D7056E - Predictive AnalyticsGroup Project – Report

Predictive Maintenance for Railway Infrastructure Using Image-Based Fault Detection

Group Members: Lars-Erik Englund, Daniel Habte-Medhanie, Victor Rigworo Kebande, Anton Sandin.

Submission Date: 2025-03-14

Summary CRISP-DM Process

The CRISP-DM methodology consists of six main components, each of which plays a vital role in the overall process. These components include business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

1. Business Understanding

Effective railway maintenance is crucial for safety and economic sustainability. Manual inspection of railway components and infrastructure is both time-consuming and risky, highlighting the need for automated AI-based solutions. Predictive maintenance, utilizing images from captured using a mounted inspection camera has the potential to significantly reduce downtime and maintenance costs by detecting faults early. The objective of this study is to evaluate the capability of machine learning models in classifying faults from

2. Data Understanding

Dataset contains 5153 images of railway tracks labeled to seven categories called cracks, flakings, joints, shellings, spellings, squats, grooves. The images were 1280x720 and had three RGB channels. Some of the challenges with the dataset was a big class imbalance, varied light conditions and angles, and some pictures coming from the same place just slightly shifted making it possible to classify the rail based on other features in the environment.

3. Data Preparation

We did four things to prepare our data. The first was splitting our dataset into a training set (60%), a validation set (20%), and a test set (20%). Then we resize the images to 224x224 to fit our models. After that we normalized it with mean values of 0.485, 0.456, 0.406 and standard deviations of 0.229, 0.224, 0.225 for each of the RGB channels respectively, the reason for choosing these numbers is that the ResNet-18 model that we will later use in the modeling phase was normalized with these values, thus it will give us a better result if we use the same normalization. Finally we applied some data augmentation techniques to reduce overfitting and enlarge our dataset like random flipping, and random rotation.

4. Modeling

Two machine learning models have been leveraged in the predictive maintenance task as follows:

Model 1: Convolutional Neural Network (CNN)

The employed CNN consists of three convolutional layers with ReLU activation, dropout, and max pooling, followed by two dense layers for classification. 3 convolutional layers, ReLU activation, dropout, max pooling.

Model 2: ResNet-18

ResNet-18 has been fine-tuned with pretrained weights on our dataset, incorporating dropout and adjusting the final layers to optimize defect classification.

5. Evaluation

For the evaluation metrics we choose accuracy, precision, recall, F1-score and a confusion matrix. When we measure our metrics we see that all are much better for the ResNet than the CNN, for example the accuracy of the CNN was 51.55% which rose to 94.27% for the ResNet. The reason we believe it's so much better is because it is deeper thus having a higher capacity and the ability to capture very subtle patterns in the dataset.

6. Deployment

It would be interesting to try to implement this model at Trafikverket and test it on more data and to compare the models predictions with some historical predictions done by humans.

If the trained model is used in actual conditions we believe the computational requirements for using the ResNet model would not exceed standard computing power.

Evaluation of the model's performance should be monitored and retrained when needed.

Time/Workload Distribution and reflections

Based on our estimates, the time or workload distribution for each of the phases in your project were: Business Understanding: 15% Data Understanding & Preparation: 35% Modeling: 45% Evaluation: 5% Deployment considerations: 10%

Reflections: Future projects would allocate slightly more time on early data understanding and data preparation to reduce later modeling complexity. It would also have been good to have a more balanced dataset, with regard to differences in fault types and also some images of “normal” conditions.

Strengths of the Project: We had a clearly defined business problem with practical relevance. There was a comprehensive data preprocessing pipeline ensuring consistent model inputs.We made an effective use of advanced deep learning models (ResNet), yielding strong predictive performance.

Weaknesses & Future Improvements: Model complexity increased computational demands (ResNet). Slight imbalance in dataset classes influenced CNN performance negatively.

To improve: Better hardware resource planning, enhanced data augmentation to mitigate class imbalance, and possibly explore lighter architectures for quicker inference without sacrificing too much accuracy.

References:

CRISP-DM

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 1–11.

Deep Residual Learning (ResNet-18)

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

https://doi.org/10.1109/CVPR.2016.90

Railway track surface faults dataset. (2024). *Data in Brief, 56*, 109958.

https://doi.org/10.1016/j.dib.2024.109958