CRYSTALGEN

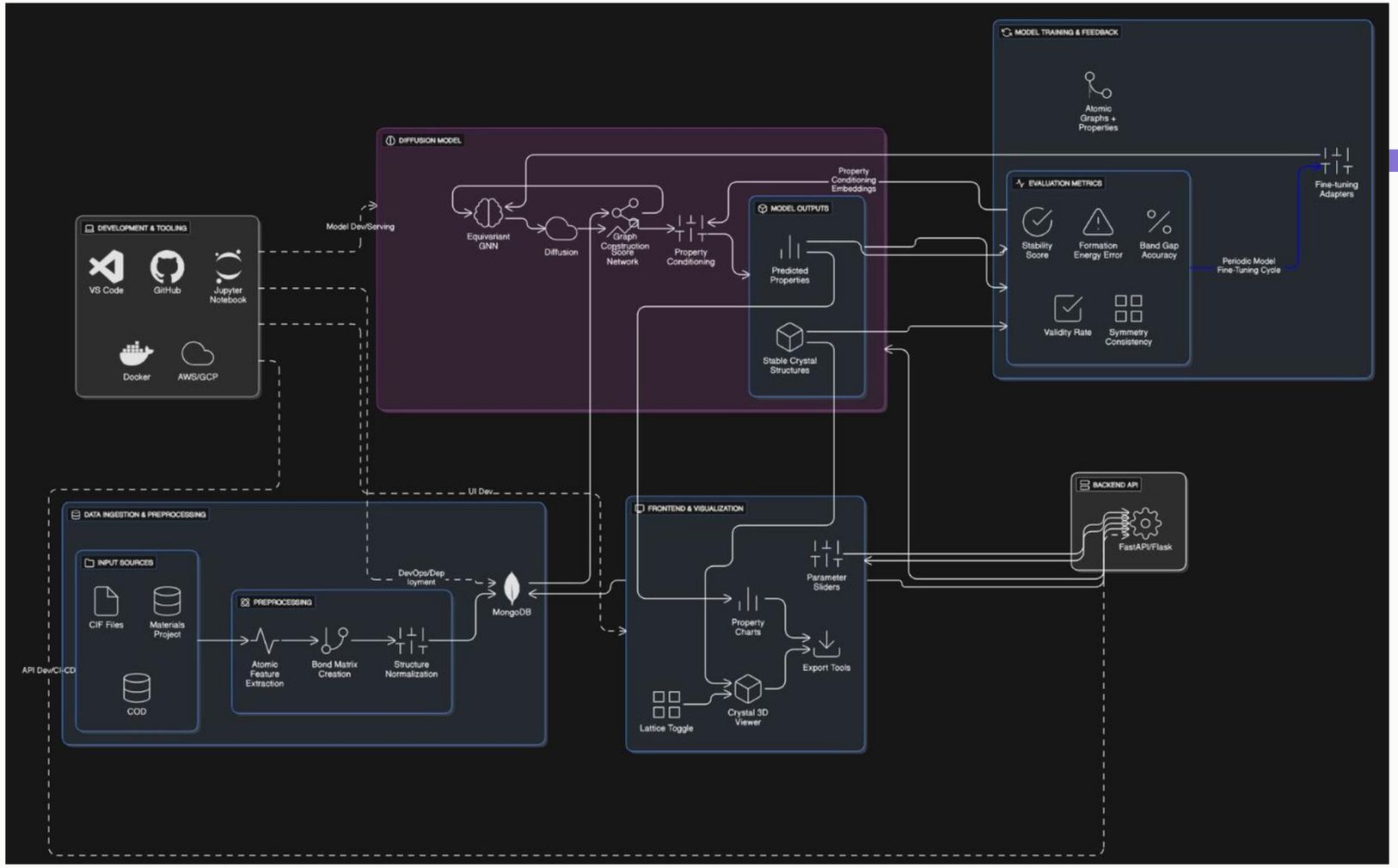
AI-Driven Crystal Structure Prediction

Presented by: Sadhvini
Roll No: 23BD1A0558
Team No: G450

Overview

CrystalGen is an AI-driven platform that predicts and analyzes crystal structures from CIF data using Graph Neural Networks (GNNs). It automates crystal analysis to deliver fast and accurate insights into structural and chemical properties, reducing manual effort and computation time.

Architecture Diagram



CIF File → Structure Parser (Pymatgen) → Graph Constructor (PyG) → GNN Feature Encoder → Diffusion-based Generator → Crystal Structure Output → 3D Visualization → Flutter User Interface

TECH Stack

Frontend

React – for an interactive and responsive user interface

Real-time 3D visualization of crystal structures

Backend

Flask – lightweight REST API framework for handling model requests

MongoDB – NoSQL database for storing CIF data and crystal predictions

AI / Model Layer

PyTorch – for deep learning and model training

Graph Neural Networks (GNNs) – to learn atomic interactions

Diffusion Models – to generate realistic crystal structures

Development Tools

VS Code, Jupyter Notebook, GitHub for development and version control

Mathematical Foundation

Reparameterization (CVAE)

Sample latent z from encoder outputs

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

KL Divergence

Measures how the learned latent distribution diverges from the standard normal distribution.

$$\mathcal{L}_{KL} = -\tfrac{1}{2} \sum (1 + \log \sigma^2 - \mu^2 - \sigma^2)$$

Total Loss Function

Combines reconstruction errors and KL regularization to train the autoencoder effectively.

$$\mathcal{L} = \mathcal{L}_{lattice} + \mathcal{L}_{frac} + \mathcal{L}_{species} + 10^{-3}\mathcal{L}_{KL}$$

Reconstruction Losses

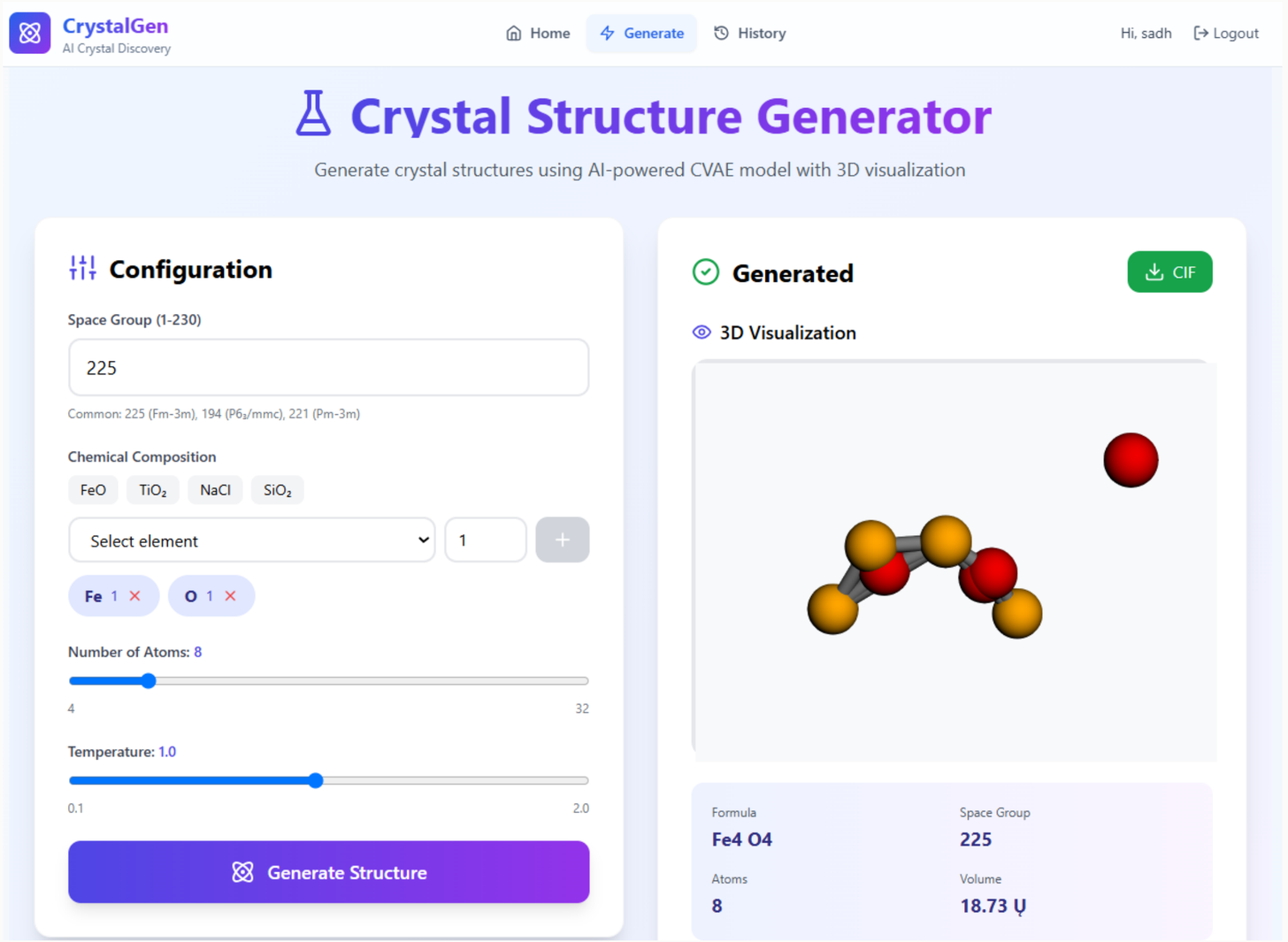
Lattice: MSE between predicted and true lattice
Fractional coords: MSE (only on real atoms)
Species: Cross-entropy between predicted and true elements

Sampling (Generation)

Generates new crystal structures by decoding random latent vectors sampled from a normal distribution.

$$z \sim \mathcal{N}(0, I); \quad (\hat{L}, \hat{F}, \hat{S}) = \text{Decoder}(z)$$

Implementation



Improvements

Our Work

Enhanced the baseline model using Graph Neural Networks for atomic interaction learning. Integrated CIF preprocessing, Flask–React UI, and 3D visualization for seamless analysis.

Key Learnings

Understood how graph-based models capture atomic relationships effectively. Gained insights into latent space encoding and hands-on experience in model tuning.

Potential Improvements

Incorporate Diffusion or Transformer models for higher accuracy.
Add stability validation and expand the dataset with complex materials.

Real-World Applications

Used for new material and crystal discovery in chemistry and physics.
Applicable to drug design, semiconductors, and energy materials.

Performance Analysis

Achieved lower reconstruction loss and improved element prediction accuracy.
Optimized preprocessing and GPU support ensured faster inference.

Challenges Faced

Faced CIF parsing and training instability, resolved using pymatgen and KL tuning.
Overcame visualization issues by adding Py3Dmol 3D rendering.

Limitations

Model trained on a limited dataset, reducing generalization.
Lacks physical validation like energy or stability checks.

Conclusion

CrystalGen successfully demonstrates how AI and Graph Neural Networks can revolutionize crystal structure prediction and generation.

By automating complex crystallographic analysis, it enables faster, data-driven material discovery with real-time 3D visualization.

THANK YOU