Optimization of Production Infrastructure processes, using Artificial Intelligence (AI) Methods (OPTIMUM)

Kristián Fodor[1], Zoltán Balogh[1,2], Martin Vozár[2], Jan Francisti[2], Štefan Koprda[2] and Marek Hrabčák[2]

[1]Kandó Kálmán Faculty of Electrical Engineering, Obuda University, Bécsi út 96/b, 1034 Budapest, Hungary, fodor.kristian@uni-obuda.hu, balogh.zoltan@uni-obuda.hu

[2]Department of Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia, zbalogh@ukf.sk, mvozar@ukf.sk, jfrancisti@ukf.sk, skoprda@ukf.sk, marek.hrabcak@ukf.sk

Abstract: This study presents the OPTIMUM framework, a machine learning-driven system, for optimizing control units in manufacturing environments. By leveraging real production data, we implemented and evaluated four key models: delay prediction, anomaly detection, efficiency optimization and prescriptive decision-making. A Random Forest classifier predicted production delays using early-stage features, while an Isolation Forest model identified operational anomalies without labeled data. A regression model accurately predicted air consumption with an R^2 of 0.9991, supporting proactive resource management. Finally, a decision tree model generated interpretable rules for minimizing delays through scheduling adjustments. The results demonstrate that machine learning techniques can not only detect inefficiencies but also recommend actionable adjustments, significantly improving responsiveness and sustainability in production systems. The proposed framework offers a modular and scalable approach to integrating AI into control units, paving the way for smarter, data-driven manufacturing operations.

Keywords: optimization; production; machine learning; artificial intelligence; anomaly detection; prescriptive modelling

1 Introduction

The integration of artificial intelligence (AI) into manufacturing processes has revolutionized the manufacturing industry, offering significant improvements in efficiency, accuracy, and operational performance. AI technologies, including machine learning, robotics, predictive analytics, and computer vision, have enabled manufacturers to optimize resources, reduce downtime, and improve product quality. It explores the multifaceted impact of AI on manufacturing efficiency, supported by insights from a variety of research papers. While AI has significantly advanced industrial engineering, challenges remain, such as data issues, ethical considerations, and the need for skilled human resources. Addressing these challenges is critical to maximizing the potential of AI in industrial applications. In addition, AI integration requires careful consideration of data governance and security to ensure successful implementation and operation [1].

Current developments in the field of industrial production and transport infrastructure are leading to a paradigmatic change in the understanding of operation and process management. This change is caused by rapid technological progress, globalization pressure on flexibility and production quality, as well as growing demands for environmental sustainability and safety. The manufacturing and logistics sectors are simultaneously facing challenges related to the effective collection of data, its transformation into meaningful knowledge and the ability to quickly make decisions based on data. In this context, there is a need for intelligent systems that integrate sensory, analytical and decision-making layers into a coherent and autonomously managed environment [2-5].

The OPTIMUM research project was created in response to this challenge, with the aim of designing and testing a system that can not only monitor and predict the condition of equipment and infrastructure, but also effectively optimize activities and manage production and transport processes using artificial intelligence models.

The starting point of the project is the observation that, despite the rapid development of available technologies, manufacturing and transport companies in the Slovak Republic are still only at the beginning of the journey towards truly data-driven operation. According to the DESI (Digital Economy and Society Index), Slovakia lags behind especially in the areas of integration of digital technologies in small and medium-sized enterprises, with only a limited share of them using big data analytics, AI, or cloud systems. An even greater challenge is the area of predictive maintenance, which is implemented by only a fraction of companies, although research clearly shows that properly deployed AI models can reduce downtime by up to 50%, extend the life of equipment by 20-30%, and reduce maintenance costs by 10-40%. Traditional approaches, such as planned or reactive maintenance, are today inefficient, costly, and prone to failure in complex environments with variable loads [6-8].

From the above, it follows that the OPTIMUM research project is conceived as a comprehensive solution that connects artificial intelligence, sensors, optimization algorithms, simulation models, elements of cyber-physical systems and the human factor. Its result should be a system capable of effectively responding to the challenges of the 21st Century: environmental, economic, technological and social.

2 Related Work

AI tools such as predictive maintenance systems, quality control algorithms, and process optimization techniques have significantly increased the efficiency, productivity, and sustainability of manufacturing environments. This response explores the effectiveness of these AI tools, supported by insights from relevant research works.

2.1 Predictive Maintenance and Equipment Optimization

Predictive maintenance powered by artificial intelligence has become a game changer in manufacturing, allowing manufacturers to anticipate equipment failures before they occur. By analyzing real-time data from sensors and historical maintenance records, AI algorithms can identify patterns and predict potential problems, reducing unplanned downtime and extending machine life [9-11]. This proactive approach not only optimizes resource utilization but also minimizes production interruptions, leading to higher overall efficiency. AI-driven predictive maintenance uses machine learning algorithms to anticipate equipment failures, thereby reducing downtime and maintenance costs. For example, AI models such as Random Forest have achieved up to 99% accuracy in predicting machine failures, making them highly effective in industrial settings [12]. Edge computing combined with AI enables real-time anomaly detection and proactive maintenance scheduling, thereby improving asset reliability and operational efficiency [13]. Predictive Maintenance (pDM) uses AI and IoT tools to monitor equipment performance and predict failures. IoT sensors collect real-time data on metrics such as vibration, temperature, and pressure, which AI algorithms analyze to predict equipment performance and identify failure patterns [14] [15] Advanced machine learning techniques, including deep learning and neural networks, are used to detect anomalies and predict equipment failures. These algorithms analyze historical and real-time data to improve maintenance planning and resource allocation [16] [17]. The pDM reports are integrated with Manufacturing Execution Systems (MES) to enable real-time decision making and process optimization, increasing overall plant efficiency and reducing unplanned downtime [15].

Advanced machine learning algorithms such as Random Forests and XGBoost are used to analyze historical and real-time data from IoT sensors, identifying trends and anomalies that may indicate impending failures [16] [18] [19]. The integration of digital twins - virtual representations of physical assets - has further improved predictive maintenance by providing real-time simulations and analysis of failures. This allows manufacturers to gain unprecedented insight into equipment performance and failure processes, ensuring timely interventions [16]. In addition, the use of edge computing and cloud-based systems provides low-latency data processing and scalable implementations, making predictive maintenance more accessible and efficient [16] [20]. Studies have demonstrated tangible benefits of

AI-driven predictive maintenance. For example, a machine learning-based approach using random forests achieved a 42% reduction in production line failures, demonstrating the potential of AI to minimize failures and improve operational reliability

2.2 Quality Control and Error Prevention

AI technologies, particularly computer vision and machine learning, have significantly improved quality control processes. Quality control is another critical area where artificial intelligence has made significant progress in manufacturing. In industries such as Printed Circuit Board (PCB) manufacturing, AI systems can detect subtle defects that are often missed by traditional inspection methods. By automating defect detection and enabling real-time monitoring, AI-driven quality control systems reduce human error, lower manufacturing costs, and improve product consistency [21] [22]. Additionally, AI-based systems can analyze historical data to predict potential failures, further reducing waste and improving yield rates. Automation, combined with deep learning, improves the accuracy and speed of defect detection. These technologies enable real-time monitoring and diagnosis of problems, preventing costly disruptions, and ensuring that products meet safety and quality standards [23]. The use of targeted quality tools helps identify and resolve manufacturing anomalies, optimize maintenance strategies, and improve overall manufacturing performance [24].

Machine learning algorithms, such as convolutional neural networks (CNN) and long short-term memory (LSTM) networks, are increasingly being used to monitor manufacturing parameters and detect defects in real time. These systems use data from IoT sensors, manufacturing systems, and digital twins to identify anomalies and ensure product quality [25] [26]. For example, a study on tool condition monitoring in milling processes used Random Forest algorithms to predict tool-related defects with 94% accuracy, eliminating the need for costly laboratory tests and increasing process efficiency [27]. Similarly, the integration of acoustic sensors with machine learning models has been used to monitor the operation of servo presses, detect early signs of wear, and optimize maintenance schedules [28]. AI-driven quality control not only reduces the likelihood of defective products, but also minimizes waste and energy consumption. By enabling real-time monitoring and adaptive control, these systems ensure that manufacturing processes remain within optimal parameters, thereby reducing the environmental footprint of production [26] [29].

2.3 Supply Chain Optimization

AI plays a key role in optimizing supply chain management, from demand forecasting to inventory management and logistics planning. By leveraging predictive analytics, AI solutions can anticipate market trends and customer demand, allowing manufacturers to adjust production schedules and inventory levels accordingly. This leads to shorter lead times, lower operating costs, and improved supply chain resilience [11] [30]. In addition, AI-driven automation streamlines supply chain processes, increases transparency, and facilitates better decision-making.

Artificial intelligence tools also play a key role in optimizing manufacturing processes and improving resource utilization. By analyzing real-time data from IoT devices and manufacturing systems, AI algorithms can identify inefficiencies and recommend adjustments to production schedules, energy consumption, and resource allocation [31] [32]. For example, a hybrid approach combining machine learning with process simulation was used to optimize milling processes, achieving an over 96% accuracy, in real-time fault detection and enabling unattended manufacturing operations [27]. Similarly, integrating overall equipment effectiveness (OEE) with predictive maintenance has enabled manufacturers to evaluate equipment performance based on availability, productivity, and quality, further refining process optimization strategies [33]. These advances have not only increased production efficiency, but also contributed to sustainability goals by reducing energy consumption and waste. For example, a study on predictive maintenance with artificial intelligence support showed a significant reduction in energy consumption, which supports greener production practices [16] [18] [40].

2.4 Process Optimization and Production Planning

AI technologies such as machine learning and neural networks are increasingly being used to optimize production planning and process parameters [41]. In Reconfigurable Manufacturing Systems (RMS), integrating AI with Petri nets and genetic algorithms has been shown to increase planning efficiency and adaptability, achieving 85% success rates in reducing lead times and improving resource utilization [34]. Similarly, AI-driven process optimization techniques allow manufacturers to dynamically adjust production parameters, ensuring higher yields and minimizing waste [21] [35]. AI facilitates real-time data analysis, improving decision-making and resource utilization. This leads to increased efficiency in manufacturing processes, as seen in AI-driven real-time quality monitoring and process optimization [26]. In mechanical manufacturing, AI applications have improved production efficiency and quality control, with predictive maintenance strategies reducing maintenance costs by up to 30% [36].

2.5 Workforce Augmentation and Human-Machine Collaboration

AI not only enhances machine capabilities, but also empowers the workforce by providing data-driven insights and automating routine tasks. AI-driven collaborative robots (cobots) and workforce augmentation tools enable seamless human-machine collaboration, highlighting the transformative synergy between humans and technology [9] [37]. This collaboration leads to improved productivity, reduced errors, and improved decision-making capabilities. The integration of AI into robotics enhances manufacturing capabilities by transforming manual processes into semi-automated processes, thereby increasing efficiency and reducing costs [38]. AI techniques such as machine learning and deep learning are being used in intelligent automation and advanced simulation, revolutionizing the manufacturing environment [37].

Digital twins have become a powerful tool for monitoring and optimizing manufacturing processes. By creating virtual representations of physical assets, manufacturers can simulate different scenarios, analyze equipment performance, and predict failures without disrupting real operations [25] [39]. A new approach using digital twins involves training CNN-based object detection models on synthetic data generated by digital twins. This method reduces the need for large training datasets and allows for real-time monitoring of manufacturing systems, even when the actual system is not available [25]. This approach has been successfully applied in industrial pilot plants, demonstrating its potential to improve process monitoring and optimization. In addition, digital twins have been used in conjunction with machine learning algorithms to monitor tool health and detect errors in milling processes. By simulating the milling process and generating analytical data, these systems achieved high accuracy in fault detection, reducing the need for costly laboratory tests [25] [27].

3 Materials and Methods

Before presenting the individual methodological components, we outline the overall analytical workflow of the OPTIMUM framework. The following subsections describe the datasets used, the preprocessing pipeline, and the modelling strategies that enable delay prediction, anomaly detection, efficiency forecasting, and prescriptive decision-making.

3.1 Data Sources

As a starting point, we have received raw production data from our partners. They produce several types of products: two types of canisters, each filled with pellets in three different colors and four different quantities. The data can be divided into 2 groups in general:

- **1st Group -** Data about the orders themselves - multiple comma-separated values (CSV) files can be linked based on ArticleID (unique identifier of a specific product).
- **2nd Group -** Information about the movement of the robotic arms of the machines

The key datasets are:

- **ManufacturingOrder.csv:** Contains records of production orders, including order types, statuses, and detailed timestamps (creation, delivery, launch, completion).
- **Order & OrderLine.csv:** Stored anonymized client order data and corresponding line items, including stock keeping units (SKUs) and planned quantities.
- **SKU.csv:** Defines individual stock keeping units with identifiers and specifications.
- **Movement.csv:** Tracks the movement of materials or products between various stages and equipment.
- **MovementAction.csv:** Logs the type and context of each movement event.
- **PAEAAirConsumption & PAEAEnergyConsumption.csv:** Capture sensor-level air and energy consumption for individual components during operation.
- **Reserve.csv:** Represents material reservations linked to production steps.
- **SPCMeasurement.csv:** Includes quality control metrics measured during or after production.

3.2 Data Integration and Preprocessing

The raw CSV files were ingested using Python and the pandas library. Initial preprocessing steps were carried out, which included merging of the data on relevant foreign keys such as ManufacturingOrderID to build a unified dataset that aligns resource usage with specific production orders and actions. Later timestamp parsing was carried out, where all datetime fields (e.g., CreationDate, LaunchedDateTime) were converted into standardized Python datetime objects.

After merging all relevant tables, the integrated dataset contained 1,965,494 records. However, only records with complete information for the selected features and the target label (FaultTarget) were retained for modelling. After this filtering, the final dataset used in the models consisted of 493,440 samples, of which 394,752 (80%) were used for training and 98,688 (20%) for testing.

Feature Engineering

Feature engineering focused on extracting actionable signals from the production data. First, *OrderProcessingTime* was computed as the time difference between the DeliveryDate and RequestDate in seconds, while *ManufacturingDuration* captured the elapsed time between LaunchFinishedDateTime and LaunchedDateTime. Temporal features were also derived, including *HourOfDay* and *DayOfWeek* extracted from the CreationDate to capture cyclical production patterns. In addition, a binary *DefectPrediction* label was generated by comparing each SPCMeasurement value to the median of all recorded SPC quality-control measurements, with values above this threshold flagged as potential defects.

3.3 Machine Learning Modeling

The goal of machine learning modeling is to optimize control units by detecting inefficiencies, forecasting delays, and recommending adjustments. The target outcomes included:

- **Delay Prediction**: Classification model to predict if a manufacturing order will finish late based on early-stage signals.
- **Anomaly Detection**: Unsupervised learning on consumption and quality metrics to flag deviations.
- **Efficiency Optimization**: Regression models to estimate energy or air usage based on order parameters.
- **Prescriptive Modeling**: Decision tree models or reinforcement learning strategies to suggest optimal movement sequences or launch timings.

Before training the Random Forest classifier, the class distribution of the target variable (*FaultTarget*) was examined to assess data balance. The dataset exhibited a substantial imbalance, with on-time orders (*FaultTarget* = 0) representing 86.39% of the samples and delayed orders (*FaultTarget* = 1) representing only 13.61%. Because the purpose of this study was to evaluate the predictive potential of naturally occurring production data, no resampling techniques (such as oversampling, undersampling, or class-weight adjustments) were applied. The dataset was then divided into training and testing subsets using an 80/20 split via random sampled scikit-learn's *train_test_split* and a fixed seed (*random_state* = 42) to ensure reproducibility. This approach preserved the natural class proportions in both subsets and enabled performance evaluation under realistic operational conditions.

3.4 Tools and Platform

All analyses and modelling were performed using Python 3 within the Google Colaboratory environment, which provided a flexible cloud-based platform for rapid experimentation and reproducible workflows. The data processing pipeline relied primarily on the *pandas* and *numpy* libraries for data ingestion, cleaning, merging, and feature engineering. Exploratory analysis and visualization were carried out using matplotlib and seaborn, enabling detailed inspection of temporal patterns, distribution shapes, and model behavior. Machine learning models, including the Random Forest classifier, Isolation Forest algorithm, and Random Forest regressor, were implemented using *scikit-learn*, which offered a consistent interface for training, evaluation, and hyperparameter configuration. This combination of tools formed a robust and efficient environment for processing heterogeneous industrial data and developing the predictive and prescriptive components of the OPTIMUM framework.

4 Results and Discussion

Building on the methodological framework described in Section 3, this section reports the performance of the applied models and provides a discussion of their relevance for real-world manufacturing environments.

4.1 Delay Prediction Using Supervised Learning

To identify potential delays in the production process, we developed a supervised machine learning model aimed at predicting whether a manufacturing order would exceed its expected production duration. The target variable, denoted as FaultTarget, was defined as a binary label indicating whether the actual manufacturing time (i.e., the difference between LaunchFinishedDateTime and LaunchedDateTime) surpassed the median manufacturing duration across all orders.

The feature set incorporated temporal and operational variables, including HourOfDay, DayOfWeek, Urgent status, OrderTypeID, and selected product quality metrics (Measurement, DefectPrediction). These features were engineered from raw timestamps, order metadata, and statistical process control (SPC) quality measurements.

We employed a **Random Forest Classifier** with 100 estimators and otherwise default scikit-learn parameters, including bootstrap sampling, the Gini impurity criterion, and unrestricted tree depth to perform the classification task. The model was trained with a fixed random seed (*random_state* = 42) to ensure reproducibility on 80% of the data and evaluated on the remaining 20%. The resulting accuracy was exceptionally high at **68.57%**, suggesting a moderate predictive signal in the selected features. The model demonstrated that early-stage signals from order metadata and quality metrics can serve as effective predictors of production delays. This provides an opportunity for integrating predictive alerts into production control systems, allowing for proactive interventions before delays materialize.

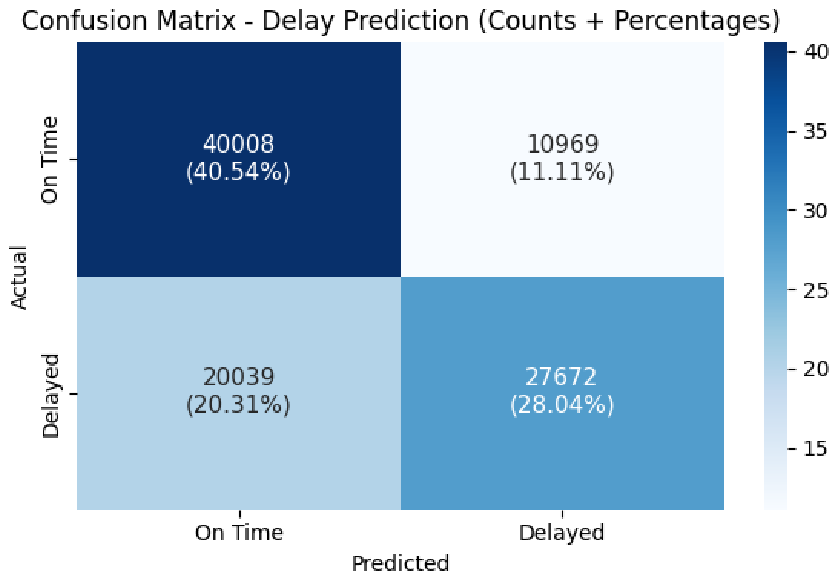

Figure 1
The confusion matrix regarding the delay prediction

The confusion matrix (Figure 1) illustrates the performance of the Random Forest classifier in predicting whether a manufacturing order will experience a delay. The matrix compares the model's predicted labels with the actual outcomes across a test set.

- The top-left cell shows **40,008** orders correctly predicted as "On Time" (true negatives).
- The bottom-right cell shows **27,672** orders correctly predicted as "Delayed" (true positives).
- The top-right cell indicates **10,969** orders that were incorrectly predicted as delayed (false positives).

- The bottom-left cell indicates **20,039** orders that were predicted as on-time but experienced a delay (false negatives).

Although the overall accuracy is high, the false negative count is non-negligible, suggesting the model occasionally fails to flag orders that will be delayed. This type of error may require additional optimization if the priority is to minimize undetected delays, especially in critical production scenarios. Adding OrderProcessingTime to the list of features when correctly classifies the target attribute with 99.97% accuracy, however, this feature was omitted as the real processing time of products are only known after the production is complete and this feature cannot be used on new, unknown data.

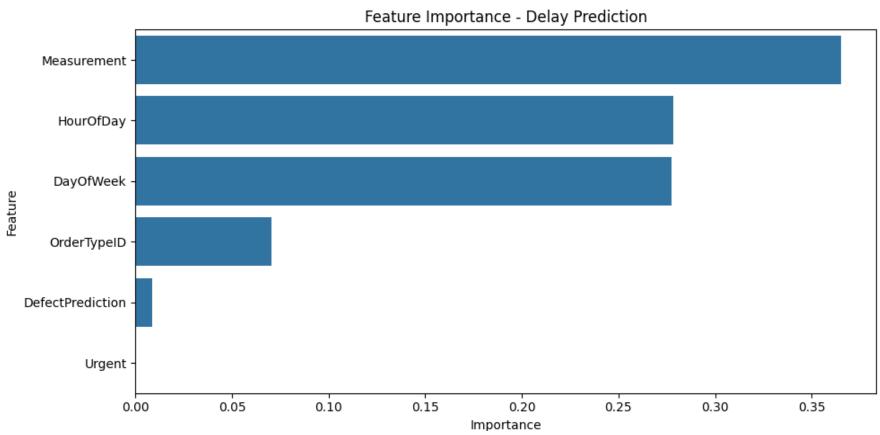

Figure 2
The feature importance regarding the delay prediction

The bar chart (Figure 2) displays the relative importance of input features used in the Random Forest model for predicting delays. The importance is based on how much each feature contributed to reducing impurity across the decision trees in the ensemble.

- **Measurement** (from Statistical Process Control data) emerged as the most influential feature, suggesting that deviations in product quality metrics are strong predictors of delays.
- **HourOfDay** and **DayOfWeek** are the next most significant features, indicating a potential temporal pattern in production performance, such as shifts or workday variations.
- **OrderTypeID** contributed moderately, reflecting that certain order categories may be more prone to delays.
- **DefectPrediction** and **Urgent** status had minimal to no impact, suggesting redundancy with other features like Measurement. An examination of the "Urgent" parameter revealed that the dataset contained no urgent orders at all.

Specifically, 100% of the records were labelled as non-urgent, and 0% were labelled as urgent. Because the feature exhibited no variation, it could not contribute any predictive signal to the Random Forest model, which explains why its feature importance was negligible. This also means that, in the current dataset, urgency cannot be evaluated as a factor influencing delay outcomes.

These insights can inform control unit design by highlighting the most predictive signals that should be monitored in real-time to anticipate delays.

4.2 Anomaly Detection

For anomaly detection, we employed the Isolation Forest algorithm from scikit-learn. The model was configured with 100 trees (*n_estimators = 100*), *max_samples = "auto", max_features = 1.0*, and a fixed random seed (*random_state = 42*) to ensure reproducibility. The contamination parameter, which controls the expected fraction of anomalies, was set to *0.05*, meaning that approximately 5% of observations are treated as outliers by the model. This value was chosen as a conservative estimate reflecting the expectation that abnormal operating conditions represent only a small minority of all production states. In an unsupervised setting where the true anomaly rate is unknown, a fixed, low contamination level helps to avoid overestimating anomalies while still highlighting statistically extreme behaviors in the joint space of *Measurement*, *OrderProcessingTime*, and *ManufacturingDuration*.

The scatter plot (Figure 3) visualizes the outcome of an unsupervised anomaly detection algorithm (Isolation Forest) applied to the manufacturing dataset. The points represent individual production orders, plotted using two key variables:

- **X-axis**: OrderProcessingTime - the total time (in seconds) to process the order.
- **Y-axis**: Measurement - a quality metric derived from Statistical Process Control (SPC), which reflects product conformity.

Each point is colored based on the anomaly classification:

- **Red points** (-1) indicate anomalies - production cases flagged as statistically deviant from the general pattern.
- **Blue points** (1) represent normal instances.

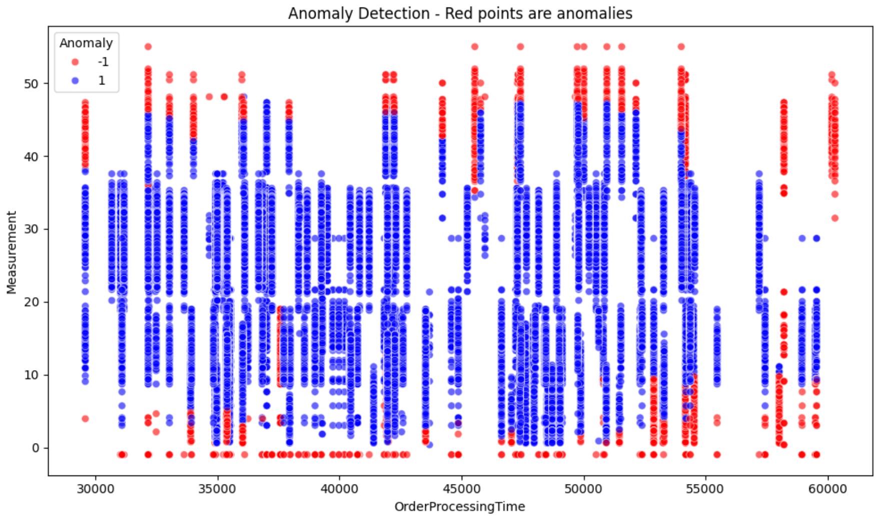

Figure 3
Anomaly Detection

We can observe a significant number of anomalies cluster at the extremes of Measurement values – both near 0 (possible underperformance or missing measurements) and near 50+ (likely exceeding acceptable quality thresholds). Anomalies also appear at specific ranges of very low or high OrderProcessingTime, suggesting a relationship between production delays and deviations in product quality. The blue (normal) points are densely concentrated around a central band of Measurement values (~10 to ~40), showing where the model identified the most consistent behavior.

The Isolation Forest model effectively isolated rare or extreme combinations of processing time and measurement that deviate from the bulk of manufacturing behavior. This approach does not require labeled data and enables early identification of outliers that could indicate process faults, measurement issues, or operational inefficiencies.

4.3 Efficiency Optimization

To evaluate the possibility of forecasting air consumption based on production metadata, we developed a regression model using a Random Forest Regressor. The model was trained to predict AirConsumption using operational features such as OrderProcessingTime, HourOfDay, DayOfWeek, and quality measurements (Measurement).

After preparing and aligning the datasets via timestamp normalization, the regression model was trained on 80% of the data and evaluated on the remaining

20%. The model achieved a Mean Squared Error (MSE) of **4,269,709** and an R^2 score of **0.9991**, indicating near-perfect predictive capability. These metrics suggest the model captured nearly all the variability in air consumption from the available features.

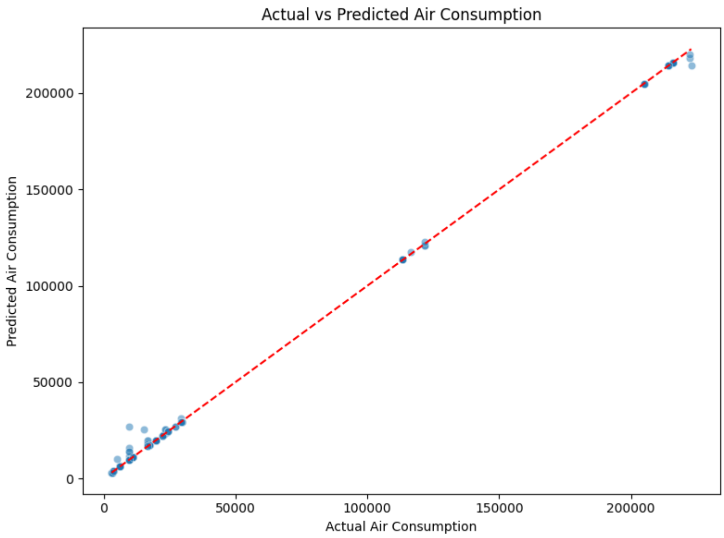

Figure 4
Actual and Predicted Air Consumption

As shown in Figure 4, the scatter plot comparing actual and predicted air consumption demonstrates a strong linear alignment along the red diagonal reference line. This confirms that the model performs consistently across the full range of observed values, including low, medium, and high-consumption cases.

The high accuracy of this model has significant implications for control unit optimization. By accurately estimating air consumption in advance, the system can:

- Schedule resource allocation more effectively
- Detect unusual consumption early (potentially flagging leaks or inefficiencies)
- And benchmark performance across shifts or products

These insights can support data-driven control logic adjustments to improve both energy efficiency and environmental sustainability in manufacturing environments.

4.4 Prescriptive Modeling

To explore actionable recommendations that could minimize delays in the production line, we developed a prescriptive model using a Decision Tree

Classifier. The model was trained to classify whether a production order would be delayed (FaultTarget) based on controllable scheduling features: HourOfDay, DayOfWeek and OrderProcessingTime. These features represent decision variables that can be adjusted by the control unit at runtime.

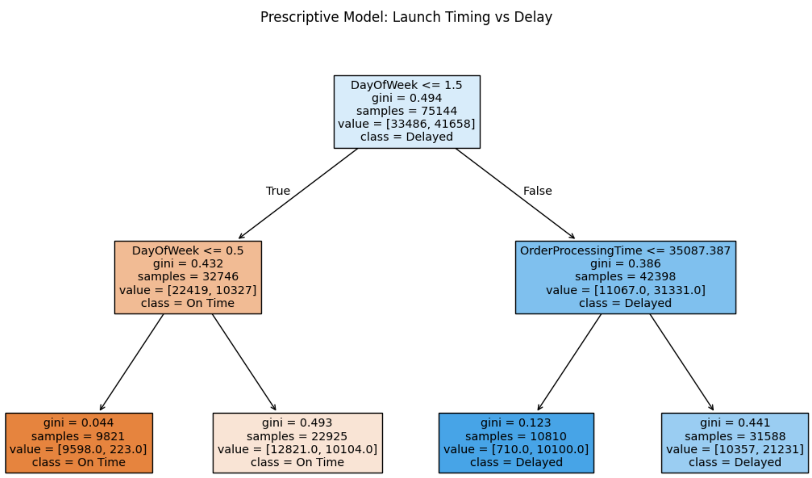

Figure 5
Decision Tree prescriptive model

The resulting decision tree, shown in Figure 5, provides a hierarchical set of rules linking production timing and scheduling decisions to the likelihood of delays. While interpretable in theory, the generated tree is relatively deep (real tree depth was 14, however, for ease of visualization Fig. 5 was limited to depth 2), which reflects the model's attempt to capture fine-grained interactions between scheduling variables and fault patterns. The decision tree created 62 leaves with 123 total nodes.

Despite its complexity, the structure reveals key insights:

- Certain time windows during the day (as defined by *HourOfDay*) are consistently associated with a lower probability of delay, potentially aligning with less congested operational shifts or more efficient workforce deployment.
- Similarly, day-of-week patterns suggest temporal bottlenecks, such as Mondays or Fridays, may require adjusted resource planning.
- The *Urgent* flag, while a logical signal of priority, did not dominate the top splits, implying that urgency alone does not drive delay likelihood unless it interacts with timing factors. Later analysis confirmed the dataset did not contain any orders flagged as urgent at all.

To improve readability and practical implementation, further pruning or surrogate rule extraction (e.g., through surrogate splits or rule-based learners) could be performed. Nonetheless, this tree structure provides an interpretable decision logic framework for control units to dynamically adjust launch timings in response to predicted delay risks.

The difference between the 98,688 samples shown in the confusion matrix (Figure 1) and the 75,144 samples displayed in Figure 5 arises from the fact that the prescriptive model operates on a more restricted subset of the data. The delay-prediction model uses all records with complete values for the features *HourOfDay*, *DayOfWeek*, *OrderTypeID*, *Measurement*, and *DefectPrediction*, resulting in 493,440 usable samples, of which 98,688 form the 20% test set.

In contrast, the prescriptive decision-tree model relies on a different feature set (*HourOfDay*, *DayOfWeek* and *OrderProcessingTime*) and requires complete values for these parameters as well as the target label. Because *OrderProcessingTime* is available for far fewer records, the resulting modelling subset is substantially smaller. After filtering, the dataset contains 75,144 valid samples, which explains why Figure 5 reports this lower value.

Thus, both values are correct; they refer to different modelling pipelines based on different data-availability constraints.

Conclusions

This study demonstrates the practical potential of integrating machine learning techniques into manufacturing environments, to enhance control unit performance and operational efficiency. Through the OPTIMUM framework, we evaluated several AI-driven strategies that collectively support the intelligent automation and optimization of production processes.

The results from the delay prediction model indicated that machine learning classifiers, particularly Random Forest, can effectively identify early indicators of delayed production. When using temporal and quality-related features such as HourOfDay, Measurement, and OrderTypeID, the model achieved a solid classification performance. Although the initial accuracy was moderate (68.57%), further feature refinement (e.g., adding OrderProcessingTime) increased the predictive accuracy significantly. However, we also demonstrated that such retrospective features must be handled with caution, when applied to future or real-time prediction scenarios.

The anomaly detection model, based on Isolation Forest, successfully flagged rare and extreme production scenarios that may otherwise go unnoticed. These anomalies correlated strongly with both high and low extremes in processing time and quality measurement, providing a valuable diagnostic layer for identifying inefficiencies, defects, or data quality issues. Importantly, this unsupervised approach required no labeled data, making it immediately applicable in industrial environments where ground truth is scarce.

The efficiency optimization model, built using Random Forest regression, showed excellent predictive capacity for estimating air consumption from production metadata, achieving an R^2 score of 0.9991. This outcome emphasizes that AI can support energy efficiency by accurately forecasting resource needs in advance, thus enabling proactive planning, anomaly detection, and sustainability-oriented process tuning.

The prescriptive decision tree model provided interpretable rules linking production scheduling decisions (e.g., *HourOfDay*, *DayOfWeek*) with the probability of delays. While the raw tree was complex, pruning and rule extraction revealed actionable insights that can guide control units in dynamically optimizing production timing. This enhances decision-making by not only predicting potential faults but also suggesting how to avoid them.

While the presented models demonstrate promising performance on the available production dataset, their generalizability to other manufacturing environments requires careful consideration. The models were trained on data originating from a specific type of production line with a well-defined process structure and sensor configuration; therefore, their performance is expected to depend on the similarity of operational conditions in other plants. Features such as *OrderProcessingTime* and *ManufacturingDuration* are highly context-dependent, and their predictive value may vary across industries where workflow stages, cycle times, and machine utilization patterns differ. For broader applicability, the models would require re-training or fine-tuning using local operational data to capture plant-specific characteristics.

From a deployment perspective, practical integration into existing control systems requires addressing data availability and quality constraints. In an operational deployment, ensuring robust timestamp logging and consistent data capture would be essential for reliable model inference. Moreover, the high-class imbalance observed in the delay prediction task (86.39% on-time vs. 13.61% delayed) implies that real-time systems must incorporate mechanisms to handle rare events, such as threshold tuning, cost-sensitive learning, or periodic model recalibration.

Another practical consideration is model drift. Industrial processes evolve due to maintenance activities, tooling changes, seasonal demand shifts, or workforce variations. Therefore, any deployed model should be accompanied by continuous monitoring, periodic revalidation, and automated retraining pipelines. Integrating the models into a digital-twin framework could further support real-time adaptation, allowing models to learn from ongoing process changes and simulate alternative scheduling or resource-allocation strategies before applying them.

Despite these challenges, the modular structure of the OPTIMUM framework supports practical deployment. Each model (delay prediction, anomaly detection, efficiency forecasting, and prescriptive decision-making) can be implemented independently or combined into a unified decision-support layer. This modularity facilitates incremental adoption in industrial settings, starting with anomaly detection or forecasting and later expanding toward more advanced prescriptive control. With appropriate data governance, monitoring, and iterative retraining, the proposed models have the potential to scale effectively across different production environments and support data-driven operational decision making.

Together, these results affirm that machine learning can meaningfully contribute to the optimization of control logic in production environments. By combining delay prediction, anomaly detection, energy consumption forecasting, and prescriptive modeling into a unified pipeline, the OPTIMUM project presents a scalable and modular AI framework applicable across manufacturing domains.

Future work will focus on deploying these models in real-time systems and extending their capability with reinforcement learning and digital twin integration.

Data Availability

The data will be available after the acceptance of the paper in a public repository with DOI: 10.17632/nz38w9vy4t.1. The data in the draft version can be accessed at: https://data.mendeley.com/preview/nz38w9vy4t

Statements and Declaration

The authors declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

This work was supported by research Research Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic under Grant 09I05-03-V02-00022: Optimization of manufacturing and transportation infrastructure processes through artificial intelligence methods (OPTIMUM).

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and the Slovak Academy of Sciences (SAS) under the contract No. VEGA 1/0385/23.

References

[1] X. Li, Y. Cheng, C. Møller, and J. Lee, "Data Issues in Industrial AI System: A Meta-Review and Research Strategy," May 2024, doi: 10.48550/arXiv.2406.15784

[2] P. S. Bilga, S. Singh, and R. Kumar, "Optimization of energy consumption response parameters for turning operation using Taguchi method," *J Clean Prod*, Vol. 137, pp. 1406-1417, 2016, doi: https://doi.org/10.1016/j.jclepro.2016.07.220

[3] C. Camposeco-Negrete, J. de Dios Calderón Nájera, and J. C. Miranda-Valenzuela, "Optimization of cutting parameters to minimize energy consumption during turning of AISI 1018 steel at constant material removal rate using robust design," *The International Journal of Advanced*

Manufacturing Technology, Vol. 83, No. 5, pp. 1341-1347, 2016, doi: 10.1007/s00170-015-7679-9

[4] R. K. Bhushan, "Optimization of cutting parameters for minimizing power consumption and maximizing tool life during machining of Al alloy SiC particle composites," *J Clean Prod*, Vol. 39, pp. 242-254, 2013, doi: https://doi.org/10.1016/j.jclepro.2012.08.008

[5] P. Albertelli, A. Keshari, and A. Matta, "Energy oriented multi cutting parameter optimization in face milling," *J Clean Prod*, Vol. 137, pp. 1602-1618, 2016, doi: https://doi.org/10.1016/j.jclepro.2016.04.012

[6] G. Campatelli, L. Lorenzini, and A. Scippa, "Optimization of process parameters using a Response Surface Method for minimizing power consumption in the milling of carbon steel," *J Clean Prod*, Vol. 66, pp. 309-316, 2014, doi: https://doi.org/10.1016/j.jclepro.2013.10.025

[7] C. Camposeco-Negrete, "Optimization of cutting parameters using Response Surface Method for minimizing energy consumption and maximizing cutting quality in turning of AISI 6061 T6 aluminum," *J Clean Prod*, Vol. 91, pp. 109-117, 2015, doi: https://doi.org/10.1016/j.jclepro.2014.12.017

[8] C. Li, L. Li, Y. Tang, Y. Zhu, and L. Li, "A comprehensive approach to parameters optimization of energy-aware CNC milling," *J Intell Manuf*, Vol. 30, No. 1, pp. 123-138, 2019, doi: 10.1007/s10845-016-1233-y

[9] X. Lin, "Artificial Intelligence in the Industrial Engineering," *Advances in Operation Research and Production Management*, Vol. 1, pp. 1-6, May 2024, doi: 10.54254/3006-1210/direct/0106

[10] A. Badrinarayanan, "AI in Manufacturing: Driving Operational Excellence," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Vol. 10, pp. 585-594, May 2024, doi: 10.32628/CSEIT24106199

[11] S. K. Lodhi, A. Y. Gill, and I. Hussain, "AI-Powered Innovations in Contemporary Manufacturing Procedures: An Extensive Analysis," *International Journal of Multidisciplinary Sciences and Arts*, Vol. 3, No. 4, pp. 15-25, Sep. 2024, doi: 10.47709/ijmdsa.v3i4.4616

[12] L. Ezzahra, Z. Hidila, A. Fentis, F. Monteiro, R. Abdoul, and A. Abdennasser, "Optimizing Industrial System From Machine Learning to Digital Twin-Driven Predictive Maintenance," May 2024, pp. 1-6, doi: 10.1109/ICM63406.2024.10815781

[13] D. Thakkar and R. Kumar, "AI-Driven Predictive Maintenance for Industrial Assets using Edge Computing and Machine Learning," May 2024, doi: 10.55544/jrasb.3.1.55

[14] P. Juneja and K. Gupta, "Application of Predictive Maintenance in Manufacturing with the Utilization of AI and IoT Tools," *INTERNATIONAL*

JOURNAL OF ADVANCED RESEARCH IN ENGINEERING & TECHNOLOGY, Vol. 16, pp. 301-309, Mar. 2025, doi: 10.34218/IJARET_16_02_018

[15] J. Iskandar, J. Moyne, K. Subrahmanyam, P. Hawkins, and M. Armacost, "Predictive Maintenance in semiconductor manufacturing," in *2015 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 2015, pp. 384-389, doi: 10.1109/ASMC.2015.7164425

[16] D. Patil, "Artificial intelligence-driven predictive maintenance in manufacturing: Enhancing operational efficiency, minimizing downtime, and optimizing resource utilization," May 2024

[17] A. Sheikh *et al.*, "Using IoT and Machine Learning Together for Manufacturing Predictive Maintenance," 2024, pp. 36-39, doi: 10.1201/9781003596721-9

[18] S. S. Sharma, V. V, and A. Malviya, "AI-Enhanced Predictive Maintenance in Intelligent Systems for Industries," in *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 2024, pp. 1-6, doi: 10.1109/ACROSET62108.2024.10743977

[19] B. Taşcı, A. Omar, and S. Ayvaz, "Remaining useful lifetime prediction for predictive maintenance in manufacturing," *Comput Ind Eng*, Vol. 184, p. 109566, 2023, doi: https://doi.org/10.1016/j.cie.2023.109566

[20] K. Sathupadi, S. Achar, S. V. Bhaskaran, N. Faruqui, M. Abdullah-Al-Wadud, and J. Uddin, "Edge-Cloud Synergy for AI-Enhanced Sensor Network Data: A Real-Time Predictive Maintenance Framework," *Sensors*, Vol. 24, No. 24, 2024, doi: 10.3390/s24247918

[21] H. Ghelani, "Advanced AI Technologies for Defect Prevention and Yield Optimization in PCB Manufacturing," *International Journal Of Engineering And Computer Science*, Vol. 10, pp. 26534-26550, Oct. 2024, doi: 10.18535/ijecs/v13i10.4924

[22] H. Ghelani, "AI-Driven Quality Control in PCB Manufacturing: Enhancing Production Efficiency and Precision," *International Journal of Scientific Research and Management (IJSRM)*, Vol. 12, pp. 1549-1564, Oct. 2024, doi: 10.18535/ijsrm/v12i10.ec06

[23] J. Chukwunweike, A. N. Anang, A. Adeniran, and J. Dike, "Enhancing manufacturing efficiency and quality through automation and deep learning: addressing redundancy, defects, vibration analysis, and material strength optimization," *World Journal of Advanced Research and Reviews*, Vol. 23, pp. 1272-1295, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2800

[24] E. Fadwa, "Advancing Multi-Level Production Defect Detection Using Quality Tools and Maintenance Optimisation," *Journal of Research, Innovation and Technologies*, Vol. 3, No. 2, pp. 129-139, 2024, doi: 10.57017/jorit.v3.2(6).04

[25] M. Urgo, W. Terkaj, and G. Simonetti, "Monitoring manufacturing systems using AI: A method based on a digital factory twin to train CNNs on synthetic data," *CIRP J Manuf Sci Technol*, Vol. 50, pp. 249-268, 2024, doi: https://doi.org/10.1016/j.cirpj.2024.03.005

[26] O. Okuyelu and O. Adaji, "AI-Driven Real-time Quality Monitoring and Process Optimization for Enhanced Manufacturing Performance," *Journal of Advances in Mathematics and Computer Science*, Vol. 39, pp. 81-89, Mar. 2024, doi: 10.9734/jamcs/2024/v39i41883

[27] A. Ebrahimi Araghizad, F. Tehranizadeh, K. Kilic, and E. Budak, "Smart Tool-Related Faults Monitoring System Using Process Simulation-Based Machine Learning Algorithms," *Journal of Machine Engineering*, Vol. 23, Oct. 2023, doi: 10.36897/jme/174018

[28] A. Aradi, A. K. Varga, and P. Takács, "Integrating Servopress Acoustic Monitoring with Machine Learning for Enhanced Predictive Maintenance: Opportunities and Limitations," in *2024 25th International Carpathian Control Conference (ICCC)*, 2024, pp. 1-6, doi: 10.1109/ICCC62069.2024.10569593

[29] H. Ö. Ünver, A. M. Özbayoğlu, C. Söyleyici, and B. B. Çelik, "Chapter 9 - Artificial intelligence for machining process monitoring," in *Artificial Intelligence in Manufacturing*, M. Soroush and R. D Braatz, Eds., Academic Press, 2024, pp. 307-350, doi: https://doi.org/10.1016/B978-0-323-99134-6.00010-4

[30] Olubunmi Adeolu Adenekan, Nko Okina Solomon, Peter Simpa, and Scholar Chinenye Obasi, "Enhancing manufacturing productivity: A review of AI-Driven supply chain management optimization and ERP systems integration," *International Journal of Management & Entrepreneurship Research*, Vol. 6, No. 5, pp. 1607-1624, May 2024, doi: 10.51594/ijmer.v6i5.1126

[31] N. Su, S. Huang, and C. Su, "Elevating Smart Manufacturing with a Unified Predictive Maintenance Platform: The Synergy between Data Warehousing, Apache Spark, and Machine Learning," *Sensors*, Vol. 24, p. 4237, Jun. 2024, doi: 10.3390/s24134237

[32] S. Syed, "Advanced Manufacturing Analytics: Optimizing Engine Performance through Real-Time Data and Predictive Maintenance," *Letters in High Energy Physics*, Vol. 2023, pp. 184-195, Nov. 2023, doi: 10.2139/ssrn.5031293

[33] B. K. Ram, N. Sharma, A. S. Joshi, and A. Vermani, "Predictive Maintenance and Production Analysis in Smart Manufacturing," in *Smart Technologies for a Sustainable Future*, M. E. Auer, R. Langmann, D. May, and K. Roos, Eds., Cham: Springer Nature Switzerland, 2024, pp. 234-244

[34] S. Hammedi *et al.*, "Optimizing Production in Reconfigurable Manufacturing Systems with Artificial Intelligence and Petri Nets," *International Journal of Advanced Computer Science and Applications*, Vol. 15, Oct. 2024, doi: 10.14569/IJACSA.2024.0151044

[35] V. N. Azarov and A. V Chekmarev, "Process Architecture as a Neural Network and Artificial Intelligence," in *2024 International Conference "Quality Management, Transport and Information Security, Information Technologies" (QM&TIS&IT)*, 2024, pp. 17-19, doi: 10.1109/QMTISIT63393.2024.10762857

[36] V. Kulynych, S. Shlyk, A. Symonova, R. Arhat, and V. Drahobetskyi, "Analysis of Modern Methods for Optimizing Technological Processes in Machine-building Production Using Artificial Intelligence," *СУЧАСНІ ТЕХНОЛОГІЇ В МАШИНОБУДУВАННІ ТА ТРАНСПОРТІ*, Vol. 1, pp. 31-37, May 2024, doi: 10.36910/automash.v1i22.1342

[37] G. Manoharan, S. Ashtikar, and M. Nivedha, "Harnessing the Power of Artificial Intelligence in Reinventing the Manufacturing Sector," 2024, pp. 113-137, doi: 10.4018/979-8-3693-2615-2.ch007

[38] J. S. Calado, J. Ferreira, J. P. Mendonça, and R. Jardim-Gonçalves, "AI Applications in the Configuration and Calibration of Industrial Machines," in *2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC)*, 2024, pp. 1-8, doi: 10.1109/ICE/ITMC61926.2024.10794282

[39] M. Yin, J. Tian, D. Zhu, Y. Wang, and J. Jiang, "A data-driven distributed process monitoring method for industry manufacturing systems," *Transactions of the Institute of Measurement and Control*, Vol. 46, No. 7, pp. 1296-1316, 2024, doi: 10.1177/01423312231195365

[40] Gy. Molnár, E. Nagy, "Current issues in effective learning: Methodological and technological challenges and opportunities based on modern ICT and artificial intelligence." *International Scientific Conference on Distance Learning in Applied Informatics*. Springer, pp. 1-11, 2024

[41] Gy. Molnar, Z. Szuts, Use of Artificial Intelligence in Electronic Learning Environments, In: Molnár, György (szerk.) *IEEE 5th International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE 2022)* Budapest, Magyarország : Institute of Electrical and Electronics Engineers (IEEE) 2022, pp. 137-140