Project Proposal

Discovery of Eutectic Compositionally Complex Alloys by Generative Machine Learning

Department of Metallurgical and Materials Engineering VNIT, Nagpur

Date: August 5, 2025

Student details: 1) Adarsh Varma (BT23MME105)
2) Keshav Gupta (BT23MME100)

1. Introduction and Project Rationale

Eutectic alloys, characterized by their unique melting behaviour and fine lamellar or acicular microstructures, offer a compelling combination of properties such as high strength, good ductility, excellent castability, and often superior wear resistance. Traditionally, eutectic compositions are found in binary or ternary alloy systems. However, the emerging field of **Compositionally Complex Alloys (CCAs)**, including High-Entropy Alloys (HEAs), presents an immense, largely unexplored compositional space for discovering novel eutectic systems. These multi-principal element alloys offer the potential for unprecedented property combinations, but identifying their specific eutectic points through conventional experimental methods is a monumental, often intractable, task due to the sheer number of possible elemental combinations and concentrations.

This project proposes a pioneering approach: leveraging **Generative Machine Learning (ML)** to intelligently explore this vast compositional landscape and accelerate the discovery of new eutectic CCAs. Unlike traditional predictive ML models that forecast properties of *known* materials, generative models can *propose* entirely new alloy compositions that are predicted to exhibit eutectic behaviour. This data-driven strategy promises to dramatically reduce the experimental burden, guide synthesis efforts, and unlock a new generation of high-performance materials with tailored microstructures and properties. This initiative represents a fusion of advanced materials science with cutting-edge artificial intelligence, aiming to revolutionize the alloy design process.

2. Problem Statement and Research Gap

The identification of eutectic compositions in multi-component alloy systems is fundamentally challenging. A eutectic point represents a specific composition at which a liquid phase transforms directly into two or more solid phases simultaneously upon cooling, resulting in a distinct microstructure and often enhanced properties. As the number of constituent elements increases in CCAs, the complexity of their phase diagrams explodes, making the experimental mapping of eutectic valleys or points computationally and experimentally prohibitive. Traditional thermodynamic calculations can assist, but they still require significant user input and may struggle with the sheer scale of the compositional space in CCAs.

Currently, the discovery of eutectic CCAs is largely serendipitous or based on limited, intuition-driven exploration. A significant research gap exists in developing an automated, intelligent framework that can:

1. **Generate novel alloy compositions** that are likely to exhibit eutectic behavior.

2. **Learn the underlying compositional rules** that lead to eutectic formation in complex systems.
3. **Guide experimental synthesis** by proposing promising candidates.

This project aims to bridge this gap by developing a generative machine learning framework. This framework will learn from existing eutectic and non-eutectic data to *propose* new, high-probability eutectic CCA compositions, thereby transforming the discovery process from a laborious search to an intelligent generation.

3. Project Objectives

This project is designed with the following specific objectives:

- **Comprehensive Data Collection and Representation:** To systematically gather and organize a diverse dataset of known eutectic (HEA's) and their associated phase formation characteristics from published literature and materials databases. This will involve representing compositions and elemental properties effectively for ML.
- **Generative Model Development:** To design, train, and validate a generative machine learning model (e.g., Variational Autoencoder (VAE) or Generative Adversarial Network (GAN)) capable of learning the distribution of existing alloy compositions and *generating* novel, chemically valid, and potentially eutectic compositions.
- **Eutectic Property Prediction Integration:** To develop or integrate a predictive machine learning model that can assess the likelihood of a generated composition being eutectic, based on its elemental properties and predicted phase behavior. This model will act as a "discriminator" or "evaluator" for the generative output.
- **Compositional Space Exploration:** To utilize the trained generative model to systematically explore the vast compositional space of CCAs, generating a list of candidate eutectic alloy compositions that are novel and have a high predicted probability of forming a eutectic microstructure.
- **Interpretability and Design Principles:** To apply Explainable AI (XAI) techniques to understand which elemental combinations and features the generative model prioritizes when proposing eutectic compositions, thereby extracting new design principles for CCAs.

4. Methodology: A Generative Data-Driven Discovery Approach

Our methodology will integrate data engineering with advanced generative and predictive machine learning algorithms to discover novel eutectic CCAs.

4.1. Data Acquisition and Representation

The project will begin by compiling a comprehensive dataset. This will involve:

- **Literature Review:** Manual extraction of data on multi-component alloy systems, specifically identifying compositions known to exhibit eutectic behavior, and also collecting data on non-eutectic compositions to provide contrast.
- **Materials Databases:** Leveraging publicly available materials databases (e.g., Materials Project, specific alloy databases) to gather elemental properties and potentially phase information.

For each alloy entry, we will collect:

- **Elemental Composition:** Atomic percentages of constituent elements.
- **Phase Information:** Categorical labels indicating whether the composition is eutectic, hypoeutectic, hypereutectic, or forms other phases (e.g., solid solution, intermetallic).
- **Elemental Descriptors:** For each element, properties like atomic number, atomic weight, melting point, electronegativity, atomic radius, valence electron count.

Representation: Compositions will be represented as vectors of elemental percentages. Additionally, we will engineer features based on the average and differences of elemental descriptors (e.g., average electronegativity, mixing entropy, mixing enthalpy) for the predictive model.

4.2. Generative Model Development

The core of this project is the generative model. We will explore:

- **Variational Autoencoders (VAEs):** A type of neural network that learns a compressed "latent space" representation of the input data. We can then sample from this latent space to generate new, valid compositions.
- **Generative Adversarial Networks (GANs):** Comprising a "generator" that creates new data and a "discriminator" that tries to distinguish real from generated data. The two networks compete, leading the generator to produce increasingly realistic (eutectic-like) compositions.

The generative model will be trained on the collected dataset of alloy compositions to learn the underlying distribution of known alloys, with a focus on compositions that tend towards eutectic formation.

4.3. Eutectic Property Prediction (Discriminator/Evaluator)

Alongside the generative model, a separate **predictive machine learning model** (e.g., a classifier like Random Forest or a neural network) will be trained. This model's role is to:

- **Predict Eutectic Likelihood:** Given a new alloy composition (either from the generative model or a real one), predict the probability of it being eutectic.
- **Guide Generative Training:** In a GAN setup, this model acts as the discriminator. For VAEs, its predictions can be used to filter or prioritize generated compositions.

4.4. Compositional Space Exploration and Optimization

Once the generative and predictive models are trained:

- **Generation of Candidates:** The generative model will be used to produce many novel alloy compositions.
- **Filtering and Scoring:** The predictive model will then evaluate these generated compositions, assigning a "eutectic score" or probability.
- **Diversity and Novelty:** We will employ techniques to ensure the generated candidates are not just slight variations of existing alloys but truly novel and diverse.

- **Iterative Refinement (Optional but Recommended):** The most promising generated compositions could potentially be fed back into the generative model's training loop (if a reinforcement learning approach is adopted) or used to refine the search space.

4.5. Interpretability and Design Principles

To provide actionable insights, we will use XAI techniques:

- **SHAP (SHapley Additive exPlanations):** To understand which elemental properties or combinations are most influential in the predictive model's assessment of eutectic likelihood.
- **Analysis of Latent Space (for VAEs):** Visualizing the latent space can reveal clusters of similar alloys and help navigate towards eutectic regions.
- **Feature Importance:** From the predictive model, identify the key elemental descriptors that correlate strongly with eutectic formation.

5. Data Sources

The project will rely on data extracted from:

1. **Peer-Reviewed Scientific Literature:** Journals focusing on phase diagrams, alloy development, and high-entropy alloys (e.g., *Acta Materialia*, *Journal of Alloys and Compounds*, *Calphad*, *Materials Science and Engineering A*).
2. **Thermodynamic Databases:** CALPHAD-type databases (e.g., TCFE, TTNI, SGTE) can provide calculated phase equilibria data, which can be used to augment experimental data or generate synthetic data for training.
3. **Open-Access Materials Databases:** Platforms like Materials Project or OQMD, which provide calculated properties and phase information for various compounds, can be a source for elemental properties and potentially some multi-component alloy data.
4. **Specialized Eutectic Databases:** If any specific databases for eutectic compositions in complex systems exist, they will be utilized.

6. Expected Outcomes and Deliverables

6.1. Expected Outcomes

Upon successful completion, this project is anticipated to yield the following significant outcomes:

- **Trained Generative Model:** A robust generative machine learning model capable of proposing novel, chemically valid alloy compositions with a high predicted probability of exhibiting eutectic behavior.
- **Eutectic Prediction Model:** An accurate predictive model that can classify whether a given complex alloy composition is likely to be eutectic or not.
- **List of Candidate Eutectic CCAs:** A prioritized list of novel, computationally discovered eutectic Compositionally Complex Alloy compositions, ready for potential experimental validation.
- **New Design Principles:** Interpretable insights into the elemental and thermodynamic factors that drive eutectic formation in multi-component alloy systems, providing fundamental knowledge for future alloy design.

- **Accelerated Materials Discovery Framework:** A validated data-driven framework that significantly accelerates the search for and design of new eutectic alloys, reducing the need for extensive trial-and-error experimentation.

6.2. Deliverables

The tangible outputs of this project will include:

- **Comprehensive Project Report:** A detailed technical report outlining the project's motivation, methodology (including generative and predictive model architectures), data collection and representation, results, discussion of generated candidates, and conclusions. This report will include relevant figures, tables, and a thorough discussion of findings.
- **Curated Alloy Dataset:** A meticulously compiled and cleaned dataset of alloy compositions and their phase information, suitable for further research.
- **Python Codebase and Notebooks:** A well-documented and organized collection of Python scripts and Jupyter notebooks containing all code for data acquisition, preprocessing, generative model training, predictive model training, composition generation, evaluation, and interpretation.
- **Presentation Slide Deck (if required):** A professional presentation summarizing the project's objectives, methodology, key results, and contributions, suitable for academic presentation.

7. Proposed Timeline

The project is planned to be completed over an 11-week period, structured into distinct phases to ensure efficient progress and timely achievement of milestones:

Phase	Duration	Milestone
Literature Review & Data Collection	Week 1-2	Initial Alloy Dataset Compiled
Data Representation & Feature Engineering	Week 3-4	Data Ready for Generative & Predictive Models
Generative Model Development & Training	Week 5-6	Initial Generative Model Capable of Composition Generation
Predictive Model Integration & Refinement	Week 7-8	Predictive Model for Eutectic Likelihood Trained & Integrated
Composition Generation & Analysis	Week 9-10	Candidate Eutectic CCAs Generated & Prioritized
Final Report & Submission Preparation	Week 11	Complete Project Submission

8. Significance and Future Directions in Materials Science

This project offers profound academic and practical significance, pushing the boundaries of materials science through the application of advanced AI:

- **Pioneering Generative Materials Design:** This project is at the forefront of materials informatics, moving beyond prediction to *generation*, which is crucial for true materials discovery. It establishes a framework for designing materials with desired properties from scratch.
- **Accelerating Eutectic Alloy Discovery:** By intelligently navigating the vast compositional space, the project can dramatically accelerate the discovery of novel eutectic CCAs, which would be nearly impossible through traditional methods.
- **Unlocking New Material Properties:** Eutectic CCAs hold immense promise for superior mechanical, thermal, and functional properties. This project directly contributes to unlocking these next-generation materials for demanding applications in aerospace, automotive, energy, and biomedical fields.
- **Fundamental Understanding:** The interpretability aspects will provide novel insights into the complex phase formation mechanisms in multi-component systems, potentially revealing new design rules for eutectic alloys.
- **Foundation for Autonomous Materials Research:** The framework developed here can serve as a foundational step towards fully autonomous materials discovery pipelines, where AI systems can propose, synthesize (virtually or robotically), and characterize new materials with minimal human intervention.
- **Broader Impact on Phase Diagram Prediction:** The methodologies for learning phase relationships can be extended to improve the accuracy and efficiency of predicting phase diagrams for other complex alloy systems.

This project offers a stimulating and highly relevant opportunity to contribute to the advancement of materials science through the innovative application of generative machine learning.