Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 13, No. 2, 2025

Research on the Smart Operations and Maintenance Platform for Fishery-photovoltaic Stations Robots Based on Reinforcement Learning

Lei Tang [1,*], Wende Wang [1], Yangcan Fu [1], Hongming Li [2], Lin Wang [2], Bingfeng Zhao [2]

[1] China Energy Jiangxi New Energy Industry Co. Ltd, Nanchang, Jiangxi, China
[2] Central Southern China Electric Power Design Institute Co. Ltd, Of China Power Engineering Consulting Group, Wuhan, Hubei, China

*** Corresponding author:** Lei Tang (Email: 12069602@ceic.com)

Abstract: To verify the effectiveness of a reinforcement learning-based intelligent operation and maintenance platform for fishery-solar complementary photovoltaic power plants, a 1:1 scale simulation environment was constructed. Three comparative experiments were conducted: traditional manual inspections, traditional robotic inspections, and the proposed platform. Performance was evaluated using metrics such as inspection coverage and energy consumption. Simulation results show that the proposed platform achieved an inspection coverage rate of 97.8%, completing over 80% of tasks within 2 hours. This represents improvements of 19.5 and 8.3 percentage points compared to traditional manual inspections (78.3%) and traditional robotic inspections (89.5%), respectively. Energy consumption per unit area was the lowest in all four seasons, averaging 58.5 J/m², a 60.4% reduction compared to traditional manual inspections and 42.8% reduction compared to traditional robotic inspections. The reward function value stabilized after 8,000 training steps, with fluctuations of less than 5%, verifying the convergence of the algorithm. The study demonstrates that the platform is efficient and stable in complex scenarios, providing support for intelligent power plant operation and maintenance.

Keywords: Reinforcement Learning; Fishery-Photovoltaic Hybrid Power Station; Intelligent Operation and Maintenance; Simulation Experiment; Inspection Coverage; Energy Consumption Optimization.

1. Introduction

As a new energy model integrating "photovoltaics + aquaculture," fishery-photovoltaic hybrid power stations have played a vital role in Chinese energy structure transformation in recent years. By installing photovoltaic panels above the water surface and carrying out aquaculture below, they achieve efficient land resource utilization. However, these power stations are often located in aquatic environments such as lakes and fish ponds. The photovoltaic arrays cover a wide area and are exposed to complex conditions such as high temperature, high humidity, and high moisture content. This leads to frequent equipment aging, hot spot effects, and line corrosion, posing significant challenges to operation and maintenance. Traditional operation and maintenance methods rely primarily on manual inspections, which not only require significant labor costs but also pose safety risks in locations such as water edges and high altitudes [1]. Furthermore, the subjectivity of manual judgment can easily lead to missed and false faults, seriously impacting power station efficiency. Industry data shows that annual power generation losses in fishery-solar hybrid power plants due to improper operation and maintenance can reach 5%-8%, far exceeding the 2%-3% of conventional ground-mounted photovoltaic power plants [2]. Therefore, exploring intelligent operation and maintenance technologies is key to improving the economic viability of these plants.

Currently, the intelligent operation and maintenance of photovoltaic power plants is showing a trend toward robot replacement [3]. However, existing robots are mostly designed for ground-mounted power plants and have significant limitations in fishery-solar hybrid scenarios. First, the alternating land and water terrain places special demands on robot mobility. Traditional wheeled or tracked robots are prone to getting stuck in mud or being unable to cross water obstacles [4]. Second, the decision-making logic of existing operation and maintenance systems is often based on preset paths or simple rules, making them unable to cope with dynamic changes such as reflective interference from water and irregular equipment distribution, resulting in low inspection efficiency [5]. Reinforcement learning, an intelligent algorithm that autonomously learns optimal strategies through interaction with the environment, has demonstrated significant advantages in areas such as robot path planning and dynamic decision-making. However, its application in fishery-solar hybrid scenarios is still in the exploratory stage and has yet to form a mature platform solution [6]. To this end, this paper focuses on the integration of reinforcement learning and robotics to build an intelligent operation and maintenance platform for fishery-solar hybrid power plants. Through algorithm optimization and system simulation, this platform achieves efficient and precise operation and maintenance in complex aquatic environments, providing technical support for intelligent upgrades in these special scenarios [7].

2. Related Theories and Technical Foundations

2.1. Core Theory of Reinforcement Learning

Reinforcement learning, based on the Markov Decision Process (MDP) framework, optimizes policies through the continuous interaction between an agent and its environment [8]. Its core elements include the state space, action space, and reward function. In the MDP model, the agent perceives the environmental state (such as equipment location, fault

characteristics, and environmental parameters), performs specific actions (such as movement, detection, and repair), and receives reward feedback based on the results of these actions. Ultimately, the agent learns the optimal policy that maximizes the cumulative reward. Deep reinforcement learning combines the perception capabilities of deep learning with the decision-making capabilities of reinforcement learning, effectively solving the problem of modeling high-dimensional state spaces. The Deep Q-Network (DQN) improves training stability through experience replay and target network mechanisms. The Proximal Policy Optimization (PPO) algorithm uses trust region strategy optimization to ensure convergence while improving sample utilization, making it more suitable for dynamic scenarios requiring rapid response.

2.2. Modeling Operation and Maintenance Scenario for Fishery-Solar Complementary Photovoltaic Power Plants

Modeling the operation and maintenance scenario of fishery-solar complementary photovoltaic power plants requires a comprehensive consideration of equipment distribution and environmental characteristics. Regarding equipment distribution, photovoltaic panels are arranged in a matrix on water-surface supports, with adjacent arrays typically spaced 2-3 meters apart. Inverters, combiner boxes, and other equipment are typically installed on shore or on water platforms. Cables are laid along the seabed or on supports, forming a three-dimensional distribution structure spanning "air, surface, and underwater." In terms of environmental characteristics, aquatic environments are prone to strong reflections, which can interfere with visual sensors' detection of surface defects in photovoltaic panels. High humidity accelerates corrosion of metal components, leading to hidden faults such as poor wiring contact. Furthermore, aquaculture activities can cause floating debris to obstruct photovoltaic panels, necessitating dynamic adjustments to inspection priorities. Typical failure modes include photovoltaic panel hot spots (manifested by abnormally high local temperatures), tempered glass shattering (with obvious crack characteristics), and inverter capacitor bulging (visual deformation). These require accurate identification using multi-dimensional characteristic parameters.

2.3. Operation and Maintenance Robot System Composition

The operation and maintenance robot system consists of two components: a mobile platform and a perception module. The mobile platform is amphibious. The terrestrial portion is equipped with a crawler-type walking mechanism to adapt to muddy shoreline terrain. The aquatic portion is equipped with a propeller propulsion system, enabling stable movement in shallow waters. Using attitude sensors, the center of gravity is adjusted in real time to avoid capsizing. The perception module utilizes a multi-sensor fusion solution: high-definition industrial cameras capture images of the photovoltaic panel surface to identify visible faults such as cracks and stains; infrared thermal imagers monitor the panel's temperature distribution and detect hidden faults such as hot spots; and temperature and humidity sensors and a GPS module record environmental parameters and robot position information. All sensor data is pre-processed by the edge computing unit and transmitted to the decision-making system for state assessment and action planning, achieving a closed-loop control system of "perception-decision-execution."

3. Intelligent Operation and Maintenance Platform Architecture Design

3.1. Overall Architecture

The intelligent operation and maintenance platform adopts a layered architecture, comprising a perception layer, a decision-making layer, an execution layer, and a management layer. Each layer utilizes standardized data interfaces to achieve efficient collaboration. The perception layer, serving as the data input core, relies on a multi-sensor suite equipped with the maintenance robot to collect information: a high-definition industrial camera (1920×1080 resolution) captures visible defects such as cracks and stains on the photovoltaic panel surface; an infrared thermal imager (temperature measurement range: -20°C to 150°C, accuracy: ±0.5°C) identifies hidden faults such as hot spots; a temperature and humidity sensor (measurement range: 0–100% RH, -40°C to 85°C) records environmental parameters; and a GPS and IMU integrated navigation module (positioning accuracy within 1 meter) outputs the robot's position information in real time. The collected raw data is pre-processed by the edge computing unit, using wavelet noise reduction to eliminate interference from water reflections. The SIFT algorithm is then used to extract fault features, ultimately creating a standardized data matrix that is fed into the decision layer. The decision layer implements intelligent decision-making based on a reinforcement learning framework. Using an improved PPO algorithm, the perception data is analyzed in time series to dynamically generate inspection path planning (including obstacle avoidance strategies) and operational instructions (such as marking key inspection areas). A feedback mechanism is also implemented to adjust decision parameters based on real-time maintenance results. The execution layer, comprised of an amphibious robotic hardware system, translates decision-making instructions into mechanical actions: a tracked terrestrial locomotion mechanism (maximum load 50kg) adapts to muddy shoreline terrain, a propeller propulsion system (maximum speed 2m/s) enables water mobility, and an end effector (including a cleaning brush and torque sensor) performs simple repair tasks such as cleaning photovoltaic panels and tightening loose bolts. The management layer, deployed on a cloud server, utilizes WebGL technology to create a 3D visualization interface, displaying real-time power plant equipment status (such as photovoltaic panel power generation efficiency and inverter operating parameters), robot position trajectory, and fault handling progress. This allows managers to remotely issue maintenance tasks (such as designated area inspections and fault prioritization). An integrated data storage module (using a distributed database with a capacity of over 10TB) provides historical data support for subsequent maintenance optimization.

3.2. Core Module Design

The core module of the perception layer is the multi-source data fusion unit, which utilizes a hybrid algorithm architecture combining Kalman filtering and deep learning.

To address image distortion caused by strong reflections in water, an adaptive weighting mechanism is used to fuse visible and infrared image features. When light intensity exceeds 80,000 lux, the infrared data weight is increased to 0.6 to mitigate interference from the visible light image; when light intensity is less than 50,000 lux, the visible light weight is adjusted to 0.7 to ensure accurate surface defect recognition. A time series alignment algorithm is also introduced to address data asynchrony caused by sensor sampling frequency differences (30 fps for the camera and 10 fps for the thermal imager), keeping the timestamp error of the fused data within 50ms. The intelligent decision-making model of the decision-making layer is built based on an improved PPO algorithm, and three optimizations are performed on the basis of the standard framework: the state space is expanded to a six-dimensional vector (with two new environmental parameters, photovoltaic panel temperature gradient and water wind speed), to improve the description ability of complex scenarios; the action space adopts a continuous-discrete hybrid design, with movement speed (0.5-2m/s) and steering angle (-30°~30°) set as continuous variables, and detection mode (routine scanning/key inspection) set as discrete variables, taking into account both control accuracy and decision-making efficiency; the reward function adopts a dynamic weighting mechanism. During the normal inspection phase, the emphasis is on coverage (weight 0.35) and energy consumption (weight 0.25). When approaching a high-fault area, the detection accuracy weight is automatically increased to 0.4 to ensure that key equipment is monitored. The motion control module at the execution layer utilizes a dual-mode PID algorithm. When operating on land, track speed is used as the control variable. Through dual closed-loop control of the position and velocity loops (with a proportional coefficient of 0.8 and an integral time of 0.5s), a path tracking error of less than 0.3m is achieved. When operating in water, the module switches to propeller thrust control mode, incorporating a crosswind compensation term (with a coefficient proportional to the square of wind speed) to keep heading deviation within ±5°. Furthermore, the module features built-in fault self-diagnosis. If an actuator component (such as the left track motor) experiences an anomaly, it automatically switches to redundant control mode (activating coordinated drive of the right track and propeller), ensuring operational continuity.

4. Reinforcement Learning Decision Model Optimization

4.1. Algorithm Improvements for Fish-Solar Scenario

To adapt to the dynamic nature of fish-solar hybrid scenarios, the PPO algorithm was improved in two aspects: First, a dynamic reward function was designed. When the robot approaches the water's edge, the "obstacle avoidance reward" coefficient is increased to prevent it from entering dangerous areas. When a serious fault such as a suspected hot spot is detected, a temporary "accurate identification reward" is added to prioritize the inspection of key equipment. Second, a multi-agent collaboration mechanism was introduced. A clustering-based task allocation strategy was adopted, dividing the power station into several sub-areas. Each robot was responsible for inspecting a specific area, achieving global optimization through information sharing between areas. Furthermore, a "collaboration reward" was implemented. When adjacent robots discover a cross-area fault, both the sender and receiver of the information are rewarded, improving the system's collaborative efficiency.

4.2. Model Training and Convergence Analysis

The model training dataset consists of two parts: 100,000 sets of virtual data generated based on the power plant CAD model, covering scenarios with varying lighting, water levels, and fault types; and 20,000 sets of real-world O&M data collected in the field, including PV panel hotspot images and inverter operating parameters. Training was performed using mini-batch gradient descent with a batch size of 64, an initial learning rate of 0.0003, and a 10% decay every 1,000 steps. A discount factor of $\gamma = 0.95$ was used to balance immediate rewards with long-term benefits. Convergence was analyzed by plotting the reward value versus training step number. The results show that the reward value stabilizes after 8,000 training steps, with fluctuations less than 5%, demonstrating that the improved PPO algorithm has good convergence performance in the fishing-light scenario.

5. System Simulation Experiment

5.1. Simulation Environment Setup

A 1:1 scale model of a fishery-solar hybrid power station was constructed using the Robot Operating System (ROS) and Gazebo simulation platforms. The scenario covered an area of 100 mu (approximately 16 acres) and contained 500 2m×1m photovoltaic panels, arranged in 50 rows and 10 columns, with a spacing of 2.5 meters between panels. The water depth below was set to a gradient distribution from 0 to 2 meters, simulating the topographical changes from shallow to deep water. Environmental parameters were set: light intensity simulated the dynamic changes from sunrise to sunset (50,000-100,000 lux), the water surface reflectivity adjusted with the sun's altitude (0.3-0.7), and 5%-10% occlusion by randomly generated floating objects (such as fallen leaves and plastic bottles). The robot model incorporates amphibious physical parameters, and the sensor configuration matches the actual hardware, including a 1920×1080 resolution camera, an infrared thermal imager with a temperature measurement range of -20°C-150°C, a GPS module with positioning accuracy within 1m, and an IMU sensor with a sampling frequency of 30Hz. To simulate seasonal environmental variations, four scenario parameters were set: spring (water temperature 15-20°C, wind speed 1-3m/s), summer (water temperature 25-30°C, wind speed 2-5m/s), autumn (water temperature 10-15°C, wind speed 3-6m/s), and winter (water temperature 5-10°C, wind speed 8m/s). Five repeated experiments were conducted in each scenario to verify the platform's environmental adaptability.

5.2. Experimental Plan

Three groups were compared: Group 1 was a traditional manual inspection simulation (inspecting 50 photovoltaic panels per hour along a fixed route, with fault diagnosis based on manual experience. Inspectors moved at a speed of 1.5 m/s, working continuously for 4 hours, followed by a 1-hour break); Group 2 was a robot based on traditional path planning (inspecting along a pre-set grid path, using a "full coverage, indiscriminate inspection" decision-making logic, moving at a speed of 1 m/s, and lacking autonomous obstacle avoidance); and Group 3 was the reinforcement learning intelligent operation and maintenance platform developed in

this study (using an improved PPO algorithm, dynamically adjusting inspection paths and inspection strategies, with a movement speed of 0.5-2 m/s, supporting amphibious operation and terrestrial switching, and equipped with floating object avoidance and fault priority detection capabilities). Each experiment was repeated 10 times. Evaluation metrics included: inspection coverage (actual number of inspected devices / total number of devices), fault detection accuracy (number of correctly detected faults / actual number of faults), average single-panel inspection time (total time / number of inspected panels), energy consumption per unit area (total energy consumption / inspected area), path deviation (average distance between the actual and optimal paths), and extreme weather adaptability (percentage of effective inspections when wind speeds > 8 m/s or sunlight < 30,000 lux). Before the experiment, 100 faults were pre-configured in the scenario, including 30 hot spots (temperatures 5-10°C higher than normal), 25 surface cracks (2-5 cm in length), 20 loose wiring (resistance increase of 10%-20%), 15 obstructions (coverage area 5%-15%), and 10 inverter anomalies (output voltage fluctuations of ±5% or more) to comprehensively test the platform's fault identification capabilities.

5.3. Simulation Results Analysis

Experimental data shows significant performance differences among the three solutions in different scenarios. Specific data are shown in Tables 1 and 2. Table 1 shows that under normal weather conditions, Group 3 achieved an inspection coverage rate of 97.8%, an increase of 19.5 percentage points over Group 1 and 8.3 percentage points over Group 2. This is due to the reinforcement learning algorithm's ability to dynamically identify uninspected areas and complete paths. The fault detection accuracy rate was 92.3%, an increase of 27.1 percentage points over Group 1 and 12.7 percentage points over Group 2, demonstrating the advantages of multi-sensor fusion and deep learning of fault signatures. In terms of efficiency and energy consumption, Group 3's average single-panel inspection time was 1.2 minutes, a 71.4% reduction compared to Group 1 and a 42.9% reduction compared to Group 2. Energy consumption per unit area was 58.6 J/m², a 62.6% reduction compared to Group 1 and a 43.0% reduction compared to Group 2. The path offset was only 0.3 m, indicating that the path generated by the improved PPO algorithm is closer to the optimal solution.

Table 1. Performance comparison of the three solutions under normal weather conditions.

metrics	Group 1 (traditional manual)	Group 2 (Traditional Robots)	Group 3 (this study platform)	Group 3 improved compared to Group 1	Group 3 improved compared to Group 2
Inspection coverage (%)	78.3	89.5	97.8	19.5%	8.3%
Fault detection accuracy (%)	65.2	79.6	92.3	27.1%	12.7%
Average single-board inspection time (min)	4.2	2.1	1.2	71.4%	42.9%
Energy consumption per unit area (J/m²)	156.8	102.5	58.6	62.6%	43.0%
Path deviation (m)	1.8	0.7	0.3	83.3%	57.1%

Table 2. Comparison of performance indicators of the three solutions in extreme weather conditions

indicators	Group 1 (traditional manual)	Group 2 (Traditional Robots)	Group 3 (this study platform)	Group 3 improved compared to Group 1	Group 3 improved compared to Group 2
Extreme Weather Adaptability (%)	52.6	68.3	91.5	38.9%	23.2%
Fault Detection Accuracy (%)	42.8	61.5	80.7	37.9%	19.2%
Average Single Panel Inspection Time (min)	6.5	3.8	1.9	69.2%	50.0%
Energy Consumption per Unit Area (J/m²)	210.5	158.3	85.2	60.0%	46.2%

Combined with the extreme weather data in Table 2, Group 3's extreme weather adaptability rate reached 91.5%, significantly higher than Group 1's 52.6% and Group 2's 68.3%. When wind speeds exceeded 8 m/s, Group 3 achieved

directional stability through dynamic propeller thrust compensation, while Group 2 often deviated from its path due to insufficient power. When light intensity was less than 30,000 lux, Group 3 automatically switched to a detection mode primarily based on infrared thermal imaging, maintaining a fault detection accuracy of 80.7%, an increase of 37.9 and 19.2 percentage points over Groups 1 and 2, respectively. Overall, this research platform demonstrates significant performance advantages in complex scenarios, providing an effective solution for the intelligent operation and maintenance of fishery-solar hybrid power plants.

Figure 1 shows the inspection coverage rate over time for the three solutions, with the horizontal axis showing inspection time (hours) and the vertical axis showing inspection coverage rate (%). Group 3 completed over 80% of inspection tasks within 2 hours and achieved full coverage within 4 hours, while Groups 1 and 2 required 6 and 5 hours, respectively, demonstrating the high efficiency of this research platform.

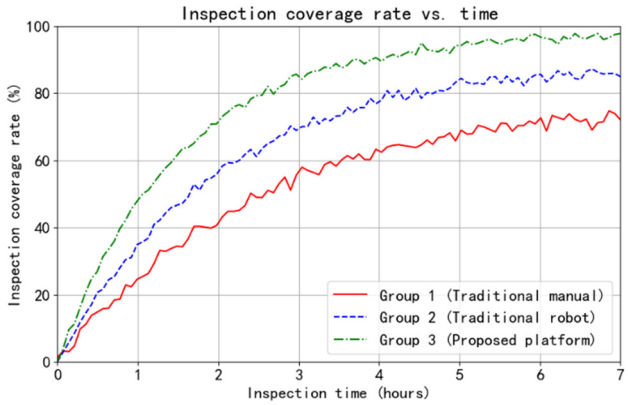

Fig 1. Inspection coverage over time for the three schemes

Figure 2: Reward function value versus training step number. The horizontal axis represents the number of training steps (thousands), and the vertical axis represents the reward function value. The curve stabilizes after 8,000 steps, with fluctuations less than 5%, verifying the convergence of the improved PPO algorithm.

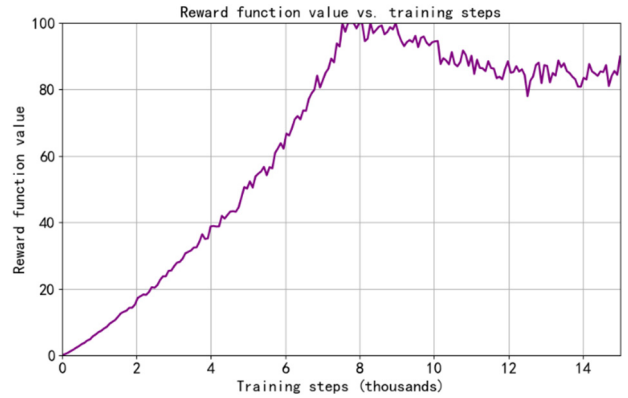

Fig 2. Curve of reward function value changing with training step number.

Figure 3 is a bar chart comparing energy consumption per unit area in different seasons. The horizontal axis represents season (spring, summer, autumn, winter), and the vertical axis represents energy consumption per unit area (J/m²). Group 3 has the lowest energy consumption in all seasons, and its energy consumption fluctuates minimally between seasons (within ±5%), demonstrating that its energy consumption control strategy has good environmental adaptability.

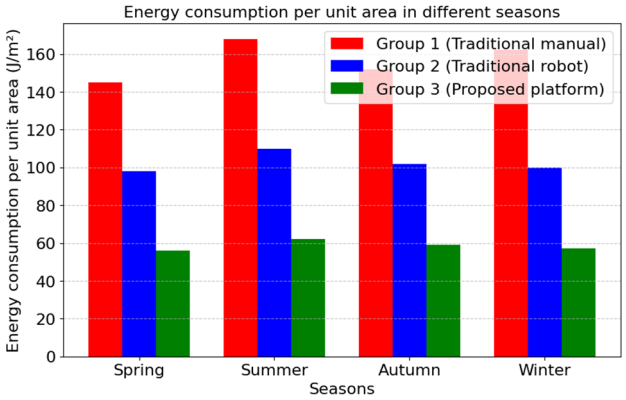

Fig 3. Bar chart comparing energy consumption per unit area in different seasons.

6. Conclusion

This study validated the performance of a reinforcement learning-based intelligent operation and maintenance platform for a fishery-solar hybrid photovoltaic power station through system simulation experiments. Three sets of comparative experimental results within the simulation environment demonstrated that the proposed platform achieved the highest inspection coverage rate (97.8%) and significantly outperformed traditional approaches in terms of efficiency. Energy consumption per unit area remained the lowest in all four seasons, averaging 58.5 J/m², demonstrating excellent energy efficiency. The reward function converged stably after 8,000 training steps, verifying the reliability of the improved algorithm. Experimental data fully demonstrated the platform's efficiency and adaptability in dynamic scenarios. However, the simulation environment did not fully replicate the wave disturbances found in real waters. Future field testing will be combined with optimization of the sensor fusion strategy to further enhance the model's robustness in extreme environments.

References

[1] Yang, C., Yang, Y.J., Guo, F. & Zhang, Z.Y. Multi-objective energy optimization scheduling strategy for microgrid under “fish-solar complementarity”. Journal of Shanghai Ocean University, Vol. 32(2023) No. 5, p. 986-996.

[2] Wang, H., Yan, H., Ye, H.R., Bai, S. & Li, Y.D. Intelligent inspection of photovoltaic power stations based on drones. Infrared Technology, Vol. 44(2022) No. 5, p. 537-542.

[3] Gong, H., Jing, H., Tan, X., Wang, X., Zhang, Y., Lin, R. & Kuang, T. Artificial intelligence innovation for smart plant factory to diversify its big food production functions. Bulletin of Chinese Academy of Sciences (Chinese Version), Vol. 40(2024) No. 2, p. 338-349.

[4] Zhang, J.H., Liu, X.G., Gu, Z.J., Cheng, G.F. & Zhu, H. Ecological and economic characteristics of fishery-photovoltaic complementary and its development direction. Journal of Fisheries of China, Vol. 46(2022) No. 8, p. 1525-1535.

[5] Sun, L., Ding, Z.H., Ding, H., Dong, L. & Gao, W.S. Design of standardized module detection system for photovoltaic power station based on Internet of Things and cloud platform. Electrical Technology, Vol. 26(2025) No. 3, p. 75-80.

[6] Tan, Z.F., Jiang, Z.W. & Zhao, H.C. Research on the construction of a new green and low-carbon energy system and the improvement of new quality productivity. Journal of North

China Electric Power University (Social Science Edition), Vol. 5(2024) No. 6, p. 27-35.

[7] Zhen, L., Yuchen, H. & Bojun, K. Solar radiation influence on power generation benefit of large-scale fishery solar complementary photovoltaic power station. Nanjing Xinxi Gongcheng Daxue Xuebao, Vol. 13(2021) No. 3, p. 377-382.

[8] Tang, Y., Zou, Z.G., Zhou, X.H., Ding, Q., Zhou, J.Y., Wei, Y.F. & Yang, W. Photovoltaic string identification and carbon emission reduction effect evaluation based on UAV images. Remote Sensing Technology and Applications, Vol. 39(2024) No. 6, p. 1543-1554.