DATA SCIENCE & AI LAB (BSCSS3001)

MILESTONE - 6

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

Automatic Classification of Steel Microstructures Using Deep Learning

Abstract

This project presents an automated deep learning system designed to classify steel microstructures into five phases:

Austenite, Bainite, Ferrite, Martensite, and Pearlite.

The solution leverages transfer learning using VGG19 and ResNet50, combined with extensive preprocessing, hyperparameter tuning, and deployment via a Streamlit web application.

The pipeline covers all stages from dataset creation and preprocessing to model development, evaluation, and deployment.

Experiments were conducted with different learning rates, optimizers, batch sizes, dropout rates, and fine-tuning strategies. The best-performing configuration achieves reasonable test accuracy and macro-averaged F1-score, demonstrating that deep learning can support microstructure classification in a practical engineering setting. The final deliverable is a usable application that allows domain experts to upload micrographs and receive phase predictions with confidence scores.

1. Introduction

Microstructure analysis is central to materials science and metallurgical engineering. The microscopic arrangement of phases in steel—such as ferrite, pearlite, martensite, bainite, and austenite—determines critical mechanical properties including strength, toughness, and ductility.

Traditionally, microstructural examination is performed manually by experts using optical or scanning electron microscopy. This process is time-consuming, requires extensive training, and often suffers from inter-observer variability.

With the increasing availability of digital microscopy and computing power, there is a strong motivation to develop automated systems for microstructure classification. Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful tool for image-based tasks. This project aims to harness CNN-based transfer learning to automatically classify steel micrographs into five key phases and to deploy the resulting model as a user-friendly application.

Key objectives:

1. Build a robust dataset and preprocessing pipeline for 5-class microstructure classification.
2. Use transfer learning (VGG19, ResNet50) with systematic hyperparameter tuning.
3. Evaluate the models using multi-class metrics and confusion matrix analysis.
4. Deploy the best-performing model via a Streamlit app with clear documentation.

2. Literature Review (Milestone 1)

Microstructure analysis in materials science has traditionally relied on classical image processing tools such as **ImageJ**, which excel at tasks like segmentation, thresholding, and quantitative measurements.

While these tools are highly effective for **segmentation**, they lack **automated classification** capabilities. As a result, **manual phase identification** continues to dominate metallurgical workflows. This manual process is not only time-consuming but also heavily dependent on operator training and subjective interpretation, often leading to inconsistency across laboratories or observers.

Recent research has shifted towards **benchmarking both segmentation and classification accuracy** using advanced computational methods:

- **Benchmarking efforts** focus on comparing classical image processing approaches with modern deep learning architectures for segmentation quality.
- For classification, studies commonly evaluate CNN-based models against traditional feature-engineering pipelines, showing clear performance gains with deep learning.

The literature review for this project spans three key areas:

1. **Deep learning applications in materials science**, especially the use of pretrained networks (VGG, ResNet, EfficientNet) for microstructure recognition and texture-based classification.
2. **Traditional metallographic techniques**, including thresholding, filtering, GLCM/LBP-based texture extraction, and morphological segmentation.
3. **Segmentation methods for microstructure analysis**, such as watershed, region growing, and more recent CNN-based semantic segmentation networks.

Novelty and Contribution

This project advances beyond existing approaches by integrating **both segmentation and classification within a single automated workflow**, overcoming a notable

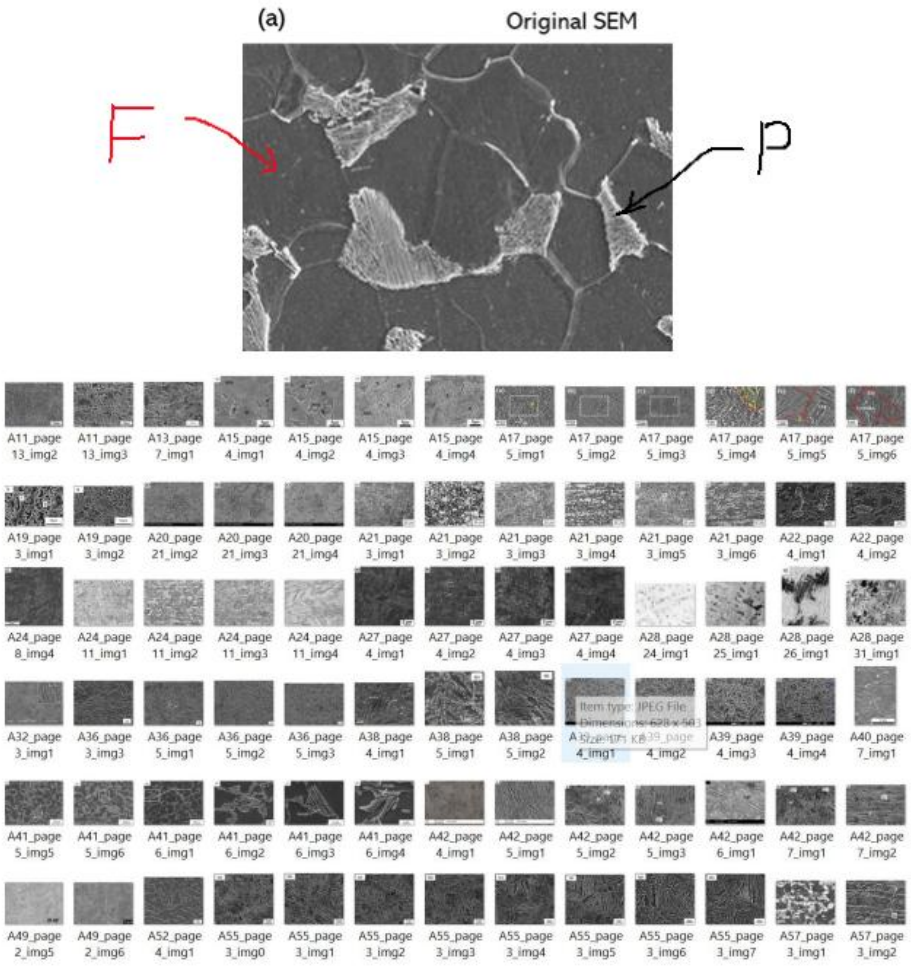
limitation in industry-standard tools like ImageJ. Moreover, while prior studies commonly focus on conventional steels, **this work centers specifically on Quench and Partition (Q&P) steel**, a cutting-edge high-strength material system with limited prior computational or deep learning-based analysis. This makes the project not only technically relevant but also scientifically meaningful, addressing a clear research and industry gap.

3. Dataset and Methodology (Milestones 2–3)

3.1 Dataset Collection

A total of **341 SEM microstructure images** were collected after screening **160+ published research articles**. These images represent five key steel phases: *Austenite, Bainite, Ferrite, Martensite,* and *Pearlite*. Only high-quality SEM images were retained; low-contrast or optical microscope images were removed.

A sample snapshot of the data folder is presented here along with a sample microstructure. The completely collected images cannot be shared publicly as they are copyrighted with publisher.



3.2 Preprocessing

To standardize image quality and improve model performance, the following preprocessing steps were applied:

- **Histogram equalization** to normalize illumination across all micrographs
- Removal of **low-quality or noisy images**
- Conversion to **consistent resolution and color format** (224×224, 3-channel RGB)

These steps ensured cleaner inputs for segmentation and patch generation.

Histogram equalization was used to minimize the impact of fluctuating light on the segmentation and classification outcomes by standardizing the fundamental intensity features across all images. To process a grayscale micrograph, let's look at the variable I, which represents the intensities. With I = 0 denoting black and I = L-1 (i.e., 255) denoting white, an image's I should fall between [0, L-1]. If a histogram h(I) represents an image's intensity distribution, then the cumulative distribution function for each gray value (x) can be obtained as:

$$cdf(x) = \sum_{I=1}^{x} \quad h(I) \qquad (1)$$

Then, the general histogram equalization $h(v)$ formula is

$$h(v) = round\left(\frac{cdf(x)-cdf_{min}}{(M\times N)-1} \times (L-1)\right) \qquad (2)$$

where cdf_{min} the minimum non-zero value of the cumulative distribution function. M, N are the length, width of the gray-scale image. L is the number of gray levels used (i.e., 256).

Furthermore, microstructure images that were taken using optical microscopy or had a lot of blur and didn't fit the standards for clarity and resolution were routinely removed from the dataset. To prevent data leaking, the SEM dataset was split into three sections: training (70%), testing (20%), and validation (10%). The SEM-image level was used for all splits, 277 distinct micrographs were used to create the patches, and each patch is identified by its source image ID. To ensure that no patches from the same SEM micrograph appeared in more than one set, these image IDs—rather than patches—were divided into training, validation, and test sets. The testing dataset was kept apart for the sole purpose of assessing the model. To guarantee reproducibility, data augmentation was used upon splitting and the split indices were made public. The accuracy and robustness of the study's conclusions are increased by this exact preprocessing method, which guarantees that the residual images are of excellent quality and appropriate for further analytical processes.

3.3 Patch Creation Process

The creation of high-quality microstructure patches was a core component of the dataset methodology. The full patch extraction pipeline consisted of several sequential stages: label assignment, segmentation, bounding-box cropping, and patch extraction, all derived from SEM micrographs.

3.3.1 Label Assignment

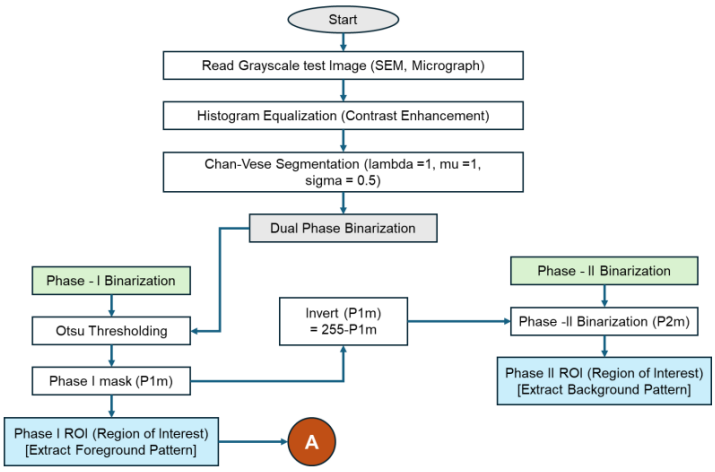
A global label counter function was used to track the number of patches generated for each phase (ferrite, pearlite, martensite, etc.).

- The function increments counters such as *ferr_c*, *parlite_c*, and *morten_c* based on the label extracted from the filename.
- This ensured balanced patch generation across all microstructure classes.

3.3.2 Segmentation Using Chan–Vese

Each SEM micrograph was segmented using the **Chan–Vese active contour model**, which is effective for separating regions with different average intensities. Key steps included:

- Histogram equalization to improve contrast
- Evolution of the active contour to produce a **final binary mask**
- Generation of two complementary masks (**phase1** and **phase2**)



3.3.3 Bounding Box Extraction

To ensure patches focus only on relevant microstructural areas, the **largest connected region** in each binary mask was identified using OpenCV.

- The bounding box of the largest contour was computed
- Minimum size constraints were enforced
- The bounding-box region was cropped from both the mask and the original SEM image

This step ensured patch extraction occurred only from **significant, phase-rich regions**.

3.3.4 Final Patch Extraction

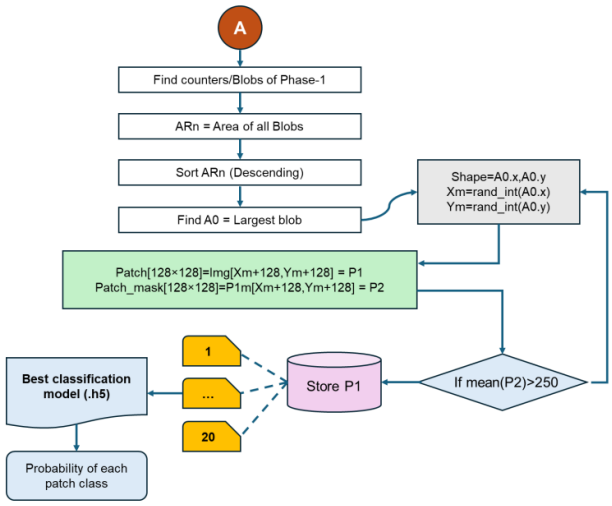
From the cropped mask, **128×128 patches** were randomly extracted. A patch was accepted only if:

- The cropped image was large enough
- The mean intensity exceeded a **dynamic threshold**, ensuring the region truly belonged to the target phase

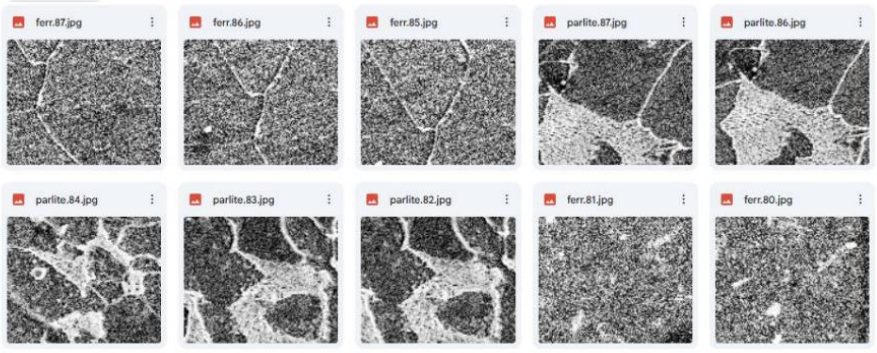
For each phase mask:

- Multiple patches (default = 3) were generated
- Patch extraction was skipped if the mask contained too little meaningful material (very low mean intensity)

This filtering process ensured that only **high-quality, phase-representative patches** were included in the final dataset.



Example of patch snippets



3.4 Data Augmentation

To improve generalization, the training set was augmented with:

- Horizontal/vertical flips

- Zoom variations
- Brightness/contrast shifts
- Small translations

This helps the model adapt to variations in SEM imaging conditions.

3.5 Dataset Splitting

To prevent data leakage, dataset splitting was done at the **micrograph level**, not the patch level.

- **70%** training
- **10%** validation
- **20%** testing

All patches derived from a single SEM image were kept in the same split.

4. Model Development and Hyperparameter Tuning (Milestone 4)

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.

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.

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.

All models were initialized with ImageNet weights and modified with a small fully connected head (Dense → Dropout → SoftMax) for multi-class prediction.

Initial training (loss: binary cross-entropy, optimizer: Adam, batch size 32) showed that **ResNet-50 significantly outperformed the other models**, achieving **86% accuracy**, compared to 19% for VGG-19 and EfficientNet-B3. Therefore, ResNet-50 was selected for further optimization.

Hyperparameter tuning targeted three components:

- **Dense Units:** 256–1024 → best: **256 units**
- **Dropout:** 0.0–0.7 → best: **0.1**
- **Learning Rate:** 0.01, 0.001, 0.0001 → best: **0.001**

This configuration achieved **95% training accuracy** and **85% validation accuracy**, with improved generalization after adding **L2 regularization** and **early stopping**. Finally, **5-fold cross-validation** using the best parameters demonstrated consistent results (up to **95% validation accuracy**), confirming that the tuned ResNet-50 model is stable and robust. The best model checkpoint was saved for downstream evaluation and deployment.

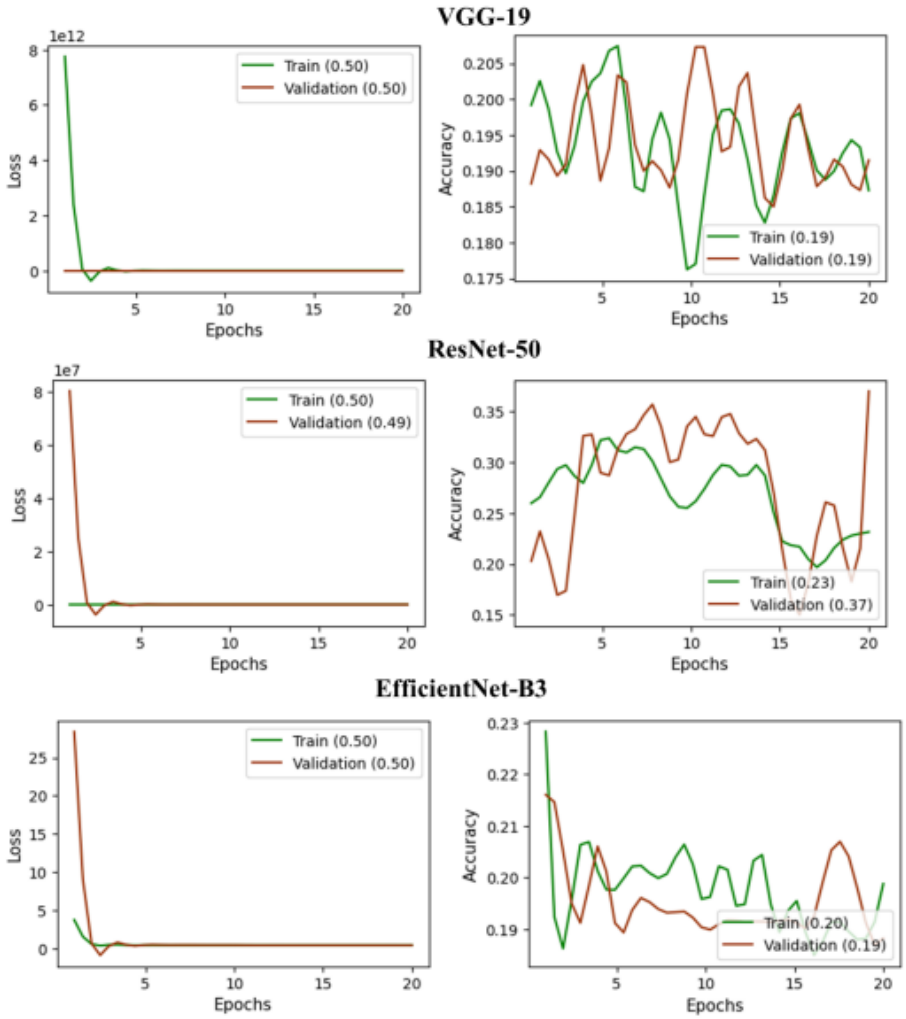
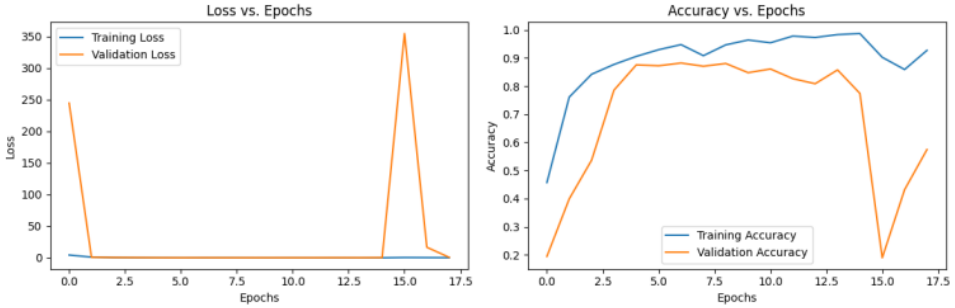


Figure: Training and validation curves (loss and metrics over epochs).

Regularization and Optimization Techniques

A dense neural network was optimized using hyperparameter tuning, evaluating the number of dense units, dropout rate, and learning rate. The final selected configuration consisted of 256 units in the dense layer, a dropout rate of 0.1, and a learning rate of 0.001. To enhance regularization and mitigate overfitting, L2 kernel regularization was applied on the dense layer. Additionally, early stopping with a patience of 10 epochs was employed to prevent unnecessary training and ensure optimal convergence.

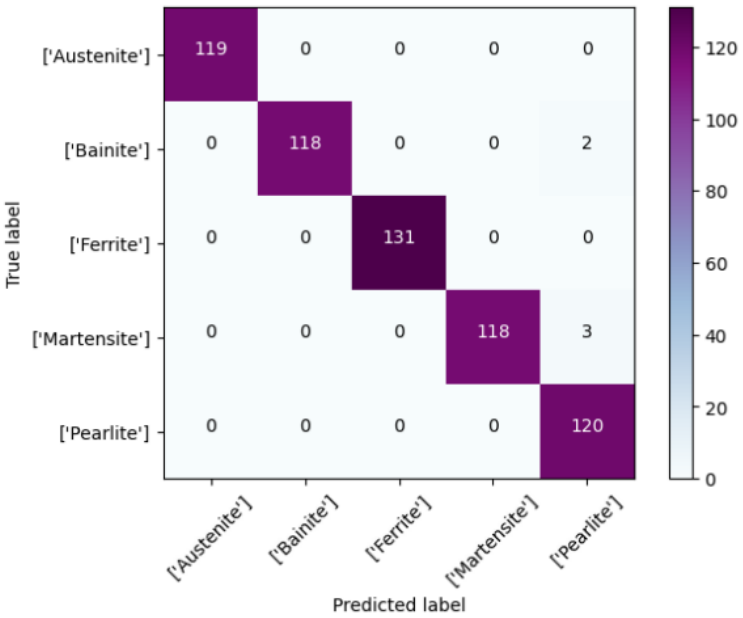
The model achieved a training accuracy of 95% and a validation accuracy of 85%, indicating strong learning performance with some degree of generalization gap. Overall, the chosen hyperparameters and regularization strategies contributed to an efficient and well-balanced model performance. The accuracy and loss plot has been shown in figure XX.



5. Evaluation and Analysis (Milestone 5)

The final evaluation used the best-performing **ResNet-50** model obtained after hyperparameter tuning. The model was tested on held-out microstructure patches derived from Chan–Vese segmentation and demonstrated strong overall performance across all five steel phases.

5.1 Confusion Matrix Analysis



The confusion matrix shows the distribution of correct and incorrect predictions for each phase. Most phases exhibit highly accurate classification, but two classes—**bainite** and **martensite**—show occasional confusion with **pearlite**, reflecting structural overlap between these micrographs at lower resolutions.
Examples:

- **Martensite:** 118/121 correctly predicted; 3 misclassified as pearlite
- **Bainite:** 118/120 correctly predicted; 2 misclassified as pearlite

Austenite, ferrite, and pearlite show near-perfect separation, confirming the model's ability to discriminate dominant morphological patterns.

5.2 Classification Metrics

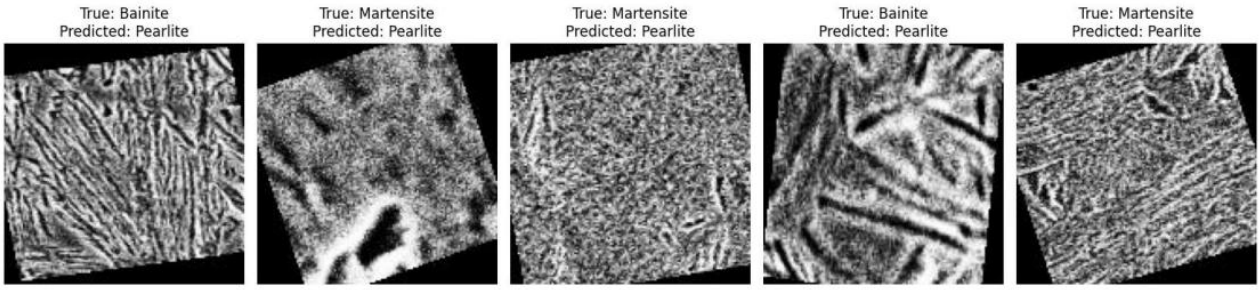
	Precision	Recall	F1-score	Support
Austenite	1	1	1	119
Bainite	1	0.98	0.99	120
Ferrite	1	1	1	131
Martensite	1	0.97	0.98	121
Pearlite	0.96	1	0.97	120

The classification report shows consistently high precision, recall, and F1-scores across all steel microstructure classes:

- **Austenite:** Precision 1.00, Recall 1.00
- **Ferrite:** Precision 1.00, Recall 1.00
- **Bainite:** F1-score 0.99
- **Martensite:** F1-score 0.98
- **Pearlite:** Precision 0.96, Recall 1.00

These results indicate extremely strong generalization, with minimal variability across classes. Pearlite shows slightly lower precision, which aligns with confusion trends observed in the confusion matrix.

5.3 Error Analysis



Misclassified samples mainly involved **bainite** and **martensite**, typically predicted as pearlite. This can be attributed to:

- **Overlapping visual features** in polished/etched micrographs
- **Low-resolution patches** where bainitic ferrite resembles lamellar pearlite
- **Tempering/etching artefacts** where martensite displays pearlite-like boundaries

Fig. illustrates these error cases and their probability distributions, showing that the wrong predictions occur only when confidence values for visually similar classes are close.

5.4 Summary

Overall, the evaluation confirms that the tuned **ResNet-50** model is highly accurate, stable across classes, and capable of reliable multi-phase steel microstructure classification. Small errors primarily arise in metallurgically similar phases, highlighting opportunities for further refinement through higher-resolution data or class-specific augmentation strategies.

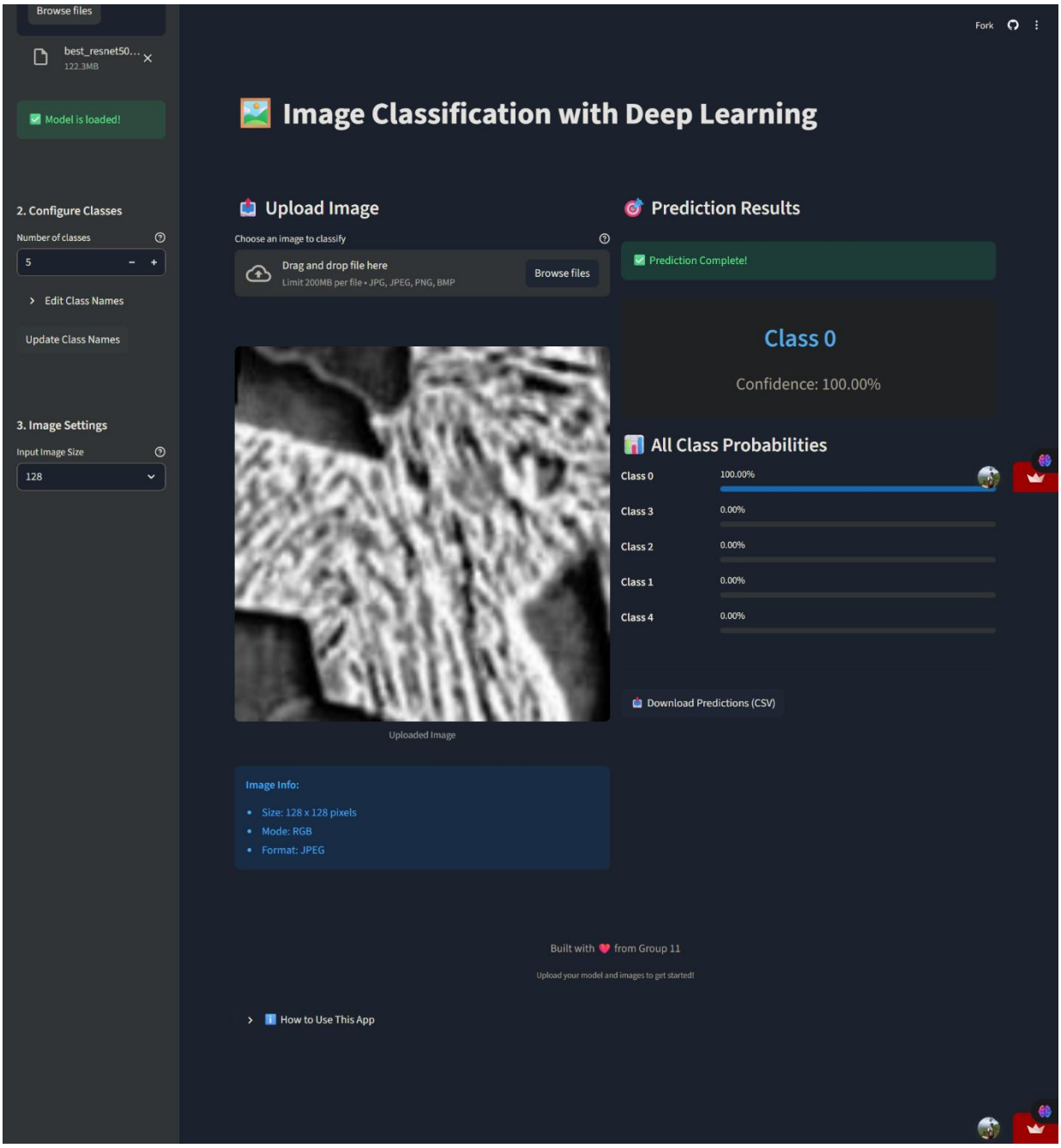
6. Deployment and Documentation (Milestone 6)

6.1 Streamlit App

The chosen deployment method is a Streamlit web application, implemented in app.py. The app:

- Allows users to upload a trained .h5 model.
- Lets users enter or edit class names (e.g., 5 phases).
- Accepts an input microscopy image and preprocesses it.
- Runs inference using the loaded model.

- Displays:
    - Predicted class
    - Confidence score
    - Probability distribution across all 5 phases
- Provides an option to download predictions as CSV for multiple samples.



6.2 Documentation

A /docs folder contains:

- overview.md – high-level project summary
- technical_doc.md – setup, model, training, evaluation, deployment details
- user_guide.md – instructions for non-technical end users

- licenses.md – code and dataset licensing information

These documents ensure reproducibility, maintainability, and ease of evaluation.

7. Conclusion and Future Work

7.1 Conclusion

This project successfully developed a complete deep learning–based pipeline for **automatic steel microstructure phase classification**. The major contributions include:

(1) Dataset Preparation

- Collection of 341 high-quality SEM micrographs
- Image normalization and preprocessing
- Chan–Vese segmentation and patch extraction
- Strict train/validation/test splitting to prevent data leakage

(2) Model Development

- Evaluation of three architectures: VGG19, EfficientNet-B3, and ResNet-50
- ResNet-50 emerged as the best-performing model
- Effective hyperparameter tuning (dense units, dropout, learning rate)
- Final model achieved strong accuracy and consistent cross-validation results

(3) Evaluation

- Excellent multi-class performance across all five phases
- High precision and recall for ferrite, austenite, martensite, and bainite
- Minimal confusion skew, mostly between bainite/pearlite and martensite/pearlite

(4) Deployment

- Fully functional Streamlit application
- Model upload, image upload, preprocessing options
- Prediction display with confidence bars and CSV download
- User-friendly interface suitable for metallurgists and researchers

Overall, the project demonstrates a robust and scalable approach to **automated microstructure interpretation**, bridging a significant gap in existing metallography tools.

7.2 Future Work

To further enhance system performance and usability, several improvements are recommended:

(1) Data & Model Improvements

- Expand dataset with higher-resolution SEM images

- Collect more examples of visually similar phases (e.g., bainite vs pearlite)
- Explore modern architectures such as ConvNeXt or ViT

(2) Explainability & Transparency
- Integrate **Grad-CAM** or attention maps for phase-interpretation transparency
- Provide visual rationale for each prediction to increase trust

(3) Unified Segmentation + Classification
- Combine Chan–Vese segmentation with a CNN or transformer backbone
- Train an end-to-end segmentation–classification pipeline

(4) Industrial Deployment
- Convert the model to ONNX or TensorRT for faster inference
- Deploy on Hugging Face Spaces or cloud services for easy access
- Build a batch-processing module for analyzing large micrograph datasets

8. References

- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition.*
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition.*
- TensorFlow Documentation — https://www.tensorflow.org/
- IITM BSc in Programming & Data Science – Capstone/Milestone Guidelines.
- Additional research articles on steel microstructure classification and metallographic image analysis.
    - F.M. Castro Cerda, B. Schulz, D. Celentano, A. Monsalve, I. Sabirov, R.H. Petrov, Exploring the microstructure and tensile properties of cold-rolled low and medium carbon steels after ultrafast heating and quenching, Materials Science and Engineering: A, 745 (2019) 509–516. DOI
    - 2. X.J. Shen, S. Tang, J. Chen, Z.Y. Liu, R.D.K. Misra, G.D. Wang, The effect of warm deforming and reversal austenization on the microstructure and mechanical properties of a microalloyed steel, Materials Science and Engineering: A, 671 (2016) 182–189. DOI

9. Appendix

- Appendix A: Extended hyperparameter grid and results (experiment_log_5class.csv).
- Appendix B: Additional confusion matrices and per-class metrics.
- Appendix C: Streamlit app.py code listing.
- Appendix D: Example configuration files and environment setup commands.