Building a Machine Learning Model to Interpret X-ray Diffraction Patterns

Mohamed Makhlouf, Priyanshu Agrawal, Matthew Caro, Ankit Cherian, Ethan Funk

UCONN
COLLEGE OF ENGINEERING
SCHOOL OF COMPUTING

Introduction

Our project develops a **machine learning framework** that **interprets X-ray diffraction (XRD) patterns** to **understand material behavior** under extreme conditions. XRD is used by materials scientists to identify crystal structures & determine material properties, such as the phase composition (solid vs liquid).

- Manual XRD analysis is time-consuming, reliant on expert interpretation, and struggles to precisely quantify solid and liquid phase percentages
- Our machine learning approach predicts phase fractions with high accuracy, enabling faster and more precise materials characterization workflows

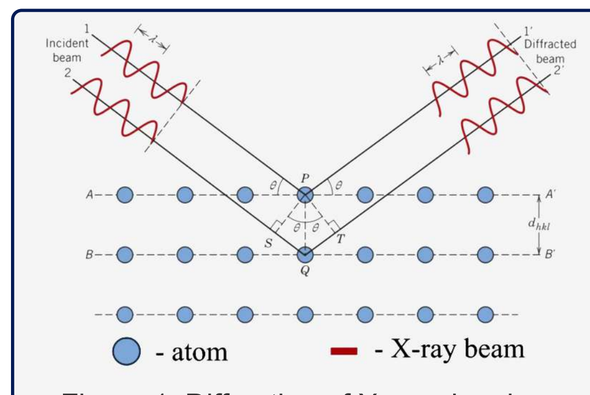
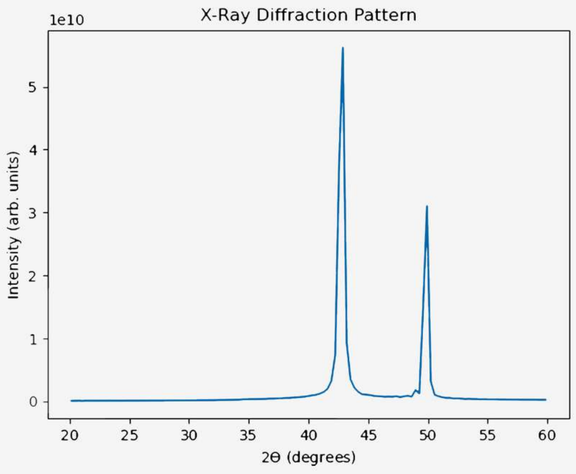


Figure 1: Diffraction of X-rays by plane (Callister & Rethwish)



Machine Learning Model

Predicted Phase Fraction (% Solid)

Figure 2: XRD pattern analysis workflow using machine learning

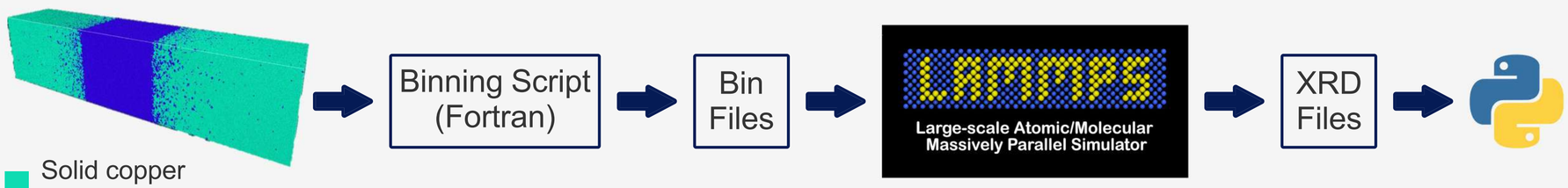
Data Generation

Molecular Dynamics Simulation

Our data comes from molecular dynamics simulations modeling **copper crystal** with temperatures between 300-1200K. A high-temperature heat source (2000K, 2500K, or 3500K) is introduced in the middle, creating varying solid-liquid fractions as heat propagates through the material over time.

XRD Simulation

We take snapshots from the molecular dynamics simulations and divide each one into five regions (bins) using Fortran. For each bin, we compute the true solid fraction and **simulate x-ray diffraction with LAMMPS**. XRD patterns and corresponding solid fractions are processed in Python to build our models.



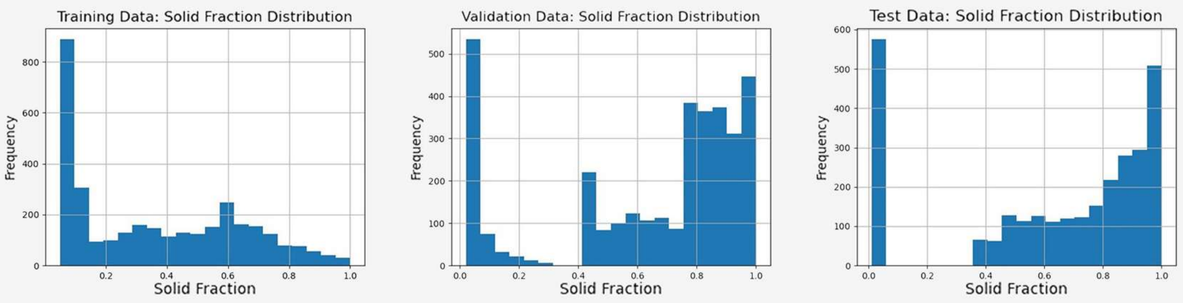
Solid copper (crystalline FCC structure)

Liquid copper (non-crystalline/amorphous structure)

Figure 3: Data generation pipeline from molecular dynamics to XRD patterns

Data Imbalance

Our training data was initially biased toward low solid fractions. To address this, we **under-sampled the top 15% of solid fraction bins** with the most data, creating a more balanced training dataset. This **improved performance** across all models and allowed better generalization to different conditions.



Training 2000K Heat Source | Validation 2500K Heat Source | Test 3500K Heat Source

Figure 4: Distribution of solid fraction values & their frequency in the dataset

Results

Support Vector Regression (SVR)

- SVR with **Radial Basis Function Kernel**
- Tuned via **grid search** on validation dataset
  - Root Mean Squared Error of 0.014 (1.4% error in solid fraction percentage)
  - **R^2 Score: 0.998**
- Most effective model for precise phase fraction predictions

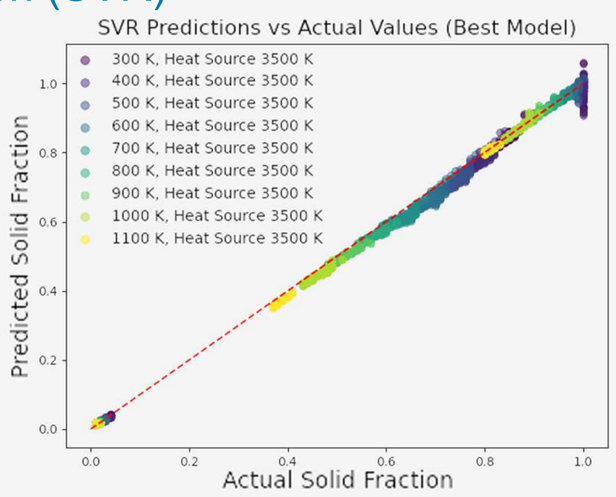


Figure 5: SVR Performance Plot

Convolutional Neural Network (CNN)

- Automatically extracts **spatial features** from diffraction patterns
- 2 convolutional blocks
- 3 fully connected layers
- Hyperparameters and architecture tuned with **keras hyperband tuner**
  - Root Mean Squared Error of 0.024 (2.4% error in solid fraction percentage)
  - **R^2 score: 0.995**

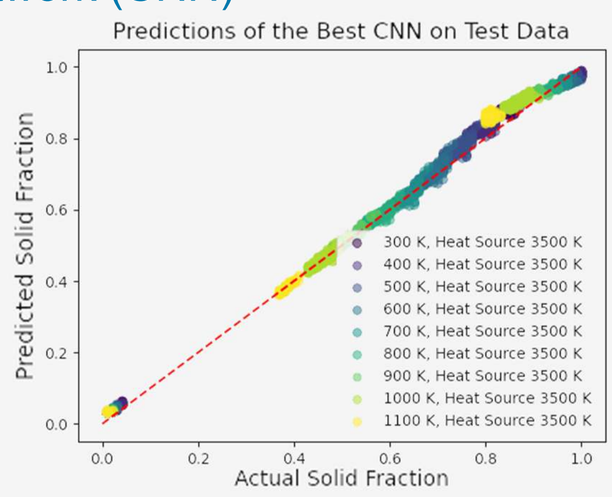


Figure 6: CNN Performance Plot

Other Models

Gradient Boosted Machines (GBM):
- RMSE of 0.042
- **R^2 score: 0.986**

Random Forests (RF):
- RMSE: 0.080
- **R^2 score: 0.948**

Both models are less effective on intermediate values (0.4-0.6 range).

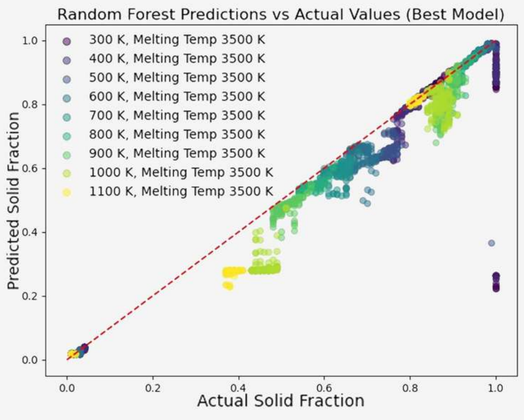


Figure 7: RF Performance Plot

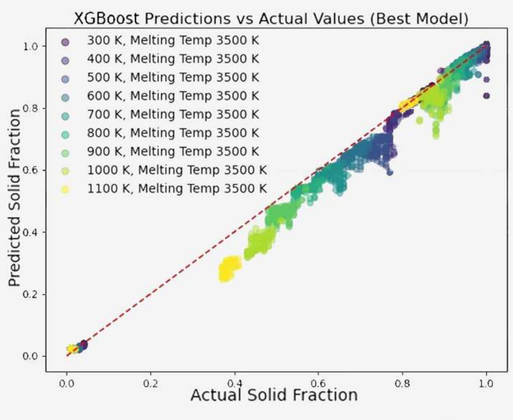


Figure 8: GBM Performance Plot

Conclusions

Developed ML models to predict solid fractions from XRD patterns. Addressing data imbalance with resampling improved model performance. Our models can:

- Generalize well across varying temperature conditions
- Eliminate subjective interpretation
- Provide precise quantitative phase fraction measurements
- Enable automated analysis

Acknowledgements

We thank **Dr. Qian Yang** (School of Computing), **Dr. Avinash M. Dongare** & **Andrew Shortridge** (Group for Extreme Mechanics and Materials Science), for their guidance and support throughout this project. We also acknowledge the computational resources provided by UConn's HPC facility.