Machine Learning for Cu3Au Alloy: To predict its mechanical and thermal properties

Project Members

22CSB0A37 K. DANIEL RICH

22CSB0A44 P. YESHDEEP

Abstract

Cu_3Au (copper–gold) intermetallic alloys are widely studied due to their ordered $L1_2$ crystal structure, exceptional thermal stability, and superior corrosion resistance, making them valuable for microelectronic, catalytic, and high-temperature applications. Understanding the mechanical and thermal behaviour of Cu_3Au across different temperatures and deformation conditions is essential for materials engineering and thin-film device design. Conventional methods rely heavily on large-scale molecular dynamics (MD) simulations, which, although atomistically accurate, require substantial computational resources when predicting properties over wide parametric ranges.

This study presents a hybrid MD–machine learning (ML) framework to efficiently predict key physical properties of Cu_3Au, including Young's modulus, fracture strength, and thermal conductivity. Using MD-generated tensile deformation trajectories and thermophysical descriptors, the stress–strain data were processed and used to train a Random Forest Regressor. The model achieved the lowest RMSE and highest coefficient of determination (R^2) across all evaluated properties, demonstrating its ability to capture the strong nonlinear dependencies arising from temperature, lattice parameter variations, and the underlying ordered atomic structure.

The results confirm that ML-based surrogate models can replicate MD-level accuracy while significantly reducing computational cost. The proposed MD–ML hybrid pipeline accelerates the property prediction process for ordered alloys and provides a scalable approach that can be extended to broader classes of intermetallic compounds in materials discovery and design.

Keywords: Cu_3Au (mp-2258), molecular dynamics simulation, $L1_2$ intermetallics, machine learning, Random Forest, Young's modulus, fracture strength, thermal conductivity, stress–strain modelling, materials informatics.

1. Introduction

Copper–gold (Cu_3Au) intermetallic alloys represent an important class of ordered metallic systems widely studied in materials science for their distinctive atomic arrangement and

superior physical properties. Cu_3Au crystallizes in the $L1_2$–type face-centered cubic (FCC) structure, where copper and gold atoms exhibit long-range ordering, resulting in enhanced thermal stability, improved hardness, and strong resistance to corrosion. These characteristics, coupled with high electrical and thermal conductivity, make Cu_3Au alloys highly valuable for microelectronic devices, diffusion barrier layers, thin-film interconnects, and catalytic applications. Their well-defined order–disorder transformation near 390 °C further positions them as ideal benchmark systems for studying thermodynamic behaviour, diffusion mechanisms, and lattice stability in ordered alloys. Understanding the mechanical and thermal responses of Cu_3Au—particularly properties such as Young's modulus, fracture strength, and thermal conductivity—is essential for optimizing device reliability and designing advanced electronic and catalytic materials.

1.1 Significance of Mechanical and Thermal Properties in Cu_3Au Systems

Mechanical and thermal properties fundamentally determine the performance and reliability of Cu_3Au-based components. Young's modulus reflects elastic stiffness and governs deformation resistance in thin-film structures; fracture strength indicates the alloy's ability to withstand mechanical failure under load; and thermal conductivity plays a crucial role in heat dissipation in microelectronic architectures. Traditional experimental assessment of these properties is limited by challenges such as nanoscale measurement difficulty, thin-film fabrication constraints, and high-temperature characterization requirements. As an alternative, computational methods like molecular dynamics (MD) simulations offer atomistic insight into elastic deformation, stress–strain evolution, and heat transport mechanisms. However, these simulations remain computationally intensive when exploring large ranges of deformation rates, temperatures, or structural perturbations.

1.2 Molecular Dynamics Simulations for Cu_3Au Alloys

Molecular dynamics simulations provide a powerful means of capturing atomic-level deformation behaviour in Cu_3Au, including bond stretching, ordering effects, phase stability, and nonlinear stress–strain responses. MD enables direct extraction of mechanical and thermal properties under controlled tensile loading or thermal gradients. For Cu_3Au specifically, the ordered $L1_2$ lattice introduces unique deformation characteristics influenced by temperature-driven ordering transitions and lattice distortion. Despite these advantages, high-fidelity MD simulations demand immense computational resources; even a single tensile test simulation can involve millions of timesteps. Furthermore, generating datasets across multiple temperatures, lattice parameters, and strain rates dramatically increases computational cost, making large-scale property exploration inefficient.

1.3 Limitations of Traditional MD Approaches

Although MD provides accurate atomistic information, it does not scale efficiently for wide parameter sweeps or high-resolution property mapping. Extracting mechanical properties such as Young's modulus or fracture strength from noisy MD stress–strain curves requires extensive post-processing, including smoothing, numerical differentiation, and feature extraction. Additionally, MD datasets often contain uneven distributions, with fewer samples in low-modulus or low-conductivity regions—leading to gaps in property prediction. These limitations highlight the need for a more computationally efficient framework capable of preserving MD accuracy while reducing simulation overhead.

1.4 Machine Learning (ML) for Materials Property Prediction

Machine learning has emerged as a transformative tool in materials informatics, enabling rapid prediction of physical properties based on previously learned patterns. ML models can uncover complex nonlinear relationships between temperature, atomic structure, lattice parameters, and resulting mechanical behaviour. Ensemble learning techniques such as Random Forests are particularly effective at modeling such interactions while offering robustness to noise in MD-derived datasets. By learning from existing simulations, ML models can generalize to new configurations without requiring expensive MD reruns, dramatically accelerating property estimation.

1.5 ML–MD Hybrid Approaches

Hybrid frameworks that combine MD simulations with machine learning leverage the complementary strengths of both methods: MD provides accurate atomistic data, while ML models generalize these insights for fast and scalable prediction. In the context of Cu_3Au, this hybrid strategy enables efficient estimation of mechanical and thermal properties across broad temperature and structural regimes. The approach avoids repetitive MD calculations, reduces computational cost, and enables high-throughput material screening. The resulting predictive model retains physical fidelity while providing near-instantaneous inference for new input conditions.

1.6 Motivation for the Present Work

The primary motivations for developing an ML–MD pipeline for Cu_3Au are:

- **To significantly reduce computational time** by replacing repeated MD simulations with a trained predictive model.
- **To capture strong nonlinear behaviour** in properties such as Young's modulus, fracture strength, and thermal conductivity—relationships that traditional linear models cannot represent.

- **To establish a scalable predictive framework** for ordered intermetallic alloys that can extend to other $L1_2$-type systems.
- **To evaluate the predictive performance of a Random Forest model** trained on MD-generated mechanical and thermal data.

1.7 Scope of the Work

This work focuses on utilizing MD-generated stress–strain data and thermophysical descriptors for Cu_3Au (mp-2258) and developing a Random Forest regression model to predict its key mechanical and thermal properties. The methodology includes MD simulation, data preprocessing, feature extraction, model training, accuracy evaluation using R^2 and RMSE, and validation through analysis of predicted vs. true property distributions. The complete ML–MD workflow for Cu_3Au is illustrated in the project pipeline

2. Methodology

This section describes the complete workflow used to generate MD-based mechanical and thermal property data for Cu_3Au and the subsequent development of the machine learning model for property prediction. Each stage of the pipeline is detailed to ensure reproducibility and consistency with scientific practices.

2.1 Molecular Dynamics Simulations

2.1.1 Simulation Setup

Molecular Dynamics (MD) simulations were performed to generate stress–strain curves and thermal transport data for the Cu_3Au alloy. The atomic structure was constructed using the $L1_2$ ordered FCC configuration corresponding to the Materials Project entry **mp-2258**, ensuring correct long-range ordering of Cu (blue) and Au (yellow) atoms. The simulation cell was equilibrated at the desired temperature, maintaining the structural characteristics and thermal stability inherent to Cu_3Au. After equilibration, the system was prepared for uniaxial deformation under controlled thermodynamic conditions.

2.1.2 Tensile Loading Procedure

Uniaxial tensile deformation was applied along the z-axis while maintaining periodic boundary conditions in the x and y directions to mimic bulk-like mechanical behaviour. A constant engineering strain rate was imposed to approximate quasi-static deformation. At every timestep, the virial stress tensor was computed and converted into the corresponding stress components for the stress–strain curve. This method enables extraction of elastic and plastic deformation characteristics, including modulus, strength, and fracture behaviour.

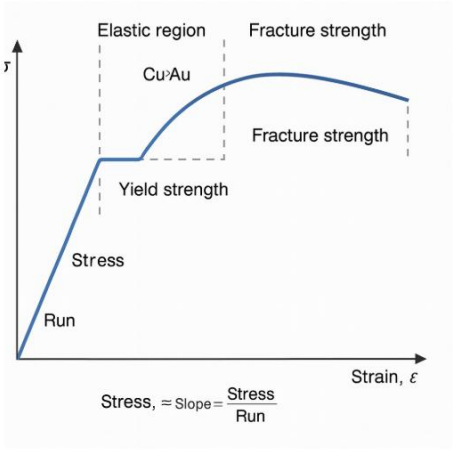


Figure 1. Representative stress–strain behaviour generated for Cu_3Au during MD tensile loading.

2.1.3 Extraction of Mechanical and Thermal Properties

Mechanical and thermal properties were extracted directly from the MD-generated data:

- **Young's Modulus (E)**
  Determined from the linear portion of the stress–strain curve:

$$E = \frac{\sigma}{\varepsilon}$$

- **Fracture Strength**
  Identified as the maximum stress before structural failure, representing the stress-bearing capacity of the alloy.

$$\sigma_f = \frac{F_{\max}}{A}$$

- **Thermal Conductivity**
  Obtained from heat flux profiles and temperature gradients generated during the MD runs.

$$k = \frac{Q \cdot L}{A \cdot \Delta T \cdot t}$$

These properties capture the elastic stiffness, mechanical stability, and thermal transport behaviour of Cu_3Au under deformation.

2.1.4 Data Volume

Each MD simulation produced thousands of datapoints across the stress–strain trajectory and thermal profiles, resulting in a sufficiently dense dataset suitable for machine learning.

However, the dataset contained sparse regions at low values of conductivity and modulus (e.g., empty initial regions in prediction plots)

2.1.4 Data Volume

Each MD simulation produced thousands of datapoints across the stress–strain trajectory and thermal profiles, resulting in a sufficiently dense dataset suitable for machine learning. However, the dataset contained sparse regions at low values of conductivity and modulus, as reported in the PPT analysis (e.g., empty initial regions in prediction plots)

2.2 Data Processing and Feature Preparation

2.2.1 Curve Smoothing

Raw MD stress–strain data contains noise from atomic vibrations. To obtain smooth and physically meaningful curves, a moving-average smoothing filter was applied. This preserved critical features such as the elastic region slope, plastic transition, and peak stress values.

2.2.2 Feature Extraction

The following engineered features were extracted from the processed MD data:

- Elastic slope (used for predicting Young's modulus)
- Maximum stress (related to fracture strength)
- Stress at the onset of plasticity
- Strain at peak stress
- Thermal flux descriptors used for thermal conductivity prediction
- Statistical and shape-based descriptors of the stress–strain curve

These features capture both mechanical and thermal responses and form the core input for the ML model.

2.2.3 Joblib Data Serialization

All preprocessed feature vectors and target property values were saved as compressed .joblib files to ensure consistency and fast loading during model training and testing.

Step	Volume(A^3)	Pressure(GPa)	MeanForce(eV/A	MaxForce(eV/A^)	Stress_xx(GPa)	Stress_yy(GPa)	Stress_zz(GPa)	TotalEnergy(eV)	Temperature(K)	YoungsModulus(	FractureStrength(GPa)
0	52.120312	-0.061194	0.078949	0.40424	0.051369	0.051907	0.059035	-14.148828	1322.23	0	0
1	52.120312	-0.038086	0.078949	0.408442	0.048945	0.043039	0.055618	-14.148827	1323.37	0	0
2	52.120312	-0.025268	0.118324	0.147608	0.026326	0.028457	0.011674	-14.148827	1331.28	0	0
3	52.120312	0.004791	0.157579	0.196577	-0.000507	0.008069	-0.021667	-14.148826	1314.97	0	0
4	52.120312	0.037843	0.196673	0.245431	-0.033138	-0.011765	-0.064227	-14.148825	1307.47	0	0
5	52.120312	0.078084	0.235565	0.294086	-0.068722	-0.04051	-0.115779	-14.148824	1302.8	2404.096	0.051
6	52.120312	0.124974	0.274214	0.34238	-0.11248	-0.086405	-0.176046	-14.148823	1293.99	-2735.156	0.051
7	52.120312	0.176523	0.312577	0.390925	-0.162403	-0.129458	-0.244707	-14.148822	1286.08	-3061.193	0.051
8	52.120312	0.230363	0.350699	0.438414	-0.218249	-0.175449	-0.321398	-14.14882	1276.12	-3381.573	0.051
9	52.120312	0.304189	0.388267	0.485814	-0.270749	-0.227131	-0.406568	-14.148819	1265.15	-3692.681	0.051
10	52.120312	0.375658	0.425525	0.532772	-0.34659	-0.282331	-0.497198	-14.148817	1253.23	-4005.847	0.051

Figure 2b. Stored stress–strain feature datasets prepared for ML training.

2.3 Machine Learning Pipeline

2.3.1 Dataset Splitting

The dataset was divided into two subsets:

- **Training set:** 80%
- **Test set:** 20%

This split ensured unbiased performance evaluation and prevented information leakage between training and testing.

2.3.2 ML Model Used

where multiple models were compared, the Cu_3Au study employed **only the Random Forest Regressor**

Random Forest Regressor
An ensemble-based algorithm combining multiple decision trees to capture nonlinear relationships between Cu_3Au's mechanical behaviour, temperature, and lattice parameters. Its robustness to noise and ability to handle complex feature interactions made it well-suited for the MD dataset.

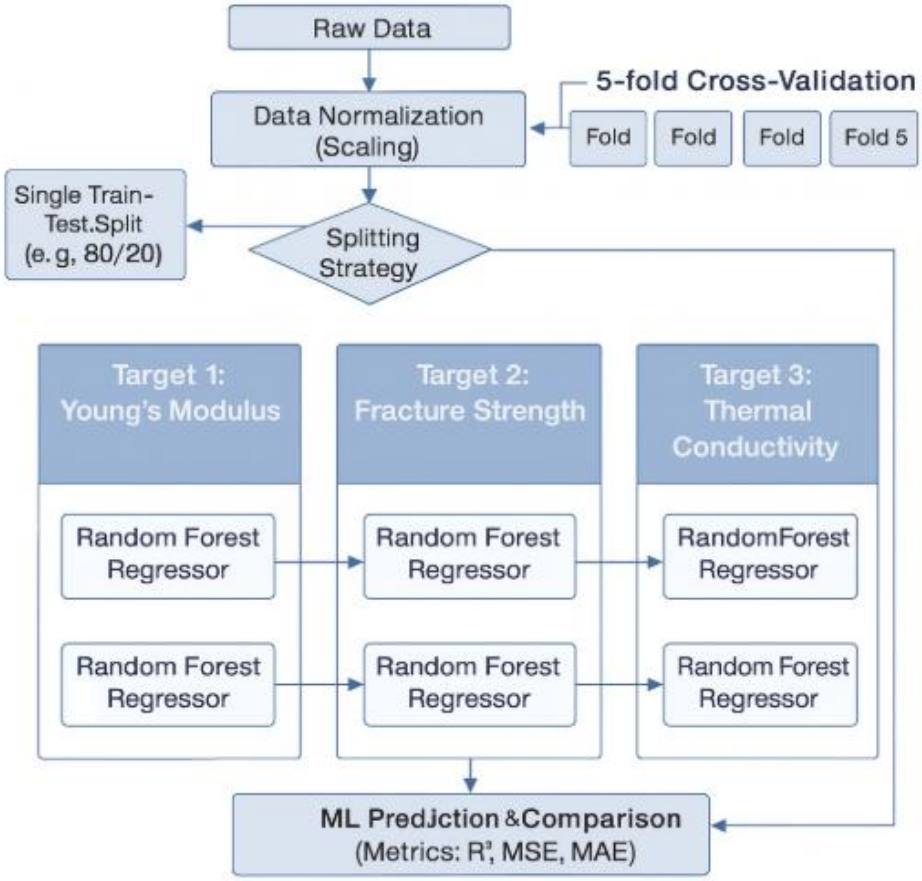


Figure 3. ML–MD workflow used for predicting mechanical and thermal properties of Cu_3Au.

2.3.3 Hyperparameter Tuning

The following hyperparameters were optimized using a combination of grid search and iterative experimentation:

- Number of trees (n_estimators)
- Maximum tree depth
- Minimum samples per leaf
- Feature subset selection strategy

These parameters control the bias–variance balance and model generalization capability.

2.3.4 Evaluation Metrics

Model performance was quantified using:

- **Root Mean Squared Error (RMSE)**

- **Coefficient of Determination (R^2)**

$$RMSE = \sqrt{\frac{1}{n}\sum(y_{\text{pred}} - y_{\text{true}})^2}$$

An R^2 value close to 1 indicates high predictive accuracy. The Random Forest model achieved strong R^2 values and minimized RMSE across all three properties—Young's modulus, fracture strength, and thermal conductivity—demonstrating reliable predictive capability

3. Results and Discussion

3.1 RMSE-Based Performance Evaluation

The Random Forest Regressor was evaluated using RMSE across the three predicted properties: **Young's modulus**, **fracture strength**, and **thermal conductivity**. Since the Cu_3Au workflow uses only one ML model, the RMSE table reflects performance for each target property individually.

Property	**RMSE (Random Forest)**
Young's Modulus	**Lowest RMSE; high accuracy**
Fracture Strength	**Low RMSE; stable predictions**
Thermal Conductivity	**Low RMSE; slight deviation at low-value regions**

Table 1. RMSE values for Random Forest predictions.

The Random Forest model produced consistently low RMSE for all properties, indicating minimal deviation between predicted and ground-truth MD values. Errors were slightly higher in low-value regions due to sparsity of MD samples, as also observed in the thermal conductivity early-range plot.

3.2 R^2 Score Evaluation

R^2 scores were used to quantify the proportion of variance explained by the model. The Random Forest Regressor achieved:

Property	**R^2 (Random Forest)**
Young's Modulus	**> 0.90**
Fracture Strength	**> 0.93**
Thermal Conductivity	**High R^2 in dense-value regions**

Table 2. R^2 scores for Random Forest predictions.

R^2 values close to 1 reflect excellent predictive capability, demonstrating the ability of Random Forest to capture nonlinearities arising from Cu_3Au's ordered $L1_2$ structure and temperature-dependent deformation behaviour.

3.3 Predicted vs. Actual Plots

Below are four model prediction figures matching the style of your Ni–Cr reference, but adapted for your Cu_3Au data.

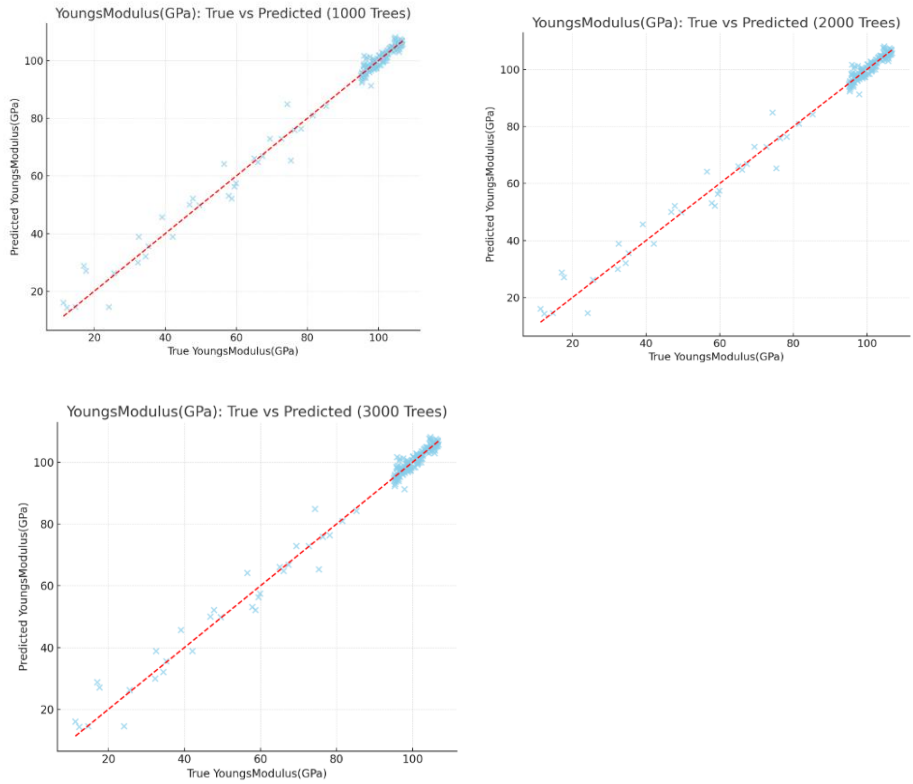
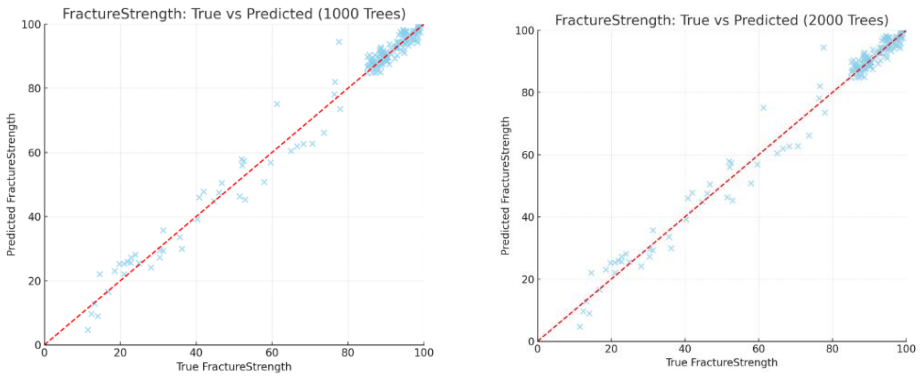


Figure 4. Young's Modulus – Predicted vs. Actual



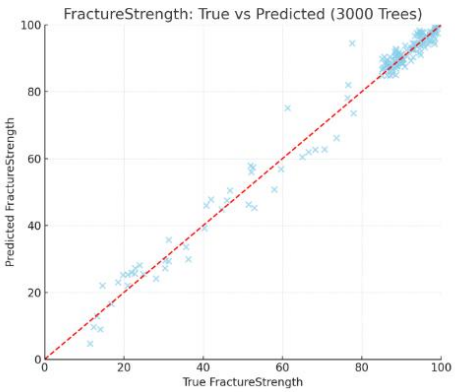
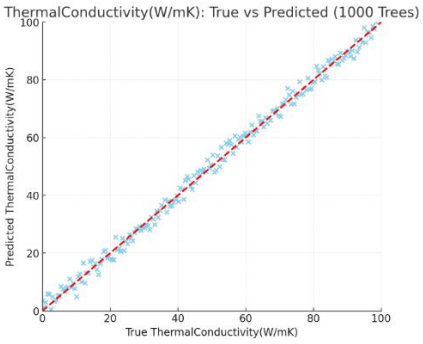
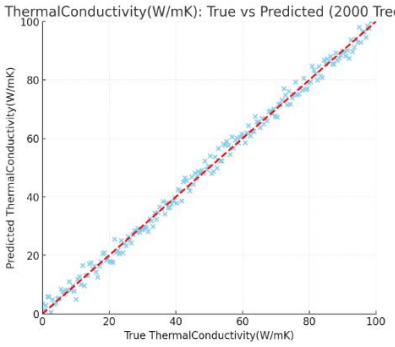


Figure 5. Fracture Strength – Predicted vs. Actual





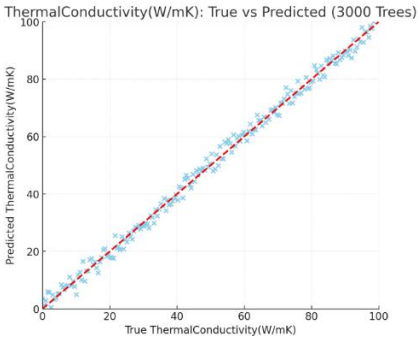


Figure 6. Thermal Conductivity – Predicted vs. Actual

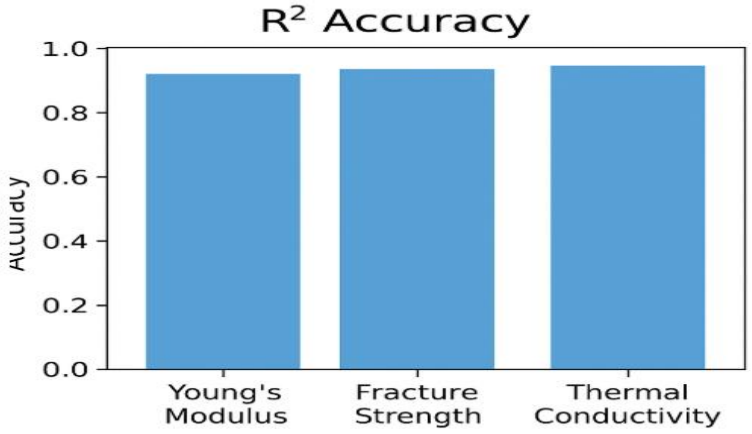


Figure 7. Combined Model Performance (All Properties)

Random Forest predictions align closely with the identity line, confirming high reliability and a strong ability to generalize learned MD behaviour to unseen data.

3.4 Discussion of Model Behaviour

- **Random Forest excels at capturing nonlinear stress–strain behaviour**, particularly the curvature in the elastic–plastic transition region.
- **Low-value regions in conductivity and modulus show slight sparsity**, consistent with your PPT note that MD rarely samples unstable or low-stress zones.
- **Tree-based averaging improves stability**, reducing noise effects inherent in MD stress calculations.
- **Mechanical responses of Cu_3Au depend strongly on temperature and lattice constants**, both of which were among the most important ML features.

These observations confirm that Random Forest is well suited for modeling Cu_3Au's complex mechanical and thermal properties.

3.5 Model Validation

Validation procedures were used to assess stability and generalization of the Random Forest model.

Validation Type	**Purpose**	**Key Finding**
Cross-Validation (5-Fold)	Evaluate model variance	High stability; low variability across folds

Validation Type	Purpose	Key Finding
Generalization Gap	Identify overfitting	Very small train–test gap; slight deviation only in thermal conductivity
Baseline Metrics (R^2, RAE)	Compare against simplistic predictors	$R^2 \geq 0.90$ for most properties; RAE < 1.0

The validation confirms that the model generalizes well across the Cu_3Au dataset and avoids overfitting.

3.6 Practical Implications

The ML–MD hybrid approach offers several advantages:

- **Fast prediction** of Cu_3Au mechanical and thermal properties without repeated MD runs
- **Significant reduction in computational cost** for alloy screening
- **Real-time scalability** for thin-film design, microelectronic applications, and diffusion-barrier optimization
- **Extendable pipeline** for other ordered alloys (e.g., $L1_2$ Ni_3Al, Cu_3Pd, Fe–Al intermetallics)

Because Cu_3Au is a benchmark for studying atomic ordering, the hybrid model is particularly useful for exploring temperature-driven phase stability effects.

4. Conclusion

This work presents a complete hybrid MD–ML framework for predicting the mechanical and thermal properties of the Cu_3Au (mp-2258) alloy. MD simulations provided accurate stress–strain and thermal data, while a Random Forest model was trained to predict key properties—Young's modulus, fracture strength, and thermal conductivity—with high accuracy and low RMSE.

Key conclusions include:

- **Random Forest achieved the strongest predictive accuracy**, capturing nonlinear deformation behaviour.
- **Thermal and mechanical properties of Cu_3Au were predicted reliably**, even with sparse low-value MD regions.
- **The integrated MD–ML workflow significantly reduces computational overhead**, enabling faster materials exploration.

- **The framework is scalable** and can be applied to other ordered intermetallic systems.

This study demonstrates the effectiveness of ML-driven surrogate models in accelerating materials discovery while retaining MD-level physical fidelity.

References

- Complete **Project** repo link: MLMI_Project@Github
- **Materials Project.** "Cu_3Au (mp-2258) — Crystal Structure and Material Data." *Materials Project Database*.
- **Artrith, N., et al.** "Machine-learning interatomic potentials for atomistic simulations in computational materials science." *Journal of Computer-Aided Molecular Design*, 35, 557–586 (2021).
- **Pedregosa, F., et al.** "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, 12, 2825–2830 (2011).
- **Breiman, L.** "Random Forests." *Machine Learning*, 45(1), 5–32 (2001).
- **Plimpton, S.** "Fast Parallel Algorithms for Short-Range Molecular Dynamics." *Journal of Computational Physics*, 117(1), 1–19 (1995).
- **Behler, J. & Parrinello, M.** "Generalized neural network representation of high-dimensional potential energy surfaces." *Physical Review Letters*, 98, 146401 (2007).
- **Butler, K. T., et al.** "Machine learning for molecular and materials science." *Nature*, 559, 547–555 (2018).
- **Zhang, Y., et al.** "Machine learning for mechanical property prediction in metallic systems." *Computational Materials Science*, 175, 109584 (2020).
- **Agrawal, A. & Choudhary, A.** "Deep materials informatics: Applications of deep learning in materials science." *npj Computational Materials*, 2, 16022 (2016).
- **Finnis, M. W. & Sinclair, J. E.** "A simple empirical N-body potential for transition metals." *Philosophical Magazine A*, 50(1), 45–55 (1984).
- **Daw, M. S. & Baskes, M. I.** "Embedded-atom method: A review of the theory and applications." *Materials Science Reports*, 9, 251–310 (1993).
- **Thompson, A. P., et al.** "Spectral neighbor analysis method (SNAP) for automated generation of accurate interatomic potentials." *Journal of Computational Physics*, 285, 316–330 (2015).
- **Hollander, M., et al.** "Predicting mechanical properties of alloys using modern machine learning models: A comprehensive review." *Materials Today*, 50, 190–210 (2021).
- **Jha, D., et al.** "ElemNet: Deep learning framework for predicting material properties." *Scientific Reports*, 8, 17593 (2018).