DATA SCIENCE & AI LAB (BSCSS3001)

MILESTONE - 3

GROUP NO. 11

Gaali Gaurav Krishna (**21f2000631**)
Dr Subhas Ganguly (**21f1000265**)
Dinesh Kumar Kumawat (**21f1001956**)

IITM BS Degree Program
Indian Institute of Technology,
Madras, Chennai,
Tamil Nadu, India, 600036

1 Context: The Problem in Microstructure Analysis

The mechanical properties of steel—such as strength, ductility, and toughness—are intrinsically linked to its **microstructure**, which is defined by the type, shape, and distribution of its constituent phases. In this project, we are focused on classifying five critical steel phases: **ferrite, pearlite, austenite, bainite, and martensite**.

1.1 The Challenge of Manual Analysis

Traditionally, the quantitative analysis of these microstructures is performed manually by trained metallurgists using optical microscopy. This process suffers from several critical drawbacks that limit high-throughput material development:

1. **Subjectivity and Ambiguity:** Phase identification, especially distinguishing between similar phases like bainite and martensite, is highly dependent on the individual metallurgist's experience and judgment, leading to **ambiguous and irreproducible results**.

2. **Time and Cost:** Manual image processing, segmentation, and phase quantification are extremely **time-consuming** and **labor-intensive**, making them impractical for the massive datasets generated in modern material research and quality control (high-throughput analysis).

3. **Scale Inefficiency:** Manual techniques are poorly equipped to analyze the thousands of micrographs needed to build statistically robust structure-property relationships.

1.2 The Proposed Deep Learning Solution

To overcome these limitations, our project proposes a **deep learning framework** to achieve automatic, objective, and high-throughput classification of steel microconstituents, moving the field of materials science toward true **computer vision of microstructure**.

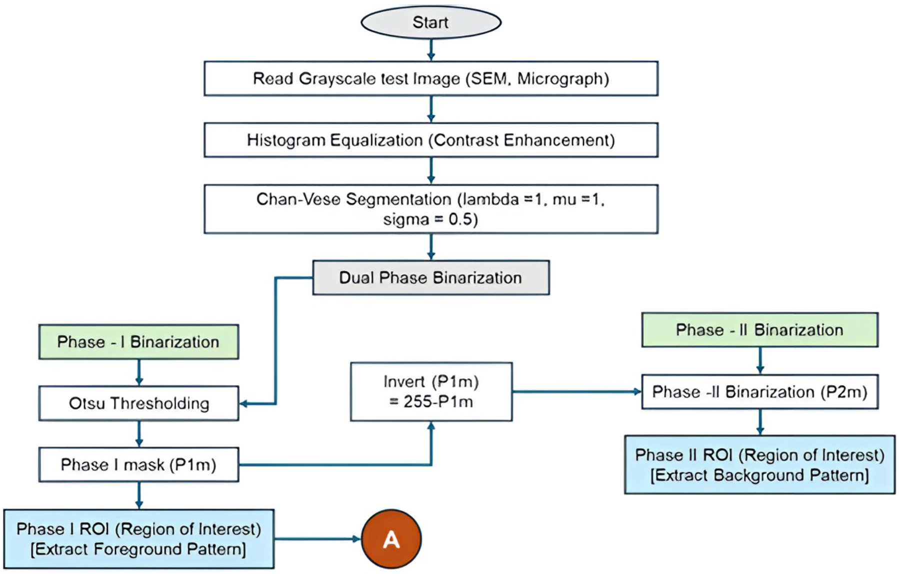

Figure 1: Schematic Procedure of the Two-Stage Deep Learning Framework.

1.3 The Role of Chan-Vese Segmentation

The initial step in our two-stage framework is to ensure that the image patches fed to the CNN models contain relevant microstructural data and not unnecessary background or noise. This is achieved using the **Chan-Vese segmentation algorithm**.

- **Model Type:** Chan-Vese is an energy minimization method, falling under classical computer vision, not deep learning. It is an active contour model (snake model) that does not rely on image gradients.

- **Working Principle:** It works by iteratively minimizing an energy function that aims to find a contour (a closed curve) that separates the image into two regions (foreground and background) based on the *average intensity* within those regions.
- **Purpose in our Framework:** The primary function of Chan-Vese is **pre-segmentation** and **data cleaning**. It is used to automatically identify the boundaries of the main phases in the micrograph, allowing us to:
    1. Extract image patches that are rich in microstructural information.
    2. Filter out irrelevant regions (e.g., sample preparation artifacts or regions of poor contrast) that could confuse the subsequent deep learning classifier.

2 Justification of Model Architecture via Literature Survey

The primary objective of this project is to develop a high-throughput framework for the automatic classification of five distinct steel microstructural phases (ferrite, pearlite, austenite, bainite, and martensite). Based on our comprehensive literature survey, the task of classifying image patches extracted from micrographs is best suited for Deep Convolutional Neural Networks (CNNs).

The survey highlighted a critical trend in materials science imaging: the move from traditional machine learning classifiers (like SVM or k-NN) to deep architectures that automatically learn intricate, hierarchical feature representations directly from raw image data.

2.1 Key Architectural Considerations from Literature

The choice of our candidate models is based on their historical impact and current state-of-the-art performance in computer vision, particularly texture and detail recognition relevant to microscopy.

1. **VGG-19 (Visual Geometry Group, Oxford, 2014)**
    - **Focus: The Necessity of Depth.** VGG models were foundational in proving that depth is a more critical factor than filter size.
    - **Uniqueness:** It uses a highly regular, uniform architecture composed exclusively of small 3×3 **convolutional kernels** stacked deeply. This simplicity and the ability to capture fine-grained, localized texture features make it an excellent candidate for the highly texture-dependent classification of steel microstructures.
2. **ResNet-50 (Deep Residual Network, Microsoft Research, 2015)**
    - **Focus: Addressing the Gradient Problem.** As models deepen past 20 layers, training often becomes unstable.
    - **Uniqueness:** ResNet-50 introduced **Skip Connections** (or residual blocks) which allow the network to bypass layers and reuse previous feature maps. This mechanism facilitates the training of ultra-deep networks (like ResNet-50) that can capture highly abstract features, which may be crucial for distinguishing subtle phase boundaries between complex phases like bainite and martensite.
3. **EfficientNetB3 (Google, 2019)**
    - **Focus: Balancing Efficiency and Accuracy.** Modern analysis demands not only high accuracy but also efficient inference time.
    - **Uniqueness:** EfficientNet introduced a novel **Compound Scaling Method** that uniformly scales all three dimensions of the network—depth, width, and image resolution—using a single set of coefficients. This optimized balance often leads to better performance with significantly fewer parameters and lower computational cost compared to ResNet or VGG, making it a strong candidate if fast, high-throughput analysis is prioritized.

2.2 Proposal for Comparative Testing

Given the balanced evidence in the literature regarding the trade-offs between architectural complexity, training speed, and ultimate accuracy, our current approach is to conduct a head-to-head comparison on a subset of our training data before finalizing the production model.

We are currently investigating three leading candidates:

Table 1: Candidate Models and Testing Status

Candidate Model	Rationale for Inclusion
VGG-19	Simplicity, proven feature extraction capability for texture analysis, and strong performance a
ResNet-50	Superior depth via residual blocks, which may be necessary to distinguish the subtle differenc
EfficientNetB3	Efficiency and robust generalization; a strong candidate if computational resources become a

3 Proposed Model: VGG-19 (Initial Focus)

Based on our initial benchmark results, the **VGG-19** architecture is the first of our candidates to show highly promising performance, making it the **initial focus** for full tuning and optimization. While comparative tests are ongoing, VGG-19's deep stack of 3×3 convolutional layers appears well-suited to capturing the fine-grained, textural features inherent in steel micrographs.

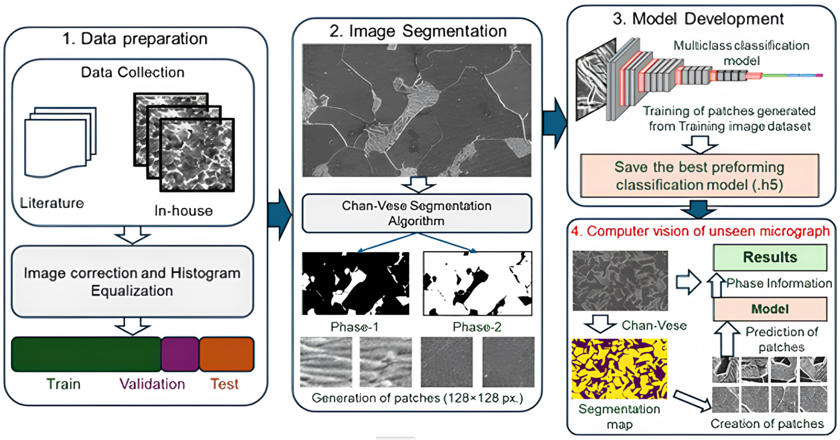

Figure 2: Schematic Procedure of the Two-Stage Deep Learning Framework.

3.1 General VGG-19 Details

The VGG-19, which will form the backbone of our classification phase, uses a uniform architectural motif:

- **Convolution Blocks:** 16 convolutional layers, grouped into five blocks separated by max-pooling layers. All convolutional layers use 3×3 filters with a stride of 1.
- **Feature Extraction:** The small filter size allows the model to capture features at multiple scales as the network deepens, a key advantage for microstructural analysis where feature scale is not uniform.
- **Classification Head:** The architecture terminates with three fully connected (FC) layers, with the final layer using a **SoftMax** activation for the five-class probability distribution.

4 Tuning and Training Plan (Next Steps)

With VGG-19 identified as our primary focus based on initial data, the next critical phase involves comprehensive hyperparameter tuning to optimize its performance before scaling the training to the entire dataset. The following strategy will be applied consistently to **all three candidate models** once they progress to this phase.

4.1 Optimization Strategy

We will employ the **Adam optimizer** due to its adaptive learning rate capabilities, minimizing the **Binary Cross-Entropy** loss function.

4.2 Hyperparameter Grid Search

A rigorous grid search will be implemented on a dedicated validation set to find the optimal configuration. The key hyperparameters to be tuned are:

Table 2: Hyperparameter Grid Search Parameters

Hyperparameter	**Values Tested**	**Purpose**
Learning Rate (α)	0.01, 0.001, 0.0001	Controls convergence speed
Batch Size	5, 10	Depends on GPU memory
Dropout Rate	0.1, 0.3, 0.6	Prevents overfitting

4.3 Data Augmentation

Data augmentation (e.g., random rotation, flips, slight zoom) will be applied *exclusively* to the training set to improve the model's robustness and ability to generalize across different image orientations and contrast variations inherent in metallographic sample preparation.

Current Status: We have completed the data preprocessing (patch extraction) and are currently in the process of running the first phase of the VGG-19 grid search.