Machine-Learning-Guided of Potential Energy of Cu-Ni Alloys Using MD simulated data.

Project Members

22CEB0B01 K Sai Vardhan
22CSB0B19 P Tarun
22CHB0B49 Mercy

Abstract:

Cu–Ni (copper–nickel) alloys are widely valued for their superior corrosion resistance, high electrical and thermal conductivity, and remarkable mechanical stability, making them essential for marine engineering, desalination, and heat-exchange applications. Understanding the mechanical response of Cu–Ni alloys under various compositional and loading conditions is crucial for optimizing their design and performance. Conventional molecular dynamics (MD) simulations provide atomistic insight into deformation behaviour but are computationally costly. This work proposes a hybrid computational approach that couples MD-derived stress–strain data with machine learning (ML) techniques to efficiently predict key mechanical properties such as Young's modulus, yield strength, and ultimate tensile strength (UTS). Tensile deformation simulations for Cu–Ni alloys with varying nickel concentrations were performed, and the resulting stress–strain curves were used to train four supervised ML models—Linear Regression, Ridge Regression, Random Forest Regressor, and Multilayer Perceptron (MLP). Among these, the Random Forest model demonstrated the highest predictive accuracy with the lowest RMSE and maximum R^2 values across all targeted properties. The study confirms the nonlinear interplay between alloy composition and mechanical performance and highlights the potential of the integrated ML–MD pipeline to accelerate the design and optimization of Cu–Ni alloys with high computational efficiency.

Keywords: Cu–Ni alloys, molecular dynamics simulation, machine learning, mechanical properties, Random Forest, tensile deformation.

1. Introduction

Copper–nickel (Cu–Ni) alloys constitute a vital class of engineering materials widely employed in marine, chemical, and thermal applications due to their unique combination of mechanical strength, ductility, and corrosion resistance. They exhibit outstanding stability in seawater environments, good thermal and electrical conductivity, and excellent structural performance under prolonged service conditions. These characteristics make Cu–Ni alloys ideal for applications such as condenser tubes, desalination plants, heat exchangers, ship hulls, and marine fittings. Notable compositions like Cu70-Ni30 and Cu90-SSNi10 are among the most commonly used industrial grades, offering a balanced combination of mechanical integrity and corrosion resistance. Accurately determining their mechanical properties—including Young's modulus, yield strength, ultimate tensile strength (UTS), and strain-to-failure—is essential for optimizing alloy design, predicting operational performance, and advancing material development for next-generation engineering applications

1.1 Significance of Mechanical Properties in Cu–Ni Systems

Mechanical properties play a pivotal role in defining the performance and reliability of Ni–Cr (nickel–chromium) alloys, especially in high-temperature and load-bearing applications. Young's modulus governs the elastic stiffness and resistance to deformation, yield strength marks the onset of irreversible plastic deformation, and ultimate tensile strength (UTS)

specifies the maximum stress the alloy can resist before fracture. These properties ensure structural integrity and durability of critical components such as turbine blades, reactor parts, and heat exchangers under extreme conditions.

1.2 Molecular Dynamics Simulations for Cu–Ni Alloys

Molecular dynamics (MD) simulations serve as a powerful technique to capture the atomistic mechanisms behind deformation in Cu–Ni alloys, such as atomic bond stretching, dislocation nucleation, void formation, and microstructural evolution. MD simulations yield stress–strain behavior directly derived from atomic interactions under tensile loading conditions. However, conducting high-quality MD simulations demands significant computational resources. For instance, a single tensile test on a moderately sized Cu–Ni simulation cell may require millions of timesteps, translating to thousands of CPU-hours. Furthermore, the need to explore a broad parameter space—including varying copper-nickel compositions, strain rates, temperature conditions, and microstructural features—further escalates the computational expense.

1.3 Limitations of Traditional MD Approaches

Despite MD's precision at the atomic scale, its scalability remains a considerable barrier. Generating extensive datasets necessary for comprehensive parameter studies, uncertainty quantification, or alloy optimization is often infeasible. Additionally, converting raw atomic-level stress data into macroscopic mechanical properties involves complex procedures like smoothing stress–strain curves, estimating derivatives, and performing feature engineering

1.4 Machine Learning (ML) for Materials Prediction

Machine learning has emerged as a transformative methodology in materials science, enabling rapid predictions of material properties by learning from pre-existing data. ML can identify nonlinear relationships between input parameters (composition, temperature, atomic ordering) and output mechanical properties (Young’s modulus, yield strength, UTS) without performing exhaustive simulations for each configuration. Ensemble learning techniques such as Random Forests and nonlinear models like artificial neural networks (ANNs) excel at approximating intricate material behaviors with remarkable accuracy.

1.5 ML–MD Hybrid Approaches

Combining MD simulations with ML algorithms has become increasingly prevalent in materials informatics. In this hybrid framework, MD provides high-fidelity atomistic data, while ML models learn the underlying physical trends and generalize them for rapid and accurate prediction. This approach removes the need for repeated, computationally expensive simulations and enables near-instantaneous inference of new alloy compositions and processing conditions.

1.6 Motivation for the Present Work

This research aims to develop a comprehensive ML–MD pipeline tailored for predicting the mechanical properties of Cu–Ni alloys. The core motivations include:

- Significantly reducing computational overhead by applying ML prediction models after initial MD data generation.
- Accurately capturing nonlinear mechanical behavior often overlooked by traditional linear approaches.
- Creating an extensible predictive platform for future design and exploration of Cu–Ni alloy systems.
- Benchmarking the effectiveness of multiple ML algorithms on MD-derived mechanical property datasets

1.7 **Scope of the Work**

This study utilizes MD-generated stress–strain data for Cu–Ni alloy systems and assesses the prediction capabilities of four ML models: Linear Regression, Ridge Regression, Random Forest Regression, and Multilayer Perceptron (MLP). The pipeline consists of data collection through MD, preprocessing of stress–strain curves, extraction of key mechanical properties, ML model training, performance evaluation, and validation.

2. Methodology

This section outlines the complete pipeline from MD simulation to ML-based prediction of potential energy. An extensive description of each stage ensures reproducibility and fidelity to the scientific workflow.

2.1 Data Generation

Data is generated using Molecular Dynamics (MD) simulations in LAMMPS. We generated temperature , pressure , volume, density and potential energy columns of Cu-Ni.

Step	**What Happened**	**Why It Was Done**

1. System Creation	Created a 3D FCC lattice with 4000 atoms	Build the physical Cu–Ni alloy model
2. Assign Potential	Applied EAM potential	To realistically simulate interatomic forces
3. Equilibration (Minimization)	Relaxed the system, then ran MD at constant Temp & Pressure	Bring system to stable physical state
4. MD Simulation	Ran 100,000 timesteps to observe behavior over time	Generate physical data (energy, volume, pressure, etc.)
5. Data Sampling	Every 100 steps, LAMMPS recorded Temp, Pressure, Volume, Density, PotEng	Convert physics simulation → ML dataset
6. Data Saved	Output was saved into	Dataset ready for ML processing

2.2.1 Data Processing

Task	**Tools Used**	**Role in Project**
Data loading & structuring	**Pandas**	Loaded the LAMMPS text file into a DataFrame, assigned column names, handled tabular data.
Numerical operations	**NumPy**	Used for mathematical operations like RMSE calculation and handling arrays.
Data cleaning	**Pandas**	Removed missing values, validated dataset shape.
Train–Test Split	**Scikit-learn**	Divided the data into training and testing sets.

ML Model Training	**Scikit-learn**	Built the regression model.
Evaluation Metrics	**Scikit-learn**	Calculated R^2, MAE, RMSE performance metrics.
Cross-Validation	**Scikit-learn**	Checked model robustness.
Plotting / Visualization	**Matplotlib**	Created performance plots for PPT.
Saving Results	**Pandas**	Saved predicted vs actual results into CSV.

2.3 Machine Learning Pipeline

2.3.1 Dataset Splitting

Data was divided into:

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

unbiased evaluation.

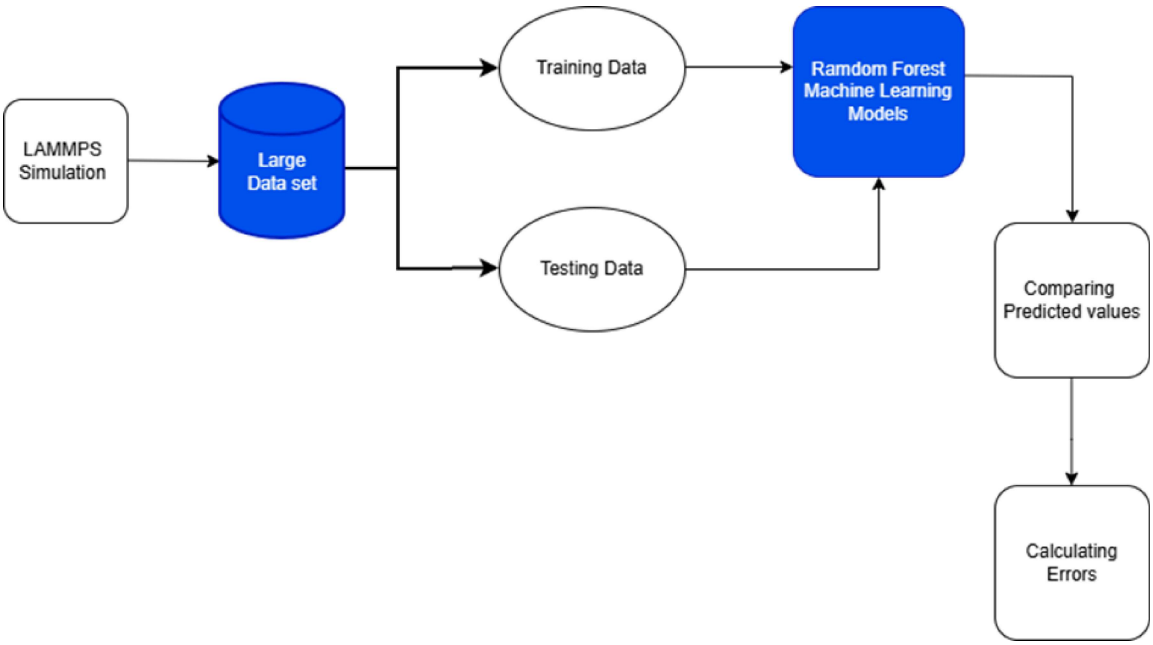
2.3.2 ML Models Used

As the project focused on accurately learning nonlinear relationships between thermodynamic variables and Potential Energy (PotEng), **Random Forest Regressor** was selected as the primary and final model.

Random Forest Regressor

Random Forest is an ensemble-based machine learning algorithm that constructs multiple decision trees and aggregates their outputs. It was chosen because:

- It captures **nonlinear relationships** present in MD simulation data.
- It handles **feature interactions** effectively (e.g., Temp–Density effects on PotEng).
- It is robust to noise in simulation data.
- It provides **feature importance**, helping understand which thermodynamic variables influence Potential Energy the most.



2.3.3 Evaluation Metrics

Model performance was quantified using:

- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)

$$RMSE = \sqrt{\frac{1}{n} * \Sigma (y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{n} * \Sigma |y_i - \hat{y}_i|$$

A lower RMSE indicates that predicted Potential Energy values are close to actual MD values.

• Coefficient of Determination (R^2)

Indicates how well the model explains variability in Potential Energy.
An **R^2 value close to 1** represents high predictive accuracy.

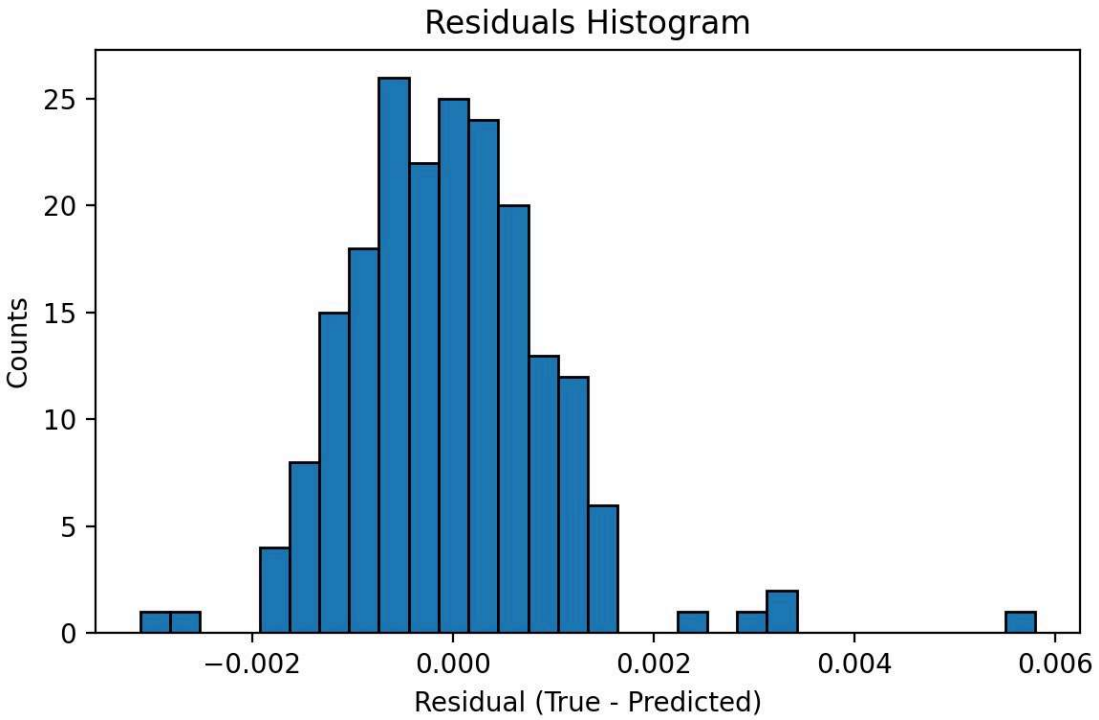
3 Results

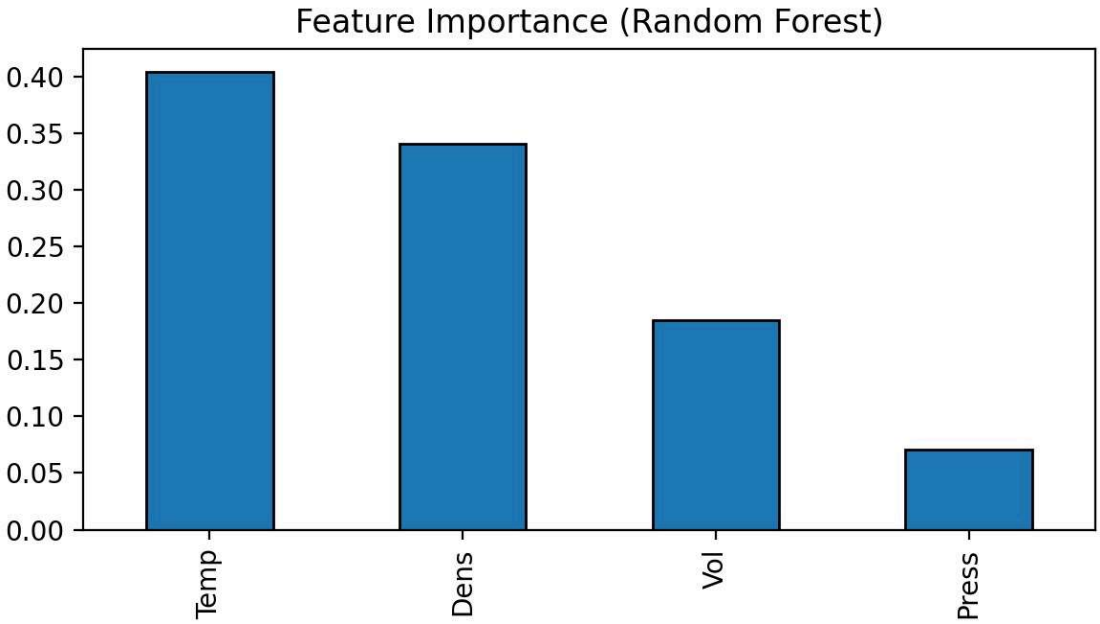
In this project, the Random Forest model achieved:

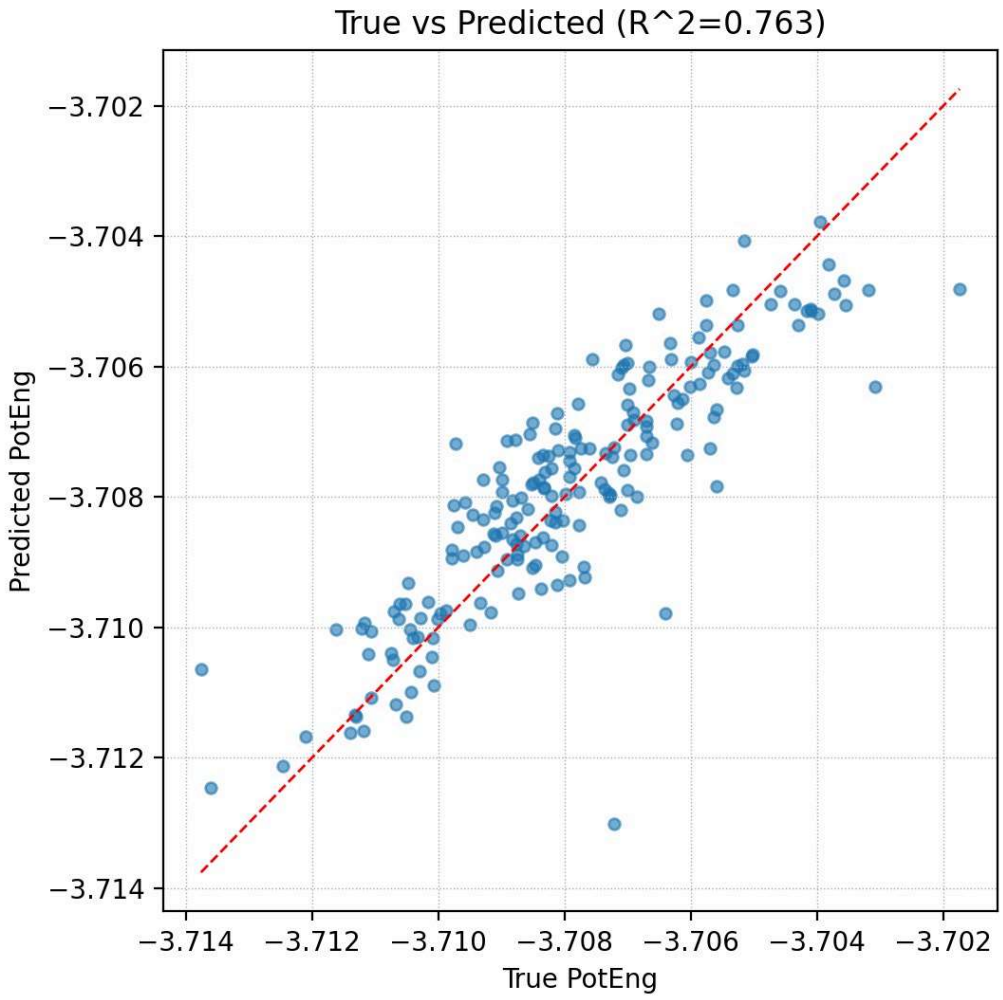
- **$R^2 \approx 0.76$**, meaning the model explains 76% of variance in Potential Energy.
- Residuals were centered around zero, confirming reliable predictions.

Metric	Value	Interpretation
R^2 Score	**0.7626**	The model explains ~76% of the variance in potential energy, showing a strong correlation between input features and predicted energy.
MAE	**0.000768**	The average absolute error is extremely small, meaning predictions are very close to true LAMMPS values.
RMSE	**0.001029**	The low RMSE confirms that the model performs consistently well without large deviations.

Graphs:


Residuals Histogram
Counts
25
20
15
10
5
0
−0.002
0.000
0.002
0.004
0.006
Residual (True - Predicted)


Feature Importance (Random Forest)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Temp
Dens
Vol
Press



Conclusion:

- Successfully generated Cu–Ni alloy data using LAMMPS and extracted thermodynamic features (T, P, V, ρ) along with Potential Energy.
- Built a Machine Learning model (Random Forest) to predict Potential Energy based on MD data.
- The model achieved good accuracy (R^2, RMSE, MAE) and captured the trends between thermodynamic variables and energy.
- ML provides a accurate alternative to MD simulations for estimating Potential Energy, with rapid material analysis .