1. Alexander SIERADZKI[1], 2. Agata PRZYBYŚ-MAŁACZEK[1], 3. Maciej JUREWICZ[1], 4. Bartłomiej BŁĄDEK[1], 5. Albina JEGOROWA[2], 6. Jarosław KUREK[1]

Department of Artificial Intelligence, Institute of Information Technology, Warsaw University of Life Sciences (1),
Department of Mechanical Processing of Wood, Institute of Wood Sciences and Furniture, Warsaw University of Life Sciences (2)
ORCID: 1. 0009-0004-9887-1054; 2. 0009-0007-8436-0271, 3. 0000-0003-4206-3723,
4. 0009-0004-5201-4671, 5. 0000-0002-8935-845X, 6. 0000-0002-2789-4732

doi:10.15199/48.2025.03.45

Comparative analysis of grad-CAM and LIME for explainable AI in CNN-based drilled hole classification in melamine faced chipboard

Streszczenie. *W przemyśle meblarskim precyzja w wierceniu otworów w płycie wiórowej laminowanej melaminą jest kluczowa dla utrzymania jakości produktu i minimalizacji strat finansowych. Ręczne monitorowanie stanu wierteł, choć do pewnego stopnia skuteczne, jest nieefektywne i mało precyzyjne. W niniejszym artykule przedstawiono porównawczą analizę dwóch technik wyjaśnialnej sztucznej inteligencji (XAI) — Gradient-weighted Class Activation Mapping (Grad-CAM) i Local Interpretable Model-Agnostic Explanations (LIME) — zastosowanych w podejściu opartym na konwolucyjnej sieci neuronowej (CNN) do automatycznego monitorowania stanu narzędzi (TCM). Wykorzystując wstępnie wytrenowaną sieć VGG16, klasyfikujemy stan wywierconych otworów na trzy kategorie: zielony, żółty i czerwony. Zarówno Grad-CAM, jak i LIME dostarczają wizualnych wyjaśnień dla predykcji modelu, zwiększając przejrzystość i niezawodność systemu. Nasze porównawcze badanie podkreśla mocne i słabe strony każdej metody w interpretacji decyzji modelu CNN, mając na celu poprawę efektywności i wiarygodności zautomatyzowanych systemów TCM. Proponowane podejście wykazuje znaczący potencjał w zastosowaniach przemysłowych, gdzie zrozumienie procesu decyzyjnego modeli AI jest tak samo krytyczne jak ich dokładność.*

Abstract. *In the furniture industry, precision in drilling holes in melamine-faced chipboard is crucial to maintaining product quality and minimizing financial losses. Manual monitoring of drill conditions, while somewhat effective, is inefficient and imprecise. This paper presents a comparative analysis of two Explainable AI (XAI) techniques-Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-Agnostic Explanations (LIME)—applied to a Convolutional Neural Network (CNN) based approach for automated tool condition monitoring (TCM). By leveraging the VGG16 pretrained network, we classify the condition of drilled holes into three categories: Green, Yellow, and Red. Both Grad-CAM and LIME provide visual explanations for the model's predictions, enhancing the transparency and reliability of the system. Our comparative study highlights the strengths and limitations of each method in interpreting the CNN model's decisions, ultimately aiming to improve the effectiveness and trustworthiness of automated TCM systems. The proposed approach shows significant potential for industrial applications, where understanding the decision-making process of AI models is as critical as their accuracy.* (***Analiza porównawcza grad-CAM i LIME pod kątem możliwej do wyjaśnienia sztucznej inteligencji w opartej na CNN klasyfikacji otworów wierconych w płycie wiórowej pokrytej melaminą***)

Słowa kluczowe: grad-cam, lime, wyjaśnialna ai, uczenie maszynowe, płyta wiórowa laminowana, monitorowanie stanu narzędzia
Keywords: grad-cam, lime, explainable ai, machine learning, melamine faced chipboard, tool condition monitoring

Introduction

The furniture manufacturing process encompasses numerous complex and precision-dependent steps. One of the most critical phases is drilling holes in melamine-faced chipboard, where inaccuracies can lead to substantial financial losses due to diminished product quality. Traditionally, the condition of drills has been monitored manually to determine the optimal time for replacement, ensuring consistent product quality. However, manual monitoring, while providing some control, lacks efficiency and precision. Therefore, there is a growing need for an automated, accurate and efficient solution.

Various Tool Condition Monitoring (TCM) methodologies have been developed to evaluate and assess the condition of different tools, including drills [3]. Although these approaches can yield accurate results, they often involve extensive preprocessing, and errors at any stage can compromise the final outcome [12], [13], [14], [15], [16], [17], [18].

Despite sophisticated features derived from vast amounts of collected data, the accuracy of such solutions rarely exceeds 90% [4], [6], [7], [8], [9].

Convolutional Neural Networks (CNNs) have the capability to automatically extract relevant features from images, reducing the need for manual feature engineering [1], [2], [5], [10], [11]. However, the black-box nature of these models raises concerns regarding their interpretability and trustworthiness, especially in critical applications. Explainable AI (XAI) strives to make machine learning models more interpretable and transparent. Prominent techniques in XAI include Gradient-weighted Class Activation Mapping (Grad-CAM) [20] and Local Interpretable Model-Agnostic Explanations (LIME) [21], [22], which provide insights into individual predictions by highlighting important regions in the input data.

This paper presents a comparative analysis of Grad-CAM and LIME applied to a CNN-based classification model for drilled holes in melamine-faced chipboard. By integrating these XAI techniques, we aim to improve the interpretability of the model, identify potential issues, and enhance trust in the automated TCM system.

Data Set

The dataset consists of images captured during controlled drilling experiments. The drilling was performed using a standard CNC vertical machining center, Busellato Jet 100, in Thiene, Italy. The material used was a typical laminated chipboard (U511SM – Swiss Krono 88 Group) widely utilized in the furniture industry, with dimensions of 2500x300x18 mm. The drilling tool applied was a 12mm Faba WP-01 drill equipped with a tungsten carbide tip.

Table 1 summarizes the data acquisition process and the final corner wear measurements for each drill at the end of the last drilling cycle.

Table 1. Summary of data collection and drill wear measurements

Drill #	Green/Yellow/Red	Total Images
1	840/420/406	2,520
2	840/700/280	2,520
3	700/560/420	2,100
4	840/560/280	2,520
5	560/560/560	1,680

CNN Model and VGG16 Pretrained Network

To classify the condition of drilled holes, we applied a Convolutional Neural Network (CNN) based on the VGG16 architecture. The VGG16 network, pretrained on the ImageNet dataset, was utilized as the base model. The final layers were modified to suit our classification task, adapting it to distinguish between the three drill conditions: Green (good condition), Yellow (moderate wear), and Red (severe wear).

The dataset was divided into five folds, each representing a different drill, to perform cross-validation. This approach ensured that the model's performance was consistently evaluated and generalized well to unseen data. In each fold, one subset was used for testing while the others were used for training.

The training process involved fine-tuning the VGG16 network on our dataset of drilled hole images. The model's performance was evaluated based on accuracy and the number of misclassified cases. Table II summarizes the evaluation results for each fold.

Table 2. Summary of data collection and drill wear measurements

Drill #	Acc %	Green as Red	Red as Gren	Total
1	71.00	1	3	4
2	70.82	8	11	19
3	63.98	1	58	59
4	56.13	61	0	61
5	70.77	1	15	16
Total	66.60	72	87	159

The overall accuracy across all folds was 66.60%. The highest accuracy was observed in Fold 1 (71.00%), while Fold 4 showed the lowest accuracy (56.13%).

Explainable AI Techniques

To interpret the predictions of our CNN model and improve its transparency, we applied two XAI techniques: Gradient weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-Agnostic Explanations (LIME). Both methods provide visual explanations by highlighting important regions in the input images that influence the model's predictions.

A. Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM uses the gradients of the target class flowing into the final convolutional layer to produce a localization map highlighting important regions in the image [22]. It operates
by:

- Performing a forward pass to compute activations.
- Computing gradients of the classification score with respect to feature maps.
- Averaging these gradients to obtain importance weights.
- Combining the weights with the feature maps to produce a heatmap.
- Overlaying the heatmap on the original image for visualization.

B. Local Interpretable Model-Agnostic Explanations (LIME)

LIME approximates the classification behavior of a complex model with a simpler, interpretable model to provide local explanations [27]. For image data, LIME:

- Segments the image into superpixels.
- Generates synthetic images by perturbing these superpixels.
- Classifies the synthetic images using the original model.
- Fits an interpretable model (e.g., linear model) to these samples.
- Highlights superpixels that contribute most to the prediction.

Numerical Experiments

We applied both Grad-CAM and LIME to interpret the predictions of our CNN model. Figures 1 and 2 show examples of correctly classified images with their respective explanations.

Figures 3 and 4 display examples of misclassified images and their explanations.

A. Analysis of Grad-CAM Explanations

Grad-CAM heatmaps indicate that the model focuses on specific regions around the drilled holes when making predictions. In correctly classified images, the highlighted areas correspond to the edges and wear patterns indicative of the drill condition. In misclassified cases, the model may focus on irrelevant regions, leading to incorrect predictions.

B. Analysis of LIME Explanations

LIME explanations highlight superpixels that contribute most to the prediction. In correctly classified images, LIME identifies segments that align with areas of wear or characteristic patterns. Misclassifications often result from the model attributing importance to superpixels that do not correspond to meaningful features, suggesting that the model may be sensitive to noise or irrelevant textures.

Discussion

The comparative analysis of Grad-CAM and LIME reveals distinct advantages and disadvantages for each method in interpreting the CNN model's decisions.

A. Advantages and Disadvantages of Grad-CAM

Grad-CAM offers several advantages:

- Class-Discriminative Localization: Grad-CAM provides class-specific heatmaps, highlighting the regions that positively influence the predicted class [22].
- Alignment with CNN Architecture: It leverages the convolutional layers' spatial information, making it particularly effective for CNNs.
- Computational Efficiency: Grad-CAM requires only a single backward pass to compute gradients, making it relatively efficient.

However, Grad-CAM has some limitations:

- Model Dependency: It is inherently tied to models with convolutional layers, limiting its applicability to other architectures.
- Coarse Explanations: The resulting heatmaps may be coarse and lack fine-grained details, potentially overlooking subtle features.
- Gradient Saturation: In some cases, gradients can saturate, leading to less informative heatmaps.

B. Advantages and Disadvantages of LIME

LIME's advantages include:

- Model-Agnostic Explanations: LIME can be applied to any classifier, regardless of its internal structure [27].
- Local Interpretability: It provides explanations for individual predictions by approximating the model locally with an interpretable one.
- Superpixel-Based Insights: By highlighting superpixels, LIME can offer insights into specific image regions influencing the prediction.

The disadvantages of LIME are:

- Dependency on Segmentation: The quality of explanations depends on the image segmentation algorithm; poor segmentation can lead to misleading explanations.
- Computational Cost: Generating explanations requires creating and classifying multiple perturbed samples, which can be computationally intensive.
- Approximation Errors: The linear model used to approximate the complex model may not capture non-linear interactions, potentially reducing explanation fidelity.

Figure 1 presents examples of LIME and Grad-CAM visualizations for correctly classified images. These visualizations help in understanding the model's focus areas and can be used to identify potential issues in the model or the dataset.

Figure 2 shows examples of LIME and Grad-CAM visualizations for incorrectly classified images. These visualizations provide insights into the model's errors, highlighting areas where the model's predictions were influenced by irrelevant or misleading features in the images.

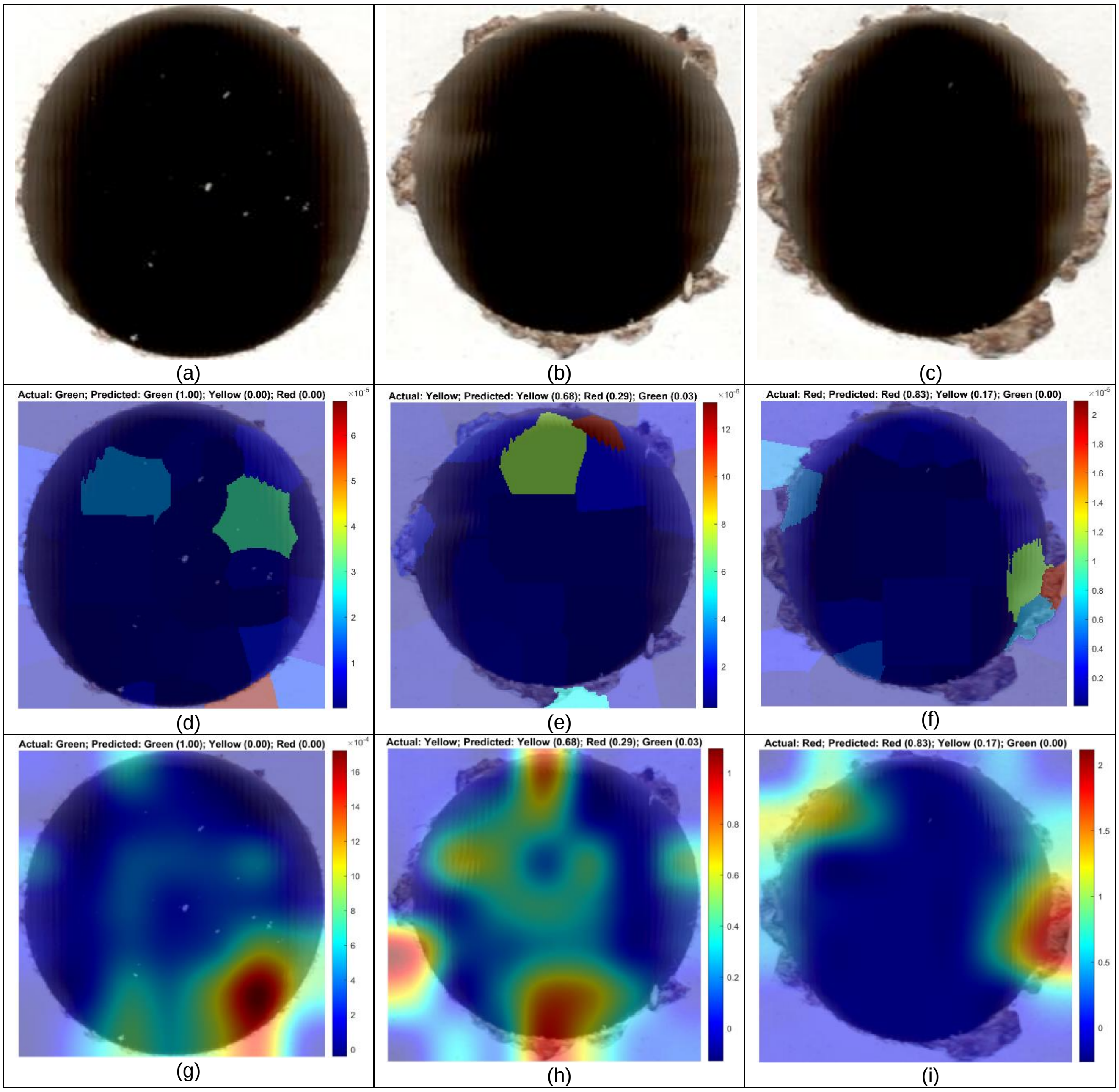

Fig 1. LIME and Grand-CAM explanations for correctly classified images: (a) Green (Predicted: Green), (b) Yellow (Predicted: Yellow), (c) Red (Predicted: Red); For LIME highlighted superpixels: (d), (e), (f) and For Grad-CAM highlighted superpixels: (g), (h), (i) respectively.

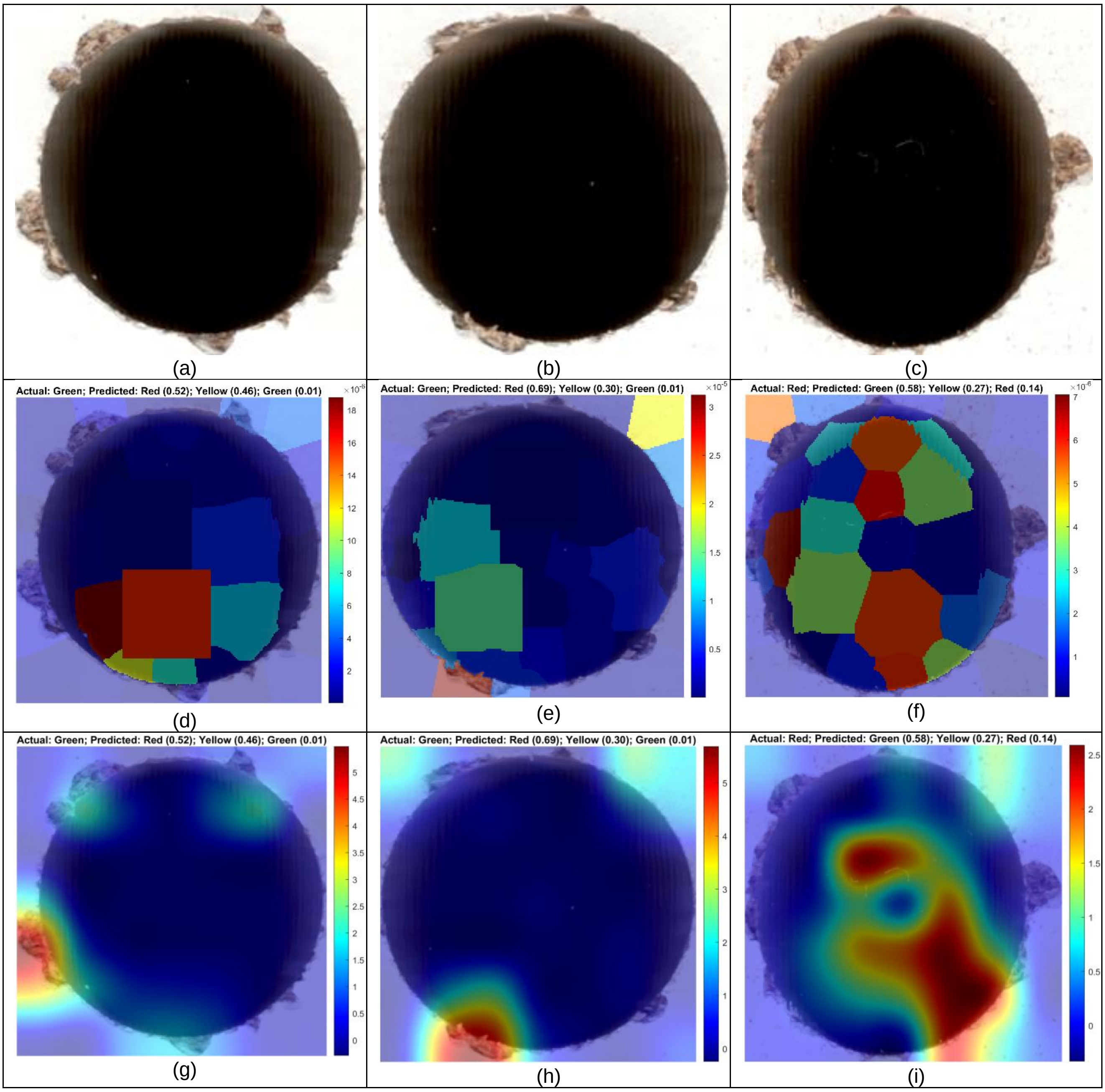

Fig 2. LIME and Grad-CAM explanations for misclassified images: (a) Green (Predicted: Red), (b) Green (Predicted: Red), (c) Red (Predicted: Green); For LIME highlighted superpixels: (d), (e), (f) and for Grad-CAM Heatmaps: (g), (h), (i) respectively

Application to Drilled Hole Classification

In our experiments, Grad-CAM provided smoother and more intuitive visualizations, effectively highlighting regions around the drilled holes that correlated with wear patterns. This helped in understanding where the model focused its attention when making predictions.

LIME offered more granular insights at the superpixel level, potentially revealing finer details. However, the reliance on segmentation meant that if the superpixels did not align well with meaningful features (e.g., edges of the holes or wear marks), the explanations could be less precise or even misleading.

Moreover, the computational cost of LIME was higher due to the need to generate numerous synthetic samples for each explanation. In a real-time industrial setting, this could be a limiting factor.

D. Recommendations

Based on the advantages and disadvantages observed:

- For CNN Models in Image Classification Tasks: Grad-CAM may be preferred due to its efficiency and alignment with convolutional architectures.
- When Model-Agnostic Explanations are Needed: LIME is suitable when explanations are required for models beyond CNNs or when comparing different model types.
- Combining Methods: Using both methods in conjunction can provide complementary insights, balancing the coarse localization of Grad-CAM with the superpixel level details of LIME.

Conclusion

This paper presented a comparative analysis of Grad-CAM and LIME for explaining the predictions of a CNN-based model classifying drilled holes in melamine-faced chipboard. By applying these XAI techniques, we enhanced the interpretability and transparency of the model, which is crucial for industrial applications where understanding the reasoning behind AI decisions is essential.

Our findings suggest that while Grad-CAM is more suitable for providing quick, class-specific visualizations in CNNs, LIME offers detailed, model-agnostic explanations at the cost of higher computational resources. Combining these methods can offer a more comprehensive understanding of the model's decisions.

Future work will focus on integrating these explanations into real-time TCM systems and exploring additional XAI techniques to further improve model interpretability and performance.

REFERENCES

[1] Hu, J., Song, W., Zhang, W., Zhao Y., Yilmaz A. (2019). Deep learning for use in lumber classification tasks. Wood Science and Technology, 53(2), 505-517. DOI: https://doi.org/10.1007/s00226-019-01086-z.

[2] Ibrahim, I., Khairuddin, A. S. M., Talip, M. S. A., Arof, H., Yusof, R. (2017). Tree species recognition system based on macroscopic image analysis. Wood Science and Technology, 51(2), 431-444.

[3] Jemielniak, K., Urbański, T., Kossakowska, J., Bombiński, S. (2012). Tool condition monitoring based on numerous signal features. International Journal of Advanced Manufacturing Technology, 59, 73-81. DOI: https://doi.org/10.1007/s00170-011-3504-2.

[4] Kuo, R. (2000). Multi-sensor integration for on-line tool wear estimation through artificial neural networks and fuzzy neural network. Engineering Applications of Artificial Intelligence, 13, 249-261. DOI:https://doi.org/10.1016/S0952-1976(00)00008-7.

[5] Kurek, J., Antoniuk, I., Górski, J., Jegorowa, A., Świderski, B., Kruk, M., Wieczorek, G., Pach, J., Orłowski, A., Aleksiejuk-Gawron, J. (2019a). Data Augmentation Techniques for Transfer Learning Improvement in Drill Wear Classification Using Convolutional Neural Network. Machine Graphics and Vision, 28, 3-12.

[6] Kurek, J., Antoniuk, I., Górski, J., Jegorowa, A., Świderski, B., Kruk, M., Wieczorek, G., Pach, J., Orłowski, A., Aleksiejuk-Gawron, J. (2019b). Classifiers ensemble of transfer learning for improved drill wear classification using convolutional neural network. Machine Graphics and Vision, 28, 13-23.

[7] Kurek, J., Kruk, M., Osowski, S., Hoser, P., Wieczorek, G., Jegorowa, A., Górski, J., Wilkowski, J., Śmietańska, K., Kossakowska, J. (2016). Developing automatic recognition system of drill wear in standard laminated chipboard drilling process. Bulletin of the Polish Academy of Sciences. Technical Sciences, 64, 633-640. DOI: https://doi.org/10.1515/bpasts-2016-0071.

[8] Kurek, J., Świderski, B., Jegorowa, A., Kruk, M., Osowski, S. (2017a). Deep learning in assessment of drill condition on the basis of images of drilled holes. In International Conference on Graphic and Image Processing. ICGIP. DOI: https://doi.org/10.1117/12.2266254.

[9] Kurek, J., Wieczorek, G., Świderski, B., Kruk, M., Jegorowa, A.,Osowski, S. (2017b). Transfer learning in recognition of drill wear using convolutional neural network. In International Conference on Computational Problems of Electrical Engineering. IEEE. DOI: https://doi.org/10.1109/CPEE.2017.8093087.

[10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations (ICLR).

[11] Tolstikhin, I., Houlsby, N., Kolesnikov, A., et al. (2021). MLP-Mixer: An all-MLP Architecture for Vision. arXiv preprint arXiv:2105.01601.

[12] Steiner, A., Kolesnikov, A., Zhai, X., et al. (2021). How to train your ViT? Data, Augmentation, and Regularization in Vision Transformers. arXiv preprint arXiv:2106.10270.

[13] Chen, X., Hsieh, C., Gong, B. (2021). When Vision Transformers Outperform ResNets without Pretraining or Strong Data Augmentations. arXiv preprint arXiv:2106.01548.

[14] Zhuang, J., Gong, B., Yuan, L., et al. (2022). Surrogate Gap Minimization Improves Sharpness-Aware Training. In International Conference on Learning Representations (ICLR).

[15] Jegorowa, A., Górski, J., Kurek, J., Kruk, M. (2020). Use of nearest neighbors (K-NN) algorithm in tool condition identification in the case of drilling in melamine faced particleboard. Maderas. Cienciay tecnología, 22, 189-196. DOI: https://doi.org/10.4067/S0718-221X2020005000205.

[16] Jegorowa, A., Kurek, J., Antoniuk, I., et al. (2021). Deep learning methods for drill wear classification based on images of holes drilled inmelamine faced chipboard. Wood Science and Technology, 55, 271-293. DOI: https://doi.org/10.1007/s00226-020-01245-7.

[17] Świderski, B., Osowski, S., Gwardys, G., et al. (2022). Random CNN structure: tool to increase generalization ability in deep learning. EURASIP Journal on Image and Video Processing, 2022(1). DOI:https://doi.org/10.1186/s13640-022-00580-y.

[18] Kurek, J., Antoniuk, I., ´ Swiderski, B., Jegorowa, A., Bukowski, M. (2020). Application of siamese networks to the recognition of the drill wear state based on images of drilled holes. Sensors, 20(23), 6978. DOI: https://doi.org/10.3390/s20236978.

[19] Selvaraju, R.R., Cogswell, M., Das, A., et al. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision, 128(2), 336-359. DOI:https://doi.org/10.1007/s11263-019-01228-7.

[20] Song, D., Yao, J., Jiang, Y., et al. (2023). A new xAI framework with feature explainability for tumors decision-making in Ultrasound data: comparing with Grad-CAM. Computer Methods and Programs in Biomedicine, 235, 107527. DOI: https://doi.org/10.1016/j.cmpb.2023.107527.

[21] Hsu, C., Morsalin, S., Reyad, M., Shakib, N. (2024). Artificial Intelligence Model Interpreting Tools: SHAP, LIME, and Anchor Implementation in CNN Model for Hand Gestures Recognition. In Technologies and Applications of Artificial Intelligence (pp. 16-29).

[22] Castillo Torres, G., Roig-Maimó, M., Mascaró-Oliver, M., et al. (2023). Understanding How CNNs Recognize Facial Expressions: A Case Study with LIME and CEM. Sensors, 23(1), 131. DOI: https://doi.org/10.3390/s23010131.