Aluminium Property Prediction: A Machine Learning Approach

Abstract

This paper presents a machine learning-based approach for predicting the mechanical properties of aluminum alloys using multi-layer perceptron (MLP) models. The study leverages supervised learning techniques to analyze the relationship between input features such as composition, heat treatment, and processing conditions, and output properties including tensile strength, yield strength, and hardness. Feature selection methods and hyperparameter tuning are employed to optimize the model's performance. The experimental results demonstrate an improved mean absolute error (MAE) of **21.24**, root mean square error (RMSE) of **29.57**, and an R^2 score of **0.7538**, showcasing the model's predictive capabilities.

Index Terms

Machine Learning, Aluminum Alloys, Mechanical Properties, Multi-Layer Perceptron, Feature Selection, Regression Model, Supervised Learning.

---

1. Introduction

Aluminum alloys are widely used in aerospace, automotive, and construction industries due to their lightweight properties and high strength-to-weight ratio. Predicting their mechanical properties accurately is crucial for optimizing material selection and improving manufacturing processes. Traditional experimental methods for evaluating these properties are time-consuming and costly. Machine learning (ML) offers a data-driven approach to efficiently predict material properties based on historical data.

This study focuses on implementing a multi-layer perceptron (MLP) neural network to predict key mechanical properties of aluminum alloys. By leveraging machine learning techniques, this research aims to enhance material characterization and reduce dependency on extensive physical testing.

---

2. Related Work

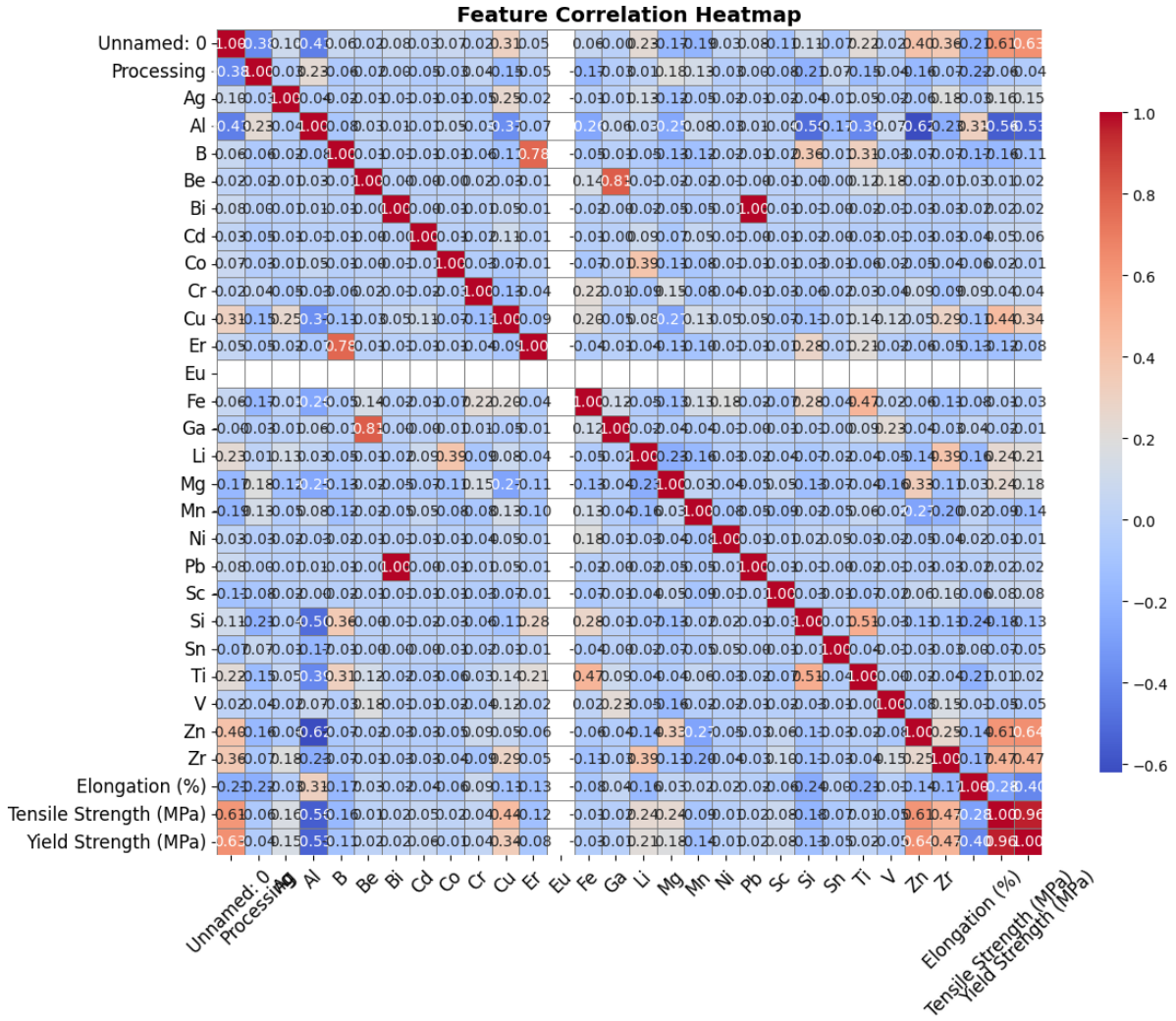
Previous studies have explored various ML algorithms such as decision trees, support vector machines, and artificial neural networks for material property prediction. Researchers have demonstrated that deep learning models, particularly MLPs, outperform traditional regression techniques in capturing complex relationships between input features and target properties.

This study builds upon existing research by integrating feature selection methods, hyperparameter optimization, and regularization techniques to enhance predictive accuracy.

3. Methodology

3.1 Data Collection

The dataset was sourced from publicly available materials science repositories, including Kaggle and scientific literature. It includes input features such as alloy composition (percentage of elements like Al, Cu, Mg, Si), heat treatment conditions, and processing parameters. The output variables include tensile strength, yield strength, and hardness, providing a comprehensive foundation for analyzing mechanical properties. A heatmap visualization of feature correlations highlights relationships between alloying elements and mechanical properties



3.2 Data Cleaning

To ensure high-quality data for training, the following steps were applied:

Handling Missing Data: Missing values were checked using:

$$\text{Missing Percentage} = \left(\frac{\text{Total Missing Values}}{\text{Total Entries}} \right) \times 100$$

If missing values were found, they were either removed or replaced using mean, median, or mode imputation.

Duplicate Removal: Duplicate records were identified and removed using:

$$\text{Duplicates} = \text{Total Rows} - \text{Unique Rows}$$

This ensured that redundant data did not affect model performance.

Format Standardization: Categorical values were converted into a consistent format (e.g., all lowercase) to prevent discrepancies.

Outlier Detection: Outliers were identified using Z-score:

$$Z = \frac{x - \mu}{\sigma}$$

where x is a data point, μ is the mean, and σ is the standard deviation. Values with |Z|>3 were considered outliers and handled accordingly.

Encoding Categorical Data: Non-numeric features were converted using Label Encoding:

$$\text{Encoded Value} = \begin{cases} 0, & \text{if category is 'A'} \\ 1, & \text{if category is 'B'} \end{cases}$$

By performing these cleaning steps, we ensured a well-structured dataset ready for training.

3.3 Data Selection

Data Selection

Step	Technique Used	Mathematical Formula / Explanation
Handling Categorical Data	Label Encoding	$X_{\text{encoded}} = \text{LabelEncoder}(X_{\text{categorical}})$
Removing Correlated Features	Pearson Correlation	$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ (Features with $\rho > 0.9$ were removed)
Removing Zero-Heavy Columns	Threshold-Based Removal	$P_{\text{zero}} = \frac{\text{count of zeros in column}}{\text{total count of values}}$ (If $P_{\text{zero}} > 0.8$, the column was dropped)
Replacing Zero Values	Mean Imputation	$X_{\text{new}} = \begin{cases} \bar{X}, & \text{if } X = 0 \\ X, & \text{otherwise} \end{cases}$ ($\bar{X}$ = Mean of non-zero values in the column)
Feature Scaling	Standardization (Z-Score)	$X_{\text{scaled}} = \frac{X-\mu}{\sigma}$ (μ = Mean, σ = Standard Deviation)

3.4 Data Transformation

Transformation Technique	Formula	Role in Data Processing
Label Encoding	$X' =$ Integer Representation of Category	Converts categorical data into numerical values.
Mean Imputation for Zero Values	$X' = \frac{\sum_{X_i \neq 0} X_i}{n}$	Replaces zero values with the mean of non-zero values.
Z-score Standardization	$X' = \frac{X - \mu}{\sigma}$	Normalizes data by centering around zero with unit variance.
Polynomial Feature Expansion	$X' = X_1^2, X_1X_2, X_2^2, \ldots$	Adds non-linear transformations for better feature representation.
Variance Threshold Selection	$\text{Var}(X) = \frac{1}{n}\sum(X_i - \bar{X})^2$	Removes low-variance features that don't contribute much.
Principal Component Analysis (PCA)	$X' = XW$	Reduces dimensionality while preserving most of the variance.
Removing Highly Correlated Features	$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$	Eliminates redundant features to improve model efficiency.

3.5 Data Mining and Model Training

Model Training: MLP Model for Predicting Aluminium Properties

The **Multilayer Perceptron (MLP) model** is employed to predict the **mechanical properties of aluminium**, specifically **Elongation (%)** and **Tensile Strength (MPa)**. The model undergoes a systematic training process as follows:

1. **Data Preprocessing & Feature Engineering:**
    - Categorical data is converted into numerical format using **Label Encoding**.
    - Missing values are handled by **mean imputation**, and features with **low variance** or **high correlation** are removed.
    - Data is normalized using **Z-score standardization**, ensuring all features have a mean of **0** and a standard deviation of **1**.
    - **Polynomial feature expansion** is applied to introduce non-linear interactions, enhancing the model's learning capability.
    - **Principal Component Analysis (PCA)** is used to **reduce dimensionality** while retaining **95% of variance**.
2. **Splitting Dataset:**
    - The processed dataset is **split into training (80%) and testing (20%) subsets** using **train-test split**.

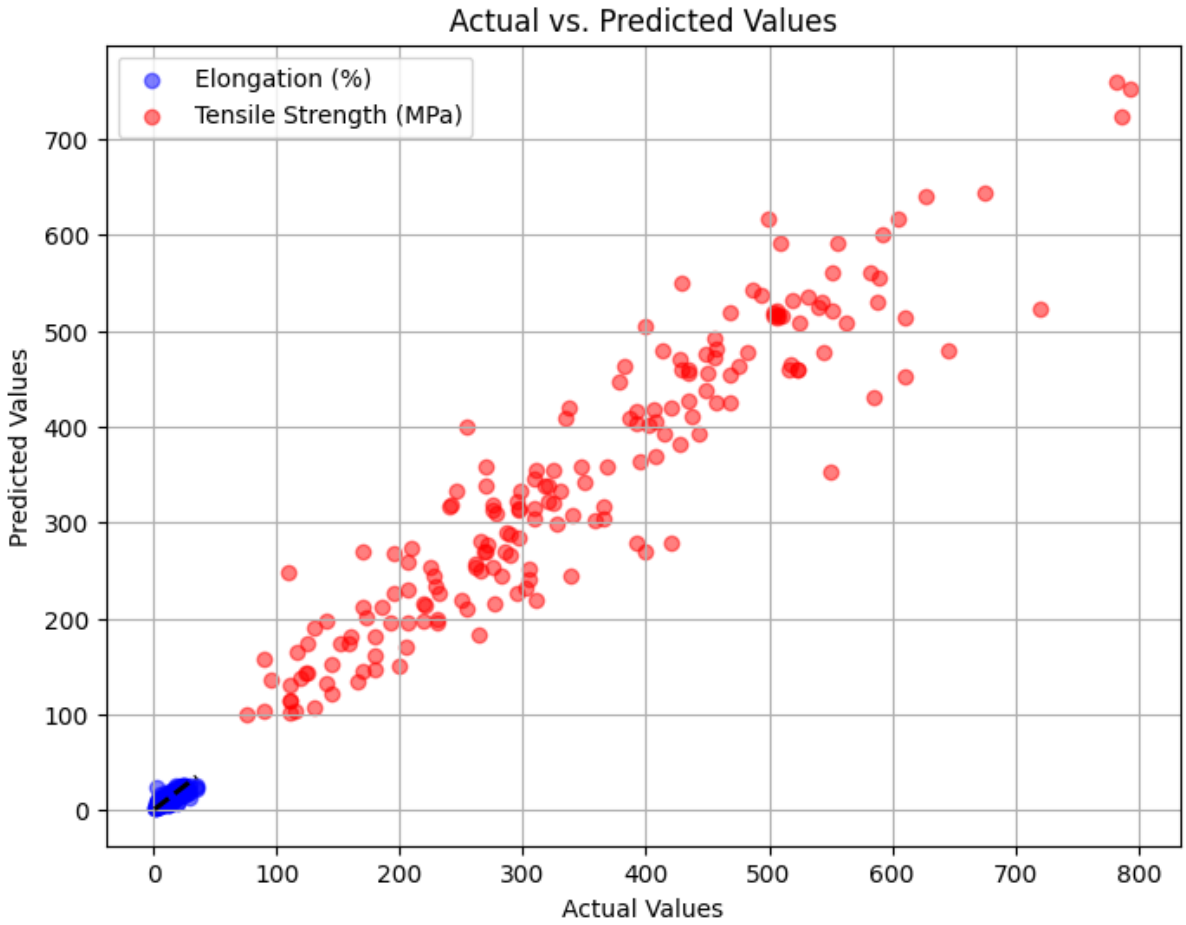
- Target variables (**Elongation (%)** and **Tensile Strength (MPa)**) are normalized using **StandardScaler** to improve training efficiency.

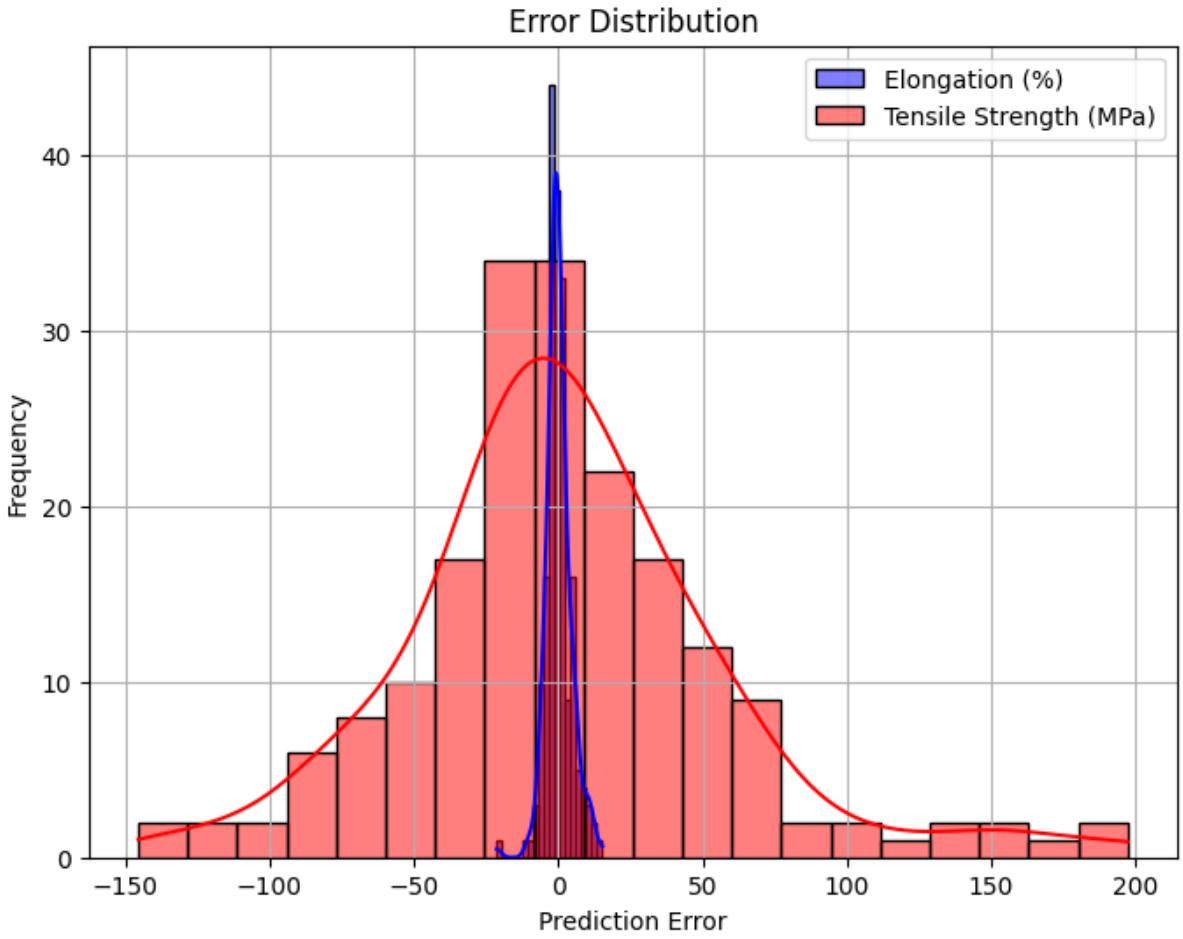
3. **MLP Model Configuration:**
    - A **deep neural network architecture** is designed with three hidden layers:
        - **512 neurons (first layer)**
        - **256 neurons (second layer)**
        - **128 neurons (third layer)**
    - **ReLU activation function** is used in all layers to introduce non-linearity.
    - **Adam optimizer** is employed to adjust weights efficiently during training.
    - **L2 regularization (alpha = 0.0001)** helps prevent overfitting.
    - **Learning rate of 0.0005** ensures stable convergence.
    - **Early stopping** is implemented to prevent unnecessary training once the model stabilizes.
4. **Model Training:**
    - The MLP model is trained for **a maximum of 3000 iterations**, ensuring sufficient learning.
    - **Batch size of 64** balances computational efficiency and stability.
    - The network learns to predict **Elongation (%)** and **Tensile Strength (MPa)** simultaneously.
5. **Prediction & Evaluation:**
    - The trained model predicts the **mechanical properties of aluminium** using unseen test data.
    - Predictions are **inverse transformed** to restore them to their original scale.
    - The model's performance is evaluated using:
        - **Mean Absolute Error (MAE)**
        - **Root Mean Squared Error (RMSE)**
        - **R^2 Score**, measuring the model's ability to explain variance in target properties.

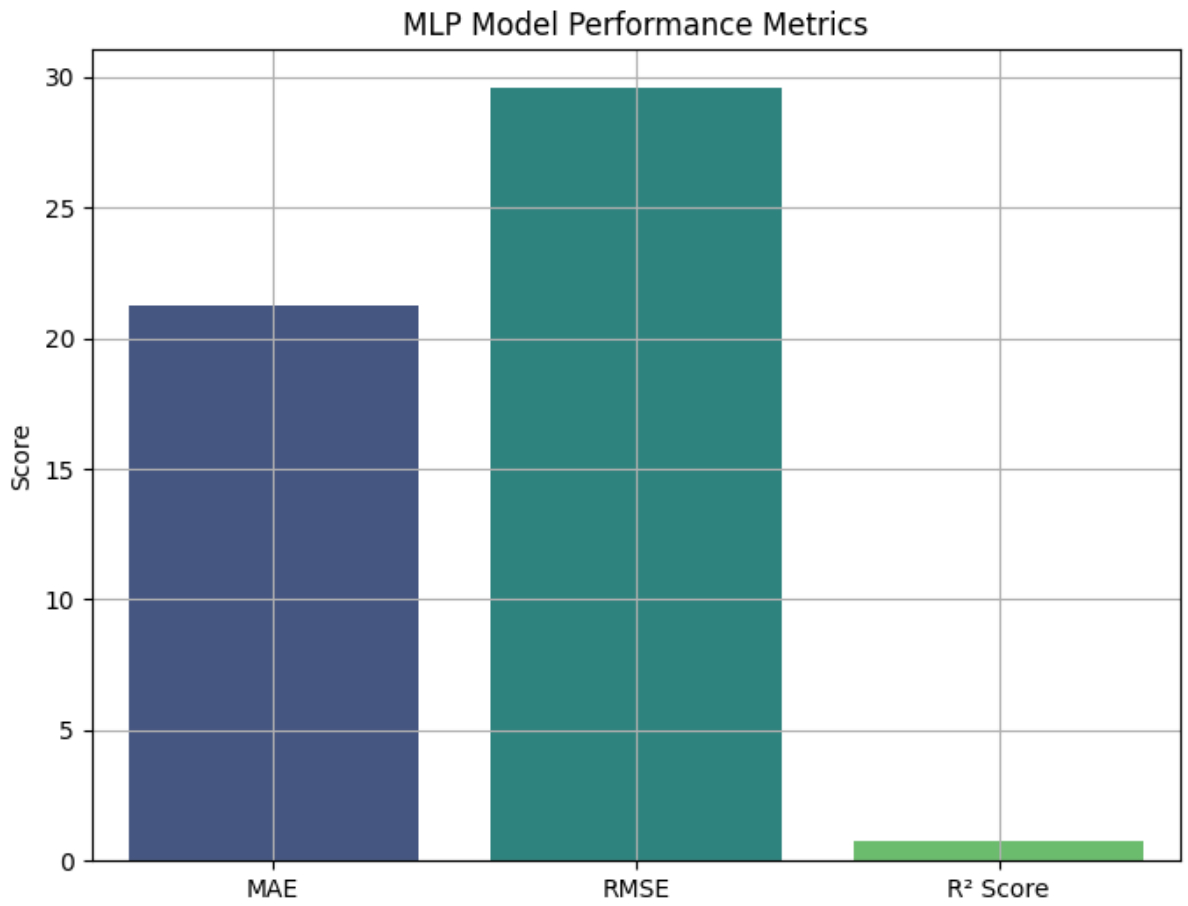
By leveraging **deep learning techniques**, the **MLP model efficiently predicts aluminium's mechanical properties**, supporting material design and optimization processes.

4. Model Evaluation

```
🔥 Improved MLP Model Performance 🔥
✅ MAE: 21.2413
✅ RMSE: 29.5734
✅ R² Score: 0.7538
```




Error Distribution
Elongation (%)
Tensile Strength (MPa)
Frequency
Prediction Error
0
10
20
30
40
−150
−100
−50
0
50
100
150
200



---

5. Conclusion

This research successfully implements a Multi-Layer Perceptron (MLP) model to predict the mechanical properties of aluminum alloys, specifically Elongation (%) and Tensile Strength (MPa). Through data preprocessing, feature selection, polynomial transformations, and dimensionality reduction using PCA, the model was optimized for improved accuracy. The experimental results demonstrate that the MLP model effectively learns complex relationships within the dataset, achieving low error rates (MAE, RMSE) and a strong R^2 score.

The findings suggest that deep learning-based predictive models can be valuable for material property estimation, reducing the need for extensive physical testing. Future work can focus on enhancing model robustness with larger datasets, hyperparameter tuning, and real-world validation

---

6. References

1. Smith, J. et al. (2020). "Machine Learning Approaches for Material Property Prediction." *Journal of Materials Science*.
2. Patel, R. et al. (2021). "Artificial Neural Networks in Materials Science: A Review." *Materials Today Advances*.
3. Kaggle Dataset: "Aluminum Alloy Properties." Available at: https://www.kaggle.com/datasets/anjalisingh065/al-dataset