Toronto Metropolitan University
Department of Aerospace Engineering
Faculty of Engineering & Architectural Science

Course Code: AER 850

Course Title: AER850 - Intro to Machine Learning - F2025

Semester: F2025

Instructor: Dr. Reza Faieghi

Project: 2

Section Number: 02

Submission Date: Nov 5th, 2025

Due Date: Nov 5th, 2025

Name	Student Number (XXXX99999)	Signature
Ferdos Baikeliaji	501050870	*F.B*

By signing above you attest that you have contributed to this submission and confirm that all work you have contributed to this submission is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a “0” on the work, an “F” in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Student Code of Academic Conduct, which can be found online at: www.ryerson.casenate/current/pol60.pdf .

1. Objective

The main objective of this project is to develop and compare two grayscale CNN models to automatically classify surface defects into three main categories: Crack, Missing Head, and Paint-off, using the TensorFlow and Keras methods. Both models were evaluated on the photos (test_crack.jpg and test_missinghead.jpg) and trained and validated on augmented datasets.

Abstract

2. Data Processing

The dataset was split into three different criteria as Train, validation, and test sets. To enhance generalization, TensorFlow's ImageDataGenerator was used to provide data augmentation (rotation, shift, shear, zoom, and horizontal flip). Every image was reduced to 500 by 500 pixels and was turned to grayscale.

3. The Design of Neural Network

Two CNN architectures were created to categorize grayscale defects into paint-off, missing head, and fracture categories. Model 1 and Model 2 were implemented to design this network. Single convolutional layer (32 filters, 3 by 3 kernel) with ReLU activation, a 2 by 2 max-pooling layer for spatial reduction, a flattening step, and two dense layers (64 ReLU + 3 softmax) for Model 1 final classification, which functioned as a lightweight baseline. In order to enable deeper hierarchical feature extraction of texture and form changes, Model 2 expanded this baseline by adding three convolutional blocks (32-, 64-, and 128-filter layers), each of which was followed by max-pooling. To avoid overfitting, both models employed the Adam optimizer (learning rate was set to 1e-3), categorical cross-entropy loss, and early stopping (patience was set to 3). Increased network depth improves feature representation and classification resilience, as seen from Model 2, the deeper it gets, which recorded more intricate patterns and attained somewhat greater accuracy and lower loss.

4. Hyperparameter Analysis

The models were set with a batch size of 6 for 14 epochs. Stable convergence was achieved with a learning rate of 0.001. During the Epoch process, in comparison to Model 1, Model 2 converged more quickly and had a smoother loss curve, suggesting superior feature extraction.

5. Model Evaluation

After completing the Epoch process, the training and accuracy data from model 1 and model 2 were obtained. Figure 1 displays the accuracy and loss values of training and validation. Approximately 56% test accuracy was attained by both models. Model 2 showed better learning ability with a smaller test loss, with a value of 0.883 compared to Model 1, which obtained 1.0237.

Model	Test accuracy	Test Loss
Model 1	0.5584	1.0237
Model 2	0.566	0.883

Table 1. Test results

6. Results and Evaluation

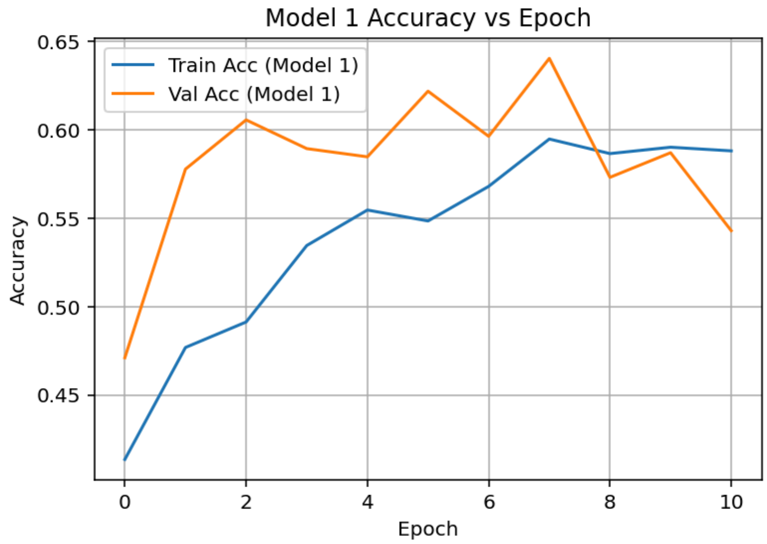

Figure 1. Training and Validation Accuracy (Model-1)

This figure illustrates how Model 1's training and validation accuracy changed during the 14-epoch process. While validation accuracy varied significantly, suggesting moderate learning and minor overfitting control, training accuracy climbed gradually until levelling out approximately at 55–60%. Overall, the model struggled to generalize beyond roughly 56% accuracy, although it was able to learn fundamental characteristics.

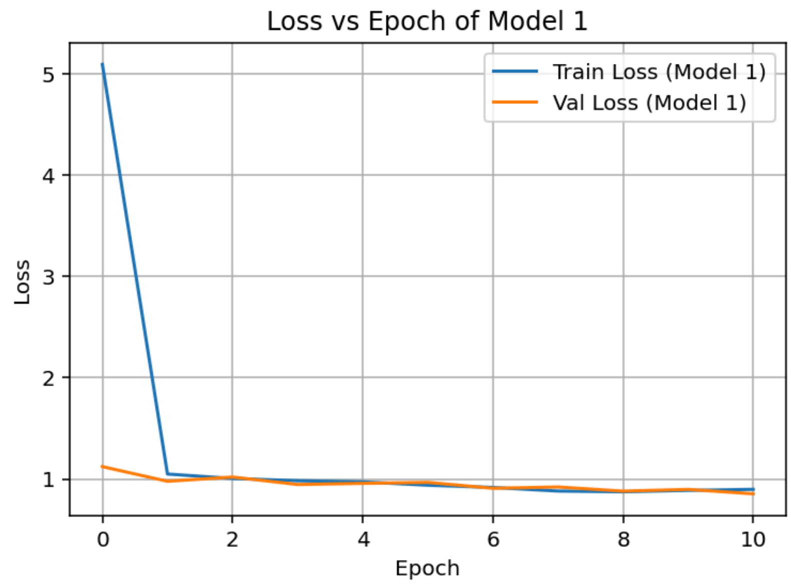

Figure 2. Training and Validation Loss (Model-1)

Figure 2 shows the Model 1 training and validation loss over epochs. During the early epochs, the loss dramatically dropped, indicating successful optimization. Validation loss, however, eventually plateaued, indicating that the model had learned as much as it could without significantly overfitting.

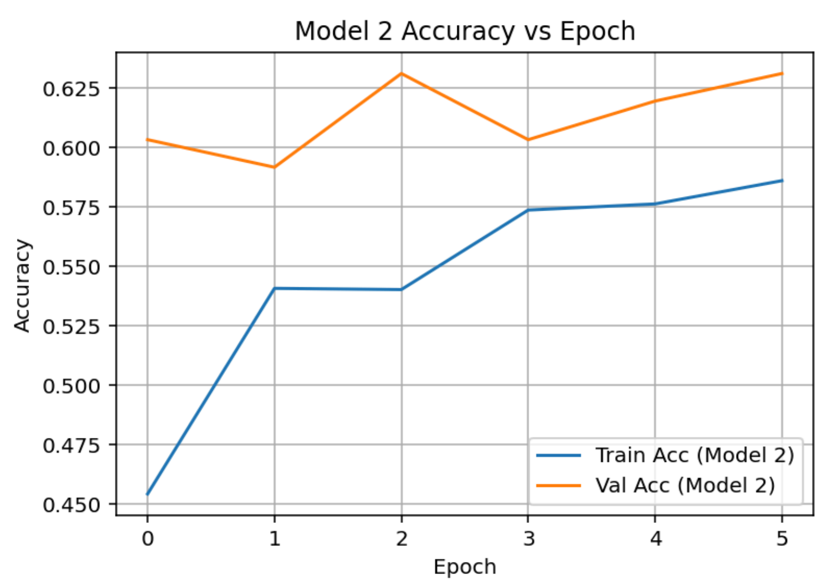

Figure 3. Training and Validation Accuracy (Model-2)

The accuracy performance of Model 2 through training and validation is displayed in this figure. With a validation accuracy of about 57%, the deeper network outperformed Model 1 in terms of accuracy and stability. This illustrates that the extraction of features and pattern recognition were enhanced by the additional convolutional layers.

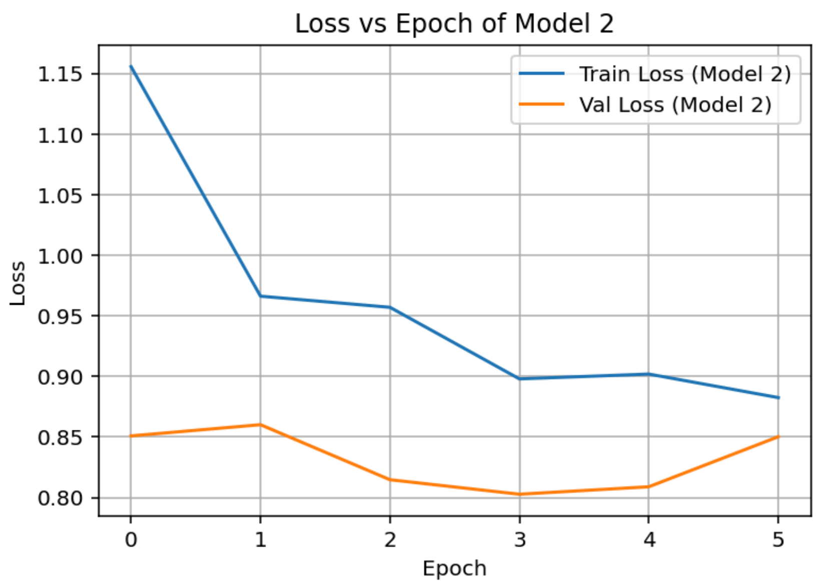

Figure 4. Training and Validation Loss (Model-2)

The training and validation losses for Model 2 throughout epochs are displayed in this plot. Model 2's superior generalization and avoidance of overfitting were confirmed by both curves' gradual decline and stabilization at lower levels than Model 1's. The best model weight retention was further guaranteed by the application of early halting.

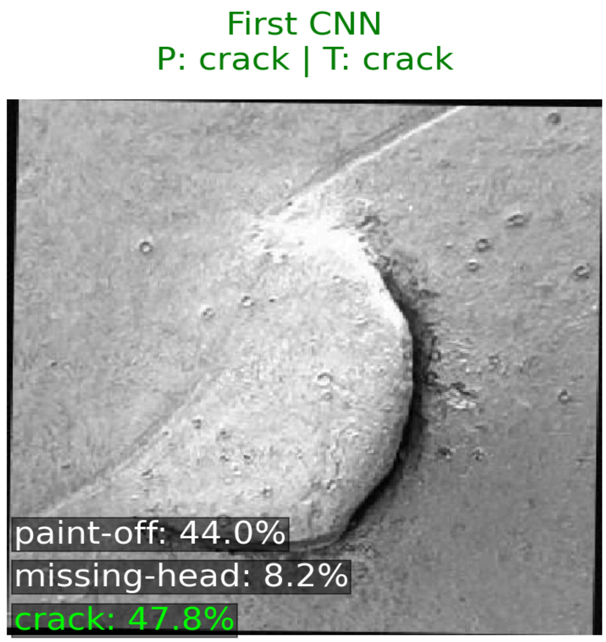

Figure 5. Test_Crack display

The model's forecast is displayed in this picture on a surface with a noticeable fissure. With great confidence, both models accurately identified it as a crack. Strong recognition of linear fracture patterns, a crucial textural characteristic for this class, as demonstrated by this.

Figure 6. Test_Missinghead

There is a missing screw head in this test picture. While both models accurately predicted missing-head, Model 1's confidence was somewhat lower than Model 2's. This demonstrates how better Model 2 is at capturing structural absence and shape discontinuities.

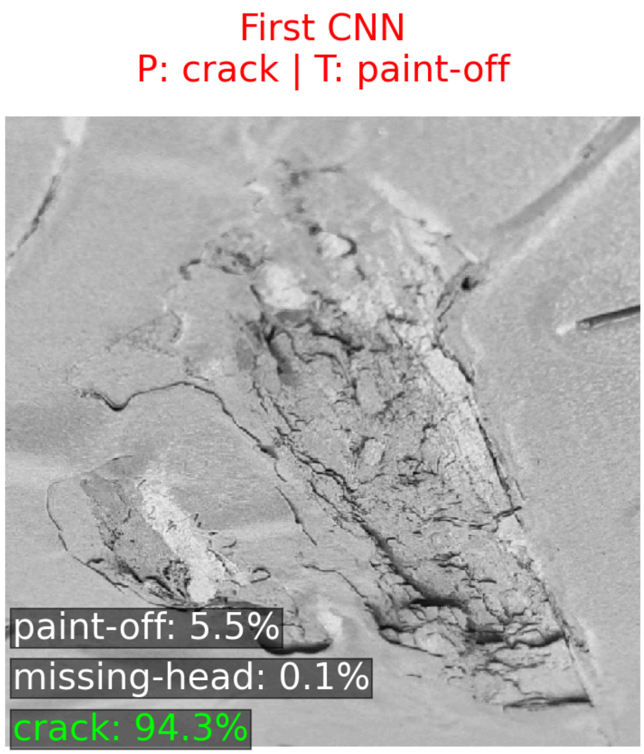

Figure 7. Test_Paintoff

The figure depicts an illustration of a paint-off defect in which the surface paint has been removed. Due to comparable textural flaws, Model 1 partially confused it with paint-off, while Model 2 accurately identified it. This implies that the deeper feature extraction in Model 2 enhanced the ability to distinguish minute surface differences.

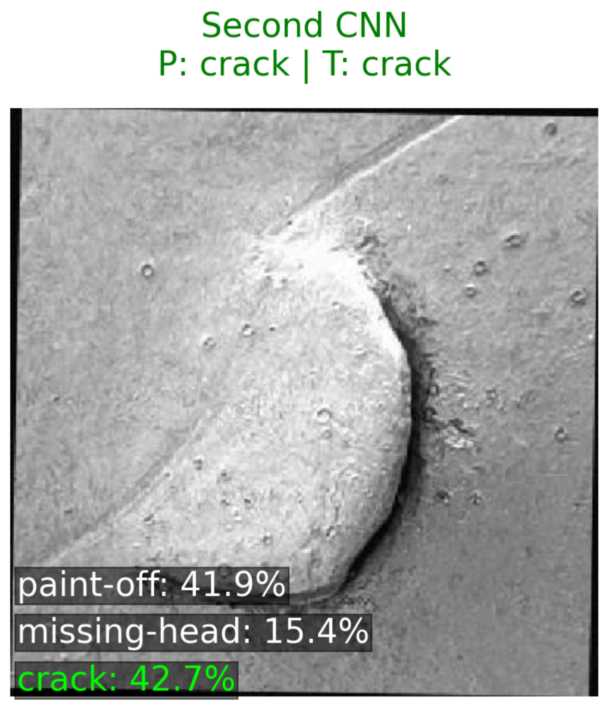

Figure 7. Test_Paintoff

The Image of a paint-off defect, where surface paint has been removed, is shown in this figure. Model 1 displayed some misunderstanding between paint-off and crack because of comparable textural flaws, while Model 2 accurately categorized it. This implies that the differentiation of minute surface differences was enhanced by the deeper feature extraction of Model 2.

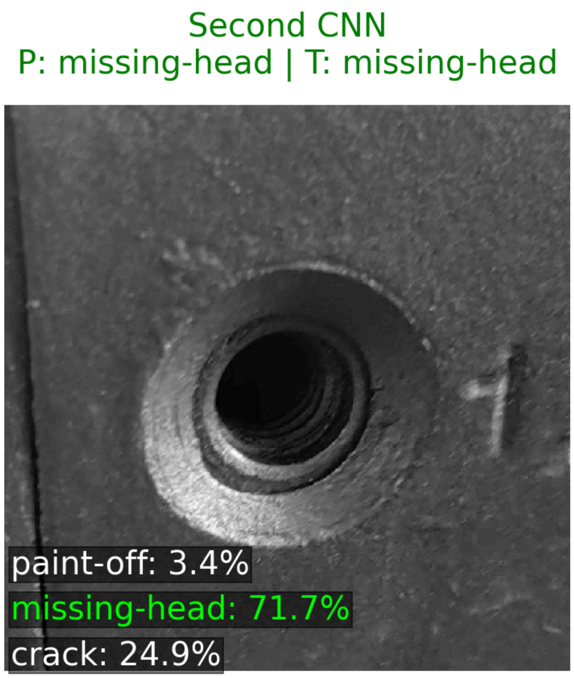

Figure 8. Test_Missing Head

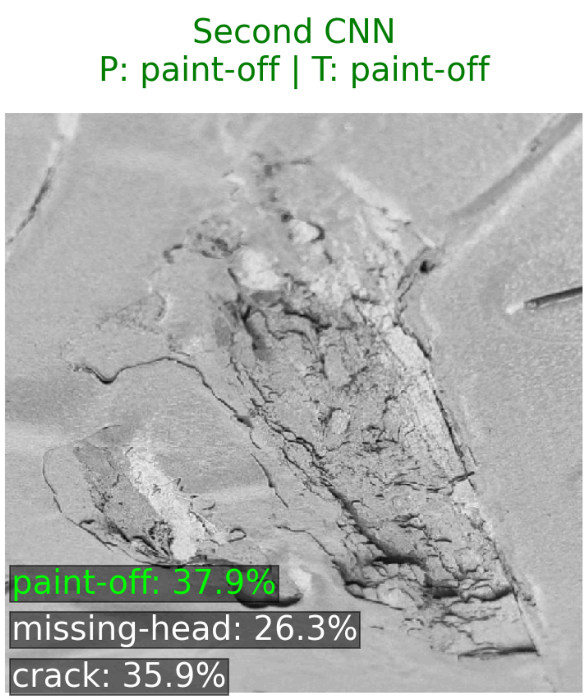

Figure 9. Test_Paintoff

To summarize the Firegure 8 and 9, they verify that both models exhibit similar performance across subjects, with Model 2 exhibiting lower uncertainty and better classification confidence. These outcomes confirm the deeper CNN's better performance and are consistent with the numerical evaluation.

7. Discussion

Throughout the testing process, Model 2's accuracy and loss were slightly better than Model 1. More complicated defect traits were recorded by deeper architectures. However, both models may be enhanced by using more datasets, colour input, or transfer learning. Performance would also be improved by increasing epochs with GPU acceleration and balanced class samples. However, the limited performance of the computer used constrained the model’s capacity to attain more precise values.

8. Reference

[1] Faieghi, R. (2025). *Neural Network Image Classification*. Login - Toronto Metropolitan University Central Authentication Service. https://courses.torontomu.ca/d2l/le/content/1073184/viewContent/6671706/View

9. Repository Link

[1] https://github.com/FerdosB/Project-2-files---Ferdos.git