Milestone 5

Overview / Objective

Microstructures are 2D images captured in microscopy, which are defined as the structure of a prepared surface of material. Nonetheless, they efficiently result in a structure-property correlation study. However, manual recognition of microscopic data is both human knowledge-intensive and potentially ambiguous, especially in the case of steel. The current study creates a computer vision of important microconstituents of steels such as ferrite, pearlite, austenite, bainite, and martensite, enabling automatic identification of microstructure. Utilizing the Chan-Vese image segmentation methods, the current work extracted the patches of phases from the collected microstructures of steels. These phase patches constituted the database for the development of a deep convolutional neural network VGG-19, ResNet-50 and EfficientnetB3. The article demonstrates the development and prediction framework of in-silico microstructure vision. Utilizing a few typical example microstructure data, the article discussed the automatic labelling knowledge intensiveness and ambiguous behavior of the created microstructure vision.

Dataset Details

A steel microstructure database was created by an extensive review of published literature. Relevant data were meticulously gathered from various scientific articles, research papers, and industrial reports that detailed the microstructural characteristics and phase compositions of steels belonging to the advanced high-strength steels. The ground reality for labeling micrographs was done as they are reported in the literature. The literature sources were carefully selected based on their relevance, rigor, and the quality of their experimental methodologies. The augmented dataset was divided into two parts, namely training (80% i.e., 2440 patches), and testing (20%, i.e., 611 patches). All splits were performed at the patch-image level, patches were derived from 188 unique micrographs, and every patch is labelled with its source image ID.

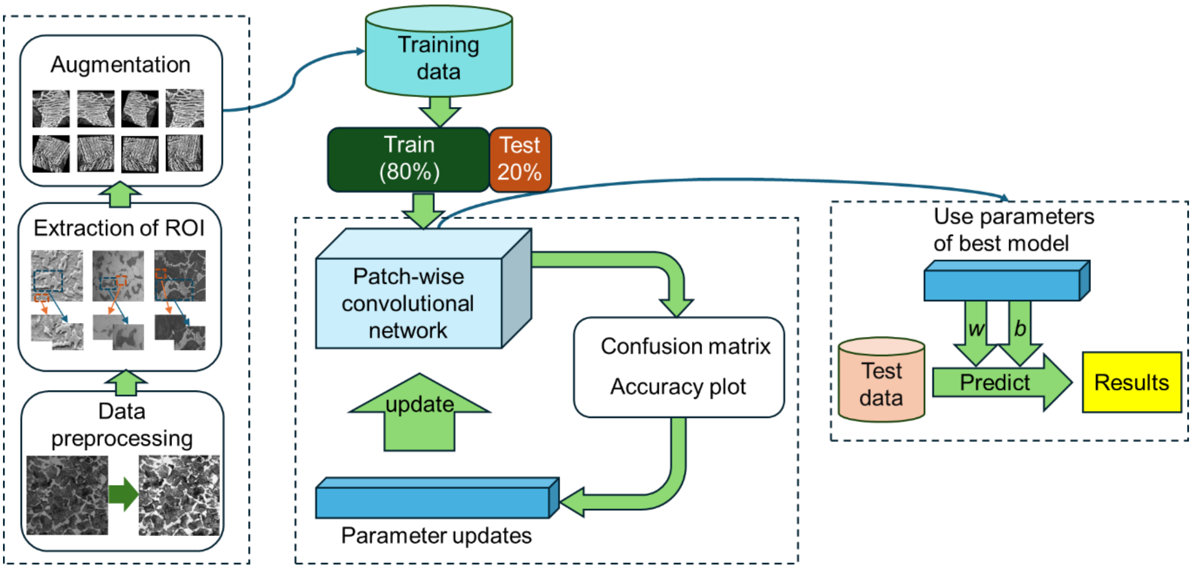

Fig 1. Schematic representation of patch classification model framework proposed for digital recognition of phase constituents such as ferrite, pearlite, austenite, bainite, and martensite in SEM micrograph image of steel.

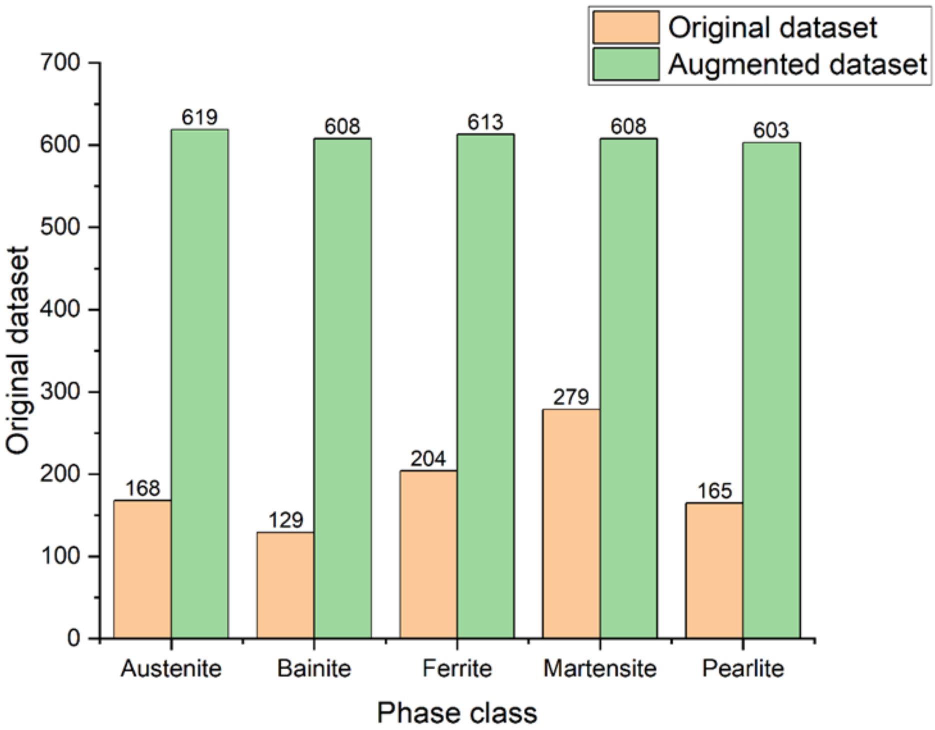

Fig 2: Class balancing of patch class created from the 188 microstructures. Data augmentation was performed using geometric transformations such as horizontal and vertical flipping, random rotations within ±20°, shearing with a shear range of 0.1, and zooming within a range of 0.2.

Model Architecture

The three commonly used CNN-based architectures, like VGG-19, Resnet-50, and EfficientnetB3, are selected to train augmented patches for a comparative analysis of the models. However, some modifications have been made to these original networks to adapt to our work. One additional fully connected layer was added to extract features with high resolution, and one dropout layer was added in the network to avoid overfitting. The last layer of the classification model was SoftMax, and the binary cross-entropy was selected as the loss function in this study. Adam was used as an optimizer for the loss function.

Training

The last layer of the classification model was SoftMax, and the binary cross-entropy was selected as the loss function in this study. Adam was used as an optimizer for the loss function with a learning rate of 0.01. Accuracy was selected as the evaluation metrics and batch size was 32. The training of the classification model was performed on Intel(R) Xeon(R) Gold 5218R CPU, and Nvidia RTX A5000 GPU in Python-based TensorFlow 2.0 platform.

The final model configuration consisting of 256 dense units, 0.1 dropout, L2 kernel regularization, and with a learning rate of 0.001 was used during the cross-validation process. Across the 5-fold folds, the model demonstrated consistent performance, achieving a best validation accuracy of 95%. This result confirms the model's stability and ability to generalize well across different subsets of the data, supporting the reliability of the chosen. The accuracy across each fold is shown in figure 6. The best performing model is stored as '.h5' for future predictions.

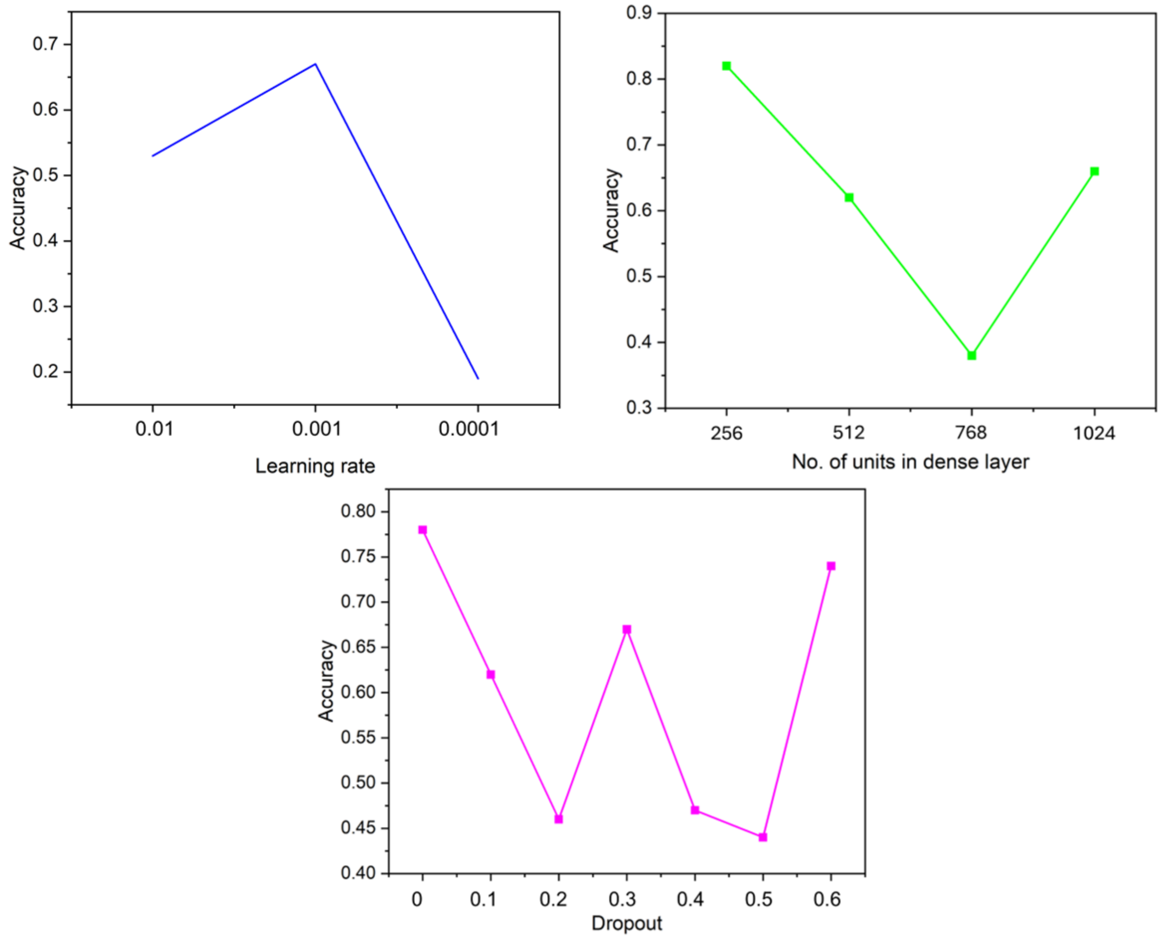

Fig. 3: Hyperparameter results

Model Evaluation & Analysis

Fig. 4 exhibits the confusion matrix of the testing data for the classification obtained from the proposed best ResNet50 model, which reveals the correct and wrong predictions for each of the microstructure phases of the steel. In general, the confusion matrix epitomizes the much deeper statistics about the type of error made by the classifier, if any. Alone, performance metrics like accuracy sometimes do not provide the complete information of the multiclass classification model's performance when the set of correct predictions in every class is not of equal size. For example, in this case, the number of correctly predicted instances for all five phases is uneven concerning their support in the training and testing datasets.

In the present context for the martensite phases, the predictability is very uncertain, unlike the other phases, namely, austenite, ferrite, and pearlite. Moreover, it can be found from Fig. 1 that the bainite is a more confused class. Out of the 121 cases which are martensite in the testing dataset, the classification model correctly predicted 118 cases, and in the rest of 3 instances the model was confused with pearlite phases. Same as in the case of the bainite phase, the model predicted the bainite 118 times correctly and 2 times it was predicted as pearlite.

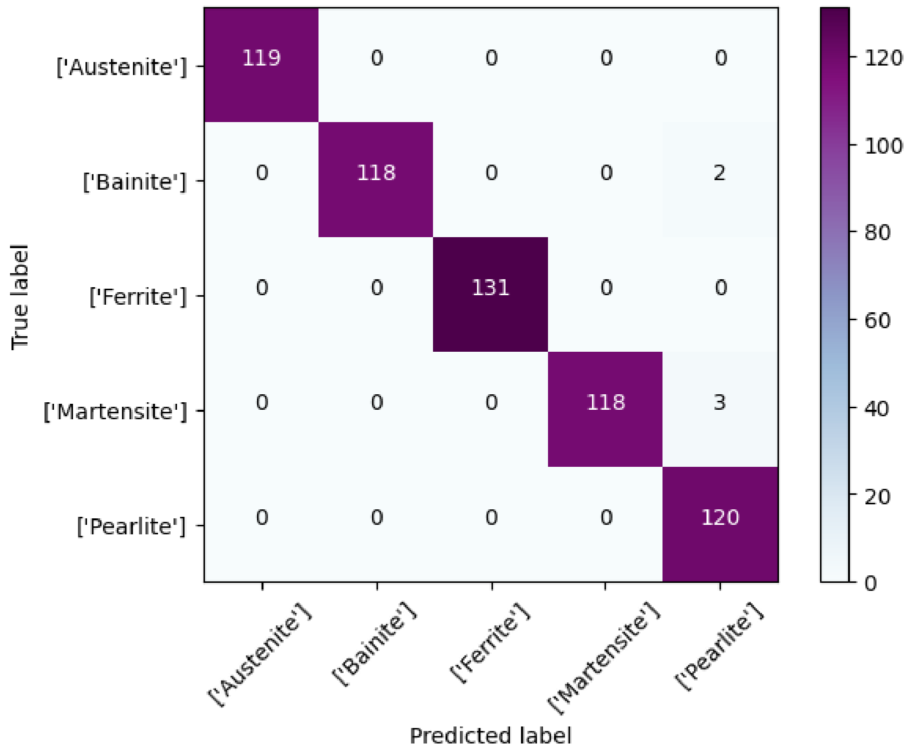

Fig. 4. Confusion matrix of the best performing model on the validation data

The trained model demonstrated strong performance across all steel microstructure classes, achieving high precision, recall, and F1-scores. Classes such as Austenite and Ferrite were perfectly classified, each achieving a precision, recall, and F1-score of 1.00. Bainite and Martensite also showed excellent results with F1-scores of 0.99 and 0.98, respectively, indicating only minor misclassifications. Pearlite exhibited slightly lower precision (0.96) but maintained perfect recall (1.00), leading to a strong F1-score of 0.97. Overall, the results indicate a highly accurate and reliable model with strong generalization capability across all microstructure classes, with minimal performance variation between them.

Table 1. The classification report of the best performing model on the validation data

	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

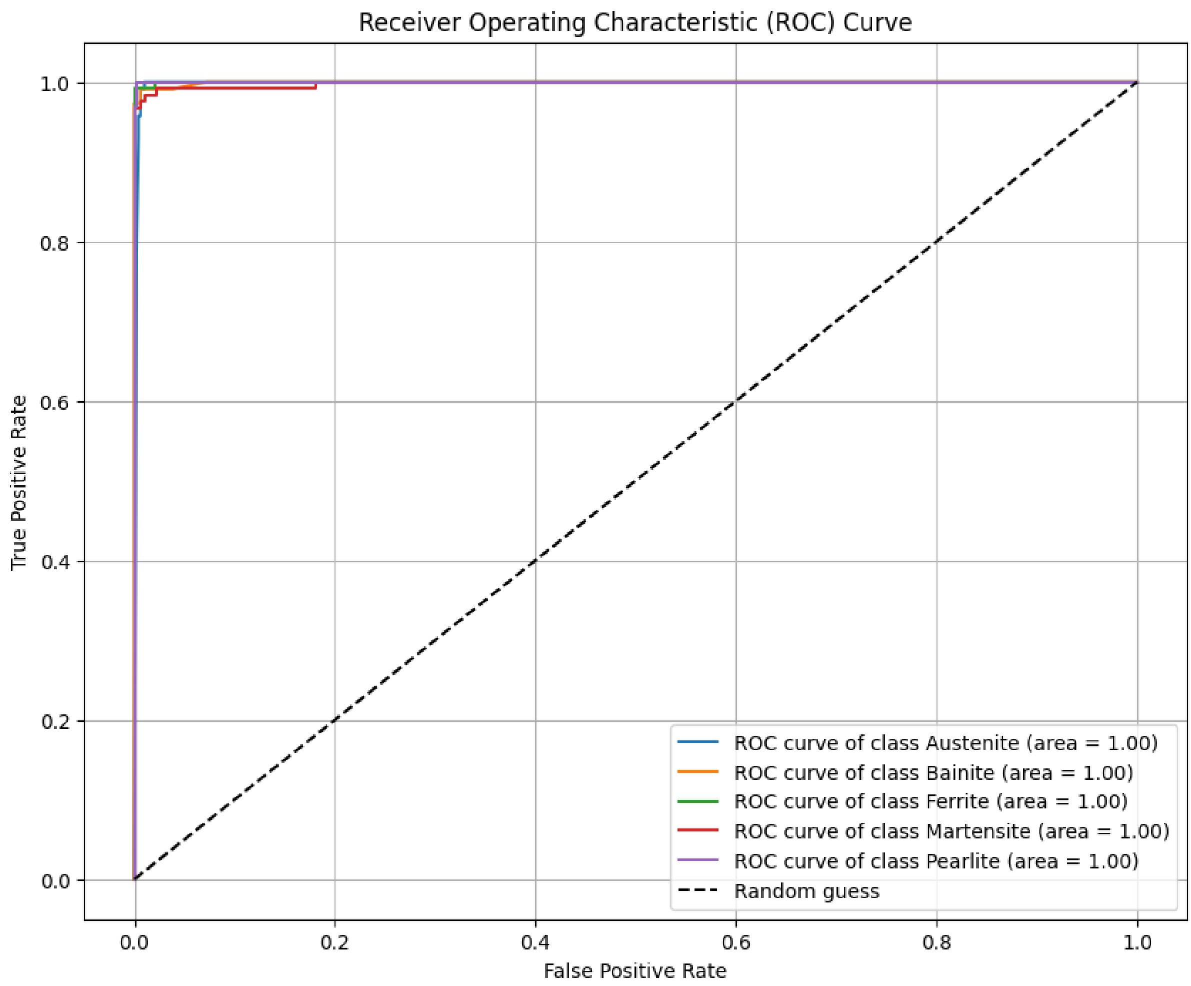

Fig 5. ROC curve

Error analysis

Fig. 6 shows the example cases of wrong prediction of the bainitic and martensite phase patches. In most cases, bainite is confused with the pearlite phase. The misclassification of bainite as pearlite can be attributed to overlapping features in their microstructures and characteristics. The model predicted bainite as pearlite because both phases exhibit relatively low carbon content and appear similar in polished and etched micrographs at low-resolution images. From a metallurgical perspective, the misclassification of martensite as pearlite can be explained by the overlapping microstructural features and transformations in steel. Pearlite-like features are due to etching or tempering, can obscure martensite's sharp boundaries, making it appear lamellar.

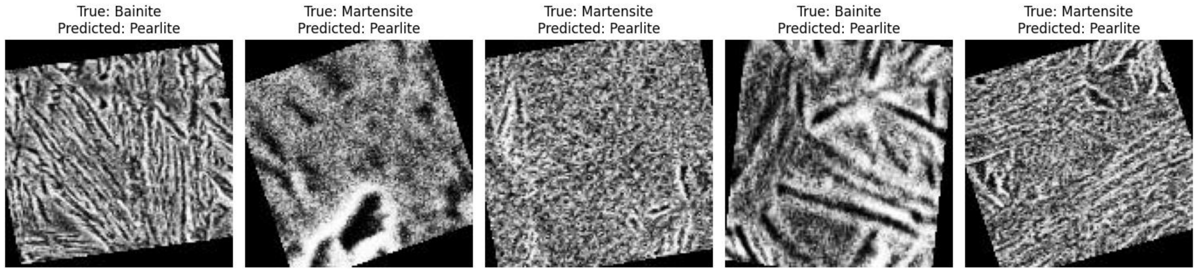

Fig. 6. The predicted probability density was obtained from the last layer of ResNet-50 classification model.

Limitation and future scope

The confusion matrix and error analysis revealed frequent misclassification among martensite, bainite, and pearlite, which is mainly consistent with their overlapping morphological features. While metallurgical explanations account for these similarities, several modeling strategies may further reduce such confusions. A hierarchical classification framework could first separate phases into broader categories (e.g., equiaxed versus lamellar/lath-like microstructures), followed by a specialized classifier to distinguish closely related phases such as bainite, martensite, and pearlite. Finally, ensemble approaches that combine deep CNN features with handcrafted texture descriptors could provide complementary perspectives and reduce phase-specific biases. Although these approaches are beyond the scope of the current work, they present promising avenues for future studies aimed at improving phase-level discrimination. In contrast, segmentation and classification task can be performed simultaneously by the models like YOLO and U-Net algorithms, however, pixel level ground truth will be required to train these models.