Railway track faults detection based on image processing using MobileNet

Zaynabe RAGALA*, Asmaâ RETBI, and Samir BENNANI

Rime Team-Networking, Modeling and e-Learning Team- Masi Laboratory- Engineering.3S Research Center-, Mohammadia School of Engineers (EMI), Mohammed V University in Rabat, Morocco, zaynaberagala@research.emi.ac.ma, retbi@emi.ac.ma, sbennani@emi.ac.ma

KEY WORDS: railway track; predictive maintenance; machine learning; Convolutional Neuronal Network; Optimizer

ABSTRACT:
The use of rail transport is increasing. Damaged rails interrupt traffic to carry out repairs. In fact, when a conductor or railway operator reports a damaged rail that interrupted the traffic in the affected area. A team of specialized agents dispatch to the site and carries out the repairs. Hence, the importance of automation of railway track faults detection to ensure track safety and reduce maintenance costs. In this work, we propose a method using image processing technologies and deep learning networks. We have studied the correlation effects of MobileNetV2 and optimization algorithms on accuracy and other performance metrics to generate a model that can achieve good performance in classifying railway track faults. The results show that the Rmsprop can improve the effectiveness of feature extraction and classification of MobileNetV2.

1. INTRODUCTION

The global rail industry spends a lot of money on the maintenance of infrastructure. For Moroccan rail, the investment has passed from 4 billion dirhams between 2000 and 2004 to 10.5 billion dirhams excluding LGV for 2010 to 2015, in fifteen years, more than 750 km of track have been renewed (Saadia Dinia, 2021).

Railways must operate with ever-increasing levels of availability, reliability, safety and security. Tracks are one of the most critical parts of a railway system. Track-Related Accidents It has consistently accounted for 30-40% of all accidents in the United States over the past 10 years (Fra, 2003).
The rail is a steel bar. Two parallel files of rails form the railway track. This rests on sleepers, generally made of concrete, guaranteeing the spacing and inclination of the rails. Beyond their function as a guide and rolling support for trains:

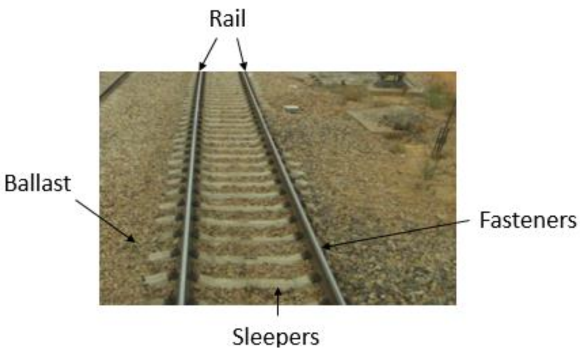


Fig. 1. Components of a railway track

Here are some of the points that it is imperative to check periodically (Lusby et al., 2011):

- Rail defects

Rail fractures are the leading cause of derailments attributable to the state of the track. Attention must pay particular to rail breaks in order to decelerate them as quickly as possible. When a rail is broken, it must be replaced before a wagon or locomotive passes above (Lusby et al., 2011).

- Sleepers

To maintain the gauge and the leveling of the track, we use the sleepers. If gap or leveling has a problem, it may be necessary to replace the sleepers (Lusby et al., 2011).

- Rail seals

To join rail ends, and their dimensions, we use the fishplates. If a fishplate is cracked or broken between two bolt holes, it need to replace it. There must be at least one bolt in each rail end.

- Ballast

The ballast that lies under and around the track performs three functions. (1) It transmits the load of our wagons and locomotives. (2) He preserves the leveling and straightening of the track. (3) It ensures good drainage of the track platform (Lam et al., 2014).
The degradation of track material quality depends not only on traffic, but also on several other factors, including (Chenariyan Nakhaee et al., 2019):

- The quality of the platform and the ballast,
- The efficiency of water flows,
- The type of armament and its age,
- The climatic conditions (bad weather, drought).

We distinguish between different defects a, which will be used later in our research

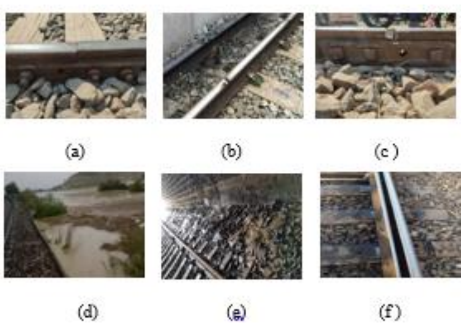


Fig. 2. A figure Sample of track fault cracked-rail (a) damage from collision, (b) Broken rail (c) Flood, (d,e) material theft(f)

Maintenance of the main track is carried out in such a way as to preserve for as long and as economically as possible, for each

line or section of line, the level of quality corresponding to the volume of traffic and the speed of traffic (Spertino et al., 2020).
This is why the maintenance of the equipment base on a cyclical and systematic check. However, the maintenance procedure goes through several phases requiring the movement of specialized inspector, which generates travel time and a blockage of rail traffic (Monitoring, 2018). This is a very subjective technique as it relies on the inspector's interpretation of the image. These inspection reports are the primary data source for managers to develop maintenance plans. Hence the interest in setting up a remote maintenance system, which consists of helping any agent near the location of the breakdown to maintain it, to save the travel time of a specialized team.
The objective is to reduce the cost of movement and the repair time for each field equipment. That said, the remote maintenance system designed must be relevant and useful in terms of early detection of failures, diagnosis and remote intervention to repair equipment and especially for the archiving of data relating to maintenance operations, and to facilitate reuse.
Our approach aims to achieve a major technological breakthrough in detection tools with the aid of artificial intelligence for vision.
Image recognition is a big topic in deep learning and the modern world because of the myriad of application areas (Wang et al., 2020). However, training an algorithm, especially to use it, is very resource-intensive, since thousands of images must be used tens of thousands of times...unless we use mobile networks (Sandler et al., 2018).

MobileNet; designed by researchers at Google; is one of the deep learning models that can be used to detect objects. MobileNet has several advantages compared to Convolutional Neural Networks (CNN) (Howard et al., 2017):

- Extremely light and small
- Infinitely faster
- Extremely accurate, especially in terms of resource consumption
- Easy to configure to improve detection accuracy

The following sections of the document presented as follows. Firstly, section 2 describes the background of MobileNet. Then, section 3 summarizes the most recent related works. After, section 4 detailed the methodology. Finally, section 5 exposes experimental results followed by discussions.

2. BACKGROUND &METHODOLOGY

2.1 MobileNet:

The CNN and MobileNet are identical, except for the convolution part. For a CNN, the standard convolution is 3×3 in size, but in a MobileNet, it is divided into two parts, which are usually called the depth wise separable convolution layer (Howard et al., 2017) :

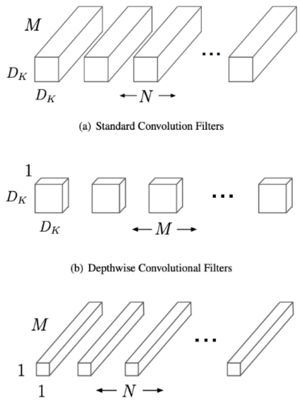


Fig. 3. Standard convolutional filters and depthwise separable filters. (a) Standard convolutional filters, (b) depthwise convolutional filters, and (c) point convolutional filters.

In MobileNet, a Depthwise Separable Convolution replaces the convolution. Depthwise Separable Convolutions is in two steps:

- Depthwise Convolution: consists of applying a filter to each channel, unlike classic convolution, which applies a filter to all channels.
- Pointwise convolution: consists of combining the outputs of the Depthwise Convolution, it called 1×1 convolution.

We can summarize things as follows: a classic convolution will take a classic image with its channels and apply its filters then combine the results (Wang et al., 2020). A MobileNet, on the contrary, will apply filters to each channel then will recombine the results (the depth of the parallelepipeds represents the channels) (Sandler et al., 2018).

2.2 MobileNet Structure:

The architecture of MobileNets is an architecture composed of 28 layers including 13 Depthwise Convolution and 13 Pointwise Convolution:

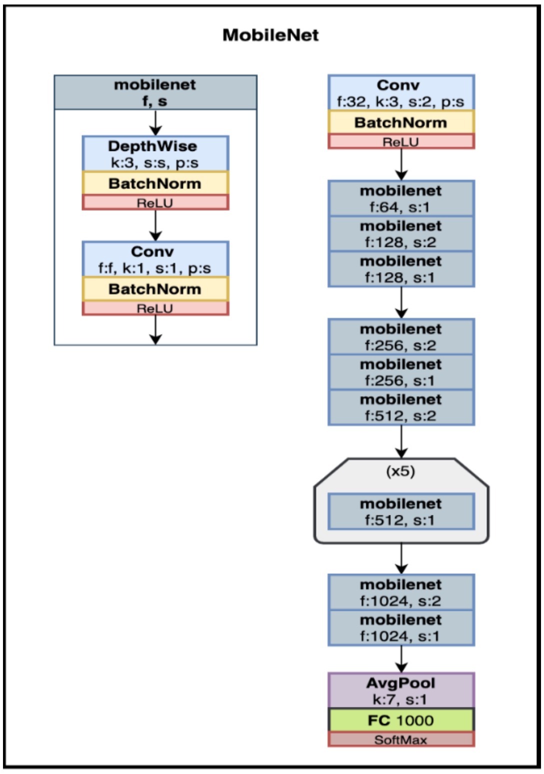


Fig. 4. MobileNet Structure

The MobileNet structure is built on depthwise separable convolutions except for the first layer which is a full convolution. By defining the network in such simple terms we are able to easily explore network topologies to find a good network (Howard et al., 2017). The base MobileNet structure is already small and low latency, many times a specific use case or application may require the model to be smaller and faster.

2.3 MobileNet V2:

MobileNet v2 is a relatively new model that can be used in a variety of devices, including embedded systems and Smartphones (Howard et al., 2017). MobileNetV2 is an inverted residual structure, with thin bottleneck layers used for Inputs and outputs. This uses a novel framework called SSDLite. It contains residual block with stride 1 while another block is present with stride 2, which is used, for downsizing. Each block consists of three types of layers (Sandler et al., 2018).

3. RELATED WORKS

The new trend in machine learning is to use more and more neural networks layers namely as deep learning algorithms. These methods rarely require data preprocessing because they can learn representations directly. These methods used in many complex applications such as image, audio, video, natural language… (Hinton, 2012). These deep learning methods are beneficial in supporting decision-making in railway engineering. Typical deep learning models applied in this field include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) as described below.
(Zauner et al., 2019) used the convolutional network to achieve semantic segmentation of images for automatic railway tamping assistance systems. The idea behind a fully foldable web is extend traditional CNNs by replacing fully connected layers with convolutional layers and add a convolutional transpose layer. Semantic segmentation enables computers to recognize it Objects in pixel-level images. In this way, a new proposed switch tamping assistance system can provide operator support and relief in complex tamping areas.
(Heidarysafa et al., 2019) used RNN to discover narrative causes of accidents in Federal Railroad Administration (US) reports. To convert accident text reports into sequence vectors, they use Term frequencies and the Word2Vec method (where each word is associated with a vector). Then use the RNN to classify the accident important inconsistencies caused and discovered based on these sequence vectors.
A CNN solution for detecting rail surface defects is proposed among them by (Faghih-Roohi et al., 2016), CNN has been used to skip the tedious image feature extraction process in other image processing methods. The accuracy of the classification of rail defects is close to 92%.
In addition to the research presented previously, the experimental studies of (Rampriya et al., 2022) using Deep Neural Network Models for Detecting Obstacles in the Real Time Aerial Railway Track Images obtained a performance rate of 96.75% with SSD Mobile Net V2.

The table 1 summarize the results obtained of the related work carried out in the literature using DNN in railway

Paper	Validation Accuracy	Methods
Zauner et al. (2019)	-	CNN
Heidarysafa et al. (2019)	75 %	Word2Vec in RNN
Faghih-Roohi et al. (2016),	92%	Large DCNN
(Rampriya et al., 2022)	96.75%	SSD Mobile Net V2

Table 1. Related work carried out of using DNN in railway

Unfortunately, the capabilities of advanced deep learning models often come at significant cost in terms of time-consuming training requirements when the data set is huge. For that, the objective of this research was to study the correlation effects of MobileNetV2 a model of MobileNet and optimization algorithms on accuracy and other performance metrics.

4. MATERIALS AND METHODS

4.1 Context of simulations:

Below is the description of the configuration of the computer hardware used for the simulation.

- CPU : Processeur i5-8250U @ 1.6 GHz (8 cpu)
- RAM : 16384MB
- Langage de programmation : Python Version 3.9
- Software : Software : Anaconda 3 / Spyder / Tensor Flow 2.5

4.2 Dataset:

This study presents a proper dataset of 400 track fault images collected manually on real conditions by inspector over a 2-year period (2020–2022).

Class	Number images
cracked rail	162
damage from collision	63
Broken rail	15
Flood	40
material theft	120

Table 2. Data distribution by classes

4.3 Methodology:

The simulation process presented in the figure 5 below gives a general idea of the process adopted for the implementation of the optimization techniques studied and the approach followed to obtain the best results. In addition, the next subsections will give more detail about the methodology with the major steps followed on the simulation process.

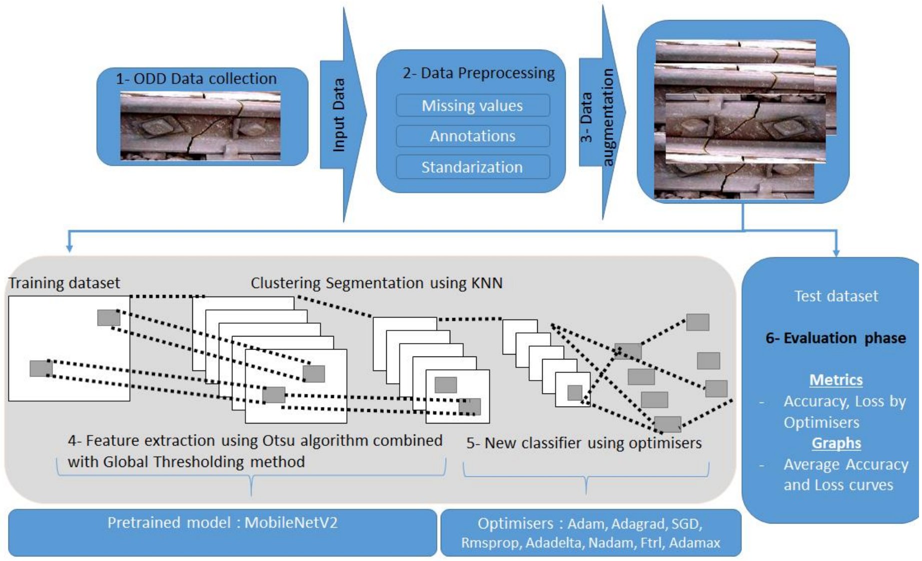


Fig. 5. General schema for experimental studies

a) Data Collection: At the beginning of our work, many human and material resources were devoted to the collection of different track fault annotated manually by experts. The dataset contains 400 images of different track fault in real conditions from railway tracks.

Train		Test		Validation	
Defective	Non defective	Defective	Non defective	Defective	Non defective
280 images	280 images	40 images	40 images	80 images	80 images

Table 3. Balanced Dataset distribution

b) Data Pre-processing: Firstly, the tasks of pre-processing the data begins by spanning the data points. Secondly, an achievement of a split on the data using the distribution detailed in table 3. Finally, the creation of an image creator object.

c) Data Augmentation: The overfitting problem in the training stage of CNNs overcome via data augmentation. In this step, we use several techniques for data augmentation operations, including rotation transformations, horizontal and vertical flips, and intensity disturbance, which include disturbances of brightness, sharpness, and contrast.

d) Feature Extraction: Some variables or characteristics are very important in forming the different models. In this case, and to keep only the most relevant variables and eliminate the harmful characteristics that can disturb the learning.

e) Classification: Classification is a function that need the use of machine learning algorithms that learn how to assign a marker to different classes from the problem sphere.

f) Evaluation: Performance metrics provide hard data to support evaluation. The results are evaluated using distinctive performance metrics like Accuracy, Error Rate, Kappa, Precision, Recall, F1 Score, Mean Absolute Error, and Log Loss to recommend the appropriate feature selection method for prediction.

5. RESULTS AND DISCUSSION

5.1 Results

Firstly, figures 6 and 7 present the performance evaluation of MobileNetV2 model using different optimization algorithms applied to the dataset. Therefore, visualizer displays the average accuracy and loss by algorithm results scores for validation.

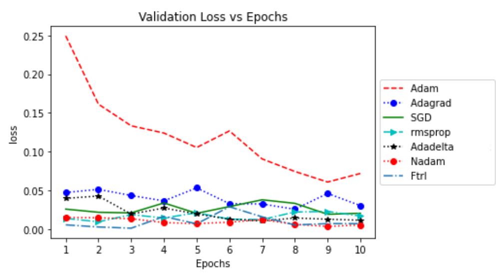


Fig. 6. Validation loss by epochs

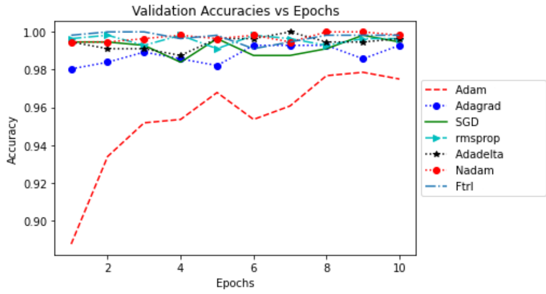


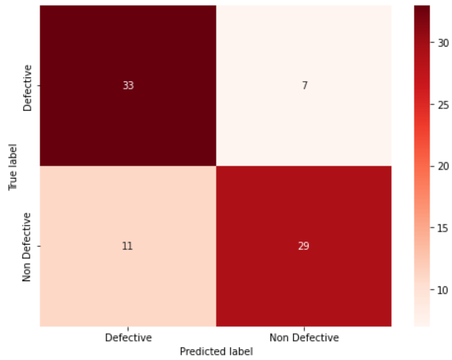
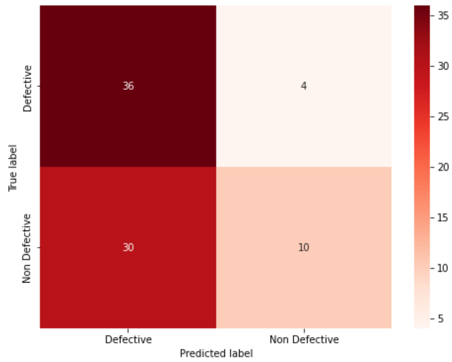
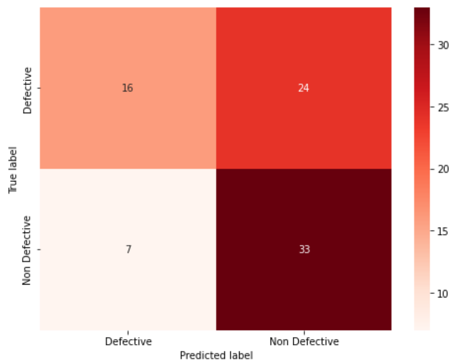
Fig. 7. Validation Accuracy by epochs

Secondly, table 4 present the classification report including a representation of the main classification metrics on defective or non-defective class. This gives a deeper intuition of the classifier behaviour over global accuracy, which can mask functional weaknesses in one class of a multiclass problem.

Classification report		precision	recall	f1-score	support
Adam	Defective	0.75	0.82	0.79	40
	Non-Defective	0.81	0.72	0.76	40
	accuracy			0.78	80
	macro avg	0.78	0.77	0.77	80
	weighted avg	0.78	0.78	0.77	80
SGD	Defective	0.55	0.90	0.68	40
	Non-Defective	0.71	0.25	0.37	40
	accuracy			0.57	80
	macro avg	0.63	0.57	0.52	80
	weighted avg	0.63	0.57	0.52	80
Adadelta	Defective	0.7	0.4	0.51	40
	Non-Defective	0.58	0.82	0.68	40
	accuracy			0.61	80
	macro avg	0.64	0.61	0.59	80
	weighted avg	0.64	0.61	0.59	80
Adagrad	Defective	0.86	0.90	0.88	40
	Non-Defective	0.89	0.85	0.87	40
	accuracy			0.88	80
	macro avg	0.88	0.88	0.87	80
	weighted avg	0.88	0.88	0.87	80
Rmsprop	Defective	0.95	0.95	0.95	40
	Non-Defective	0.95	0.95	0.95	40
	accuracy			0.95	80
	macro avg	0.95	0.95	0.95	80
	weighted avg	0.95	0.95	0.95	80
Ftrl	Defective	0.96	0.65	0.78	40
	Non-Defective	0.74	0.97	0.84	40
	accuracy			0.81	80
	macro avg	0.85	0.81	0.81	80
	weighted avg	0.85	0.81	0.81	80

Table 4. Classification report

Finally, table 5 illustrates the validate accuracy of the model in the confusion matrix.
The model has misclassified several cases; the main reason for this is due to the similarity in appearance between defective ad non-defective images; especially in very small track fault; therefore, the performance of the Mobile Net model is overall good for this experiment.

Model	confusion matrix
Adam	
SGD	
Adadelta	

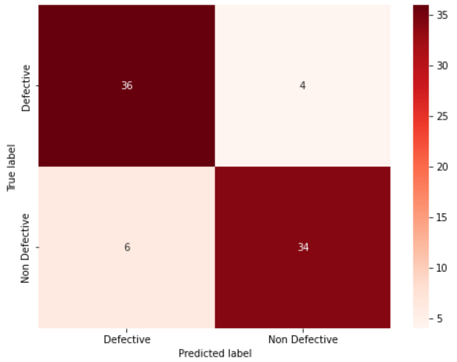
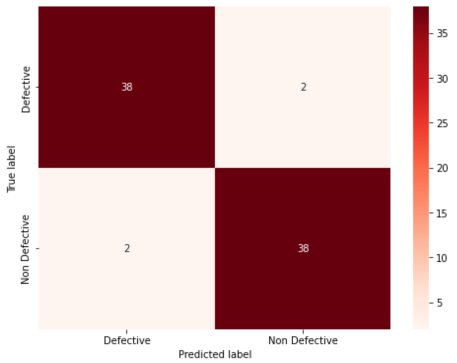
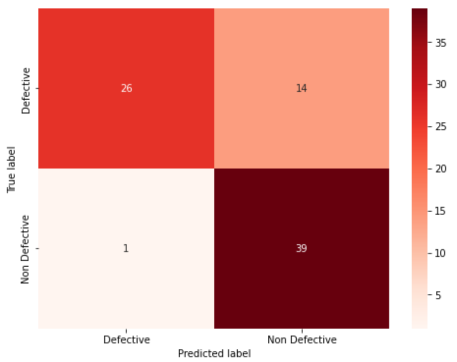
Adagrad	
Rmsprop	
Ftrl	

Table 5. Confusion matrix of models

5.2 Discussion

Firstly, to induce image batches in real-time during the training process, by applying the following variations of the training images: Rotation, Width and/or height shifting of the image dimensions, Zoom the image and Horizontal and/or vertical flipping to enhance the model performance in low data regime. Another challenge during this simulation is the similarity between defective and non-defective images; this explains the error prediction shows in the confusion matrix of Table 5.

Secondly, the main contribution of this work is to design a classification model that achieves good performance in a balanced two class datasets; the classification results from all the optimization algorithms illustrated in figure 6, 7, and detailed for the algorithms in table 4. This allows comparing optimization algorithm; but also, the choice of evaluation metrics depends on the objective of the system; and there are some standard metrics used all over the world such as accuracy and loss. However, when the data skewed; selecting the model on classification accuracy could be severely misleading.
The RMSprop optimizer outperforms the other algorithms and is like the gradient descent algorithm with momentum. The restriction on the oscillations in the vertical direction explained the results. It is a very robust optimizer, which has pseudo-curvature information. Additionally, it can deal with stochastic objectives very nicely, making it applicable to mini batch learning. It converges faster than the other optimizers do.

Finally, the results obtained can be interpreted in two ways, the first one evaluates the impact of the optimization algorithm on the classification of the railway track fault, but also the performance evaluated on datasets whose size is very reduced thus joins the standard problems to which data scientists use different data management techniques. That said, the contribution is very important to improve the conditions for detecting railway track fault and to optimize the detector based on an improved CNN model, which is Mobile Net. To this extent, the number of parameters used in this model predictor present an important use in mobile terminal, which is very suited to this application. In addition, the integration of new CNN optimization techniques will serve to improve the performance but also the sensitivity of the model with a view to integrating new classes and predictors and thereby allowing having tools for more and more efficient and evolving in different context of railway maintenance.

6. CONCLUSION

In this paper, an intelligent system detector allowing to classify defective and non-defective track fault was proposed and evaluated; based on CNN MobileNetV2; In this work; six optimization algorithms are used; namely: Adam; Adagrad; SGD; Rmsprop, Adadelta and Ftrl; to generate a CNN model that can achieve good performance in classifying railway track fault. The results show that the rmsprop can improve the effectiveness of feature extraction and classification. The model constructed based on the best-suited algorithm have boosted the results and achieved best accuracy. The proposed MobileNet achieves great performance in track fault classification and mobile deployment open very interesting perspectives in term of applications.
In fact, we will integrate this work into an inspection tool equipped with a decision support module, to facilitate the synthesis of information and the creation of the inspection report. The exploitation of massive inspection data by MobileNet will also allow predictive maintenance.

References

Chenariyan Nakhaee, M., Hiemstra, D., Stoelinga, M., van Noort, M., 2019. *The Recent Applications of Machine Learning in Rail Track Maintenance*: A Survey, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 91–105. https://doi.org/10.1007/978-3-030-18744-6_6

Faghih-Roohi, S., Hajizadeh, S., Nunez, A., Babuska, R., De Schutter, B., 2016. *Deep convolutional neural networks for detection of rail surface defects*, in: Proceedings of the International Joint Conference on Neural Networks. pp. 2584–2589. https://doi.org/10.1109/IJCNN.2016.7727522

Fra, 2003. FRA Guide for Preparing Accident / Incident Reports.

Heidarysafa, M., Kowsari, K., Barnes, L., Brown, D., 2019. *Analysis of Railway Accidents' Narratives Using Deep Learning*, in: Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018. IEEE, pp. 1446–1453. https://doi.org/10.1109/ICMLA.2018.00235

Hinton, G., 2012. IPAM Summer School 2012 Tutorial on: Deep Learning.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*.

Lam, H.F., Hu, Q., Wong, M.T., 2014. *The Bayesian methodology for the detection of railway ballast damage under a concrete sleeper*. Eng. Struct. 81, 289–301. https://doi.org/10.1016/j.engstruct.2014.08.035

Lusby, R.M., Larsen, J., Ehrgott, M., Ryan, D., 2011. *Railway track allocation*: Models and methods. OR Spectr. https://doi.org/10.1007/s00291-009-0189-0

Monitoring, R., 2018. *State-of-the-Art Review of Railway* Track 1–18. https://doi.org/10.3390/infrastructures3010003

Rampriya, R.S., Suganya, R., Nathan, S., Perumal, P.S., Suganya, R., Nathan, S., Perumal, P.S., 2022. *A Comparative Assessment of Deep Neural Network Models for Detecting Obstacles in the Real Time Aerial Railway Track Images A Comparative Assessment of Deep Neural Network Model*s, Applied Artificial Intelligence. Taylor & Francis. https://doi.org/10.1080/08839514.2021.2018184

Saadia Dinia, 2021. *Morocco's Railways Office to Invest $848 Million in Infrastructure Rehabilitation* [WWW Document]. https://www.moroccoworldnews.com/2021/11/345322/moroccos-railways-office-to-invest-848-million-in-infrastructure-rehabilitation.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. *MobileNetV2: Inverted Residuals and Linear Bottlenecks*, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE,pp.4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Spertino, F., Member, S., Chiodo, E., Member, S., Ciocia, A., Ratclif, A., 2020. *Maintenance Activity , Reliability , Availability and Related Energy Losses in Ten Operating Photovoltaic Systems up to 1 . 8 MW 9994*. https://doi.org/10.1109/TIA.2020.3031547

Wang, W., Li, Y., Zou, T., Wang, X., You, J., Luo, Y., 2020. *A novel image classification approach via dense-mobilenet models*. Mob. Inf. Syst. 2020. https://doi.org/10.1155/2020/7602384

Zauner, G., Mueller, T., Theiss, A., Buerger, M., Auer, F., 2019. *Application of semantic segmentation for an autonomous rail tamping assistance system,* in: IS and T International Symposium on Electronic Imaging Science and Technology. pp. 1–6. https://doi.org/10.2352/ISSN.2470-1173.2019.7.IRIACV-462