ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250327093142

Preliminary communication
Received: 2025-03-27, Accepted: 2025-04-30

A Machine Vision Approach to Assessing Steel Properties through Spark Imaging

Goran Munđar*, Miha Kovačič, Uroš Župerl

Abstract: Accurate and efficient evaluation of steel properties is crucial for modern manufacturing. This study presents a novel approach that combines spark imaging and deep learning to predict carbon content in steel. By capturing and analyzing sparks generated during grinding, the method offers a fast and cost-effective alternative to conventional testing. Using convolutional neural networks (CNNs), the proposed models demonstrate high reliability and adaptability across different steel types. Among the tested architectures, MobileNet-v2 achieved the best performance, balancing accuracy and computational efficiency. The findings highlight the potential of machine vision and artificial intelligence in non-destructive steel analysis, providing rapid and precise insights for industrial applications.

Keywords: Carbon Content Prediction; Convolutional Neural Networks; Deep Learning; Machine Vision; Spark Imaging; Steel Analysis

1 INTRODUCTION

Rapid and reliable determination of carbon content in steel is crucial for quality control in industrial manufacturing, as carbon significantly influences steel properties such as strength, hardness, and durability. Conventional techniques for assessing carbon content, including chemical analysis methods like optical emission spectroscopy and combustion analysis, are accurate but often require extensive sample preparation, specialized laboratories, and substantial time, making them less suitable for rapid, in-line inspection in manufacturing environments [1, 2].

An alternative, traditional method known as the spark test involves observing the characteristics of sparks generated when steel is ground against an abrasive wheel. Spark patterns, such as brightness, shape and stream length, vary according to carbon content, allowing experienced technicians to make qualitative assessments quickly. However, this manual method is highly subjective and lacks quantitative precision [2].

To overcome these limitations, researchers have explored the potential of spark imaging coupled with automated image analysis techniques. Prior studies demonstrated promising results in classifying steels into broad carbon categories using automated image processing methods. Nakata et al. [1] automated spark testing to measure carbon content through controlled imaging and machine-learning analysis. Similarly, Benjawilaikul and Kaewwichit (2022) [3] reported high accuracy in categorizing carbon steels by extracting visual features from spark images using traditional machine learning approaches. These initial efforts highlighted the potential for automating spark analysis but were often limited by the reliance on manually selected image features.

Recent advancements in deep learning, particularly convolutional neural networks (CNNs), offer a powerful solution to automatically capture complex visual features directly from raw image data, eliminating subjective human interpretation and manual feature engineering. CNNs have been successfully applied in various material science tasks, including identifying microstructural features and classifying alloys and metals [4-6]. Specifically, deep learning has been shown to enhance accuracy and robustness significantly over traditional machine-learning methods in image-based material characterization [7].

In line with these advancements, our study proposes a machine vision approach that leverages spark imaging and CNNs to predict steel carbon content rapidly and non-destructively. We employ three CNN architectures: ResNet, MobileNet, and VGG. We evaluated their performance in predicting precise carbon content from spark images. Each CNN architecture provides distinct advantages:

- ResNet (Residual Network) includes skip connections that facilitate training deeper models, significantly improving accuracy [8].
- MobileNet is designed for computational efficiency, using depth wise separable convolutions, making it suitable for real-time industrial applications [9].
- VGG, known for its straightforward deep convolutional layer structure, offers robust image recognition performance, albeit at increased computational cost [10].

By integrating these CNN models into a spark imaging workflow, our approach provides an efficient, accurate, and reproducible method for assessing steel carbon content.

2 METHODOLOGIES

2.1 Experimental Setup

The experiment was performed using 75 different steel samples provided by Štore Steel d.o.o. Fig. 1 illustrates all the steel samples utilized during the experimental procedure. Each sample underwent grinding tests to generate sparks, which were captured using a customized CNC grinding setup. A ceramic grinding pin (40 mm diameter, 10 mm height, grain size K30) suitable for grinding steel and cast iron was used for grinding. To ensure consistent grinding force, a pneumatic cylinder with a proportional electro-pneumatic pressure regulator (Enfield TR-010-g10-s) was utilized. A constant pressure of 0.6 bar was applied, producing a steady grinding force of approximately 75.4 N.

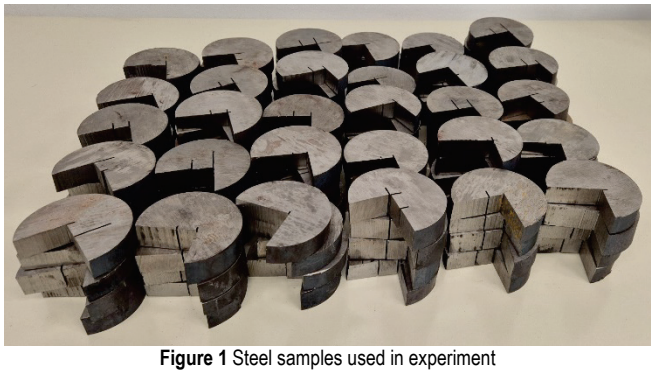

Figure 1 Steel samples used in experiment

The experimental setup during grinding of steel samples is shown on Fig. 2. This figure also clearly indicates the positions of the grinding tool, steel sample, camera, and pneumatic cylinder within the setup.

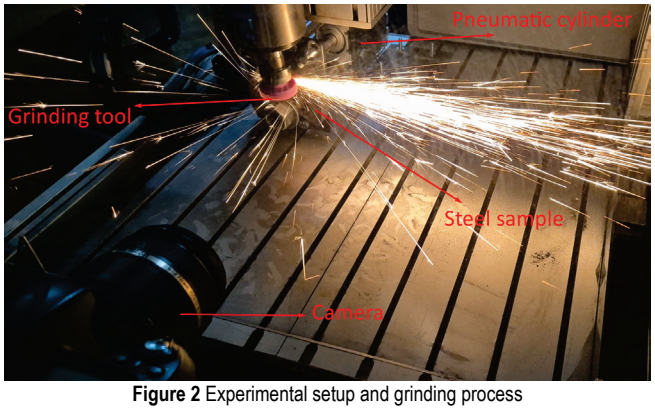

Figure 2 Experimental setup and grinding process

Spark images were captured using a Nikon D90 camera equipped with a 23 mm focal length lens, positioned strategically behind the grinding wheel. The camera was set to capture images at a resolution of 3216 × 2136 pixels.

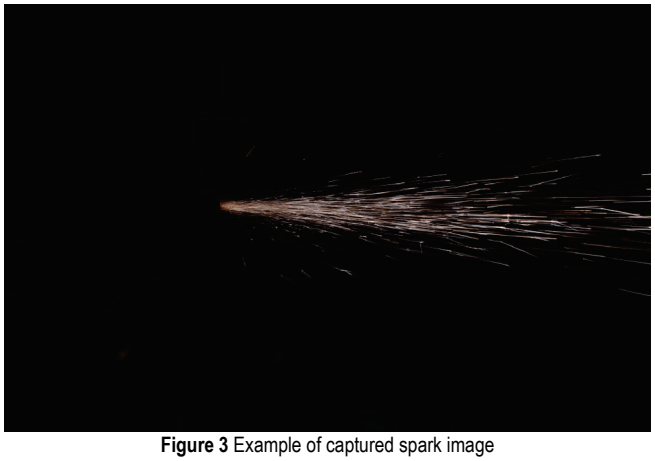

Figure 3 Example of captured spark image

Camera parameters were carefully selected (shutter speed of 1/160 s, aperture f/8, ISO 200, and white balance set at 2500 K) to ensure optimal image quality. These settings effectively eliminated the influence of external lighting and environmental conditions, as sparks emitted during grinding were the brightest elements captured by the camera. Consequently, the background, tools, and other surroundings remained invisible, ensuring consistent image capture conditions and reliable spark feature extraction.

For each steel sample, sparks were recorded across four separate grinding cycles, resulting in approximately 60 spark images per sample. An example of a captured spark image is presented in Fig. 3.

2.2 Image Preprocessing

Captured spark images required preprocessing before use in CNN model training. Initial images (3216 × 2136 pixels) contained substantial irrelevant black background information. Consequently, images were cropped to dimensions of 1400 × 900 pixels, capturing only the relevant spark regions. Further, images were resized to 350 × 225 pixels to enhance computational efficiency. MATLAB R2023b software was employed for these preprocessing steps, enabling accurate isolation of spark characteristics relevant for carbon content prediction. Fig. 4 presents the preprocessing steps applied to spark images, including cropping to remove irrelevant background and resizing for efficient CNN model training.

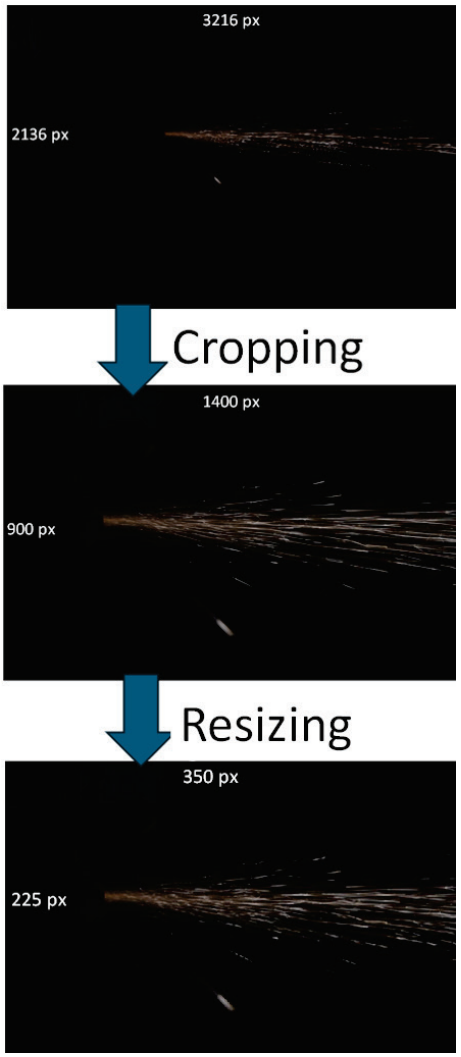

Figure 4 Steps of image preprocessing – cropping and resizing of spark images

2.3 CNN Model Development and Training

To predict steel carbon content based on spark image characteristics, three convolutional neural network (CNN) architectures were specifically chosen: ResNet-50, MobileNet-v2, and VGG-16. These architectures were selected primarily due to their immediate availability within MATLAB's Deep Learning Toolbox [11], facilitating efficient model implementation and ensuring robust support for training and validation procedures.

Moreover, these architectures were strategically chosen to represent diverse computational complexities, network depths, and parameter counts, providing a comprehensive assessment of the trade-offs between accuracy and computational resources:

- ResNet-50: With 50 layers and approximately 25.6 million learnable parameters, ResNet-50 includes residual (skip) connections that effectively alleviate issues related to training deeper networks. This architecture is particularly known for its high accuracy and robustness in complex visual tasks.
- MobileNet-v2: With 53 layers but only around 3.5 million learnable parameters, MobileNet-v2 emphasizes computational efficiency through depthwise separable convolutions. Its lightweight design makes it highly suitable for real-time industrial applications where computational resources may be constrained.
- VGG-16: Featuring 16 layers and about 138 million learnable parameters, VGG-16 is characterized by its straightforward yet deep convolutional structure, providing robust image recognition performance. However, its higher parameter count entails greater computational costs and resource requirements compared to the other two models.

These selections enabled a thorough comparative analysis across architectures of significantly different depths and computational footprints, ultimately guiding the identification of a balanced and practical solution for industrial deployment.

CNN models were adapted through transfer learning from pre-trained models, modifying both the input and output layers to transform the architectures from classification into regression models suitable for predicting carbon content. Specifically, the image input layer was modified to accept images with dimensions of 225 × 350 pixels. Additionally, the final three layers of each CNN model—fc1000, fc1000_softmax and the classification layer were removed. In their place, six new layers were added: three fully connected layers with 256, 128, and 1 neuron, respectively, interleaved with three ReLU activation layers. These modifications enabled the CNNs to output a continuous value corresponding to carbon content. The structural changes are illustrated in Fig. 5.

The training utilized MATLAB R2023b with the Deep Learning Toolbox [11]. Training parameters were standardized across all architectures:

- Epochs: 50
- Mini-batch size: 64
- Learning rate: 0.001
- Optimizer: Adam.

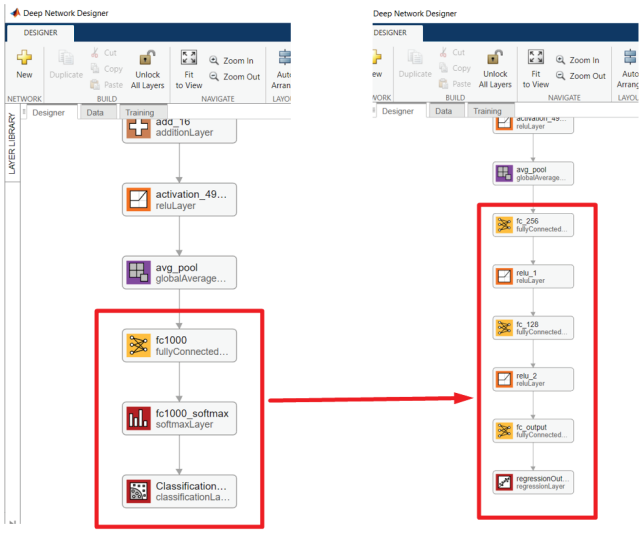

Figure 5 Modification of CNN architectures from classification to regression models

Models were validated using five-fold cross-validation [12] to ensure robustness and reduce bias. Tab. 1 shows the distribution of 75 steel samples used in five-fold cross-validation, listed by their batch (sarge) values. Each fold contained 15 unique samples, ensuring representative coverage of various carbon contents across the training and testing datasets. This division enabled robust evaluation and minimized bias in assessing the performance of the CNN models.

Table 1 Distribution of steel samples in five-fold cross-validation

Fold1	Fold2	Fold3	Fold4	Fold5
91563	93677	93646	93322	93668
92456	91461	92213	93647	92326
91457	92415	93541	92250	92713
91959	93746	92252	93797	93802
92423	92412	92147	92065	93644
92369	91802	92781	91942	91677
91724	93139	92331	93137	91721
92062	93645	92329	91943	93696
93278	92715	93511	92613	92251
91688	92718	92624	92710	93516
91958	93648	92047	92716	92623
93545	92253	91678	93544	93518
93706	91792	92943	92249	93873
92424	92622	93642	92145	92796
93675	92330	92328	92471	92215

2.4 Evaluation Metrics

The performance of each CNN model was evaluated using Mean Absolute Percentage Error (*MAPE*), Root Mean Squared Error (*RMSE*) and the coefficient of determination (R^2). These metrics provided comprehensive quantitative measures of prediction accuracy and model explanatory power. Lower *MAPE* and *RMSE* values indicate higher prediction accuracy, whereas R^2 values closer to 1 indicate a better fit of the regression model to the actual data. Training times were also recorded to assess computational efficiency, crucial for practical applicability in industrial scenarios. Detailed comparisons of these metrics are presented in Chapter 3.

3 RESULTS

The predictive performance of each CNN model was evaluated based on *MAPE*, *RMSE*, and R^2 metrics, summarized across five-fold cross-validation.

MAPE results are presented in Tab. 2 and Fig. 6. This metric represents the relative error percentage between the predicted and actual carbon content values. MobileNet-v2 achieved the lowest average *MAPE* (14.52%), making it the most accurate model. ResNet-50 followed with an average *MAPE* of 17.01%, while VGG-16 exhibited the highest error at 21.09%, indicating weaker predictive performance.

Table 2 Mean Absolute Percentage Error (MAPE) for each model across five folds

MAPE (%)	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
ResNet50	14.85	11.55	18.82	19.38	20.43	17.01
MobileNet-v2	15.39	12.10	13.13	19.37	12.62	14.52
VGG-16	24.52	19.11	17.10	25.74	18.99	21.09

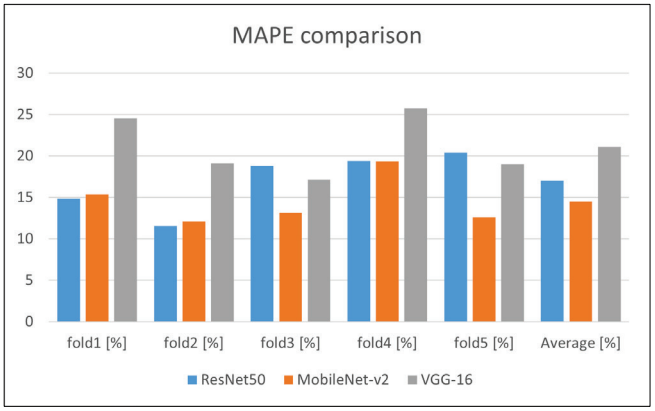

Figure 6 Comparison of Mean Absolute Percentage Error (*MAPE*) across different CNN models

RMSE values are summarized in Tab. 3 and Fig. 7, providing an absolute measure of prediction error. Once again, MobileNet-v2 showed the lowest RMSE (0.0623), followed by ResNet-50 (0.0665) and VGG-16 (0.0705). These results align with the *MAPE* findings, further confirming that MobileNet-v2 was the most effective model in reducing absolute prediction errors for carbon content.

Table 3 Root Mean Squared Error (*RMSE*) for each model across five folds

RMSE	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
ResNet50	0.0887	0.0631	0.0650	0.0616	0.0543	0.0665
MobileNet-v2	0.089	0.0659	0.0480	0.0595	0.0492	0.0623
VGG-16	0.0983	0.07	0.0588	0.0685	0.0568	0.0705

R^2 metric, shown in Fig. 8, clearly demonstrated that MobileNet-v2 provided the highest average explanatory power (R^2 = 0.786), signifying strong model performance. ResNet-50 demonstrated a moderate average R^2 of 0.683, whereas VGG-16 had the lowest explanatory power with an average R^2 of 0.495.

The computational efficiency of each model was assessed based on training time, as shown in Fig. 9.

MobileNet-v2 had the shortest training time at 5.13 minutes, making it not only the most accurate but also the fastest to train. VGG-16 was the slowest, taking 9.18 minutes, due to its deeper architecture and higher parameter count.

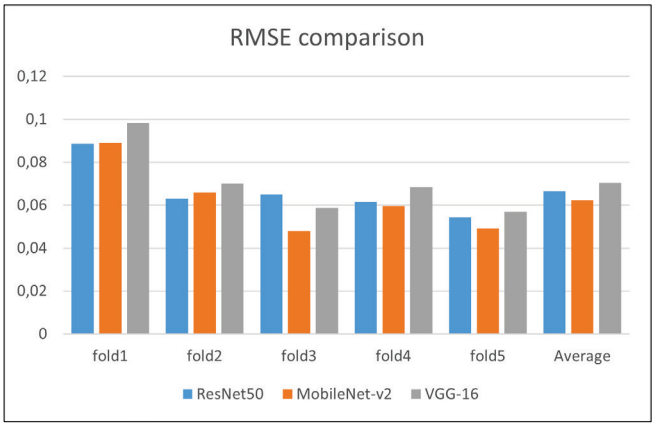

Figure 7 Comparison of Root Mean Squared Error (*RMSE*) across different CNN models

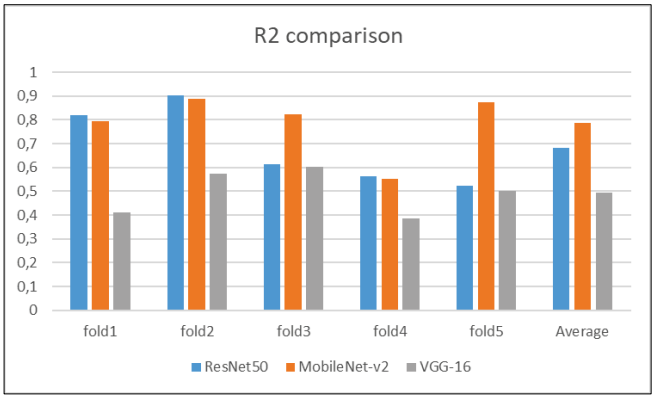

Figure 8 Comparison of Coefficient of determination (R^2) across different CNN models

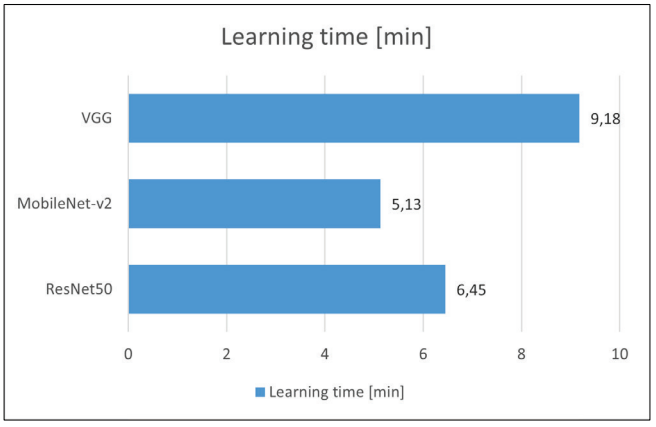

Figure 9 Training time comparison for different CNN models

4 CONCLUSIONS

This study explored the application of CNNs for predicting carbon content in steel based on spark images. Three different architectures (ResNet-50, MobileNet-v2, and VGG-16) were evaluated using a five-fold cross-validation approach, assessing their prediction accuracy and computational efficiency.

The results demonstrated that MobileNet-v2 outperformed both ResNet-50 and VGG-16, achieving the lowest *MAPE* of 14.52%, lowest *RMSE* of 0.0623 and highest

coefficient of determination (R^2) of 0.7861. Additionally, MobileNet-v2 required the shortest training time of 5.13 min, making it the most efficient model. While ResNet-50 provided competitive accuracy (*MAPE*: 17.01%, *RMSE*: 0.0665, R^2: 0.68372), it was slightly slower (6.45 min). In contrast, VGG-16 showed the highest errors (*MAPE*: 21.09%, *RMSE*: 0.0705, R^2: 0.49508) and the longest training time (9.18 min), indicating that it may not be the most suitable model for this task. MobileNet-v2 is the most suitable model for industrial applications, offering efficiency and accuracy for real-time steel classification.

The current dataset of 75 steel samples yielded promising predictive accuracy. However, to enhance model robustness and generalization for broader industrial application, future work will involve expanding our database by including more steel samples and a wider variety of steel materials. Additionally, incorporating advanced CNN architectures, hybrid machine-learning methods, and attention visualization techniques, such as attention maps or saliency analyses, will be considered. These visualization methods would enhance the decision-making process of CNN models by identifying critical spark image regions influencing predictions. This study primarily aimed to demonstrate the feasibility of predicting steel carbon content using CNNs and spark imaging; further refinements will provide more detailed insights into model interpretability and practical implementation.

Overall, this research highlights the significant potential of machine vision combined with deep learning for rapid, accurate, and non-destructive steel property analysis in industrial settings.

Acknowledgements

This research was funded by the Slovenian Research Agency; grant numbers ARIS P2-0162 and P2-0157.

5 REFERENCES

[1] Nakata, T. (2012). Development of automated spark testing technique by image processing to measure carbon content in steel materials. *IFAC Proceedings Volumes*, *45*(23), 118-119. https://doi.org/10.3182/20120910-3-JP-4023.00029

[2] See https://www.ispatguru.com/spark-testing-of-steels/.

[3] Benjawilaikul, T., & Kaewwichit, T. (2022). Classification of Carbon Steels by Automated Spark Test Technique Using Feature Extraction Based on Machine Learning Image Processing: DOI: 10.14416/j.ind.tech.2022.08.005. *The Journal of Industrial Technology*, *18*(2), Article 2.

[4] Azimi, S. M., Britz, D., Engstler, M., Fritz, M., & Mücklich, F. (2018). Advanced steel microstructural classification by deep learning methods. *Scientific reports*, *8*(1), 2128. https://doi.org/10.1038/s41598-018-20037-5

[5] DeCost, B. L., Francis, T., & Holm, E. A. (2017). Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. *Acta Materialia*, *133*, 30-40. https://doi.org/10.1016/j.actamat.2017.05.014

[6] Zhang, Z., Mativenga, P., Zhang, W., & Huang, S. Q. (2024). Deep Learning-Driven Prediction of Mechanical Properties of 316L Stainless Steel Metallographic by Laser Powder Bed Fusion. *Micromachines*, *15*(9), 1167. https://doi.org/10.3390/mi15091167

[7] Xu, H., Xu, Z., & Zhang, K. (2019, November). Mechanical properties prediction for hot roll steel using convolutional neural network. In *International Conference on Bio-Inspired Computing: Theories and Applications* (pp. 565-575). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-15-3415-7_47

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.4128/9781606492147

[9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[10] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[11] MathWorks (2023). MATLAB Deep Learning Toolbox, R2023b. The MathWorks Inc., Natick, Massachusetts, USA. Available: https://www.mathworks.com/products/deep-learning.html

[12] Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, *21*, 137-146. https://doi.org/10.1007/s11222-009-9153-8

Authors' contacts:

Goran Munđar, M.Sc. Mech. Eng.
(Corresponding author)
University of Maribor, Faculty of Mechanical Engineering,
Smetanova 17, 2000 Maribor, Slovenia
goran.mundjar@um.si

Miha Kovačič, PhD, Assoc. Prof.
Štore Steel d.o.o.
Železarska 3, 3220 Štore, Slovenia
miha.kovacic@store-steel.si

Uroš Župerl, PhD, Assoc. Prof.
University of Maribor, Faculty of Mechanical Engineering,
Smetanova 17, 2000 Maribor, Slovenia
uros.zuperl@um.si