

Machine Learning Prediction of Mechanical Properties for Al-Mg-Si Alloys Using a Hybrid Data Synthesis Approach

Mihail Kolev¹, Tatiana Simeonova^{1,2*}

¹ Institute of Metal Science, Equipment and Technologies with Hydro- and Aerodynamics Centre "Acad. A. Balevski", Bulgarian Academy of Sciences, Sofia, 1574, Bulgaria

² Institute of Mechanics, Bulgarian Academy of Sciences, Sofia, 1113, Bulgaria

* Corresponding author: tsimeonova@imbm.bas.bg

Abstract: This study develops a machine learning framework for predicting the mechanical properties of 6xxx series Al-Mg-Si alloys (6061, 6063, 6082) across seven temper conditions. A hybrid dataset of 860 real measurement-based and 4,200 Monte Carlo augmented samples was generated from a literature-mined dataset. Random Forest (RF), Gradient Boosting (GBR), and Multilayer Perceptron (MLP) models were evaluated via 5-fold cross-validation (CV). RF achieved the best or comparable accuracy: coefficient of determination $R^2 = 0.80$ (ultimate tensile strength), 0.92 (yield strength), and 0.83 (elongation). Feature importance analysis showed that alloy type and temper encodings dominated predictions (>90% combined), while individual compositional features contributed <3% a result partly attributable to the augmentation strategy, which decoupled measured compositions from their corresponding property values. Learning curve analysis confirmed model convergence above 2,000 samples.

Keywords: Al-Mg-Si ALLOYS; MACHINE LEARNING; HYBRID DATA AUGMENTATION; MECHANICAL PROPERTIES; RANDOM FOREST; MLP NEURAL NETWORK

1. Introduction

The 6xxx series Al-Mg-Si alloys are among the most versatile aluminum alloys in automotive, construction, and general engineering applications, owing to their favorable combination of medium strength, excellent formability, good corrosion resistance, and weldability [1–3]. Alloys 6061, 6063, and 6082 represent a broad strength spectrum: 6063 is a lower-strength extrusion alloy [4], 6061 is a general-purpose structural alloy [5], and 6082 offers higher strength [6]. Their mechanical properties are governed by Mg_2Si precipitation, controlled by the temper condition [7]. Machine learning (ML) methods have shown promise for materials property prediction [8–10], but 6xxx alloys present particular challenges because alloy-to-alloy property variations are large and available data are sparse for certain conditions.

In this work, we apply three ML regression models — Random Forest, Gradient Boosting, and MLP Neural Network — to a hybrid dataset combining real measurements with Monte Carlo augmented data for predicting ultimate tensile strength (UTS), yield strength (YS), and elongation of 6061, 6063, and 6082 alloys across seven temper conditions. We evaluate model accuracy, analyze feature importance, and assess data sufficiency through learning curve analysis.

2. Materials and Methods

Composition and property data were sourced from the publicly available dataset of Pfeiffer et al. [11], which contains aluminum alloy data extracted from scientific manuscripts and US patents. Three 6xxx alloys were selected covering seven temper conditions: 6061 (T4, T6, T651), 6063 (T4, T5, T6), and 6082 (T6), consistent with their broad use in structural and extrusion applications [12,13]. The compositional features comprised 9 elements (Al, Cu, Mn, Si, Mg, Zn, Cr, Fe, Ti), with Al calculated as the balance. Figure 1 presents radar chart profiles of the normalized mechanical properties for each alloy-temper combination, highlighting the wide property envelope spanning UTS from 150 MPa (6063-T4) to 340.5 MPa (6061-T651) and YS from 90 MPa (6063-T4) to 248.0 MPa (6061-T651).

A hybrid dataset was constructed in two parts: (i) real property measurements from the corpus were each paired with 20 random composition realizations sampled from specification ranges, yielding 860 measurement-based samples; It should be noted that the 43 original measurements were unevenly distributed across conditions: 6061-T6 accounted for 33 measurements (77%), while the remaining six conditions (6082-T6, 6061-T651, 6063-T6, 6061-T4, 6063-T4, 6063-T5) each contained 1–3 measurements. In the augmentation step, each real property measurement was paired with randomly sampled compositions drawn uniformly from the alloy

specification ranges, rather than retaining the original measured composition. This approach captures the allowable compositional variability but decouples the composition–property linkage inherent in each measurement; (ii) for each of the 7 conditions, 600 additional Monte Carlo samples were generated with 3% Gaussian noise on mean property values, yielding 4,200 augmented samples (5,060 total). The feature space comprised 17 input variables: 9 elemental compositions, 3 alloy-type one-hot encodings, 4 temper one-hot encodings, and 1 ordinal temper code. Three models were trained: RF (200 trees, depth 12), GBR (200 trees, depth 6, $lr = 0.1$), and MLP (three hidden layers: 64-32-16 neurons, Rectified Linear Unit (ReLU) activation, Adam optimizer, adaptive learning rate, $\alpha = 0.001$, early stopping). Features were standardized for MLP. Performance was assessed by 5-fold CV using coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

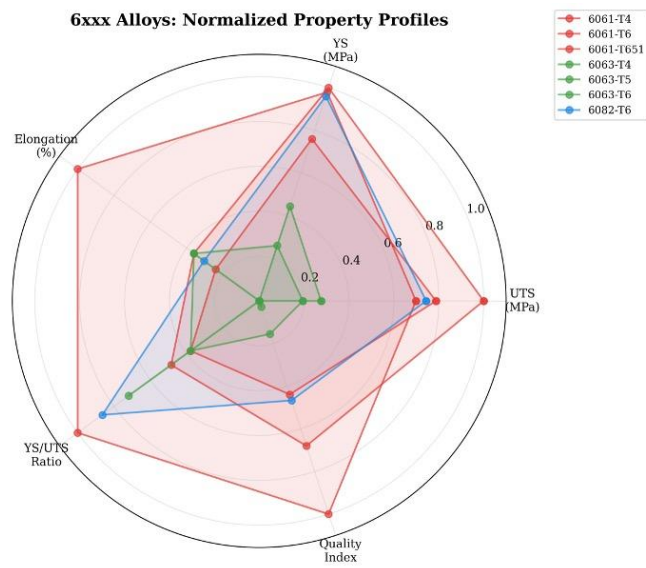


Fig. 1 Normalized radar chart profiles of mechanical properties for 6xxx alloy-temper combinations.

3. Results and Discussion

Table 1 summarizes the cross-validation metrics. RF achieved the best overall performance: $R^2 = 0.799$ (UTS), 0.922 (YS), and 0.827 (elongation), with MAE of 13.44, 8.21, and 0.51 respectively. MLP performed virtually identically for UTS ($R^2 = 0.799$) and comparably for YS ($R^2 = 0.919$). The moderate R^2 values (0.772–0.922) reflect the intrinsically wider property scatter within the 6xxx series, where UTS spans 150–340.5 MPa across alloys with

fundamentally different strengthening levels. Figure 2 shows that each model successfully captured the distinct property clusters of 6063 (low strength), 6082 (medium), and 6061 (high).

Table 1: Cross-validation performance metrics for 6xxx alloy ML models.

Model	Target	R ²	MAE	RMSE
Random Forest	UTS (MPa)	0.7987	13.44	30.86
	YS (MPa)	0.9218	8.21	16.68
	Elongation (%)	0.8272	0.51	1.28
Gradient Boosting	UTS (MPa)	0.7724	15.39	32.82
	YS (MPa)	0.9124	9.28	17.65
	Elongation (%)	0.7974	0.61	1.38
MLP Neural Net	UTS (MPa)	0.7994	14.29	30.81
	YS (MPa)	0.9187	8.90	17.01
	Elongation (%)	0.8219	0.56	1.30

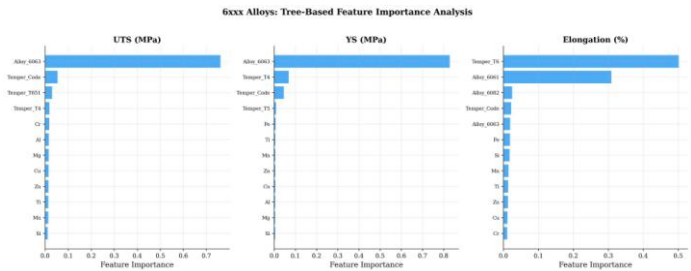


Fig. 3 Tree-based feature importance analysis for UTS, YS, and elongation prediction of 6xxx alloys.

The hierarchically clustered correlation heatmap (Fig. 4) shows strong UTS–YS correlation ($r = 0.85$), while both strength properties correlated positively with the 6061 indicator ($r = 0.67-0.68$) and negatively with 6063 ($r = -0.84$ to -0.90). The T651 temper showed moderate positive correlation with UTS ($r = 0.47$). Hierarchical clustering grouped compositional features separately from encoding and property variables, confirming that the two information domains are largely orthogonal in the 6xxx system.

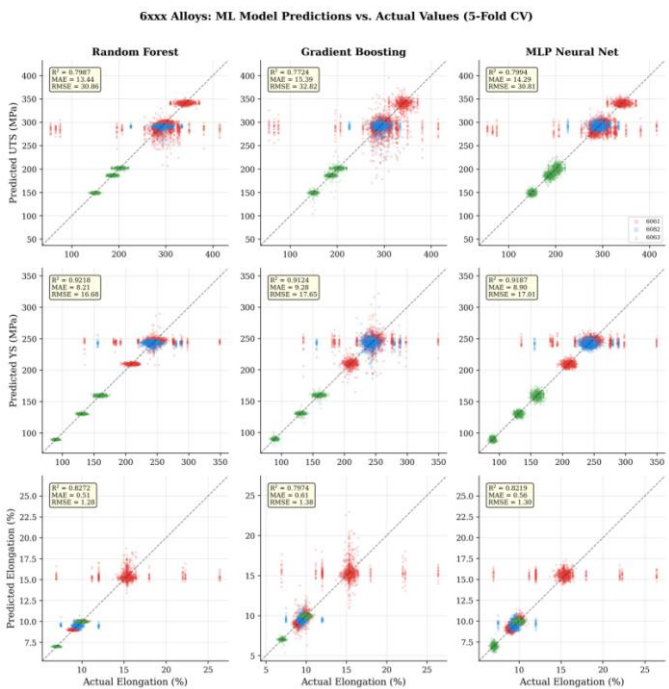


Fig. 2 Predicted vs. actual values for RF, GBR, and MLP models (5-fold CV), colored by alloy type: 6061 - red, 6082 - blue, 6063 - green.

Feature importance analysis (Fig. 3), assessed as the average of Random Forest and Gradient Boosting tree-based importances, revealed clear target-specific patterns. For UTS, the 6063 alloy indicator was the dominant feature (~76%), with the temper code contributing ~5% and the T651 indicator ~3%. For YS, the 6063 indicator was overwhelmingly dominant (~83%), reflecting the large strength gap between 6063 and the other alloys. For elongation, the T6 temper (50%) and 6061 alloy (31%) indicators dominated. Individual compositional features each contributed <3%, indicating that alloy designation and temper condition encodings dominate predictions. However, this low compositional contribution is partly a consequence of the augmentation strategy, in which property measurements were paired with randomly sampled compositions from specification ranges rather than the actual measured compositions. Because the composition–property linkage was decoupled during augmentation, the models cannot learn genuine composition–property relationships within each alloy-temper class, and the low importance of individual elements should therefore be interpreted as a methodological limitation rather than a purely physical finding.

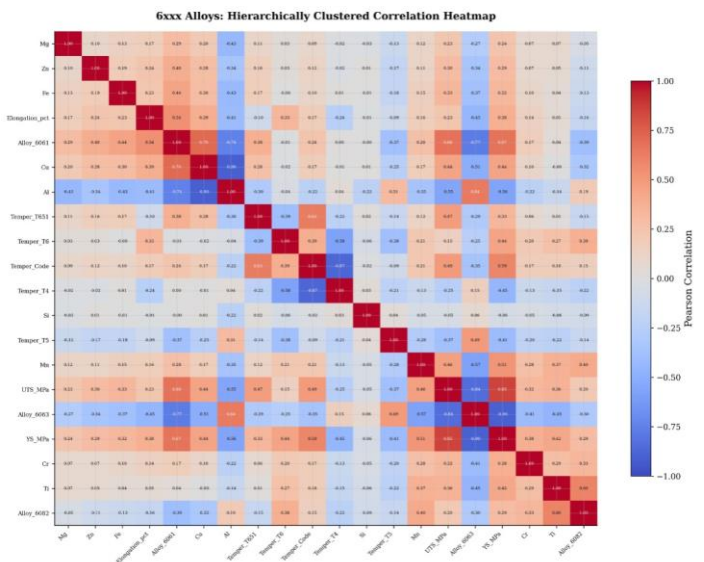


Fig. 4 Hierarchically clustered Pearson correlation heatmap of compositional, encoding, and property features

Learning curve analysis (Fig. 5), generated using simplified model configurations to reduce computational cost (RF: 100 trees, depth 10; GBR: 100 trees, depth 5; MLP: two hidden layers 64–32, 300 iterations; 3-fold CV), showed that RF achieved stable UTS prediction ($R^2 \approx 0.65-0.66$) above ~2,000 training samples. GBR exhibited slower convergence, while MLP matched RF at larger training sizes but showed higher variance at small sample sizes. The convergence at ~2,000 samples suggest that the hybrid augmentation strategy provides sufficient data volume for model training, although the lower R^2 values compared to Table 1 reflect the reduced model complexity used for the learning curve analysis.

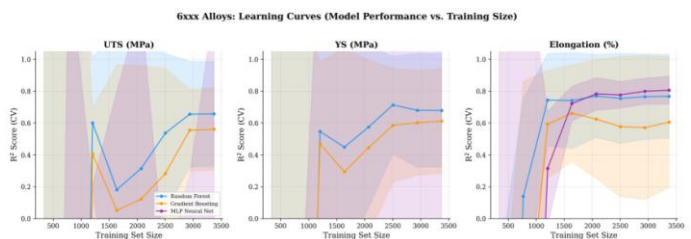


Fig. 5 Learning curves showing R^2 score vs. training set size for RF, GBR, and MLP models (simplified configurations; see text for details).

4. Conclusions

A hybrid ML framework combining real measurements with Monte Carlo augmentation was developed for predicting UTS, YS, and elongation of 6061, 6063, and 6082 Al-Mg-Si alloys across seven temper conditions. RF provided the best or comparable performance across all three targets (UTS: $R^2 = 0.80$, MAE = 13.44 MPa; YS: $R^2 = 0.92$, MAE = 8.21 MPa; elongation: $R^2 = 0.83$, MAE = 0.51 percentage points). Alloy type and temper condition encodings dominated predictions (>90% combined importance), while compositional features contributed <3% each; the low compositional contribution is partly attributable to the augmentation strategy, in which random compositions from specification ranges were paired with property measurements, thereby decoupling the composition–property linkage. Learning curve analysis confirmed convergence above ~2,000 samples. The moderate R^2 values reflect the wide property envelope of the 6xxx series. The framework demonstrates that corpus-mined data with hybrid augmentation provides a viable route for rapid property screening of engineering aluminum alloys, although future work should explore augmentation strategies that preserve composition–property correlations within each alloy-temper class.

5. Funding

This research was funded by the Bulgarian National Science Fund, Project **KII-06-H97/10** “Fabrication and Investigation of Antifriction Hybrid Nanocomposites Based on Aluminum Alloys Using an Integrated Approach Combining Experimental Methods and Machine Learning”.

6. References

1. Chauhan, K.P.S.; FUTURE INSTITUTE OF ENGINEERING AND TECHNOLOGY, BAREILLY (U.P.) Influence of Heat Treatment on the Mechanical Properties of Aluminium Alloys (6xxx Series): A Literature Review. *Int. J. Eng. Res. Technol.* (Ahmedabad) **2017**, *V6*, doi:10.17577/ijertv6is030301.
2. Gao, Y.-H.; Zhang, X.-X.; Zhou, J.-Z.; Xu, P.; Xin, Z.-J.; Gao, J.-H.; Shi, Y.-H.; Lü, Y.; Zhao, Y.-F.; Li, J.-Y. Engineering Heterogeneous Microstructures for Enhanced Strength and Ductility in Air-Cooled Al-Mg-Si Alloys. *Trans. Nonferrous Met. Soc. China* **2025**, *35*, 1017–1031, doi:10.1016/s1003-6326(24)66731-2.
3. Hagen, A.B.; Wenner, S.; Bjørge, R.; Wan, D.; Marioara, C.; Holmestad, R.; Ringdalen, I.G. Local Mechanical Properties and Precipitation Inhomogeneity in Large-Grained Al-Mg-Si Alloys. *SSRN Electron. J.* **2022**, doi:10.2139/ssrn.4216356.
4. *Effect of Copper/Alumina Hybrid Reinforcement on the Microstructure and Mechanical Properties of Stir Cast Aluminum Alloy AA6063*;
5. Wang, H.; Zheng, H.; Hu, M.; Ma, Z.; Liu, H. Synergistic Effect of Al₂O₃-Decorated Reduced Graphene Oxide on Microstructure and Mechanical Properties of 6061 Aluminium Alloy. *Sci. Rep.* **2024**, *14*, 16213, doi:10.1038/s41598-024-67004-x.
6. Jiang, W.; Zhu, J.; Li, G.; Guan, F.; Yu, Y.; Fan, Z. Enhanced Mechanical Properties of 6082 Aluminum Alloy via SiC Addition Combined with Squeeze Casting. *J. Mater. Sci. Technol.* **2021**, *88*, 119–131, doi:10.1016/j.jmst.2021.01.077.
7. Qian, X.; Parson, N.; Chen, X.-G. Effect of Post-Homogenisation Cooling Rate and Mn Addition on Mg₂Si Precipitation and Hot Workability of AA6060 Alloys. *Can. Metall. Q.* **2020**, *59*, 189–200, doi:10.1080/00084433.2020.1719332.
8. Shawon, A.R.; Ghosh, R.; Islam, M.A. Analysis of Thermal Conductivity of Aluminum Alloys by Compositions and Tempering Process Using Machine Learning. *Sci. Rep.* **2025**, *15*, 33352, doi:10.1038/s41598-025-15868-y.
9. Li, J.; Zhang, Y.; Cao, X.; Zeng, Q.; Zhuang, Y.; Qian, X.; Chen, H. Accelerated Discovery of High-Strength Aluminum Alloys by Machine Learning. *Commun. Mater.* **2020**, *1*, doi:10.1038/s43246-020-00074-2.

10. J. Soofi, Y.; Rahman, M.A.; Gu, Y.; Liu, J. A Feasibility Study of Machine Learning-Assisted Alloy Design Using Wrought Aluminum Alloys as an Example. *Comput. Mater. Sci.* **2022**, *215*, 111783, doi:10.1016/j.commatsci.2022.111783.
11. Pfeiffer, O.P.; Liu, H.; Montanelli, L.; Latypov, M.I.; Sen, F.G.; Hegadekotte, V.; Olivetti, E.A.; Homer, E.R. Aluminum Alloy Compositions and Properties Extracted from a Corpus of Scientific Manuscripts and US Patents 2021.
12. Silva, M.S.; Barbosa, C.; Acselrad, O.; Pereira, L.C. Effect of Chemical Composition Variation on Microstructure and Mechanical Properties of a 6060 Aluminum Alloy. *J. Mater. Eng. Perform.* **2004**, *13*, 129–134, doi:10.1361/10599490418307.
13. Trindade, W.G.; Padovezzi, R.O.; Pinto, E.P.G.; Marques, G.A.; Amorese, R.A.; Silva, M.C.; Damasio, T.S.; Andrade, G.O. de; Luca, R. de; Nascimento, P.H.T. do Influence of Alloying Elements on the Mechanical Properties and Electrical Conductivity of 6000 Series Aluminum Alloy. A Review. *CONTRIBUCIONES A LAS CIENCIAS SOCIALES* **2025**, *18*, e18763, doi:10.55905/revconv.18n.6-193.