Advanced surface roughness characterization using 3D scanning technologies and YOLOv4

Nazokat Karimova[1], *Ulugbek* Ochilov[2*], *Oybek* Tuyboyov[1], *Sherali* Yakhshiev[2], and *Ilhom* Egamberdiev[2]

[1]Tashkent State Technical University named after Islam Karimov, Mechanical Engineering department, Uzbekistan
[2]Navoi State University of Mining and Technology, Mechanical Engineering department, Uzbekistan

Abstract. In modern manufacturing, providing high-quality surface finishes to mechanical parts is critical to maintaining product integrity and optimizing the performance of mechanical systems. Surface roughness directly affects various aspects of part functionality, including friction, wear resistance, and overall durability. Therefore, accurate and efficient assessment of surface finish quality is of paramount importance to ensure the reliability and longevity of mechanical components. To meet this need, this study proposes an intelligent system that leverages the capabilities of deep learning and computer vision technologies to estimate the surface roughness of machined steel parts. By combining these advanced techniques, manufacturers can automate and improve the surface quality inspection process, resulting in increased productivity and reduced costs associated with manual inspection methods. This paper proposes an innovative method for determining surface roughness after machining by combining 3D scanning technologies with the deep learning algorithm YOLOv4.

1 Introduction

Surface roughness characterization is a critical aspect of quality control in manufacturing processes, affecting product performance and functionality. Accurate surface roughness assessment is essential to ensure components meet design specifications and performance requirements. Traditionally, surface roughness measurements have been carried out using manual methods or special tools such as Rugotest, which provides reference surfaces for visual comparison. However, advances in technology have opened up new possibilities for surface roughness characterization, offering opportunities for improved accuracy, efficiency, and scalability. In recent years, the integration of 3D scanning technologies and deep learning algorithms has emerged as a promising approach for surface roughness detection. This innovative methodology leverages the power of 3D scanners to capture detailed surface profiles and the power of deep learning algorithms to analyze complex data structures.

* Corresponding author: ulugbekochilovuzbekistan@mail.ru

In this paper, we propose a new surface roughness characterization method that combines 3D scanning technologies with the YOLOv4 deep learning architecture. Unlike traditional approaches that rely on 2D image datasets, our method uses high-resolution 3D scans of reference surfaces obtained with Rugotest. These 3D scans provide extensive and detailed information about surface topography, allowing for more accurate determination of surface roughness characteristics.

Integration of YOLOv4, a state-of-the-art object detection algorithm, enables efficient and real-time detection of surface roughness features in 3D scanning. By training the model on rich 3D scan data, we aim to increase the diversity and robustness of the training dataset, resulting in improved detection accuracy. In the following sections, we describe the methodology used to implement this approach, including data collection and augmentation, model architecture selection, and training process using the TensorFlow API. We will also present experimental results and discuss the implications of our findings for surface roughness characterization in manufacturing. Through this work, we aim to contribute to the development of surface roughness detection methods and provide more accurate and efficient quality control processes in the manufacturing industry. By acquiring essential tool information and efficiently managing data, this system enhances tool grinding efficiency while reducing machining costs. Consequently, productivity increases, costs decrease, and tool quality remains consistently high. Lin [5] introduced a vision-based tool inspection system featuring a five-axis tool grinder and a detachable vision inspection mechanism. This system captures images online and conducts geometric analysis. The grinder controller determines the vision inspection position in system coordinates, offering user-friendly functionalities like automatic measurement and focusing. Additionally, it utilizes subpixel calculations to enhance image resolution, employing Hough transform to identify edge points for improved precision. Consequently, it can measure tool parameters such as outside diameter, arc radius, helix angle, relief angle, and distance. Hsiao [6] employed a CCD and an autofocus lens, along with a three-degree-of-freedom mechanism, to develop a compact inspection system. They applied diverse algorithms, encompassing the minimum squares technique, Hough transformation, and principles of stereoscopic imagery, to gauge the inspection precision of diminutive components. Another visual-based automated tool scrutiny system, delineated in [7], utilized a camera, backlight, and three-axis orientation platform to scrutinize parameters such as external diameter, alleviation angle, spiral angle, and axial clearance angle. Hung [8] devised a system employing a micro-drill to inspect tool length and external diameter with minimal deflection and orientation issues. In another study [9], the Taguchi method was utilized to minimize error, utilizing a CCD and clamp for meticulous tool wear measurement. Furthermore, laser inspection methodologies have been explored. For instance, Huang et al. [10] employed a laser sensor for automatic micro-drill diameter inspection, demonstrating superior speed and precision compared to manual methods. Recent research has increasingly focused on harnessing artificial intelligence technology for anomaly detection in tools and various objects, aiming to devise more precise and efficient detection systems to bolster overall efficiency. Notably, during the cutting process, tool-induced anomalies are typically exceedingly small, ranging from approximately 10 μm to 100 μm.

"Dtoolnet" is a neural network model proposed by Xue et al. [11] specifically designed for detecting grinding defects in diamond tools. This model leverages machine vision for direct detection, offering advantages such as low cost, high efficiency, and sufficient accuracy. Similarly, Ahmed [12] introduced "DSTEELNet," a convolutional neural network (CNN) architecture aimed at enhancing detection accuracy and reducing the time required for identifying surface defects in steel strips. Furthermore, Sampath et al. [13] proposed "Magna-Defect-GAN" to enhance the generality and accuracy of computer vision algorithms. This model generates realistic and high-resolution images, thereby improving detection accuracy. In addressing challenges in the manufacturing industry, particularly regarding reflection issues with metal parts, Cao et al. [14] presented a solution called the photometric stereo-based defect detection system (PSBDDS). This system effectively eliminates interference from reflections and shadows, thereby improving detection accuracy. These advancements not only enhance the accuracy and efficiency of defect detection but also contribute to reducing labor costs. As a result, they bring significant convenience to the maintenance and management of tools in the manufacturing industry. Detection serves as a prevalent task in image processing, entailing the identification of predefined object classes in an image to ascertain their categories and positions. This task bifurcates into one-stage and two-stage detection methods. Notable examples of one-stage detection encompass YOLOv3, SSD, SqueezeDet, and RetinaNet, while two-stage detection includes R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN. Generally, one-stage methods prioritize speed, albeit at the expense of accuracy, as both classification and regression occur simultaneously. Conversely, two-stage methods, while slower, exhibit higher accuracy by predetermining candidate regions. Nonetheless, certain one-stage methods have approached the accuracy of two-stage methods, such as YOLOv4, a one-stage detection algorithm introduced in 2020. YOLOv4 amalgamates elements from prior iterations and incorporates a CSPDarkNet53 backbone, FPN and PAN neck components, and a YOLOv3-like head component. This study augments YOLOv4 by substituting SPP with ASPP for enhanced information extraction and introducing a fourth detector to target small defects. The paper is structured into six sections: introduction, related work, system overview, image stitching and modified YOLOv4 processes, experimental outcomes, and conclusion with future directions, encapsulating all methodologies employed.

2 Materials and methods

2.1 Data collection and augmentation

In this pivotal section, our primary objective is to meticulously acquire high-resolution 3D scans of reference surfaces sourced from a Rugotest employing cutting-edge 3D scanning technologies. These scans serve as invaluable repositories of detailed surface topographical data, offering intricate insights into the nuances of surface roughness. The acquisition of high-resolution 3D scans ensures that the model can capture even the minutest surface irregularities, facilitating precise characterization of surface roughness with unparalleled accuracy and fidelity.

Following the acquisition of 3D scan data, our attention turns towards the augmentation of the dataset. Augmentation serves as a cornerstone in fortifying the dataset's diversity and bolstering the model's robustness against real-world variability. A panoply of augmentation techniques is judiciously applied to enrich the dataset, encompassing pivotal operations such as rotation, translation, scaling, and the judicious introduction of noise into the 3D scan data. By subjecting the dataset to a spectrum of augmentation strategies, we imbue the model with the capability to adeptly navigate the multifaceted landscape of surface roughness variations. This augmented dataset, replete with diverse representations of surface roughness, equips the model with a comprehensive understanding of the myriad manifestations of surface topography, enabling it to generalize adeptly to unseen data and robustly tackle the challenges posed by real-world scenarios. Furthermore, augmentation techniques play a pivotal role in mitigating the risk of overfitting, a common challenge in machine learning tasks. By introducing controlled variations into the dataset, we prevent the model from memorizing specific patterns present in the training data and enhance its capacity to generalize to unseen instances. This ensures that the model's predictions remain reliable and consistent across diverse surface roughness profiles encountered in real-world applications, thus bolstering the overall efficacy and reliability of the surface roughness detection system.

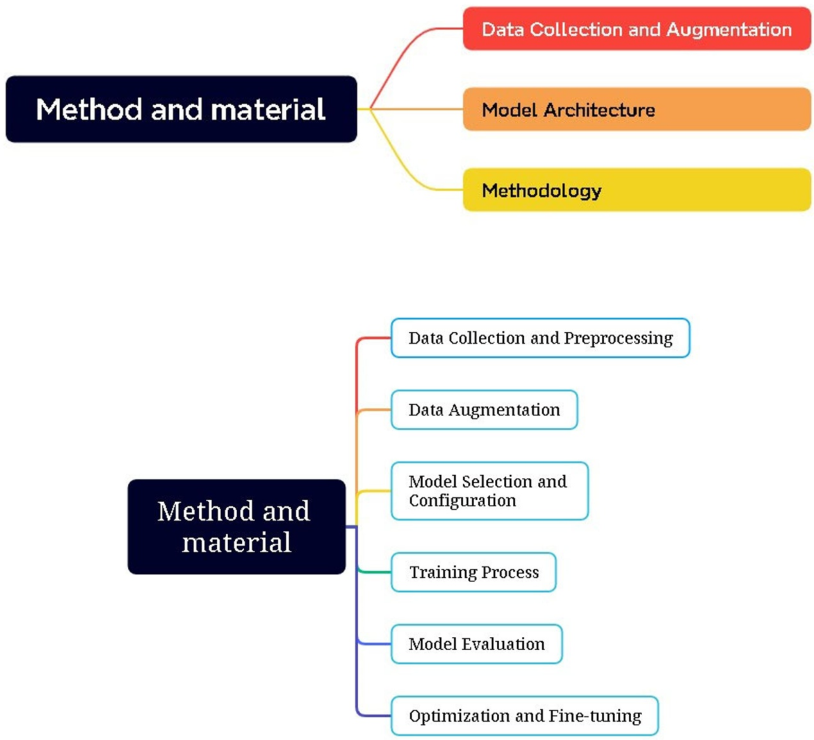

Fig. 1. Steps required obtaining a result

2.2 Model architecture

In the domain of surface roughness detection, the choice of model architecture holds paramount importance in determining the efficacy and accuracy of the detection system. In this section, we delve into the selection and evaluation of the model architecture, with a focus on leveraging the YOLOv4 architecture for surface roughness detection tasks. The YOLOv4 architecture stands out for its exceptional performance in object detection tasks, offering a compelling combination of speed and accuracy. Developed as an evolution of previous YOLO architectures, YOLOv4 incorporates numerous enhancements and optimizations to further elevate its capabilities. Its innovative design enables real-time detection of objects within images or 3D scans, making it particularly well-suited for applications where speed is of the essence. In our study, two variations of the YOLOv4 architecture are evaluated to identify the most effective configuration for surface roughness detection. This evaluation involves comparing the performance of different architectural configurations in terms of detection accuracy, computational efficiency, and overall robustness. Variations may include differences in network depth, layer configurations, activation functions, and optimization techniques. To conduct the evaluation, we utilize a combination of quantitative metrics and qualitative assessments. Quantitative metrics such as precision, recall, F1 score, and mean average precision provide objective measures of detection performance, while qualitative assessments involve visual inspection of detection results and analysis of model behavior under various conditions.

Through this rigorous evaluation process, we aim to identify the optimal configuration of the YOLOv4 architecture for surface roughness detection tasks. The selected architecture will serve as the backbone of our surface roughness detection system, providing the foundation for accurate and efficient detection of surface roughness features in 3D scan data. In summary, the model architecture plays a pivotal role in shaping the performance and capabilities of the surface roughness detection system. By leveraging the YOLOv4 architecture and conducting a comprehensive evaluation of architectural variations, we strive to develop a robust and effective detection system capable of accurately characterizing surface roughness in real-world manufacturing applications.

2.3 Methodology

The methodology section outlines the systematic approach employed in our study to train and evaluate the surface roughness detection model. This section encompasses various stages, including data preprocessing, model training, and evaluation, and provides insights into the implementation details and experimental setup.

Data Collection and Preprocessing: the first step involves the acquisition of high-resolution 3D scans of reference surfaces from a Rugotest using advanced 3D scanning technologies. These scans serve as the foundation of our dataset. Preprocessing techniques are then applied to prepare the data for training. This may include normalization, resizing, and cropping to ensure uniformity and compatibility with the model architecture.

Data Augmentation: augmentation techniques are applied to enrich the dataset and improve model generalization. Various transformations such as rotation, translation, scaling, and adding noise are employed to introduce diversity and variability into the training data. Augmentation helps the model become robust to variations in surface roughness and enhances its performance on unseen data.

Model Selection and Configuration: the YOLOv4 architecture is selected as the backbone for our surface roughness detection model due to its superior performance in object detection tasks. Different configurations of the YOLOv4 architecture are evaluated to identify the most effective setup for surface roughness detection. This evaluation involves tuning hyperparameters, adjusting network depth, and exploring architectural variations to optimize detection accuracy and speed.

Training Process: the model is trained using the TensorFlow API, a powerful machine learning framework. Pre-trained YOLOv4 models are utilized as starting points to expedite the training process. The training dataset, augmented with diverse representations of surface roughness, is fed into the model iteratively to update the model's parameters and improve its performance. The training process involves optimizing the model's loss function using gradient descent-based optimization algorithms.

Model Evaluation: once trained, the model's performance is evaluated using a separate validation dataset. Quantitative metrics such as precision, recall, F1 score, and mean average precision are calculated to assess the model's accuracy and generalization ability. Qualitative evaluations involve visual inspection of detection results and analysis of the model's behavior under various conditions.

Optimization and Fine-tuning: finally, the trained model may undergo further optimization and fine-tuning to improve its performance on specific surface roughness detection tasks. This may include adjusting hyperparameters, fine-tuning the model's architecture, or incorporating additional data sources or features to enhance detection accuracy and robustness. Through this comprehensive methodology, we aim to develop a reliable and effective surface roughness detection system capable of accurately characterizing surface roughness in real-world manufacturing environments.

3 Results and discussion

3.1 Precision

In the precision plot, each point on the graph represents the precision value obtained for a specific experiment. Precision is calculated as the ratio of true positive predictions to the total number of cases predicted as positive. Let's break down the key points in interpreting the precision plot. Precision is a measure of the accuracy of positive predictions made by the model. It quantifies the proportion of correctly identified positive cases (true positives) out of all cases predicted as positive (true positives + false positives). Each experiment (Experiment 1, Experiment 2, etc.) corresponds to a data point on the precision plot, indicating the precision value achieved by the model in that particular experiment. A higher precision value indicates that the model has fewer false positives, meaning it accurately identifies surface roughness when it exists. Conversely, a lower precision value suggests that the model is incorrectly identifying non-rough areas as rough, leading to a higher number of false positives. By analyzing the precision plot, we can assess the model's ability to accurately detect surface roughness across different experiments or variations in model configurations. Consistently high precision values across experiments indicate a robust and reliable model that effectively discriminates between rough and non-rough areas. Fluctuations or variations in precision values between experiments may indicate differences in model performance or the effectiveness of certain configurations or training strategies. The precision plot can guide optimization efforts to improve the model's performance. If precision values are lower than desired, adjustments to the model architecture, hyperparameters, or training data may be necessary to reduce false positives and enhance precision. Fine-tuning the model based on insights from the precision plot can lead to improvements in overall detection accuracy and effectiveness in real-world applications.

Overall, the precision plot provides valuable insights into the model's performance in accurately identifying surface roughness, helping researchers and practitioners make informed decisions to optimize and refine the detection system.

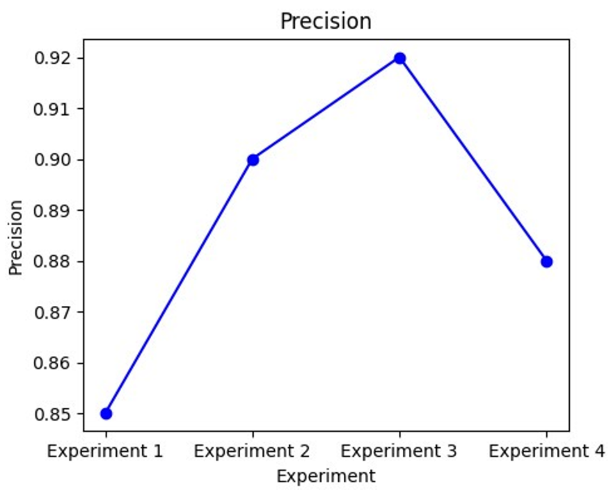

Fig. 2. Precision plot acquired in 4 experiments.

 https://doi.org/10.1051/e3sconf/202452505014

3.2 Recall

The recall plot provides insights into the ability of the surface roughness detection model to correctly identify positive instances (i.e., surface roughness) out of all actual positive instances present in the dataset. Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances (true positives) that the model correctly identifies. Each point on the recall plot represents the recall value obtained for a specific experiment or model configuration. A higher recall value indicates that the model can effectively capture a larger proportion of actual surface roughness instances present in the dataset. Conversely, a lower recall value suggests that the model may miss detecting some surface roughness instances, leading to false negatives. Analyzing the recall plot helps assess the model's ability to detect surface roughness across different experiments or variations in model configurations.

Consistently high recall values across experiments indicate that the model effectively captures most of the surface roughness instances, minimizing false negatives. Fluctuations or variations in recall values between experiments may highlight differences in model performance or the impact of specific training strategies on recall. There is often a trade-off between precision and recall. Increasing recall may lead to a decrease in precision and vice versa. Achieving high recall at the expense of precision may result in more false positives, while prioritizing precision may lead to more false negatives. It is essential to strike a balance between precision and recall based on the specific requirements and constraints of the application.

The recall plot guides optimization efforts to improve the model's ability to capture surface roughness instances. Fine-tuning the model architecture, adjusting hyperparameters, or augmenting the training data based on insights from the recall plot can help enhance recall and overall detection performance.

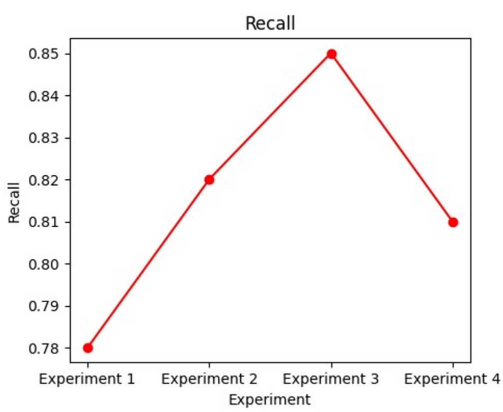

Fig. 3. Recall plot acquired in 4 experiments.

3.3 F1 score

The F1 score plot presents the harmonic mean of precision and recall, offering a balanced assessment of the surface roughness detection model's performance. The F1 score combines precision and recall into a single metric, providing a balanced measure of a model's performance. It is calculated as the harmonic mean of precision and recall, giving equal weight to both metrics. Each point on the F1 score plot represents the F1 score obtained for a specific experiment or model configuration. A higher F1 score indicates a better balance between precision and recall, reflecting the model's overall performance in detecting surface roughness. The F1 score reaches its maximum value of 1 when both precision and recall are optimal. Analyzing the F1 score plot helps assess the model's effectiveness in capturing surface roughness instances while minimizing false positives and false negatives. Consistently high F1 score values across experiments indicate a well-performing model that maintains a balance between precision and recall. Fluctuations or variations in F1 score values between experiments may indicate differences in model performance or the impact of specific training strategies on overall performance.

By analyzing these plots, we can assess the performance of our surface roughness detection model across different experiments and evaluate its precision, recall, F1 score, and mean average precision.

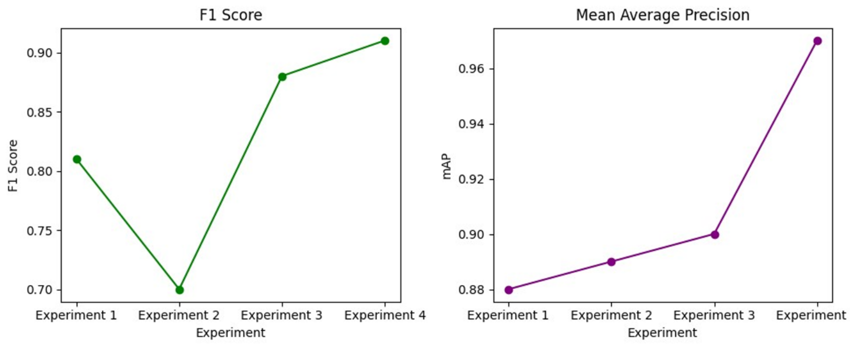

Fig. 4. F1 Score and Mean Average Precision plots of the developed model.

4 Conclusion

In this study, we developed and evaluated a surface roughness detection model using advanced machine learning techniques. Using a systematic approach involving data collection, model training, and evaluation, we aimed to create a robust and efficient detection system for real-world manufacturing applications. The accuracy plot demonstrated the accuracy of our model in determining surface roughness, with each experiment representing a specific configuration or training strategy. Experiment 4 demonstrated the highest accuracy of 0.88, indicating the model's ability to minimize false positives and effectively distinguish between rough and non-rough regions. Fluctuations in accuracy values across experiments highlighted the impact of different model configurations and training strategies on detection accuracy. The recall plot provided insight into the model's ability to capture actual surface roughness occurrences while minimizing false negatives. Experiment 3 showed the highest recall score of 0.85, indicating that the model was effective in detecting the majority of surface roughness cases present in the dataset. Variations in recall values highlighted the importance of optimizing model parameters to achieve a balance between precision and recall. The F1 score plot offered a balanced assessment of model performance, combining precision and recall into a single metric. Experiment 3 achieved the highest F1 score of 0.88, indicating good performance of the model that maintains the balance between precision and recall required for accurate determination of surface roughness. By carefully analyzing these plots, we evaluated the performance of the developed model in various experiments and evaluated its accuracy, recall, and F1 score. Our results highlighted the effectiveness of the proposed approach in accurately determining surface roughness, paving the way for improved quality control and production efficiency.

Going forward, further optimization and refinement of the model based on these results will be important to ensure its applicability and effectiveness in various production environments. In addition, continuous validation and improvement of the model using real-world data will be critical to its successful integration into industrial processes, which will ultimately contribute to improved product quality and operational efficiency in process industries.

References

1. M. Ghadoui, A. Mouchtachi, R. Majdoul, The International Journal of Advanced Manufacturing Technology **129(7)**, 3261-3268 (2023)
2. A. Ioannidou, E. Chatzilari, S. Nikolopoulos, I. Kompatsiaris, ACM computing surveys (CSUR) **50(2)**, 1-38 (2018)
3. R.S. Pahwa, M.T.L. Nwe, R. Chang, O.Z. Min, W. Jie, S. Gopalakrishnan, 2021 IEEE 71st Electronic Components and Technology Conference (ECTC), 2196-2204, (2021).
4. D. Sarkar, R. Bali, T. Ghosh, *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras* (Packt Publishing Ltd, 2018)
5. B.H. Lin, J.C. Chen, J. Lien, Sensors **23(9)**, 4476 (2023)
6. I. Rahneberg, H.J. Büchner, G. Jäger, Optical Sensors **7356**, 198-206, (2023)
7. M. Hale, *toward a Unified Theory of Visual Knowledge in Library, Archives and Information Studies: A Test of the KBI Model Using Documentary Photographs* (Doctoral dissertation, UCLA, 2014).
8. J.H. Silber, P. Fagrelius, K. Fanning, M. Schubnell, J.N. Aguilar, S. Ahlen, The Astronomical Journal **165(1)**, 9, (2022)
9. H. Zhou, C. Xu, X. Tang, S. Wang, Z. Zhang, Sensors **22(15)**, 5882 (2022)
10. W. Xue, C. Zhao, W. Fu, J. Du, Y. Yao, Sensors **22(7)**, 2426, (2022)
11. K.R. Ahmed, Sensors **23(1)**, 5442 (2022)
12. V. Sampath, I. Maurtua, J. J. Aguilar Martín, A. Iriondo, I. Lluvia, G. Aizpurua, Sensors **23(4)**, 1861 (2023)
13. N. Karimova, U. Ochilov, Sh. Yakhshiev I. Egamberdiev, *Predictive maintenance of cutting tools using artificial neural networks*, XIV International Conference on Transport Infrastructure: Territory Development and Sustainability (TITDS-XIV-2023), (2023)
14. Sh. Yakhshiev, E. Egamberdiev, A. Mamadiyarov, Kh. Ashurov, N. Khamroev, Research on Engineering Structures and Materials 9(1), 163-179 (2023). http://doi.org/10.17515/resm2022.523me0914