UDC 004.896
IRSTI 28.23.27

https://doi.org/10.55452/1998-6688-2026-23-2-187-204

[1]Samigulina Z.I.,
PhD, ORCID ID: 0000-0002-5862-6415,
e-mail: z.samigulina@kbtu.kz
[1]*Dyussenkulova B.Z.,
PhD student, ORCID ID: 0009-0001-2788-6521,
*e-mail: bu_dyussenkulova@kbtu.kz
[1]Butakova D.A.
Master's student, ORCID ID: 0009-0006-8151-4928,
e-mail: d.butakova@kbtu.kz

[1]Kazakh-British Technical University, Almaty, Kazakhstan

REINFORCEMENT LEARNING–DRIVEN CONTROL STRATEGIES FOR SMART MANUFACTURING SYSTEMS

Abstract

Nowadays, the creation of smart manufacturing systems has high importance. Neural networks have been widely applied to solve complex manufacturing challenges. The paper is devoted to the study of neural networks with reinforcement learning as PPO (Proximal Policy Optimization), DQN (Deep Q-network) for state diagnosis of industrial equipment within the GEMMA (Guide d'Etude des Modes de Marche er d'Arret) model. The GEMMA French approach is established on the SFC (Sequential Function Charts) language and includes standards for controlling technical processes. An application of neural networks in area D of the GEMMA model is introduced. Modelling and experimental results were conducted based on synthetic and experimental datasets. The implementation of the architecture considered allows us to achieve reliable results for industrial data.

Keywords: Smart manufacturing system, reinforcement learning, proximal policy optimization (PPO), deep Q-network (DQN), informer transformer, Guide d'Etude des Modes de Marche er d'Arret (GEMMA) model, diagnostic of industrial equipment.

Received May 5, 2025; revised October 1, 2025, February 21, 2026; accepted April 7, 2026.

Introduction

Nowadays, smart manufacturing systems play a significant role in production. In the concept of smart manufacturing system Artificial Intelligence (AI), Internet of Things, Machine Learning (ML) and Digital Twin (DT). Smart Factory is determined as a manufacturing system, in which AI, IOT and automation are applied for enhancing efficiency and flexibility.

Methods of artificial intelligence and machine learning are widely applied, especially reinforcement learning (RL) is used for defining appropriate strategy, classification and other tasks. In RL agent, environment and action are applied in completing appropriate tasks. RL methods have broad applications in engineering, electronics, etc. [1]. Deep reinforcement learning is defined as a method in which are combined methods of deep learning and reinforcement learning. As was stated in [2], a data-driven approach was proposed for manufacturing automation and digital twin was used for modelling production cells, forecasting failures. For industrial control was used deep Q-learning in the context of smart manufacturing and Manufacturing Intelligence was trained on the constructed DT. Moreover, in the study [3] was introduced production control method, based on DT and RL. The proposed method has demonstrated high efficiency.

In addition to methods of neural networks and reinforcement learning methods, federated learning methods are widely used in smart manufacturing. As was stated in the research [4], multi-level architecture, which consists of the federated learning, IoT and crowdsourcing. The proposed approach can handle in terms of development of intelligent systems for predictive maintenance and diagnosis of faults.

Mezair, Djenouri and et.al proposed an advanced deep learning scheme for fault forecasting. The new framework integrates LSTM, CNN and graph neural network for working with heterogenous data. According to the results of the experiment, the proposed approach is superior in comparison with existing solutions in terms of training accuracy and speed and less power in consumed [5]. In the paper [6] were reviewed recent studies about fault detection and working algorithm have been presented. The dataset from Case Western Reserve University was taken for implementing Machine learning approaches. A novel anomaly detection approach was proposed and adapted to traditional manufacturing environments. This approach is combined with dynamic network theory and unsupervised machine learning. According to results, the new proposed method demonstrated f1-score more than 84 percents, precision with ninety-nine percents and recall more than 74 percents [7].

In the study [8], AI-driven predictive maintenance system, which is established on Digital Twin technology for monitoring industrial faults, diagnosis, and forecasting. For maintenance optimization and downtime reduction, Machine and Deep learning approaches, edge-computing and reinforcement learning were applied in the system. According to the validation results on smart manufacturing, prediction accuracy was increased by thirty-five percents, unplanned downtown was reduced by 40 percents and maintenance cost by twenty-five percents. The transition from Industry 4.0 to Industry 5.0, which is aimed at reliance, flexibility and interaction between human and machine. To address adaptation and problem about resource allocation, Digital twin system with a Soft Actor-Critic approach was proposed. The new proposed method was based on the Reinforcement learning [9]. The paper [10] examines the use of predictive maintenance in Industry 4.0 and emphasizes main challenges such as the need for a large amount of failure data and the choice of an adaptive maintenance approach. A multi-agent method based on reinforcement learning (RL) is proposed, where agents partially observe equipment status and coordinate task allocation among technicians with different skills. Experiments show that this method reduces failures and downtime, improving maintenance efficiency by ≈75% compared to traditional strategies. In the study [11] was proposed a novel algorithm for maintenance optimization in Industry 4.0 using Reinforcement learning. The system degradation process considered. In other words, an “imperfect pair” model was introduced, where efficiency is fallen in case of each subsequent repair. For development maintenance strategies, Double Deep Q-network based agent was applied, the agent was adapted to various scenarios without predefined maintenance threshold.

In the smart manufacturing system job shop scheduling is considered, as problem about job shop scheduling related to resource optimization, minimization of time differences, tasks allocation in machines. Machine learning algorithms, especially Reinforcement Learning, is widely used for job shop scheduling problems. In the research [12] was considered Smart Manufacturing Scheduling (SMS) and for optimizing job shop scheduling was proposed conceptual mode. The model is established around three key components: semantic ontology, hierarchical agent structure and deep reinforcement learning (DRL). The proposed model increased flexibility through improved scheduling. Additionally, according to the study [13], mathematical description of the smart manufacturing service scheduling problem. Novel method, which was established on the Deep reinforcement learning, was introduced to reduce maximum completion time. The efficiency was tested in accordance with two case studies. In the study [14] was considered problem about dynamic scheduling in job shop scheduling under unforeseen equipment failures. As solution Proximal Policy Optimization (PPO) algorithm was offered a for improving performance, PPO was tested in a real production environment and demonstrated better results in comparison with state-of-the-art algorithms. Moreover, Yang and Xu have developed a system architecture, mathematical model for minimization delay costs and Deep Reinforcement

learning system with agents with scheduling and reconfiguration. The Advantage-Actor-Critic algorithm was applied and according to the experiments, decision quality and computation time have been demonstrated as approximately 55 and 88 percents. Also, shown decision time made the method appropriate for real-time optimization [15]. The study [16] introduced a reinforcement learning (RL) based approach for dynamic route planning of automated guided vehicles (AGVs) in smart warehouses and factories. The novelty of the research was in the integration of RL with AGVs in a dynamic environment considering workstations, charging stations and storage areas. Experiments in a smart manufacturing environment have validated the effectiveness of the method for complex logistics problems. The solution was scalable and contributed to improving the efficiency of modern industrial systems. Zhou, Tang, et. al in study [17] proposed AI-scheduler, based on RL to avoid complications about planning task. The introduced AI-scheduler applies composite reward functions for dynamic planning optimization in condition of uncertainty. Experiments have shown that the new AI scheduler improves key performance indicators (reduces production time, reduces costs, balances equipment load) and effectively handles unexpected events such as urgent orders and equipment breakdowns.

Furthermore, in smart manufacturing systems, hybrid algorithms are being developed. For example, in the study [18] was proposed a hybrid ML model, which combines RL and Genetic Algorithm (GA) for production scheduling optimization. The proposed model trained for 500 iterations. The experimental results have demonstrated that production efficiency increased by 39 percents, resource utilization was reached to 91 percents, equipment downtime was reduced by 34 percents, energy consumption per task decreased by 17 percents and indicator for delivery on time was increased to 94 percents. This study demonstrates the potential of integrating ML into production planning, providing flexibility, resilience and efficiency within Industry 4.0. In the research [19] was considered the concept of a federated digital twin (FDT), expanding the use of digital twins (DT) to manage complex interconnected systems. The main issue was that the variable processing speed of DT functions, which makes it difficult for optimal plan application. The system, which is called Federated Digital Twin, was introduced. The proposed system models Digital Twin relationships by temporal heterogeneous graphs and applied Deep Learning and Graph Neural networks for flexible planning. The proposed approach is superior to traditional algorithms and demonstrates efficiency in managing digital twins.

Hybrid flow shops play a key role in modern smart factories, where various unforeseen situations arise. Dynamic planning in such systems (HFSP) becomes a challenging task due to the uncertainty of processing times, dynamic deadlines for receiving orders and flexible maintenance of equipment. In the paper [20] the Neuro Evolution of Augmenting Topologies (NEAT) algorithm was introduced to minimize the maximum order fulfillment time. Experiments have shown that NEAT surpasses Deep Q-Network (DQN) and classical priority planning rules (PDRs), providing a faster and better response to dynamic changes. Testing on new data has confirmed his ability to effectively adapt to previously unknown tasks.

The article is structured as follows: Section 2 is devoted to the formulation of the research problem; Section 3 describes the methods applied in the study; Section 4 presents the dataset description; Section 5 contains the simulation and experimental results. The conclusion summarizes the main findings and research outcomes.

Problem statement

Modern industrial automation systems consist of high-priced equipment, the timely maintenance of which can reduce the economic costs of the enterprise, maintenance and repair. Industrial automation systems are built considering international quality standards such as ISA, IEC, etc. The integration of artificial intelligence methods into automatic process control systems should consider the features of the operation of industrial equipment, the specifics of industrial data and the principles of operation of programmable logic controllers.

Current research is devoted to the modernization of the Guide d'Etude des Modes de Marche er d'Arret (GEMMA) model, to implement different modes of operation of equipment, as well as

considering disaster recovery procedures. It is proposed to integrate a training-based neural network with reinforcement to diagnose equipment health in Area D (post-accident repair procedures) to prevent plant failures. To solve this problem, the following architectures were considered: Proximal Policy Optimization (PPO), Deep Q-network (DQN), and Informer transformers. Software and hardware implementation was carried out based on "Industrial Automation Lab", Schneider Electric in JSC KBTU.

Materials and methods

3.1 GEMMA Guide paradigm

In the study, smart manufacturing system for machine diagnosis was developed. The proposed system was established on the basis of reinforcement of learning neural networks and GEMMA Guide paradigm. The GEMMA Guide paradigm was represented as a block-diagram that describes different modes of the system's operation and transitions between modes. The primary purpose of the GEMMA Guide can be stated as correct determination of the process' operating modes and improvement of the interaction between the operator and the system. The improved interaction between the operator and the system contributes to prevention errors and delays in production, The GEMMA Guide paradigm was cooperated in 1984 by the National Agency for the development of Automated Production [21].

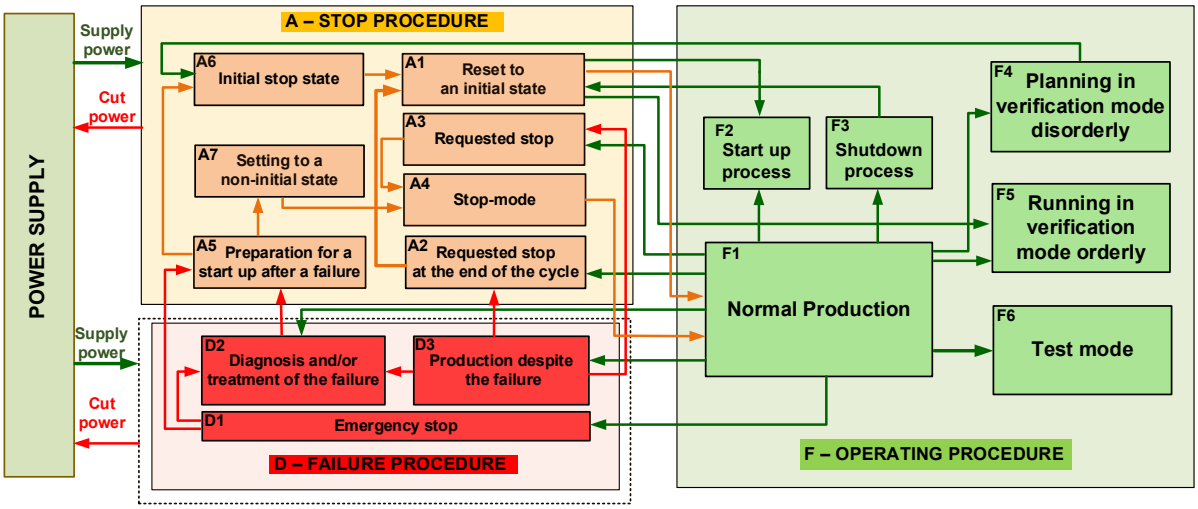


Figure 1 – Architecture of the GEMMA model for complex object control

As was illustrated in Figure 1, the GEMMA Guide paradigm consists of the A, F and D modes. The modes are organized as follows:

- A (Stop procedure) – stopping processes, including a completion cycle, stopping in a definite state and starting of the system to return to its original state;
- F (Operating procedure) – processes of normal production, including start, stop, test and checking modes;
- D (Failure procedure) – process of activation of the failures and pressing emergence stops.

In Figure 2 the architecture of a Smart manufacturing system with faults diagnosis was demonstrated. This system was based on reinforcement learning neural networks and GEMMA Guide paradigm. The main purpose was stated as faults diagnosis in zone D using Reinforcement learning approaches. The realization consists of the stages below:

Algorithm 1. Intelligent industrial data processing based on GEMMA model and neural networks with reinforcement learning.

Step 1. Collecting data from programmable logic controllers (Modicon M340) and saving in appropriate format.
Step 2. Data splitting for training and testing.
Step 3. Data preprocessing.
Step 4. Creation of the environment for realization of Reinforcement Learning Methods.
Step 5. Implementation of the RL approaches.
Step 6. Evaluation of the results using metrics.
Step 7. Decision making.

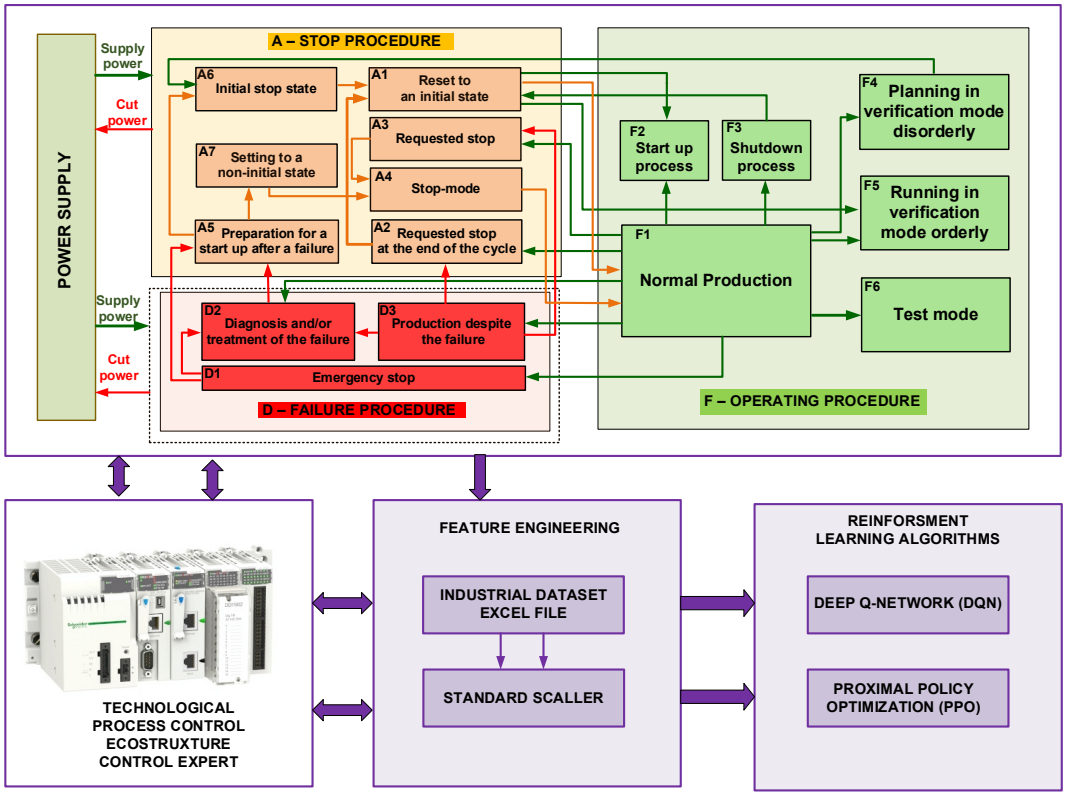


Figure 2 – Architecture of a smart manufacturing system with diagnostics of equipment based on reinforcement learning neural networks and GEMMA Guide Paradigm

Algorithms of reinforcement learning (RL), based on neural networks, were considered as application of machine learning algorithms to the GEMMA model. Architectures of the Proximal Policy Optimization (PPO) and Deep Q-network (DQN) for industrial data allow solving the problem of forecasting failures in the equipment and process optimization. The singularity of the application of these networks is the ability to handle with a large dataset. The disadvantages are the implementation complexity and long training process, however for the analysis of archival data and predictive analytics these neural networks are represented with interest. The efficiency of application of these networks was evaluated by using comparative analysis with neural networks of the class of transformers such as Informer and Time Series transformers.

3.2 Hardware description

The results of simulations and experiments on real equipment were conducted in the Industrial Automation Lab (Schneider Electric) (Figure 3) based on the Modicon M340 series programmable logic controller (PLC).

Figure 3 – Controller Modicon M340

Modicon M340 supports 1024 discrete I/O, can process up to seven thousand instructions per millisecond, has a large memory capacity of 4 MB for programs and 256 KB for data storage, supports communication protocols Ethernet TCP/IP, CANopen, Modbus, etc. The GEMMA model is built on the Modicon M340 controller in the EcoStruxture Control Expert software product.

3.3 Proximal Policy Optimization

Proximal Policy Optimization (PPO) is type of RL algorithm, which is based on policies and experiences. The experiences are collected by the PPO agent by interaction with the environment. Eventually, the experience is saved in the memory of the agent [22]. As a classification task, fault diagnosis is included; the agent should guess the class of a sample. As class is assumed by an agent, an environment distributes a reward immediately to the agent. For instance, the reward is awarded positively in case of correct determination of the class; otherwise, the negative reward is assigned [23]. Below pseudocode of PPO is represented for fault diagnosis.

Pseudocode 1 – PPO algorithm for industrial data

Input parameters: features
Output parameter: class

1. Initialization:
   Initialization of the environment for industrial data (for example, awarding rewards)
   Data normalization
   Balancing of the classes
   Initialization of neural networks and policies
   Assigning hyperparameters (epsilon, episodes, learning rate)
2. Training of the model
3. Calculating the rewards
4. Updating of the model
5. Returning the feedback from the model

Data from sensors and generated data are accepted to environment; state is formulated and is transferred to PPO agent. As in the PPO agent, states are processed by input and hidden layers; PPO agent chooses actions that state for faults or normal production. Distribution of actions probabilities is generated by actor network and output signal in the form of arrangement of 2 classes as fault or normal production.

3.4 Deep Q-network (DQN)

The DQN algorithm is identified as an approach of deep reinforcement learning. The premise behind the application of the DQN algorithm is that deep learning (DL) and Reinforcement learning (RL) methods are integrated. A function that should be maximized or minimized is built by the Q-learning method to apply for DL. The Q-value is approximated by the application of value function in the DQN algorithm [24]. Below is demonstrated the pseudocode of DQN.

Pseudocode 2 – DQN algorithm for industrial data

Input parameters: features
Output parameters: class

1. Initialization
   Setting hyperparameters: learning rate, discount factor, epsilon, beta, number of episodes *n*
   Setting experience buffer, mini-batch size
2. Data preprocessing
   Standard scaling
   Data splitting for training and testing
3. Neural network initialization
   Q-network
   Experience replay buffer
   Adam optimizer
4. For each episode:
   Initialization of the environment
   Get (state, reward, done)
   Error calculation
   Save (previous state, reward, done, state)
   If memory is full:
      Setting mini batch
      Update Q-network
5. For each step:
   Update Q-target
6. Model evaluation

As in PPO approach, the state is interconnected with environment and access input parameters. After receiving the current condition, the decision will be taken in the form of output parameters as 0 or 1. In experience replay are saved the past conditions as state, etc. DQN is applied to be a training agent in terms of making decisions on the base of state conditions.

3.5 Transformers

In the modern world, the architecture of transformers is getting popular. Similar architecture can be applied to time series analysis. Also, the application of technological process in the industry, information from sensors is identified as in the form of series. However, not all transformers can be implemented to solve industrial problems. According to the metrics, high indicators are shown after implementation of transformers.

3.5.1 Informer Transformer

Informer Transformer is modified version of transformer, which is developed for time series analysis. In another words, informer transformers are determined as advancement on the transformer. The structure of the Informer Transformer is identical with transformer that it consists of multi-layers [25]. In the transformer ProbSparse Self Attention is applied for analyzing significant points. ProbSparse attention accepts key (K), query (Q) and value (V) as input and computes the dot product using the formula:

$$A(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

where $Q \in R^{LQ\times d}$, $K \in R^{LK\times d}$,$V \in R^{LV\times d}$, the input dimension is denoted by d[26].

Architecture of the informer transformer is illustrated on Figure 4.

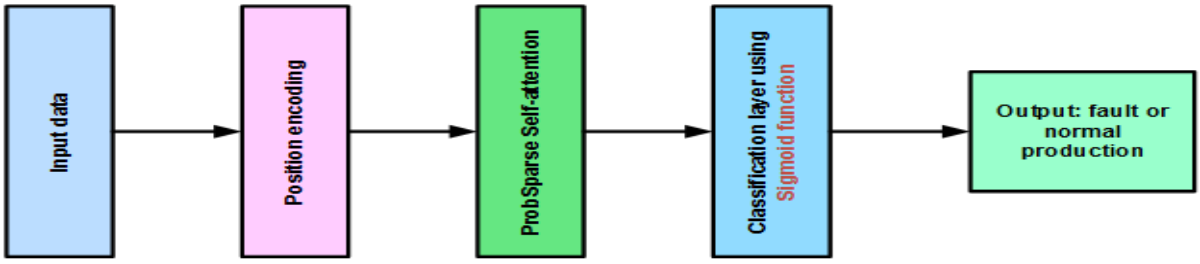


Figure 4 - Informer transformer architecture

As the input data are received normalized time series data. Position encoding is applied as order of input parameters is not considered. At the prospers self-attention stage, significant elements are chosen to concentrate on vital characteristics. The sigmoid activation function is applied for transformation of outputs into class probabilities.

3.6 Prioritized Experience Replay (PER)

Prioritized Experience Replay (PER) is applied to RL systems with high capacity. A buffer in Experience Replay has functionality in terms of saving visited experience. This approach demonstrates high effectiveness in training deep RL models [27]. PER was created with direct and simple prioritizing the experiences with high values of temporal differences (TD) in sampling of experience for training process. TD errors are applied to introduce the experience priorities and produce bias. Stochastic prioritization with experience probability was presented to avoid issues about producing of bias as was stated in the formula (2), where p_i is denoted as priority of τ_i, α denotes a coefficient, which describes the process of identifying priority in sampling.

$$P(i) = \frac{{p_i}^\alpha}{\sum_k {p_k}^\alpha} \quad (2)$$

The formulas (3) and (4) are stated for proportional and rank-based prioritizations as for detailed priorities:

$$p_i = |\delta_i| + \varepsilon \quad (3)$$

$$p_i = \frac{1}{rank(i)} \quad (4)$$

where $\delta_i$temporal difference error of is τ_i, ε is denoted as minimum positive number and $rank\,(i)$ is the index of τ_i [28].

3.7 L2-regularization

L2 regularization is a Machine Learning method for model controlling without avoiding features. The main idea is penalization of large coefficients and reducing the risk of overfitting. The loss function is adapted by L2 regularization using penalty addition to the sum of model parameters coefficients and the formula of the L2 regularization is determined as shown in the formula (5):

$$Loss = Loss_{orig} + \lambda \sum_{i=1}^{n} {w_i}^2 \quad (5)$$

where $Loss_{orig}$ the original loss function for regression is, λ is identified as regularization parameters for penalty control and wi is determined as model coefficients. In neural networks, L2 regularization is included in training using weight decay and the formula is changed, as was illustrated in formula (6):

$$Loss = Loss_{orig} + \lambda \sum_{j=1}^{m} \sum_{k=1}^{n} {w_{jk}}^2 \quad (6)$$

where w_{jk} is identified as the weights, which connects the neurons between layers j and k, k, m are the numbers of layers and n is the number of linkages [29].

Dataset description

Two datasets were considered in this research. The first artificial dataset generated for the development of approaches has dimensionality R1=1500x6. For second dataset, which has dimensionality R2=150x6, are represented measurements from sensors from industrial stands from Schneider Electric (Figure 4).

Synthetic data

Synthetic data was generated for the implementation of experiments. Synthetic data is artificially generated data, which is applied for imitation of real data. The end-users can control the degree of similarity level of generated and real data. It helps researchers to develop and test methods, examine and validate machine learning algorithms with provision of efficiency [30]. In table 1 specification of the dataset is illustrated. Generated data consisted of six columns and 1500 rows: the target column was identified with two classes as failure or normal production.

Table 1 – Specification of the artificial dataset

Number	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Target
0	0.374540	0.950714	0.731994	0.598658	0.156019	1
1	0.155995	0.058084	0.866176	0.601115	0.708073	0
2	0.020584	0.969910	0.832443	0.212339	0.181825	0
3	0.020584	0.969910	0.832443	0.212339	0.181825	0
4	0.183405	0.304242	0.524756	0.431945	0.291229	0
…	…	…	…	…	…	…
1500	0.897397	0.119381	0.327843	0.815745	0.597312	1

Experimental data from Modicon M340 PLC (Industrial Automation Lab)

Data collection is shown on the example of an industrial object of control of an absorption plant. The technological process is implemented at the stand with programmable logic controllers of the Modicon M241 series, Modicon M340, HMI displays of the Harmony series, frequency converters of the Altivar series. Figure 5 shows the 3rd model of the bench in Fusion 360 software.

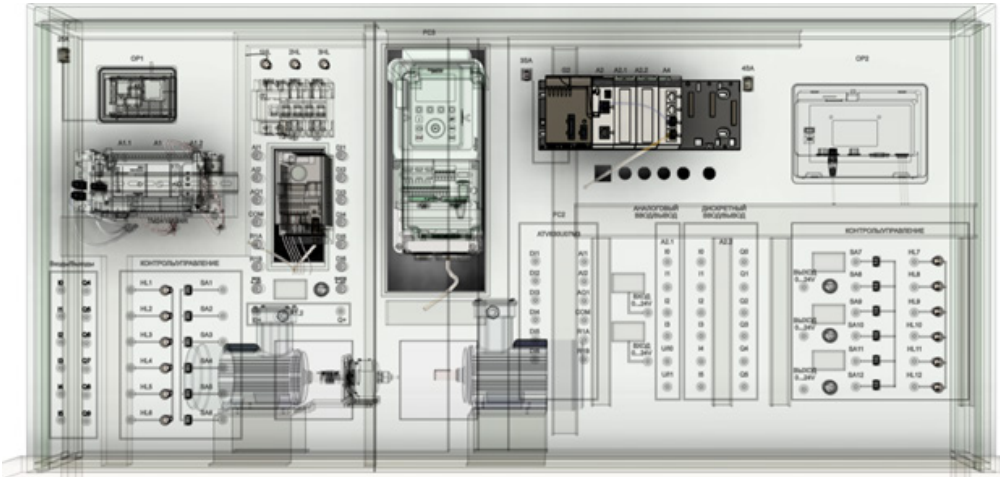


Figure 5 – 3D model test bench with industrial PLCs

Figure 6 shows the software implementation of the absorption plant [37].

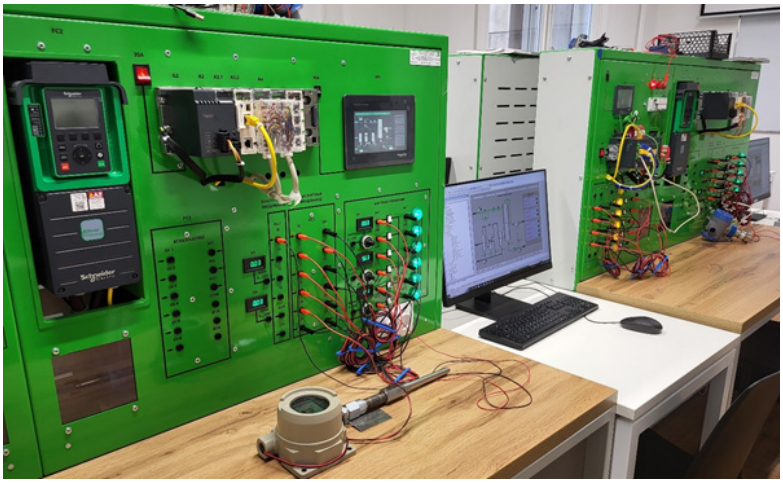


Figure 6– Data collection based on Modicon M340
programmable logic controller

Figure 7 illustrates the mnemonic diagram of process control.

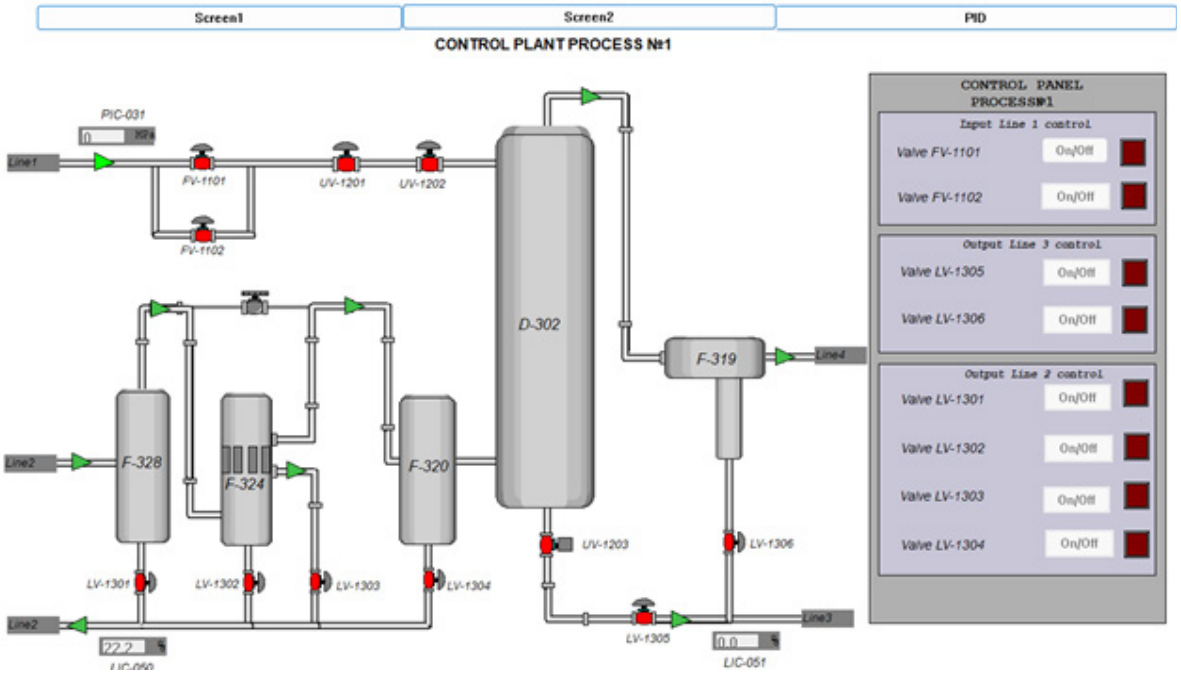


Figure 7 – Mnemonic diagram of Absorption process control
Installation of D-302

The experimental data was obtained from sensors. In this dataset contains information about the indicators of pressure and temperature. As the target column was marked column about classes. The classes were classified as “0” and “1”. In the table the view of dataset was demonstrated. In the dataset indicators are saved from experiments, as recordings about temperature and pressure from sensors.

Table 2 – Specification of the experimental data

№	Temp_sensor (C)	PT1	PT2	PT3	PT4	Class
0	24,66	0,3375	0,285	0,315	0,27	1
1	24,66	0,3375	0,285	0,315	0,27	1
2	24,66	0,3375	0,285	0,315	0,27	1
3	24,66	0,3375	0,285	0,315	0,27	1
4	24,5	0,3375	0,285	0,30375	0,36	1
	…	…	…	…	…	…
97	24,79	0,3	0,3	0,315	0,315	1

The columns “PT1, PT2, PT3, PT4” indicate data from 4 various stands.

Data preprocessing

Data preprocessing is determined as a vital stage before training model, as data from sensors may have noise, missing values and outliers. In data preprocessing tasks are included data cleaning, data reduction, scaling, transformation and dividing data into the partitions [31]. Low-quality data can lead to consequences such as predictions with low accuracy and ineffectiveness in terms of time. Industrial data are facing challenges such as:

- complex data structures;
- difficulties in terms of integration and data aggregation;
- issues about data quality evaluation due to high dimension;
- lack of unified standards of data quality;
- prominent level of requirements to data processing quality [32].

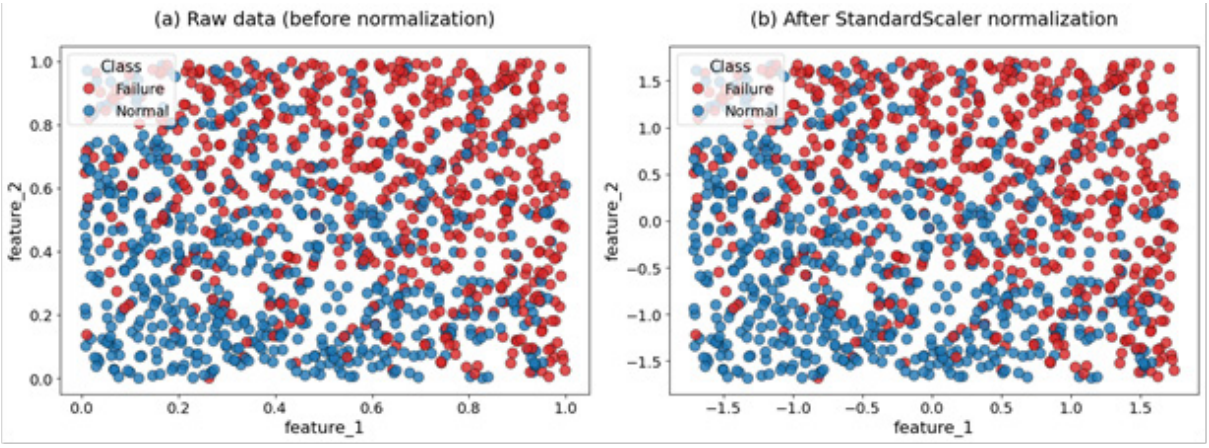
Emission removal techniques, balancing of classes, feature scaling are applied to improve quality of models. These steps help to increase performance of metrics, avoid hyper training and generatability of the model.

3.3.1 Standard Scaling

The standard scaling is presented in equation (1):

$$z = \frac{x-u}{s} \tag{7}$$

where z is determined as output vector of scaled numerical features of item, x is the input vector and s are standard deviation [33]. Figure 8 demonstrates distribution of parameters in synthetic and experimental datasets before and after processes of normalization.



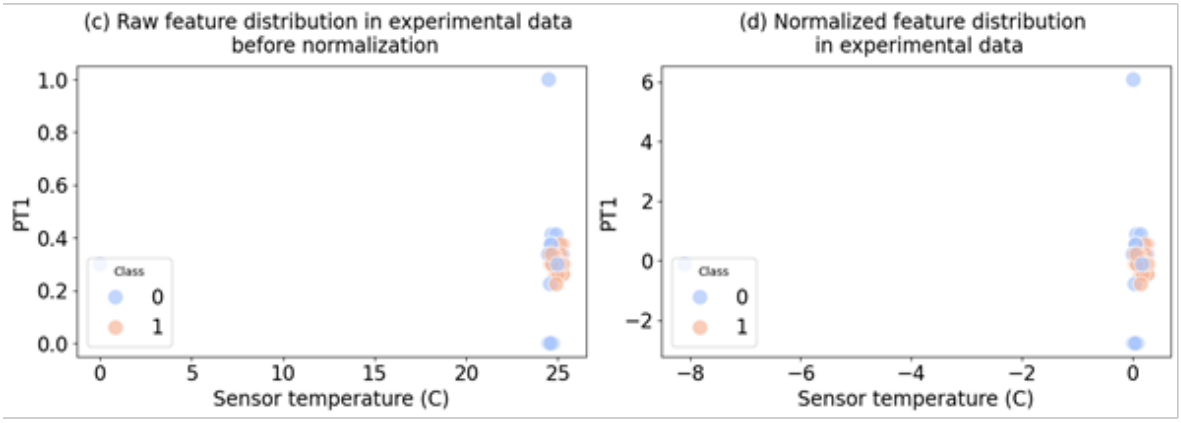


Figure 8 - Distribution of experimental data before and after normalization

After normalization using StandardScaler method, values of standard deviation were equal to 1. It can be explained by centering of the data around zero after subtraction of the averages from each value.

Results and discussion

Performance evaluation of PPO, DQN, Informer Transformer

After implementing the sampling techniques on the initial dataset, the classifiers were developed based on PSO and ensemble methods, which results are provided in this section. Precision evaluates the accuracy of model in terms of guessing positive predictions. In classification tasks accuracy is applied whether the model forecasted the label. Recall is determined as tool for measuring the model's ability to identify positive instances among all positives in the dataset. F1-score is a statistical measure, which consists of precision and recall, it evaluates ability determining positive examples. The formulas of metrics as accuracy, precision, recall and f1-score are illustrated in (7) – (11), where TP is true positive, FP is false positive, TN is true negative, and FN are defined as False negative values [34].

$$precision = \frac{TP}{TP+FP} \tag{8}$$

$$recall = \frac{TP}{TP+FN} \tag{9}$$

$$f1-score = \frac{2*precision*recall}{precision+recall} \tag{10}$$

$$accuracy = \frac{TP+TN}{TP+TN+FN+FP} \tag{11}$$

The ROC-curve is identified as receiver operating characteristic with the formula:

$$ROC(\sigma) = \left(1 - TN(\sigma), TP(\sigma)\right) \tag{12}$$

where $1 - TN(\sigma)$ can be defined as false-positive (FP) of the threshold (σ)[35].

Evaluation of synthetic data using performance metrics

Synthetic data was generated by the function for creating datasets using Python programming language. After, PPO, DQN with PER and Informer transformer methods were applied for synthetic data. Then applied methods were evaluated metrics as accuracy, precision, recall, f1-score, and ROC-AUC scores. Table 3 demonstrates the indicators of performance metrics. PPO has the highest

indicator in terms of accuracy, precision and f1-score, while DQN with PER has DQN with PER has recall with 100 percents. Informer transformers have the best performance in terms of ROC-AUC. According to the percentage in performance metrics, it is meant that PPO model was able to predict classes and has less quantity of false positives. The quantity of false positives plays significant role in smart manufacturing because it will help in avoiding premature line stops and penalties for idle time, which may be caused by false alarm. The maximum percentage of recall means that DQN with PER never miss's never real failures and a smaller number of false negatives. It can be explained with that fact Prioritized Experience Replay (PER) has buffer for assigning priorities in terms of training. In manufacturing processes, a failure to detect a fault is dangerous or may lead to serious losses such as equipment breakdown, accident, and defective batch. ROC-AUC score has the highest percentage in all models.

Table 3 – Performance metrics for PPO, DQN with PER, and Informer Transformer approaches on synthetic data (data breakdown: 80% training, 20% test)

Metrics	PPO	DQN with PER	Informer transformer
Accuracy	99%	94,67	96,67%
Precision	99,31%	90,12%	95,33%
Recall	98,63%	100%	97,95%
F1-score	98,97%	94,81%	96,62%
ROC-AUC score	98,99%	94,81%	99,55%

All metrics are reported as percentages (%). In Figure 9 the ROC curves are shown. All realized models have ROC-AUC more than 90 percents.

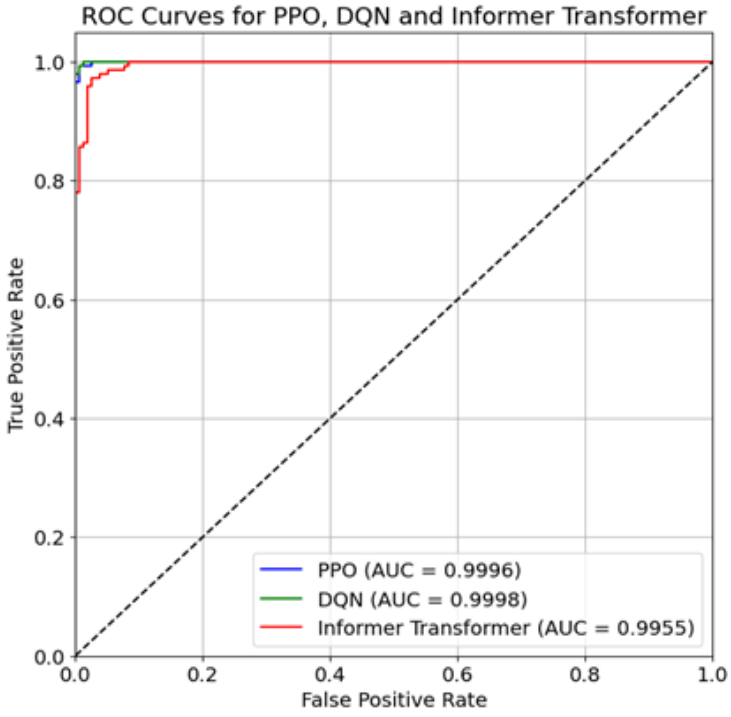


Figure 9 – ROC curves for PPO, DQN, and Informer Transformer models on synthetic data (80/20 — train/test)

The average AUC values for 3 runs are shown in Figure 7 (PPO: lr=2e-5, batch=512, gamma=0.919, clip=0.21; DQN: lr=5e-4, batch=64, gamma=0.99, train_freq=4; Informer: input=5,

hidden=128, layers=2, dropout=0.1, lr=9.586e-4. Axes: FPR (X) and TPR (Y), from 0 to 1. AUC — dimensionless metric). The ROC-AUC curves have percentages close to 100 percents due to training models for generated data. It can be explained with overfitting due to noise in the data.

Evaluation of experimental data using performance metrics

The experimental data was collected from temperature sensors from 6 stands from Schneider Electric laboratory. As in (4.2), the experimental data was evaluated by metrics as accuracy, precision, recall, f-score and ROC-AUC score. In Table 4 performance metrics are presented. As experimental data differed in terms of maintenance, the realized models were indicated low values in comparison with indicators of synthetic data performance. DQN demonstrated the highest percentages of accuracy and precision. Additionally, DQN and Informer transformers had equal percentages of recall and f1-score. Informer transformers show the highest values of ROC-AUC score, it means that the model handled with classification. In contrast with DQN and Informer transformers, PPO model had the lowest values of performance metrics. In can be explained and supposed with basement on trials and errors.

Experimental method has demonstrated the less values of performance metrics in comparison with synthetic data. It can be explained that data was dynamic, and overfitting hasn't been experienced.

Table 4 – Performance metrics for PPO, DQN with PER, and Informer Transformer on synthetic data (80% training, 20% test)

Metrics	PPO	DQN	Informer transformer
Accuracy	70%	88%	85%
Precision	80%	88%	88%
Recall	70%	85%	85%
F1-score	64%	84,1%	84,1%
ROC-AUC score	79%	75%	83,85%

All metrics are reported as percentages (%) (Table 4). In Figure 10 ROC-AUC curves for realized models are presented. The realized models have AUC indicator values less than 0.9.

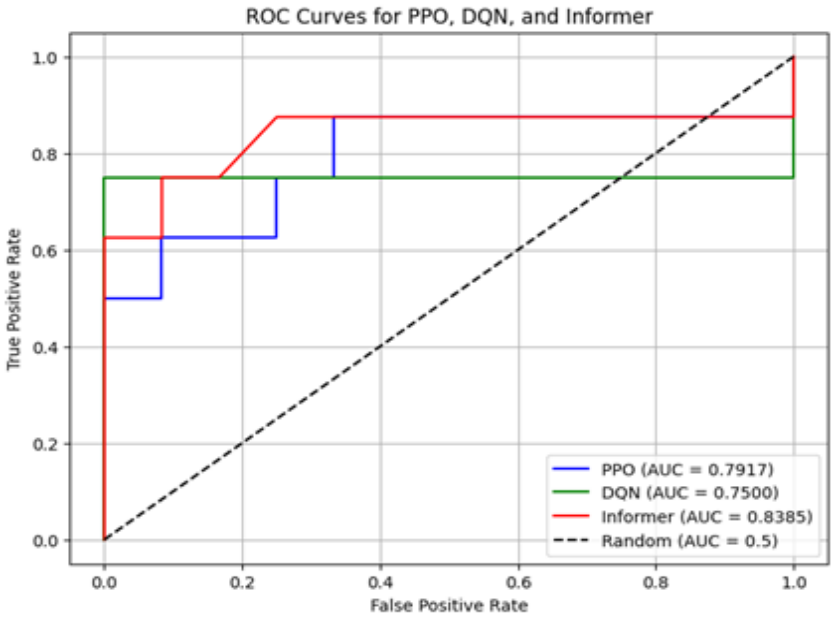


Figure 10 – ROC curves of PPO, DQN, and Informer Transformer models on experimental data (split: 80/20— train/test)

The following parameters were used: PPO (seed=42): lr=1e-5, batch=256, n_steps=4096, gamma=0.95, clip=0.25, ent_coef=0.55, vf_coef=2.0; DQN (seed=45): lr=1e-5, batch=64, buffer=150000, gamma=0.95, train_freq=20, target_update=800; Informer: l2_lambda=1e-5. Axes: FPR (X) and TPR (Y), range [0, 1]. AUC — dimensionless metric.

The DQN has the smallest value of AUC, as illustrated in Figure 8. The Informer transformer has the highest values of AUC in comparison with Reinforcement learning models such as PPO and DQN.

Conclusion

The aim of this research was to implement Reinforcement learning methods fault classification in smart manufacturing systems. PPO, DQN with PER were realized for synthetic, experimental data and Reinforcement learning methods were compared with Informer transformers. To compare and analyze effectiveness of the model's performance metrics and loss functions are applied. According to the results of the evaluation and analysis, Reinforcement learning, and Informer transformer models have pros and cons in terms of their tasks. For synthetic data DQN with PER and Informer transformer were appropriate as these approaches demonstrated high ability to classify faults and normal production, which contributes to minimizing the risk of false alarms. Also, in synthetic data, all implemented approaches the highest percentage in terms of all evaluated metrics, more than ninety percents. As artificial data was generated, overfitting might be caused and metrics should be checked properly for experimental data. For experimental data DQN with PER was suitable. In both cases as synthetic data with high dimension and experimental data, DQN with PER shows the best results as this approach are linked with setting priorities in training. Developed models can be applied in manufacturing processes to make decisions based on faults forecasting. In further research, attention mechanisms will be combined with developed models.

REFERENCES

1 Jayakumar, S., and Nandakumar, S. Distributed resource optimization using the Q-learning algorithm in device-to-device communication: A reinforcement learning paradigm. Results in Engineering, 23, 102462 (2024). https://doi.org/10.1016/j.rineng.2024.102462

2 Xia, K., Sacco, C., Kirkpatrick, M., et al. A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. Journal of Manufacturing Systems, 58, 210–230 (2021). https://doi.org/10.1016/j.jmsy.2020.06.012

3 Park, K.T., Son, Y.H., Ko, S.W., and Noh, S.D. Digital twin and reinforcement learning-based resilient production control for micro smart factory. Applied Sciences, 11 (7) (2021). https://doi.org/10.3390/app11072977

4 Ullah, I., Hassan, U.U., and Ali, M.I. Multi-level federated learning for Industry 4.0: A crowdsourcing approach. Procedia Computer Science, 217, 423–435 (2022). https://doi.org/10.1016/j.procs.2022.12.238

5 Mezair, T., Djenouri, Y., and Belhadi, A. A sustainable deep learning framework for fault detection in 6G Industry 4.0 heterogeneous data environments. Computer Communications, 187, 164–171 (2022). https://doi.org/10.1016/j.comcom.2022.02.010

6 Neupane, D., and Seok, J. Bearing fault detection and diagnosis using Case Western Reserve University dataset with deep learning approaches: A review. IEEE Access, 8, 93155–93178 (2020). https://doi.org/10.1109/ACCESS.2020.2990528

7 Mattera, G., Mattera, R., Vespoli, S., and Salatiello, E. Anomaly detection in manufacturing systems with temporal networks and unsupervised machine learning. Computers & Industrial Engineering, 203, 111023 (2025). https://doi.org/10.1016/j.cie.2025.111023

8 Prabu, S., Senthilraja, R., Ali, A.M., Jayapoorani, S., and Arun, M. AI-driven predictive maintenance for smart manufacturing systems using digital twin technology. International Journal of Computational and Experimental Science and Engineering, 11 (1), 1350–1355 (2025). https://doi.org/10.22399/ijcesen.1099

9 Asmat, H., Din, I., Almogren, A., and Khan, M. Digital twin with soft actor-critic reinforcement learning for transitioning from Industry 4.0 to 5.0. IEEE Access (2025). https://doi.org/10.1109/ACCESS.2025.3546085

10 Ruiz Rodríguez, M.L., Kubler, S., de Giorgio, A., Cordy, M., Robert, J., and Le Traon, Y. Multi-agent deep reinforcement learning based predictive maintenance on parallel machines. Robotics and Computer-Integrated Manufacturing, 78, 102406 (2022). https://doi.org/10.1016/j.rcim.2022.102406

11 Pliego Marugán, A., Pinar-Pérez, J.M., and García Márquez, F.P. A reinforcement learning agent for maintenance of deteriorating systems with increasingly imperfect repairs. Reliability Engineering & System Safety, 252, 110466 (2024). https://doi.org/10.1016/j.ress.2024.110466

12 Serrano-Ruiz, J.C., Mula, J., and Poler, R. Development of a multidimensional conceptual model for job shop smart manufacturing scheduling from the Industry 4.0 perspective. Journal of Manufacturing Systems, 63, 185–202 (2022). https://doi.org/10.1016/j.jmsy.2022.03.011

13 Zhou, L., Zhang, L., and Horn, B. Deep reinforcement learning-based dynamic scheduling in smart manufacturing. Procedia CIRP, 93, 383–388 (2020). https://doi.org/10.1016/j.procir.2020.05.163

14 Zhang, M., Lu, Y., Hu, Y., et al. Dynamic scheduling method for job-shop manufacturing systems by deep reinforcement learning with proximal policy optimization. Sustainability, 14 (9) (2022). https://doi.org/10.3390/su14095177

15 Yang, S., and Xu, Z. Intelligent scheduling and reconfiguration via deep reinforcement learning in smart manufacturing. International Journal of Production Research, 60 (16), 4936–4953 (2022). https://doi.org/10.1080/00207543.2021.1943037

16 Ho, G., Tang, Y., Leung, E., and Tong, P. Integrated reinforcement learning of automated guided vehicles dynamic path planning for smart logistics and operations. Transportation Research Part E: Logistics and Transportation Review, 196 (2025). https://doi.org/10.1016/j.tre.2025.104008

17 Zhou, T., Tang, D., and Wang, L. Reinforcement learning with composite rewards for production scheduling in a smart factory. IEEE Access, 9, 752–766 (2021). https://doi.org/10.1109/ACCESS.2020.3046784

18 Bin Shaikat, F., and Ahmed Faysal, S. Optimization of production scheduling in smart manufacturing environments using machine learning algorithms (2025).

19 Kim, Y., Kim, H., Ha, B., and Kim, W. Federated digital twins: A scheduling approach based on temporal graph neural network and deep reinforcement learning. IEEE Access (2025). https://doi.org/10.1109/ACCESS.2025.3530558

20 Chen, Y., Zhang, J., and Rauf, M. Dynamic scheduling of hybrid flow shop problem with uncertain process time and flexible maintenance using NeuroEvolution of Augmenting Topologies. IET Collaborative Intelligent Manufacturing, 6 (3) (2024). https://doi.org/10.1049/cim2.12119

21 Schoepp, S., Taghian, M., and Miwa, S. Enhancing hardware fault tolerance in machines with reinforcement learning policy gradient algorithms (2024). https://doi.org/10.48550/arXiv.2407.15283

22 Modirrousta, M.H., Aliyari Shoorehdeli, M., and Yari, M. Imbalanced classification in faulty turbine data: New proximal policy optimisation. IET Collaborative Intelligent Manufacturing, 6 (3) (2024). https://doi.org/10.1049/cim2.12114

23 Yang, Y., Juntao, L., and Lingling, P. Multi-robot path planning based on a deep reinforcement learning DQN algorithm. CAAI Transactions on Intelligence Technology, 5 (3), 177–183 (2020). https://doi.org/10.1049/trit.2020.0024

24 Zhu, Q., Han, J., Chai, K., and Zhao, C. Time series analysis based on Informer algorithms: A survey. Symmetry, 15 (4) (2023). https://doi.org/10.3390/sym15040951

25 Zhang, X., Yang, K., and Zheng, L. Transformer fault diagnosis method based on TimesNet and Informer. Actuators, 13 (2) (2024). https://doi.org/10.3390/act13020074

26 Pan, Y., et al. Understanding and mitigating the limitations of prioritized experience replay (2023). https://doi.org/10.48550/arXiv.2007.09569

27 Bu, F., and Chang, D.E. Double prioritized state recycled experience replay. In: Proceedings of the IEEE International Conference on Consumer Electronics - Asia (2020). https://doi.org/10.1109/ICCE-Asia49877.2020.9276975

28 Chandrinos, N., Loi, I., Zachos, P., et al. Effectiveness of L2 regularization in privacy-preserving machine learning (2024). https://doi.org/10.48550/arXiv.2412.01541

29 Alinda Rahmi, N., and Defit, S. The use of hyperparameter tuning in model classification: A scientific work area identification. International Journal of Applied Sciences and Smart Technologies, 8 (4) (2025). http://dx.doi.org/10.62527/joiv.8.4.3092

30 Shekhar, S., Bansode, A., and Salim, A. A comparative study of hyper-parameter optimization tools. In: 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (2021). https://doi.org/10.1109/CSDE53843.2021.9718485

31 Jordon, J., Szpruch, L., Houssiau, F., et al. Synthetic data – what, why and how? (2022). https://doi.org/10.48550/arXiv.2205.03257

32 Fan, C., Chen, M., Wang, X., et al. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. Frontiers in Energy Research, 9 (2021). https://doi.org/10.3389/fenrg.2021.652801

33 Bekar, E.T., Nyqvist, P., and Skoogh, A. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. Advances in Mechanical Engineering, 12 (5) (2020). https://doi.org/10.1177/168781402091

34 Ou, F., Wang, H., Zhang, C., et al. Industrial data-driven machine learning soft sensing for optimal operation of etching tools. Digital Chemical Engineering, 13 (2024). https://doi.org/10.1016/j.dche.2024.100195

35 Terven, J., Cordova-Esparza, D., Ramirez-Pedraza, A., et al. Loss functions and metrics in deep learning (2023). https://doi.org/10.48550/arXiv.2307.02694

36 Le, P.B., and Nguyen, Z.T. ROC curves, loss functions, and distorted probabilities in binary classification. Mathematics, 10 (9) (2022). https://doi.org/10.3390/math10091410

37 Samigulina, G.A., Samigulina, Z.I., Bekeshev, D., and Butakova, D. Data-driven machinery faults detection techniques using artificial intelligence in Industry 4.0 concept. Procedia Computer Science, 257, 404–411 (2025). https://doi.org/10.1016/j.procs.2025.03.053

[1]Самигулина З.И.,
PhD, ORCID ID: 0000-0002-5862-6415,
e-mail: z.samigulina@kbtu.kz
[1]*Дюсенкулова Б.Ж.
докторант, ORCID ID: 0009-0001-2788-6521,
*e-mail: bu_dyussenkulova@kbtu.kz
[1]Бутакова Д.А.
магистрант, ORCID ID: 0009-0006-8151-4928,
e-mail: d.butakova@kbtu.kz

[1]Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

КҮШЕЙТУ АРҚЫЛЫ ОҚЫТУДЫ ПАЙДАЛАНА ОТЫРЫП, ЗИЯТКЕРЛІК ӨНДІРІСТІ БАСҚАРУ ӘДІСТЕРІ

Аңдатпа

Бүгінде ақылды өндіріс жүйелерін құру – өзекті міндет. Күрделі өндірістік мәселелерді шешуге мүмкіндік беретін нейрондық желілер кеңінен қолданылады. Мақала GEMMA моделінің бөлігі ретінде өнеркәсіптік жабдықтардың күйін анықтауға арналған DQN, PPO сияқты нейрондық желілер негізінде құрылған арматурамен оқыту алгоритмдеріне арналған. Француздық GEMMA моделі SFC (Sequential Function Charts) тіліне негізделген және процестерді басқару стандарттарын қамтиды. GEMMA моделінің D аймағына нейрондық желілерді енгізу ұсынылады. Модельдеу және эксперимент нәтижелері екі түрлі деректер базасы негізінде жүзеге асырылды: біреуі жасанды түрде жасалды, екіншісі өнеркәсіптік өндірістен алынды. Қарастырылған архитектураларды қолдану өнеркәсіптік деректермен жұмыс істеу үшін жақсы нәтижелерге қол жеткізуге мүмкіндік береді.

Түйін сөздер: ақылды өндірістік жүйе, нығайтумен оқыту, proximal policy optimization, deep Q-network, Guide d'Etude des Modes de Marche er d'Arret (GEMMA).

[1]Самигулина З.И.,
PhD, ORCID ID: 0000-0002-5862-6415,
e-mail: z.samigulina@kbtu.kz
[1]*Дюсенкулова Б.Ж.,
PhD докторант, ORCID ID: 0009-0001-2788-6521,
*e-mail: bu_dyussenkulova@kbtu.kz
[1]Бутакова Д.А.
магистрант, ORCID ID: 0009-0006-8151-4928,
e-mail: d.butakova@kbtu.kz

[1]Казахстанско-Британский технический университет, г. Алматы, Казахстан

МЕТОДЫ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫМ ПРОИЗВОДСТВОМ С ИСПОЛЬЗОВАНИЕМ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Аннотация

В настоящее время создание умных производственных систем является актуальной задачей. Широкое распространение получили нейронные сети, которые позволяют решать сложные производственные проблемы. Статья посвящена исследованию нейронных сетей с подкреплением DQN, PPO для диагностики состояния промышленного оборудования в рамках модели GEMMA. Французский подход GEMMA построен на основе языка SFC (Sequential Function Charts) и содержит стандарты по управлению технологическими процессами. Предлагается внедрение нейронных сетей в зону Д, модели GEMMA. Результаты моделирования и экспериментов осуществлялись на основе двух баз данных, одна сгенерирована искусственным путем, вторая взята с промышленного производства. Применение рассмотренных архитектур позволяет добиться хороших результатов для работы с индустриальными данными.

Ключевые слова: умная производственная система, обучение с подкреплением, proximal policy optimization, deep Q-network, Guide d'Etude des Modes de Marche er d'Arret (GEMMA) модель.