Eduvest – Journal of Universal Studies
Volume 5 Number 2, February, 2025
p- ISSN 2775-3735- e-ISSN 2775-3727

CLASSIFICATION OF VERTICAL AND LATERAL TRACK IRREGULARITIES USING GOOGLENET FROM GRAMIAN ANGULAR SUMMATION FIELD ENCODING

Gemuruh Geo Pratama, Vani Virdyawan, Yunendar Aryo Handoko
Institut Teknologi Banding, Indonesia
Email: 23122022@mahasiswa.itb.ac.id

ABSTRACT

Abstract— The ability to classify track conditions has become a critical issue in the railway industry, as delayed detection or unaddressed adverse track conditions can profoundly impact railway safety. Current track maintenance primarily relies on manual inspections and specialized monitoring vehicles, which are constrained by their inspection frequency. Deploying models that correlate vehicle dynamic responses with track conditions in in-service trains could significantly enhance fault detection. However, existing studies utilizing machine learning approaches are notably limited in capturing complex time-series information from vehicle dynamic responses, especially when the data are derived from real measurements rather than simulations. To address these challenges, we propose the application of GoogleNet and Gramian Angular Summation Field (GASF) transformation for classifying track conditions using vehicle dynamic responses. For comparison, we will demonstrate the limitations of traditional machine learning approaches, specifically Logistic Regression and XGBoost, where only the standard deviation and peak value are extracted as features. Subsequently, we propose our approach using the GoogleNet architecture, combined with GASF to transform the time-series data into image representations. Our proposed model achieves high accuracy, in classifying vertical and lateral track conditions, significantly outperforming the machine learning model. The results of this study demonstrate that our proposed method can learn complex nonlinear features, and make accurate classifications. Additionally, the study highlights the inability of the machine learning model, to classify track conditions accurately, and provides evidence that standard deviation and peak value are insufficient as features for complex systems like vehicle dynamic responses.

KEYWORDS *vehicle dynamic response, googlenet, gramian angular summation field, logistic regression, xgboost*

How to cite: Pratama G.G et all. (2025). Classification of Vertical and Lateral Track Irregularities using GoogleNet from Gramian Angular Summation Field Encoding. Journal Eduvest. *Vol 5*(Number 2): 2969-2979
E-ISSN: 2775-3727

INTRODUCTION

Adverse track conditions, such as vertical and lateral adversity, can significantly impact vehicle safety and passenger comfort by causing increased vibration and instability (Liu et al., 2020). A subtle defect on the track even minor ones may result in severe adverse outcomes for the train running at high velocity. Maintaining track conditions is crucial for railway safety. Currently, the process for maintaining these conditions involves preventive maintenance, which is commonly conducted through manual inspection and specialized track-monitoring vehicles. However manual inspection cannot serve as the primary tool for detecting rail damage, as it is conducted based on a predetermined schedule or in response to reports of existing damage. Such delayed responses can result in deferred repairs, potentially leading to undesirable incidents. On the other hand, modern track-monitoring vehicles are equipped with advanced sensors, including lasers, sensitive accelerometers, gyroscopes, and GPS. Despite this technological sophistication, the data collected are used to identify track irregularities based solely on predefined geometry thresholds (Karunianingrum & Widyastuti, 2020), which may not accurately reflect the actual vehicle responses on the track. Several approaches have been undertaken to identify track conditions based on the dynamic response of train vehicles. For instance (Koziol, 2016; Tsunashima & Hirose, 2022) employed a time-frequency analysis approach, specifically using wavelet transform, to analyze car body acceleration data and its correlation with track geometry. With the advancement of information technology, electronics, and the Internet of Things (IoT), sensors installed on train vehicles generate data that can be utilized for maintenance purposes and fault detection. One such application is the identification of track irregularities. Several studies have employed this data-driven approach to develop machine-learning models for classifying track irregularities based on acceleration data (De Rosa et al., 2021; Hoang & Kang, 2019; Tsunashima, 2019; Wikaranadhi et al., 2024). Generally, the detection of track irregularities is conducted using dynamic response data obtained from simulations. Standard deviation and peak values calculated from acceleration are typically used as features, with track geometry that is often predetermined and may not accurately reflect actual track conditions. However, in other domains, research on fault detection has further explored alternative approaches, such as deep learning (Hong et al., 2020; Ma et al., 2022). Deep learning offers several advantages over frequency analysis or traditional machine learning. Among these advantages are its suitability for handling unstructured data, such as the vibration data generated by the dynamic response of trains. Additionally, deep learning automates the process of feature extraction (Wicaksono et al., 2019).

In this paper, we will utilize a variant of deep learning architecture called GoogleNet. Using GoogleNet allows us to create a network that is both deep and wide while maintaining constant computational resources (Seibold et al., 2022). Additionally, GoogleNet has been successfully used in the fault detection domain (Hatami et al., 2018). We will also transform our vehicle dynamic response data from its time series into a 2D image representation using the GASF method (Wang et al., 2023). This method enables us to convert temporal data into 2D images without sacrificing the temporal relations of the data (Li et al., 2020; Yang et al., 2024). Using image representations not only increases the interpretability of the data but also helps us to overcome the problem of imbalanced data, which is inherently present in real-world datasets.

RESEARCH METHODS

The dataset was obtained from special inspections of track vehicles. The acquired data will undergo preprocessing in two stages. The first stage involves labeling the data according to QN1 conditions, guided by the UIC 518 criteria. The second stage involves feature development, which branches into two parts. The first part focuses on feature extraction, specifically the standard deviation and peak value, to create features for our Linear Regression and XGBoost models. The second part, for our GoogleNet model, involves time-series transformation using the Gramian Angular Summation Field.

Datasets

Our Vehicle dynamic response and track irregularities data were acquired from a sensor mounted on an INKA special vehicle Fig. 1. During its operation on a track near Blitar on November 17, 2021 in track with normal gauge 1067 mm and at speed of 97 km/hours

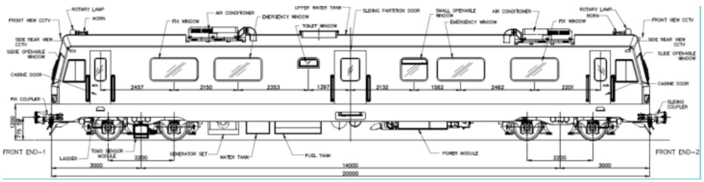

Fig. 1 INKA Track Monitoring Vehicle

Several variables related to the vehicle's dynamic response, used in the classification process, were recorded. These variables include lateral acceleration, and vertical acceleration. For track profile data, the recorded information includes both left and right track profiles, as well as left and right track alignment.

The data collected from the special inspection vehicle comprises track geometry profiles and vehicle dynamic responses. Track conditions measurements can be obtained through the dynamic response of the vehicle, using acceleration measurements. Inspection vehicles are equipped with accelerometers to capture vertical and lateral dynamic responses. These accelerometers are installed on the carbody, bogie frame or wheelsets. The data from the accelerometers can be utilized to analyze track conditions in the range D1. The choice to use accelerometer data from the carbody is driven by the primary goal of developing a model and reading data from accelerometers installed on in-service trains Fig. 2.

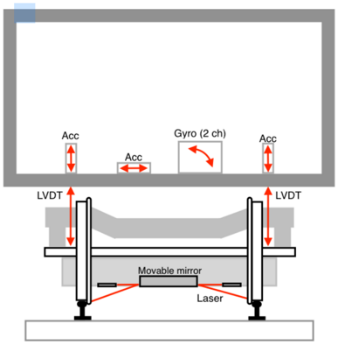

Fig. 2 Principle of inertia-based track measurement system

Three wavelength categories for track adverse condtion, D1, D2, and D3, are considered, as defined in Table 1

Table 1. Track Irregularities Wavelength Range

Category	Wavelength	Type
D1	$3\,m < \lambda \leq 25\,m$	
D2	$25\,m < \lambda \leq 70\,m$	
D3	$70\,m < \lambda \leq 150\,m$	Vertical
	$70\,m < \lambda \leq 200\,m$	Lateral

Particular focus must be given to the D1 wavelength range of track adverse condition, as they have the potential to impact operational safety; for, irregularities within the D2 and D3 ranges are primarily associated with passenger comfort, as outlined in the EN13858:2019.

Method

There are three main processes involved in training our deep learning model: the first involves data preparation, including label and feature development; the second involves model architecture design; and the third is model training.

Label Development

The first step begins with labeling of geometry data, labeling process will encompass vertical and lateral data This data is refined based on the D1 irregularities' spatial wavelength range, spanning 3 to 25 meters, as outlined in EN 13848 using band-pass filter. Corresponding to an operating speed of 97 km/h, this wavelength range is equivalent to a frequency range between 1.07 and 8.98 Hz. furthermore, the data will be segmented using a random start-slicing method, where the starting point of the data is chosen randomly, as illustrated in Fig. 3.

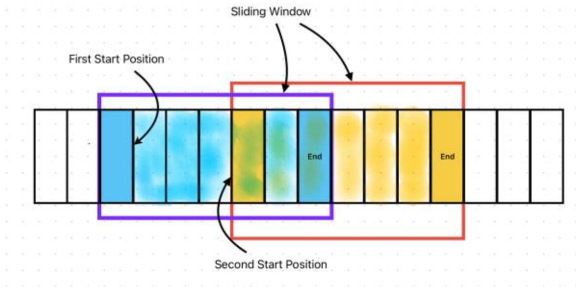

Fig. 3 Random Start Slicing

The sliding window size will consist of 400 points, representing a 100-meter track section, with each data point spaced 0.25 meters apart. Table 2 shows the detailed parameters used in this process.

Table 2. Random Start Slicing Parameters

Parameter	KM 124	KM 125
Data Spacing	0.25 m	0.25 m
Window	400 points	400 points
Section	100 m	100 m
Min Value	180.50	0
Max Value	999.75	499.50
Iteration	7000	
Speed	97 km / hours	

We will iterate the slicing process for 7000 iterations to maximize the results. Subsequently, a filtering process will be implemented to remove duplicate data from our result sets. After sets of segmented data are acquired, the standard deviation and peak value for each segment is calculated as the threshold to determine the track quality index in each segment. The limit values for the lateral level and vertical level are consistent with the QN1 limit specified by UIC 518. A track section is classified as 'Class 0' (indicating acceptability or normal) if both its standard deviation and peak value are within the prescribed limits for the respective feature. Conversely, it is designated as 'Class 1' (signifying unacceptability or adverse condition) if both the standard deviation and peak value exceed these limits. Table 3 shows standard deviation values as provided by UIC518.

Table 3. Standard Deviation Reference

Variable QN 1	Limit Value Vertical	Limit Value Lateral

Standard Deviation	1.8 mm	1.2 mm
Peak Value	8.0 mm	8.0 mm

Table 4 shows the results of the labeling process for lateral and vertical track conditions.

Table 4. Labeling Results

Type	**Track adverse condition (class 1)**	**Track Good Condition (class 0)**
Vertical	3509	715
Lateral	1555	2669

Feature Development

Concerning feature data, we also implement a windowing technique to segment our data, as depicted in Fig. 4

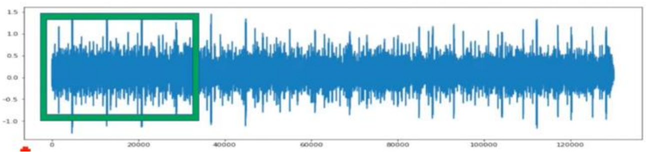

Fig. 4 Chunk the data using the selected window

The selected window size must correspond to the dimensions of the previously defined label, ensuring alignment between the input feature data and the label data. For the features used in our Logistic Regression and XGBoost models, we begin by calculating the standard deviation and determining the peak value for each window of data.

Gramian Angular Summation Field encoding of dynamic response data

For the features of our GoogleNet model the subsequent step, after windowing, involves the application of Piecewise Aggregation Approximation (PAA). This technique serves to diminish the dimensionality of time series data. It operates by segmenting the time series into equal-sized portions, each of which is then represented by its mean value. This method effectively truncates the length of the time series, which is crucial for reducing the dimensions of the Gramian Angular Summation Field (GASF) matrix. We proceeded to transform the feature data into a Gramian Summation Angular Field (GSAF) matrix. Suppose observations of n time series data given by $X = \{x_1, x_2, x_3, \ldots, x_n\}$, rescale X to fall between [-1,1] by using:

$$\tilde{x}_i = \frac{(x_i - max(X)) + (x_i + max(X))}{max(x) - min(X)} \quad (1)$$

In the second step, the rescaled values are transformed into polar coordinates. This is achieved by encoding the data value as the angular cosine and the temporal component as the radius, utilizing the following equation:

$$\{\emptyset = arccos(\tilde{x}_i), -1 < \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X}\; r = \frac{t_i}{N}, t_i \in N \quad (2)$$

The (2) represents the time at i and N as a factor to regulate the span of the polar coordinate. The outcome derived from (2) possesses crucial characteristics. Firstly, the temporal relationship is maintained in the polar coordinates, as an increase in time results in the corresponding value being wrapped among various coordinate points. Second, the point becomes bijective monotonic at coordinate $\tilde{x}_i \in [0, \pi]$. After transformation applied Gramian matrix for each coordinate value.

$$G = (\langle \emptyset_1, \emptyset_1 \rangle \langle \emptyset_1, \emptyset_2 \rangle \cdots \langle \emptyset_1, \emptyset_n \rangle \langle \emptyset_2 \vdots \vdots \ddots \vdots \langle \emptyset_n, \emptyset_1 \rangle \langle \emptyset_n, \emptyset_2 \rangle \cdots \langle \emptyset_n, \emptyset_n \rangle) \quad (3)$$

In the Gramian matrix author (Z. Wang et al., 2015) suggest a modified dot product operation, as the norm of each vector has been adapted to account for time dependency as shown in (3). Specifically: (i) The inner product computation involving two separate observations demonstrates a bias towards the more recent observation, as the norm increases with time; (ii) in calculating the inner product of observation with itself, the resulting norm exhibits bias as well.

$$G = (cos(\emptyset_1 + \emptyset_1)\; cos(\emptyset_1 + \emptyset_2) \cdots cos(\emptyset_1 + \emptyset_n)\; cos(\emptyset_2 + \emptyset_1)\; cos(\emptyset_2 + \emptyset_2) \ldots cos(\emptyset_2 + \emptyset_n) \vdots \vdots \ddots \vdots\; cos(\emptyset_n + \emptyset_1)\; cos(\emptyset_n + \emptyset_2) \cdots cos(\emptyset_n + \emptyset_n)) \quad (4)$$

Equation (4) illustrates the time series being transformed into a polar coordinate and finally into a GASF matrix. Subsequently, this matrix was converted into an image format as shown in Fig. 5, and Fig. 6 shows several examples of image representation resulting from GASF transformation. The image will serve as the input for the training data of our GoogleNet model.

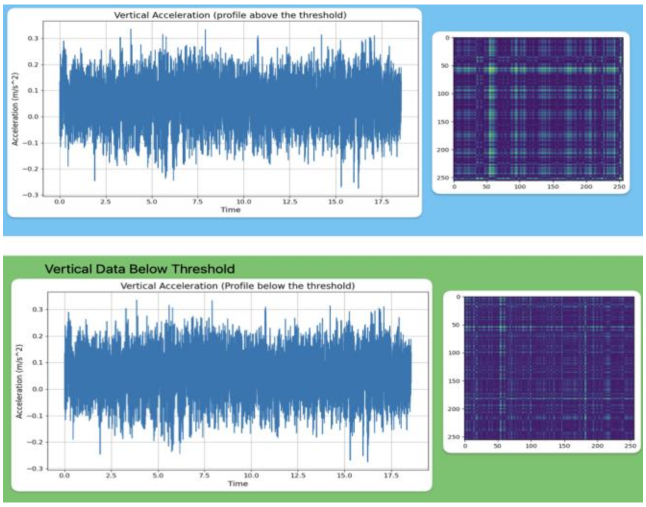

Fig. 5 Vertical Dynamic Response Data Transformation

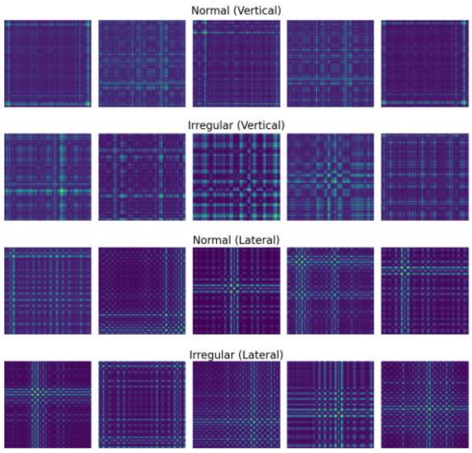

Fig. 6 Images from vehicle dynamic response

Model Training

We begin with Logistic Regression and XGBoost. For both methods, we perform downsampling on our label and feature data to create a balanced dataset. For vertical classification, we downsample to match class 0 track conditions, which consist of 715 data points. For lateral track condition classification, we downsample to match class 1 track conditions, which consist of 1,555 data points.

For GoogleNet we also perform downsampling similar to the machine learning method to rebalance the dataset and We will setup our GoogleNet model according to [13]. All convolutions in the network, including those within the Inception modules Fig. 7, utilize rectified linear activation functions (ReLU). The network's receptive field size is 256x256, and it processes greyscale color channels with mean subtraction. We will use "#3×3 reduce" and "#5×5 reduce" to refer to the number of 1×1 filters in the reduction layer applied before the 3×3 and 5×5 convolutions, respectively. The pool projection column indicates the number of 1×1 filters in the projection layer following the built-in max-pooling. All these reduction and projection layers also use rectified linear activation functions.

Our GoogleNet architecture has 5,971,889 trainable parameters. For both the machine learning and GoogleNet approaches, we use an 80:20 training and testing data split. To address our limited dataset, to train the GoogleNet model we employ cross-validation to obtain optimal hyperparameters. The optimal hyperparameters for vertical data are 55 epochs, batch size 32, learning rate 0.01, weight decay 0.815, and dropout rate 0.4. For lateral data, the best hyperparameters are 45 epochs, batch size 32, learning rate 0.01, weight decay 0.815, and dropout rate 0.4. We applied dropout to prevent overfitting, and weight decay to create a dynamic learning rate to improve convergence even.

RESULT AND DISCUSSION

Result metrics from Logistic Regression and the XGBoost method are shown in Table 5.

Table 5. Logistic Regression And Xgboost Metrics

Model	class	Precision	Recall	Accuracy
Log-Reg (Vertical)	0	0.54	0.57	0.54
	1	0.54	0.51	
XGBoost (Vertical)	0	0.50	0.50	0.50
	1	0.50	0.50	
Log-Reg (Lateral)	0	0.54	0.12	0.49
	1	0.48	0.89	
XGBoost (Lateral)	0	0.52	0.45	0.50
	1	0.48	0.56	

The metrics indicate that Logistic Regression and XGBoost are unable to classify track conditions effectively. The results for class 0 (good) and class 1 (adverse) are no better than random guessing, as the accuracy is only 50%. As shown in Fig. 8, when we plot the relationship between the standard deviation and peak value calculated from the dynamic response, for vertical acceleration or lateral acceleration, the correlation is very weak, with a Pearson correlation coefficient of only 0.03. This suggests that simply calculating the standard deviation and peak value from the dynamic response is insufficient to classify track conditions accurately.

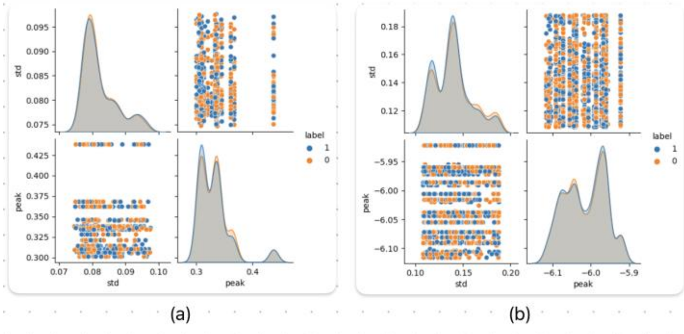

Fig. 7 (a) pair plot of standard deviation and peak value from vertical acceleration, (b) Standard deviation and peak value from lateral acceleration

Cross-validation provides a method to obtain optimal hyperparameters for training our GoogleNet model. As shown in TABLE VII, we achieved a robust model for lateral and vertical classification with accuracies of 98% and 97%, respectively. For the vertical model, the recall for the positive class (adverse condition) is lower than the recall for the negative class (good condition), indicating that the model is less accurate in detecting adverse conditions compared to normal conditions. Despite this, the recall for adverse conditions is still above 90%, specifically 95%, demonstrating strong model performance.

Table 6. Googlenet Model Vertical And Lateral Metrics

Model	**class**	**Precision**	**Recall**	**Accuracy**
GoogleNet (Vertical)	0	0.96	0.99	0.97
	1	0.98	0.95	
GoogleNet (Lateral)	0	0.97	0.99	0.98
	1	0.99	0.97	

CONCLUSION

Combining the GoogleNet model with the GASF method enables highly accurate classification of track conditions, as demonstrated by the metrics evaluation, which shows an accuracy above 95%. This proves that the GASF transformation effectively captures complex features from dynamic response data, allowing GoogleNet to perform feature extraction and classification successfully. In contrast, the machine learning models, Logistic Regression and XGBoost, were unable to learn the data effectively, resulting in only about 50% accuracy. This poor performance is also influenced by the feature extraction method we used—standard deviation and peak value from acceleration data—which is insufficient to represent the complexity of vertical and lateral dynamic responses. Additionally, the almost nonexistent correlation between the peak value and standard deviation in the acceleration data suggests that we need other features to better represent train dynamic responses.

REFERENCES

De Rosa, A., Kulkarni, R., Qazizadeh, A., Berg, M., Di Gialleonardo, E., Facchinetti, A., & Bruni, S. (2021). Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 235(1), 107–120.

Hatami, N., Gavet, Y., & Debayle, J. (2018). Classification of time-series images using deep convolutional neural networks. Tenth International Conference on Machine Vision (ICMV 2017), 10696, 242–249.

Hoang, D.-T., & Kang, H.-J. (2019). A survey on deep learning based bearing fault diagnosis. Neurocomputing, 335, 327–335.

Hong, K., Jin, M., & Huang, H. (2020). Transformer winding fault diagnosis using vibration image and deep learning. IEEE Transactions on Power Delivery, 36(2), 676–685.

Karunianingrum, D. I., & Widyastuti, H. (2020). Penilaian Indeks Kualitas Jalan Rel (Track Quality Index) berdasarkan Standar Perkeretaapian Indonesia (Studi Kasus: Cirebon-Cikampek). Jurnal Aplikasi Teknik Sipil, 18(1), 81–86.

Koziol, P. (2016). Experimental validation of wavelet based solution for dynamic response of railway track subjected to a moving train. Mechanical Systems and Signal Processing, 79, 174–181.

Li, X., Kang, Y., & Li, F. (2020). Forecasting with time series imaging. Expert Systems with Applications, 160, 113680.

Liu, C., Thompson, D., Griffin, M. J., & Entezami, M. (2020). Effect of train speed and track geometry on the ride comfort in high-speed railways based on ISO 2631-1. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 234(7), 765–778.

Ma, K., Chang'an, A. Z., & Yang, F. (2022). Multi-classification of arrhythmias using ResNet with CBAM on CWGAN-GP augmented ECG Gramian Angular Summation Field. Biomedical Signal Processing and Control, 77, 103684.

Seibold, M., Hoch, A., Farshad, M., Navab, N., & Fürnstahl, P. (2022). Conditional generative data augmentation for clinical audio datasets. International Conference on Medical Image Computing and Computer-Assisted Intervention, 345–354.

Tsunashima, H. (2019). Condition monitoring of railway tracks from car-body vibration using a machine learning technique. Applied Sciences, 9(13), 2734.

Tsunashima, H., & Hirose, R. (2022). Condition monitoring of railway track from car-body vibration using time–frequency analysis. Vehicle System Dynamics, 60(4), 1170–1187.

Wang, Y., Zou, Y., Hu, W., Chen, J., & Xiao, Z. (2023). Intelligent fault diagnosis of hydroelectric units based on radar maps and improved GoogleNet by depthwise separate convolution. Measurement Science and Technology, 35(2), 25103.

Wicaksono, S., Husodo, H. T., & Mahyuddin, A. I. (2019). Dynamic analysis of

Bangladesh’s broad gauge train using multibody system. AIP Conference Proceedings, 2187(1).

Wikaranadhi, P., Alfian, S. D., Handoko, Y. A., & Palar, P. S. (2024). Railway track vertical irregularities classification based on vehicle response using machine learning method. Journal of Physics: Conference Series, 2739(1), 12054.

Yang, J., Sun, Y., Chen, Y., Mao, M., Bai, L., & Zhang, S. (2024). Time series-to-image encoding for saturation line prediction using channel and spatial-wise attention network. Expert Systems with Applications, 237, 121440.