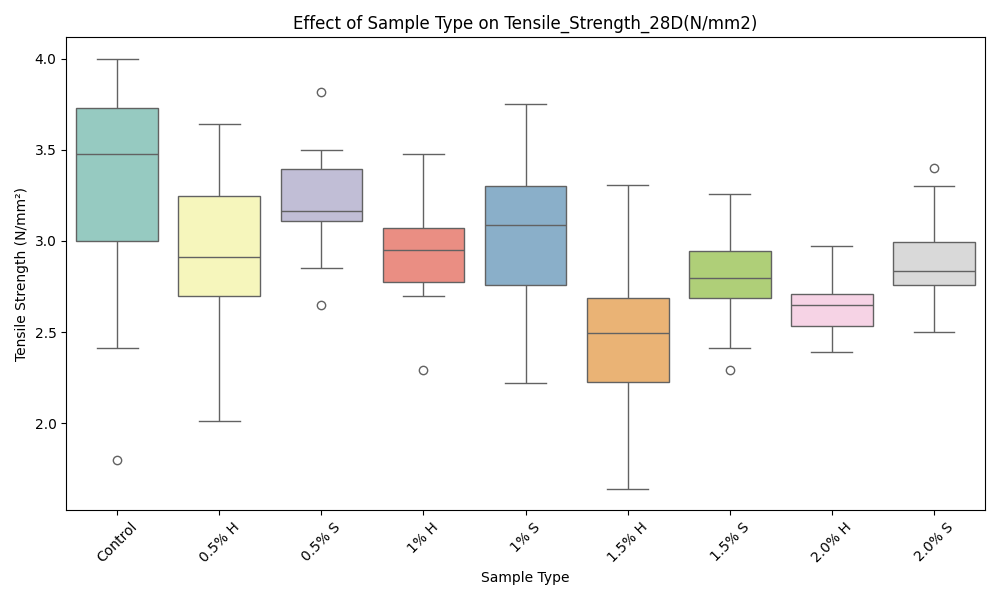
* **Histograms**: Tensile strength at all time points approximated a normal distribution, with the 28-day values centered around ~2.92 N/mm².
* **Boxplots**: These visualizations revealed outliers, particularly at 21 and 28 days, indicating potential variability in the late-age measurements.
* **Violin Plots**: 28-day tensile strength was plotted by sample type to assess intra-group spread and density.
* **Correlation Matrix**: A Pearson correlation map was generated to quantify the linear relationship between tensile strength values across all time intervals.

*Note*: While plots were generated for all time points (7D–28D), this report includes only 28-day visuals to maintain clarity and focus on the prediction target.

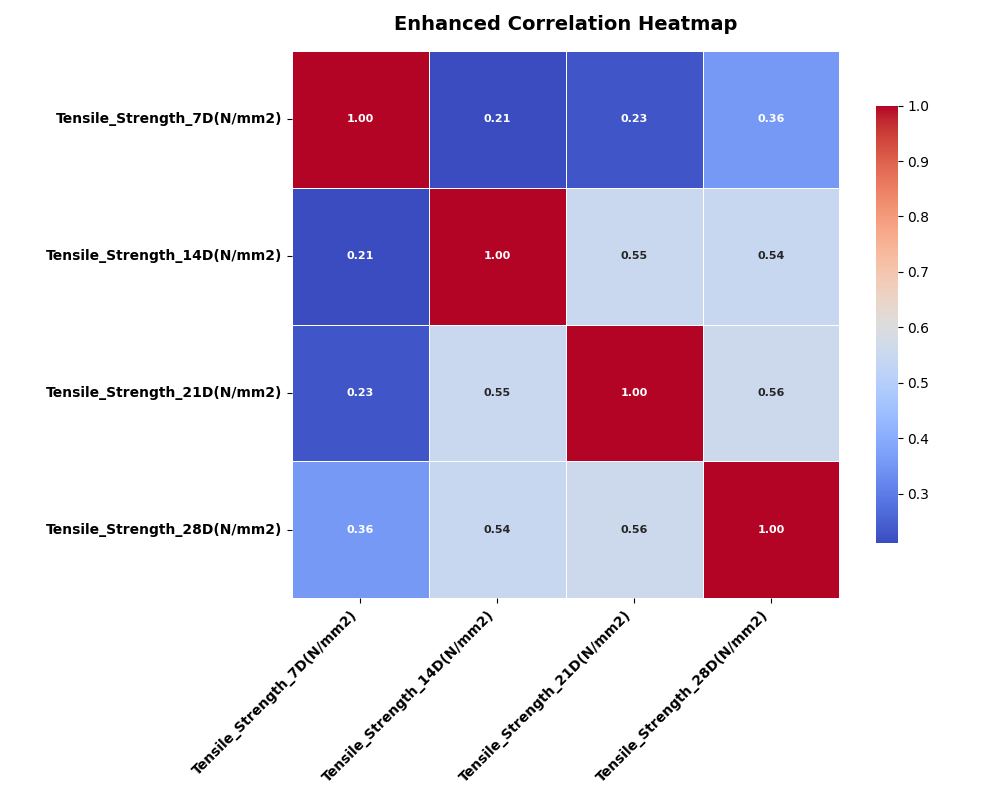
**Figures**:

**Figure 1**: Histogram of 28-Day Tensile Strength

 **Figure 2**: Violin Plot of 28-Day Strength by Sample Type



**Figure 3**: Boxplot of 28-Day Strength (Outliers Highlighted)

 **Figure 4**: Correlation Heatmap of Tensile Strength Measurements

**4. Data Preprocessing**

To prepare the data for modeling:

* **Reordering Columns**: Columns were reordered to prioritize *Sample Type* and group load and strength measurements.
* **Handling Multicollinearity**: Tensile load and strength features showed near-perfect correlations (≈0.9999) (Kutner et al., 2005), leading to the removal of load columns to avoid multicollinearity, which can destabilize models (Hair et al., 2019).
* **One-Hot Encoding**: The categorical *Sample Type* was encoded into dummy variables (e.g., *Sample Type\_Control*).
* **Outlier Handling**: Numeric features were capped using the Interquartile Range (IQR) method (1.5 × IQR bounds).
* **Scaling**: Features were standardized using *RobustScaler* to mitigate outlier effects.
* **Data Splitting**: The dataset was split into 80% training (144 samples) and 20% testing (36 samples) sets, with 14 features after preprocessing.

**5. Model Development: Artificial Neural Network (ANN)**

**5.1 Architecture**

* Input Layer: 14 features
* Hidden Layers: 2 layers × 32 neurons (ReLU), Batch Normalization, Dropout (20%)
* Output Layer: Single neuron with linear activation

**5.2 Training Parameters**

* Loss Function: Mean Squared Error (MSE)
* Optimizer: Adam (learning rate = 0.001)
* Epochs: 200
* Batch Size: 16
* Validation Split: 15%
* EarlyStopping: Patience = 15

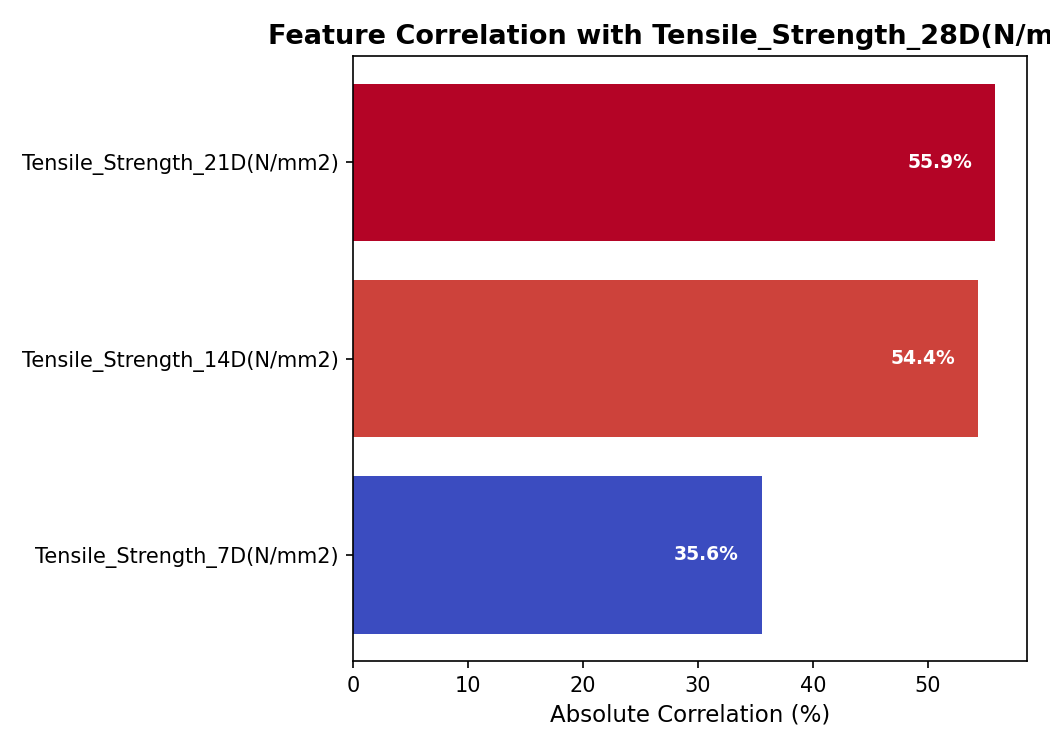
**6. Model Evaluation**

**6.1 Performance Metrics**

|  |  |
| --- | --- |
| **Metric** | **Value** |
| MSE | 0.3855 |
| R² Score | 0.8785 |

This indicates that approximately 88% of the variance in 28-day compressive strength is explained by the model.

**6.2 Feature Importance**  
Permutation-based feature importance identified 21-day tensile strength as the most influential predictor. Earlier time points and sample type features also contributed but to a lesser extent.

****

**Figure 5**: Feature Importance Bar Chart

**7. Interpretation and Insights**

* **High R² Value (0.8785):** The model demonstrates strong explanatory power, indicating that approximately 87.85% of the variance in the 28-day compressive strength is accounted for by the input features. This suggests a reliable fit between the model and the observed data.
* **Low MSE (0.3855):** The mean squared error indicates a relatively low average squared difference between the predicted and actual compressive strength values. This reflects a good level of prediction accuracy.
* **Target Value Distribution:** Although the model performs well, it is important to note that the target variable (compressive strength) falls within a moderately narrow range. This can cause small absolute errors to have a larger proportional impact, which may affect model sensitivity.

**8. Personal Reflection**  
This project reinforced the need for careful feature selection and highlighted the limitations of applying complex models to small datasets. While ANNs are powerful, their success depends on data richness and diversity. The project provided valuable experience in data preprocessing, visualization, and neural network modeling within the context of civil engineering.

**9. Conclusion and Recommendations**  
The artificial neural network (ANN) model developed to predict 28-day compressive strength demonstrated strong predictive performance, with a high coefficient of determination (R² = 0.8785) and a low mean squared error (MSE = 0.3855). These metrics indicate that the model effectively captures the relationship between the input features and the target variable, making it a valuable tool for preliminary strength estimation.

Despite the promising results, the model's generalizability could be improved by incorporating additional influential variables such as environmental conditions (e.g., temperature, humidity) and material-specific properties. Future work should consider expanding the dataset and applying cross-validation or ensemble methods to further validate and enhance model robustness.

Overall, this ANN model provides a solid foundation for predicting compressive strength and can be refined and extended in subsequent studies for more comprehensive applications in concrete mix design and quality control.

**Recommendations**:

* Include environmental and material properties (e.g., humidity, curing conditions)
* Try simpler models like Random Forest or Gradient Boosting
* Expand dataset for better generalizability

**References**

* Ashby, M. F., & Jones, D. R. H. (2018). *Engineering materials 1: An introduction to properties, applications, and design* (5th ed.). Butterworth-Heinemann.
* Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
* Chollet, F. (2015). *Keras: Deep learning library for Theano and TensorFlow*. [https://keras.io](https://keras.io/)
* Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
* Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
* Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
* Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
* Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.
* Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI*, *14*(2), 1137–1145.
* Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
* Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill.
* Montgomery, D. C., & Runger, G. C. (2012). *Applied statistics and probability for engineers* (5th ed.). Wiley.
* Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
* Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.