Machine Learning Prediction of Mechanical Properties for Al-Cu Alloys Using Monte Carlo Data Augmentation

Mihail Kolev[1,5], Tatiana Simeonova [1,2,5*] Ivo Lozanov [3], Vilyana Kazanlaklieva[4], Viktoria Gerdzhikova [4]

[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
[3] Department of Inorganic Chemistry, Faculty of Chemistry and Pharmacy, Sofia University “St. Kliment Ohridski”, Sofia, Bulgaria
[4] University of Chemical Technology and Metallurgy Sofia, Bulgaria
[5] National Center for Mechatronics and Clean Technologies, 8 “Kliment Ohridski” Blvd., Building 8, 1756 Sofia, Bulgaria
* Corresponding author: tsimeonova@imbm.bas.bg

Abstract: *This study presents a machine learning framework for predicting the mechanical properties of 2xxx series Al-Cu alloys (2024, 2219, 2524) across 11 temper conditions. Monte Carlo augmentation generated 8,800 synthetic samples from compositional specification ranges of a literature-mined dataset. Three regression models, Random Forest, Gradient Boosting, and SVR-RBF were evaluated via 5-fold cross-validation (CV) to predict ultimate tensile strength (UTS), yield strength (YS), and elongation. All models achieved coefficient of determination $R^2 > 0.991$, with Mean Absolute Error MAE ≤ 7.4 MPa for UTS, ≤ 5.4 MPa for YS, and ≤ 0.51% for elongation. Feature importance analysis revealed that temper condition encoding dominated predictions (>75% importance), while individual compositional features contributed <5% each. The high predictive accuracy reflects the effectiveness of the augmentation scheme in capturing within-group property–composition–temper relationships, though generalization to unseen alloy–temper conditions remains to be validated. The results illustrate the potential of combining corpus-mined data with Monte Carlo augmentation for rapid alloy property screening.*

Keywords*: Al-Cu ALLOYS; MACHINE LEARNING; MECHANICAL PROPERTIES; MONTE CARLO AUGMENTATION; RANDOM FOREST; GRADIENT BOOSTING*

1. Introduction

The 2xxx series Al-Cu alloys (2024, 2219, 2524) are critical structural materials in aerospace applications, valued for their high strength, fatigue resistance, and damage tolerance [1,2]. Their mechanical properties depend strongly on both composition and temper condition, making property prediction across alloy-temper combinations a non-trivial task. Machine learning (ML) methods have shown promise for materials property prediction [3–5], but limited dataset sizes per alloy-temper condition remain a challenge [6,7]. Monte Carlo augmentation, which generates synthetic samples by sampling from compositional specification ranges, offers a physically motivated strategy to expand training datasets [8]. In this work, we apply three ML regression models to predict UTS, YS, and elongation of three 2xxx alloys across 11 temper conditions, using Monte Carlo-augmented data derived from a publicly available dataset.

2. Materials and Methods

Composition and property data were obtained from the dataset compiled by Pfeiffer et al. [9] and uploaded in Materials Cloud Archive, which contains aluminum alloy data extracted from scientific literature and US patents. Three alloys were selected: 2024 (T3, T351, T4, T6, T8), 2219 (O, T6, T62, T851), and 2524 (T3, T351). The feature space comprised 10 elemental compositions (Al, Cu, Mn, Si, Mg, Zn, Cr, Fe, Ti, Zr), 3 alloy-type and 8 temper-type one-hot encodings, and 1 ordinal temper code (22 features total). Figure 1 presents the mean experimental properties for each alloy-temper combination.

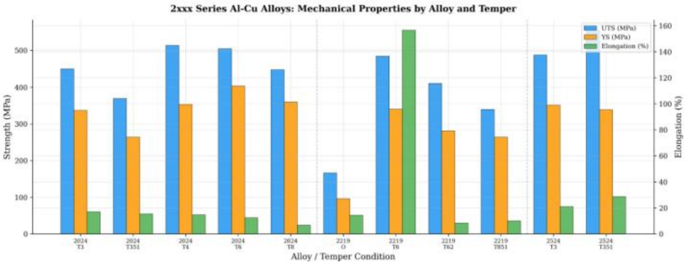

Fig. 1 *Mechanical properties (UTS, YS, elongation) by alloy and temper for 2xxx series Al-Cu alloys.*

For each of the 11 conditions, 800 synthetic samples were generated (8,800 total) by uniformly sampling element concentrations from specification ranges, computing Al as the balance, and adding 2% Gaussian noise to mean experimental property values. Three models were trained: Random Forest (RF, 200 trees, max depth 12, min samples split 5), Gradient Boosting (GBR, 200 trees, max depth 6, lr = 0.1, min samples split 5), and SVR-RBF (C = 100, $\varepsilon = 0.1$, γ = scale, features standardized). Performance was assessed by 5-fold cross-validation using coefficient of determination R^2, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

3. Results and Discussion

Table 1 summarizes model performance. All models achieved $R^2 > 0.991$ for all targets. RF provided the best overall accuracy (UTS MAE = 7.04 MPa, YS MAE = 5.08 MPa, elongation MAE = 0.46%). The marginal differences among models indicate that the augmented dataset effectively captured property-composition-temper relationships. Figure 2 shows predicted vs. actual plots confirming tight clustering along the 1:1 diagonal across all alloy clusters.

Table 1: *Cross-validation metrics for ML models.*

Model	Target	R^2	MAE	RMSE
Random Forest	UTS (MPa)	0.9924	7.04	9.06
	YS (MPa)	0.9931	5.08	6.56
	Elongation (%)	0.9994	0.46	1.03
Gradient Boosting	UTS (MPa)	0.9921	7.20	9.24
	YS (MPa)	0.9928	5.19	6.69
	Elongation (%)	0.9993	0.49	1.11
SVR (RBF)	UTS (MPa)	0.9916	7.39	9.52
	YS (MPa)	0.9923	5.32	6.90
	Elongation (%)	0.9993	0.50	1.06

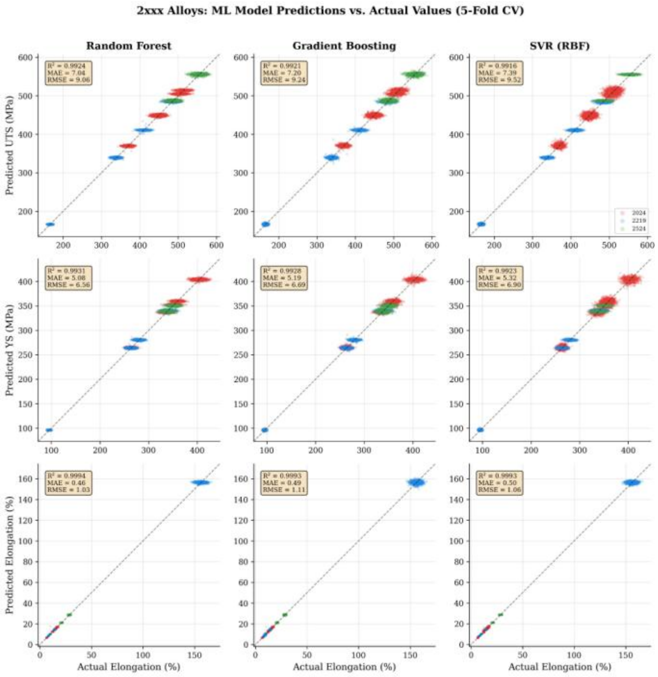

Fig. 2 *Predicted vs. actual values for RF, GBR, and SVR-RBF models (5-fold CV), colored by alloy: 2024 - red, 2219 - blue, and 2524 - green.*

Feature importance rankings (Fig. 3) show that temper-related features dominated all property predictions. For UTS, the temper code and O temper indicator together accounted for ~80% of importance. For elongation, the T6 indicator was most important (RF: 0.42), reflecting the anomalous 2219-T6 elongation (156.5%). This value should be interpreted with caution: the source dataset contains only two elongation records for 2219-T6 (11.0% and 302.0%), and the 302% value is physically unrealistic for a conventional precipitation-hardened Al-Cu alloy, where typical elongation ranges from 10 to 17%. The resulting mean (156.5%) is therefore likely dominated by a data extraction error in the original corpus and may inflate the elongation R² through a lever-arm effect. Individual compositional features contributed <5% each, consistent with the fact that within the 2xxx system, property differences arise primarily from thermomechanical processing history rather than minor compositional variations [10,11].

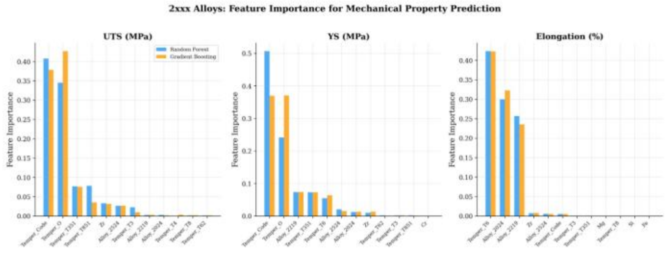

Fig. 3 *Feature importance from RF and GBR models for UTS, YS, and elongation prediction.*

The correlation heatmap (Fig. 4) confirms strong UTS-YS correlation (r = 0.93), a strong negative correlation between O temper and strength (r = −0.81 to −0.85), and weak negative Cu-strength associations (r ≈ −0.16). The augmented property distributions (Fig. 5) show that Monte Carlo sampling preserved realistic property ranges for all conditions, with the noted exception of the 2219-T6 elongation discussed above.

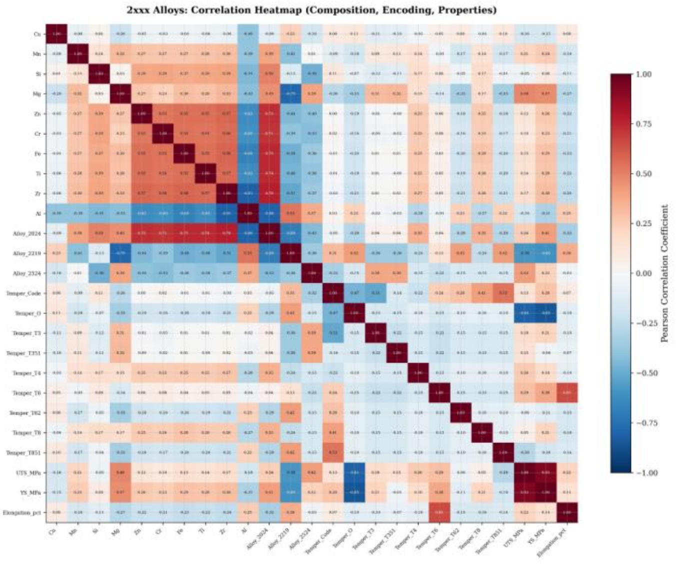

Fig. 4 *Pearson correlation heatmap of compositional features, encoding variables, and mechanical properties*

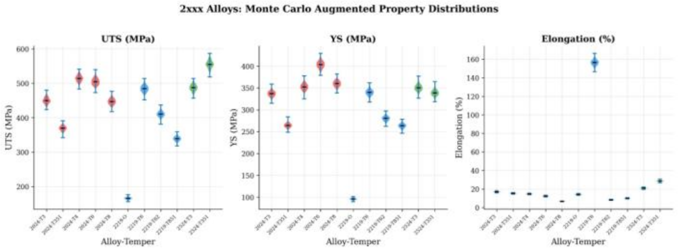

Fig. 5 *Monte Carlo augmented property distributions (violin plots) for each alloy-temper combination.*

4. Conclusions

A Monte Carlo augmented ML framework was developed for predicting UTS, YS, and elongation of 2024, 2219, and 2524 Al-Cu alloys across 11 temper conditions. All three models (RF, GBR, SVR-RBF) achieved R² > 0.991 for all targets. RF provided the best performance with MAE of 7.04 MPa (UTS), 5.08 MPa (YS), and 0.46% (elongation). Temper condition dominated feature importance (>75%), while compositional features contributed <5% each. It should be noted that the high R² values partly reflect the augmentation design: since all synthetic samples within a given alloy–temper condition share the same mean property values with only 2% noise, the cross-validation task effectively reduces to recovering group means from categorical encodings. The present results therefore demonstrate strong interpolative accuracy within the training domain but do not establish extrapolative generalization to unseen alloy–temper conditions. Additionally, the 2219-T6 elongation data should be treated with caution due to a likely outlier in the source dataset. Future work should evaluate leave-one-condition-out cross-validation to assess true predictive capability, and incorporate broader alloy systems to improve compositional sensitivity. Despite these limitations, the framework illustrates how combining literature-mined datasets with Monte Carlo augmentation can serve as a practical tool for rapid property screening of aluminum alloys within known compositional and processing domains.

5. Funding

This research was funded by the Bulgarian National Science Fund, Project **КП-06-Н77/5** “Self-lubricating hybrid aluminum metal matrix composites: synthesis, experi-mental and computer modeling of mechanical and tribological properties”.

6. Acknowledgement

The work in this publication was performed using equipment funded by project BG16RFPR002-1.014-0006 “National Center for Mechatronics and Clean Technologies”.

7. References

1. Li, S.-S.; Yue, X.; Li, Q.-Y.; Peng, H.-L.; Dong, B.-X.; Liu, T.-S.; Yang, H.-Y.; Fan, J.; Shu, S.-L.; Qiu, F.; et al. Development and Applications of Aluminum Alloys for Aerospace Industry. *J. Mater. Res. Technol.* **2023**, *27*, 944–983, doi:10.1016/j.jmrt.2023.09.274.
2. García-Aguirre, K.A.; Holguín-Momaca, J.T.; Ruiz-Esparza-Rodriguez, M.A.; Guía-Tello, J.C.; Garay-Reyes, C.G.; Estrada-Guel, I.; Martínez-Sánchez, R. Microstructural Evolution and Strengthening Mechanisms in a 2xxx Series Modified Al Alloy. *Microsc. Microanal.* **2022**, *28*, 2854–2856, doi:10.1017/s1431927622010753.
3. Elkatatny, S.; Alsharekh, M.F.; Alateyah, A.I.; El-Sanabary, S.; Nassef, A.; Kamel, M.; Alawad, M.O.; BaQais, A.; El-Garaihy, W.H.; Kouta, H. Optimizing the Powder Metallurgy Parameters to Enhance the Mechanical Properties of Al-4Cu/XAl2O3 Composites Using Machine Learning and Response Surface Approaches. *Appl. Sci. (Basel)* **2023**, *13*, 7483, doi:10.3390/app13137483.
4. Hu, H.; Zhao, F.; Yong, W.; Jiang, L.; Zhang, Z.; Xie, J. A Machine Learning Strategy to Achieve Dual-Synchronous Property Improvement of Aviation Al-Cu-Mg Alloy. *J. Mater. Sci. Technol.* **2026**, *244*, 208–230, doi:10.1016/j.jmst.2025.04.036.
5. Fu, C.; Lin, X.; Huang, H.; Yue, C.; Zuo, X.; Zheng, B.; Du, K.; Zheng, W. Synergistic Enhancement of Strength and Plasticity in Al-Cu Alloys Using Interpretable Machine Learning Algorithms. *Mater. Sci. Eng. A Struct. Mater.* **2025**, *942*, 148702, doi:10.1016/j.msea.2025.148702.
6. 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.
7. 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.
8. Park, J.; Yoon, H.-K. Methodology for Generating Diverse Geotechnical Datasets Using Monte Carlo Simulation and Genetic Algorithms. *Comput.-aided Civ. Infrastruct. Eng.* **2025**, *40*, 5494–5511, doi:10.1111/mice.70106.
9. Pfeiffer, O.P.; Liu, H.; Montanelli, L.; Latypov, M.I.; Sen, F.G.; Hegadekatte, V.; Olivetti, E.A.; Homer, E.R. Aluminum Alloy Compositions and Properties Extracted from a Corpus of Scientific Manuscripts and US Patents 2021.
10. Zhou, Y.; Lin, X.; Kang, N.; Huang, W.; Wang, Z. Mechanical Properties and Precipitation Behavior of the Heat-Treated Wire + Arc Additively Manufactured 2219 Aluminum Alloy. *Mater. Charact.* **2021**, *171*, 110735, doi:10.1016/j.matchar.2020.110735.
11. Zhang, L.; Huang, G.; Wang, L.; Lu, G.; Guan, S. Microstructure-Strengthening Correlation of 2219 Al Alloy Subjected to Ultrasonic Melt Treatment, Hot Rolling and Heat Treatment. *J. Mater. Res. Technol.* **2023**, *27*, 7470–7481, doi:10.1016/j.jmrt.2023.11.167.