Düzce University
Journal of Science & Technology

Research Article

Towards Transparent Control Systems: The Role of Explainable AI in Iterative Learning Control

Mustafa KUTLU [a], Mohammed MANSOUR [a, *]

[a] *Department of Mechatronics Engineering, Faculty of Technology, Sakarya Applied Sciences University, Sakarya, TURKEY*
** Corresponding author's e-mail address: mohammedmansour@subu.edu.tr*
DOI: 10.29130/dubited.1535271

ABSTRACT

This paper presents a novel approach to improving the performance and interpretability of Iterative Learning Control (ILC) systems through the integration of Explainable Artificial Intelligence (XAI) techniques. ILC is a powerful method used across various domains, including robotics, process control, and traffic management, where it iteratively refines control inputs based on past performance to minimize errors in system output. However, traditional ILC methods often operate as "black boxes," making it difficult for users to understand the decision-making process. To address this challenge, we incorporate XAI, specifically SHapley Additive exPlanations (SHAP), into the ILC framework to provide transparent and interpretable insights into the algorithm's behavior. The study begins by detailing the evolution of ILC, highlighting key advancements such as predictive optimal control and adaptive schemes, and then transitions into the methodology for integrating XAI into ILC. The integrated system was evaluated through extensive simulations, focusing on robotic arm trajectory tracking and traffic flow management scenarios. Results indicate that the XAI-enhanced ILC not only achieved rapid convergence and high control accuracy but also maintained robustness in the face of external disturbances. SHAP analyses revealed that parameters such as the proportional gain (Kp) and derivative gain (Kd) were critical in driving system performance, with detailed visualizations providing actionable insights for system refinement. A crucial metric for control precision was the root mean square error (RMSE), which was reduced to as low as 0.02 radians in the robotic arm case, indicating extremely precise tracking of the intended route. Similarly, the ILC algorithm effectively maintained the ideal traffic density within the predetermined bounds in the traffic management scenario, resulting in a 40% reduction in congestion compared to baseline control measures. The resilience of the ILC algorithm was also examined by introducing changes to the system model, external disturbances, and sensor noise. The algorithm demonstrated a high degree of stability and accuracy in the face of these disruptions. For instance, in the robotic arm case, adding noise to the sensor readings had a negligible effect on the algorithm's performance, increasing the RMSE by less than 5%. This integration of XAI into ILC addresses a significant gap in control system design by offering both high performance and transparency, particularly in safety critical applications. The findings suggest that future research could further enhance this approach by exploring additional XAI techniques and applying the integrated system to more complex, real-world scenarios.

Keywords: *Control System, Iterative Learning, AI, Explainable AI*

Received: 30/08/2024, Revised: 17/09/2024, Accepted: 23/09/2024

Şeffaf Kontrol Sistemlerine Doğru: Tekrarlı Öğrenme Kontrolünde Açıklanabilir Yapay Zekanın Rolü

Öz

Bu makale, Açıklanabilir Yapay Zeka (XAI) tekniklerinin entegrasyonu yoluyla Tekrarlı Öğrenme Kontrolü (ILC) sistemlerinin performansını ve yorumlanabilirliğini iyileştirmek için yeni bir yaklaşım sunmaktadır. ILC, robotik, süreç kontrolü ve trafik yönetimi dahil olmak üzere çeşitli alanlarda kullanılan güçlü bir yöntemdir ve burada sistem çıktısındaki hataları en aza indirmek için geçmiş performansa dayalı olarak kontrol girdilerini tekrarlı olarak iyileştirir. Ancak, geleneksel ILC yöntemleri genellikle "kara kutular" olarak çalışır ve kullanıcıların karar alma sürecini anlamasını zorlaştırır. Bu zorluğun üstesinden gelmek için, algoritmanın davranışına ilişkin şeffaf ve yorumlanabilir içgörüler sağlamak üzere XAI'yi, özellikle SHapley Eklemeli Açıklamaları (SHAP) ILC çerçevesine dahil ediyoruz. Çalışma, ILC'nin evrimini ayrıntılı olarak açıklayarak, öngörücü optimal kontrol ve uyarlanabilir şemalar gibi önemli gelişmeleri vurgulayarak başlıyor ve ardından XAI'yi ILC'ye entegre etme metodolojisine geçiyor. Entegre sistem, robotik kol yörünge takibi ve trafik akışı yönetimi senaryolarına odaklanarak kapsamlı simülasyonlar yoluyla değerlendirildi. Sonuçlar, XAI ile geliştirilmiş ILC'nin yalnızca hızlı yakınsama ve yüksek kontrol doğruluğu elde etmekle kalmayıp aynı zamanda harici bozulmalar karşısında sağlamlığını da koruduğunu göstermektedir. SHAP analizleri, orantılı kazanç (Kp) ve türev kazancı (Kd) gibi parametrelerin sistem performansını yönlendirmede kritik olduğunu ve detaylı görselleştirmelerin sistem iyileştirmesi için eyleme geçirilebilir içgörüler sağladığını ortaya koymuştur. Kontrol hassasiyeti için kritik bir istatistik, kök ortalama kare hatasıydı (RMSE). RMSE, robotik kol durumunda 0,02 radyana kadar düşürüldü ve bu, amaçlanan rotanın son derece hassas bir şekilde izlendiğini göstermektedir. Karşılaştırıldığında, ILC algoritması, trafik yönetimi senaryosunda ideal trafik yoğunluğunu önceden belirlenmiş sınırlar içinde etkili bir şekilde korudu ve bunun sonucunda temel kontrol önlemleriyle karşılaştırıldığında tıkanıklıkta %40'lık bir azalma sağlandı. Sistem modeline değişiklikler, dış bozulmalar ve sensör gürültüsü eklenerek ILC algoritmasının dayanıklılığı incelendi. Algoritma, bu bozulmalar karşısında yüksek derecede kararlılık ve doğruluk gösterdi. Örneğin, robotik kol durumunda, sensör okumalarına gürültü eklemek algoritmanın performansı üzerinde ihmal edilebilir bir etkiye sahipti ve RMSE'yi %5'ten daha az artırdı. XAI'nin ILC'ye bu şekilde entegre edilmesi, özellikle güvenlik açısından kritik uygulamalarda hem yüksek performans hem de şeffaflık sunarak kontrol sistemi tasarımındaki önemli bir boşluğu giderir. Bulgular, gelecekteki araştırmaların ek XAI tekniklerini araştırarak ve entegre sistemi daha karmaşık, gerçek dünya senaryolarına uygulayarak bu yaklaşımı daha da geliştirebileceğini göstermektedir.

Anahtar Kelimeler: *Kontrol Sistemi, Tekrarlı Öğrenme, Yapay Zeka, Açıklanabilir Yapay Zeka*

I. INTRODUCTION

Iterative Learning Control (ILC) has emerged as a significant method in the realm of control systems, offering a systematic approach to improving system performance through the repetitive execution of tasks [1]. The foundational principle of ILC lies in its ability to learn from past iterations and adjust control inputs accordingly, thereby progressively minimizing errors and enhancing overall system efficiency. This approach is particularly effective in scenarios where tasks are repetitive, and the desired output trajectory is predefined, making ILC a powerful tool in various domains, including robotics, process control, and traffic management [2].

The initial development of ILC can be traced back to the late 1980s, with the pioneering work of Oh, Bien, and Suh [2], who introduced an ILC method designed specifically for robot manipulators. Their research demonstrated the algorithm's ability to achieve convergence under specific conditions, particularly in systems experiencing minor perturbations from a nominal trajectory. This groundbreaking work laid the foundation for subsequent advancements in the field, sparking a wave of research focused on refining and extending the basic ILC framework to address increasingly complex control challenges. As the field of ILC evolved, researchers sought to enhance the robustness and adaptability of ILC algorithms. Lee and Bien proposed "iterative learning control with multi-modal input," a significant advancement that improved the algorithm's ability to handle variable initial conditions [3]. This method synthesized control inputs based on the initial condition state, thereby enhancing convergence properties and expanding the applicability of ILC to a broader range of scenarios. Such developments underscored the importance of adaptability and robustness in modern control systems. Further innovations in ILC were introduced by Amann et al. [4], who developed a predictive optimal ILC algorithm that integrated present and future predicted errors to calculate the current control input. This approach, akin to model-based predictive control, demonstrated significant improvements in control performance, particularly in dynamic environments where anticipating future states is crucial for achieving accurate and stable outcomes. These predictive aspects expanded the potential applications of ILC, making it a more versatile and powerful tool in various domains.

The integration of Explainable Artificial Intelligence (XAI) into control systems has broad significance across diverse fields such as power systems, air-traffic management, healthcare, IoT, and prosthetic technologies. By enhancing transparency and human understanding, XAI improves decision-making, system optimization, and user interaction, leading to more effective and user-friendly control solutions. XAI in control systems has been a topic of interest in various fields, including power systems, air-traffic management, medical applications, human-machine interfaces, IoT systems, and visual quality control. Zhang et. al., introduced the use of the SHAP method in deep reinforcement learning models for power system emergency control, providing clear explanations for under-voltage load shedding [5]. Xie et. al., utilized the XGBoost library explanations in air traffic management decision support systems to enhance human understanding and analysis [6]. Sheu et. al., conducted a survey on medical XAI, highlighting model enhancements, evaluation methods, and future improvements in healthcare explainability [7]. Kang et. al., proposed an XAI approach to optimize sensor disposition in EMG-IMU multimodal fusion systems for prosthetic hand control, aiming to reduce system redundancies and improve patient quality of life [8]. Dobrovolskis et. al., 2023 developed an explainable rule-based smart home system for IoT applications, emphasizing the importance of XAI in user-friendly systems [9]. Maxwell et. al., 2023 discussed the significance of user centric design methodology in developing meaningful XAI solutions for various operational contexts, including human-on-the-loop control and ex-post investigations [10].

The integration of XAI into ILC represents a pivotal advancement in the field, particularly in enhancing the interpretability and transparency of control systems. XAI refers to a set of methods and techniques designed to make the decision-making processes of AI systems more understandable to humans. In the context of ILC, XAI plays a crucial role in providing insights into how control decisions are made, enabling operators and engineers to better understand the behavior of the ILC algorithms and the rationale behind their adjustments. Control systems are improved by integrating XAI with ILC to make them more visible, flexible, and effective. Interpretability is brought to ILC by XAI, which promotes confidence and helps users comprehend the decision-making process especially in vital applications like industrial automation and autonomous systems. This openness helps with diagnostics, accelerating the tuning of the ILC system and simplifying the identification and resolution of performance problems. Additionally, by making the learning process intelligible and permitting human inspection and intervention, XAI facilitates improved collaboration between human operators and machines. By elucidating the steps done by the ILC system, XAI guarantees accountability in safety-critical applications, which is essential for adhering to safety and regulatory regulations. Furthermore, by emphasizing transferable learning components, XAI promotes applications in novel situations and aids in the generalization of ILC across tasks. Through optimization of the learning process, this integration speeds up convergence and enhances system performance, allowing the control system to adjust to dynamic changes or customized requirements. In the end, XAI improves ILC's responsiveness, interpretability, safety, and efficiency across a variety of applications.

The importance of XAI in ILC becomes evident when considering the complexity of modern control systems, where the interactions between various components can be highly intricate. Traditional ILC methods, while effective, often operate as "black boxes," making it difficult for users to discern the underlying processes that lead to specific control decisions. By integrating XAI techniques, such as model interpretability and explainability frameworks, the decision-making process within ILC can be made more transparent, thereby increasing trust and reliability in these systems [11]. XAI's role in ILC is particularly relevant in applications where safety and reliability are paramount, such as in autonomous vehicles, industrial automation, and healthcare robotics. For example, in the domain of robotics, where ILC is widely used for trajectory tracking and precision control, XAI can help elucidate why certain control inputs are chosen over others, especially in situations where the system deviates from expected behavior. This transparency not only aids in troubleshooting and refining control algorithms but also enhances user confidence in the system's operations. Moreover, the integration of XAI into ILC aligns with the broader trend in AI research toward developing systems that are not only powerful but also interpretable and accountable. As highlighted by [12], incorporating insights from disciplines such as philosophy, psychology, and cognitive science can significantly enhance the effectiveness of XAI, leading to more human-centered and user-friendly AI systems. In the context of ILC, this approach facilitates a deeper understanding of how control decisions are made, thereby improving the overall efficiency and safety of the system.

This increasing focus is a reflection of the demand for AI systems that, particularly in crucial real applications, not only offer optimal performance but also transparency in their decision-making processes. Krajna et al. (2022) underlined in their study the concrete advantages of implementing XAI in real world contexts, specifically stressing how explainability added to AI systems can greatly improve user comprehension, confidence, and adoption rates [13]. These practical uses show that explainability is an important component of making AI technology more widely applicable and influential, not just an academic endeavor. The study emphasizes the importance of explainability in light of the growing use of AI systems in delicate and risky situations. Furthermore, Bacco, Luca, et al. has research that the use of XAI for natural language processing tasks is becoming more popular, as seen by the latest developments in extractive summarizing approaches, especially for sentiment analysis [14]. This research underscores the need of transparency in AI-driven text interpretation, which is critical for domains like market analysis, social media monitoring, and customer feedback systems. It does this by leveraging XAI to provide insights into how AI models arrive at sentiment analysis results. An AI's value and reliability can be increased by providing an explicable reasoning for its sentiment classification.

The idea to assess explainable Machine Learning (ML) models using an application grounded evaluation framework is another significant advancement in XAI research. This concept is particularly relevant in the clinical arena, where the adoption of AI depends on its capacity to yield understandable, practical results for practitioners, as proposed by [15]. XAI models can enhance patient outcomes by building trust in the technology while also assisting doctors in making well-informed decisions by integrating explainability into clinical AI systems. Adaptive control systems, particularly robotics related ones, show great promise when XAI is combined with a parallel ILC architecture. The study by Chotikunnan et al. serves as an example of how XAI can be integrated to improve robotic systems' capacity for learning and adaptability [16]. According to this research, XAI may play a significant role in enhancing control systems, which would enable robots to more effectively explain their actions and modifications in dynamic situations. This would enhance safety, effectiveness, and human robot cooperation. Furthermore, the potential of XAI in tackling high-stakes, multidisciplinary challenges is demonstrated by its application to complex geospatial problems, such as earthquake spatial probability and hazard estimation in the Arabian Peninsula [17]. Researchers are better able to explain the logic behind hazard projections when they use explainable AI techniques in environmental modeling, which increases the usefulness of the findings for emergency managers and policy makers. This increases AI's usefulness in catastrophe management, as explainability can mean the difference between taking preventative action and responding insufficiently.

Considering the moral questions raised by bias and justice in automated decision-making systems, this is especially pertinent. Explainability in hiring algorithms can guarantee openness, lessen prejudices, and promote a fairer procedure [18]. XAI's promise in the life sciences is further expanded by the increasing interest in explainable multi-task learning for multi-modality biological data, as demonstrated by recent study [19]. In this context, XAI is viewed as a crucial element for deciphering the intricacy of biological systems, providing scholars and professionals with enhanced comprehension of how AI models handle and comprehend diverse data kinds. This research has important ramifications for personalized medicine, because patient trust and adherence to treatment recommendations are largely dependent on comprehending the reasoning behind AI-driven diagnostic or therapeutic suggestions. A forward-looking view for the nexus between AI and environmental stewardship is also presented by the incorporation of XAI into conservation initiatives. According to Hessami, Mateen A., et al., there is a growing need to modernize conservation models in order to take advantage of XAI and ILC systems in a more comprehensive manner [20]. A symbiotic relationship between AI technologies and environmental preservation is created when conservationists can better understand and optimize their methods by utilizing explainability in AI models used for ecosystem management and animal monitoring.

The literature on ILC and XAI reflects a dynamic and evolving field, with significant contributions from researchers who have sought to optimize control algorithms while also making them more interpretable. For instance, recent advancements in fractional-order ILC for fractional-order systems [21] and neural network-based ILC for nonlinear systems [22] illustrate the potential of integrating AI techniques to enhance both performance and interpretability in complex control systems. These developments highlight the ongoing efforts to bridge the gap between powerful control algorithms and the need for transparency in their operations. The integration of XAI into ILC represents a significant step forward in the field of control systems. By making the decision making processes within ILC more transparent and interpretable, XAI enhances the usability, safety, and reliability of these systems across various domains. The ongoing research in this area suggests that future directions may focus on further optimizing ILC algorithms for specific applications, while simultaneously enhancing their interpretability through advanced XAI techniques. Figure 1 illustrates the fundamental difference between Traditional Artificial Intelligence (AI) and XAI in the context of industrial robotics, specifically focusing on a pick-and-place task. In the top half of the figure, Traditional AI is depicted as a process that transforms training data into a learned function through a ML process. While this approach can yield effective decision-making capabilities, it often operates as a "black box," leaving users with unanswered questions about the reasoning behind specific decisions. This lack of transparency can lead to mistrust, particularly in safety-critical applications where understanding the rationale behind AI decisions is crucial. In contrast, the bottom half of the figure demonstrates how XAI enhances the AI process by making the decision-making process transparent and interpretable. By incorporating XAI techniques, users can gain insights into why the AI made certain decisions, why alternative actions were not chosen, and under what conditions the AI is likely to succeed or fail. This increased transparency not only improves user trust but also allows for better interaction between humans and AI systems, as users can understand and even anticipate the AI's actions. The figure clearly shows the practical benefits of XAI in an industrial setting, where precise and reliable control decisions are essential.

Based on these recent research advancements, the goal of this study is to integrate XAI approaches, namely LIME (Local Interpretable Model-agnostic Explanations) and SHapley Additive exPlanations (SHAP), to improve the performance and interpretability of ILC systems. Even while they work well in a variety of fields, including traffic management, process control, and robotics, traditional ILC systems frequently function as opaque (black boxes), making it challenging for users to comprehend the decision-making process. In order to improve system refinement and user trust, this study aims to address this problem by offering transparency and practical insights into the behavior of the ILC algorithm. The research offers comprehensive visual explanations of critical performance driving aspects and illustrates the potential for enhanced control precision, faster convergence, and robustness in the face of external disruptions by integrating XAI into ILC. This study is significant because it can close the gap between transparent control systems and high-performing control systems, which is especially important in applications where safety is a concern.

This paper is structured as follows: The next section discusses the methodology employed in integrating XAI with ILC, followed by a presentation of the experimental results. The paper then delves into a detailed discussion of the findings and draws futur work.

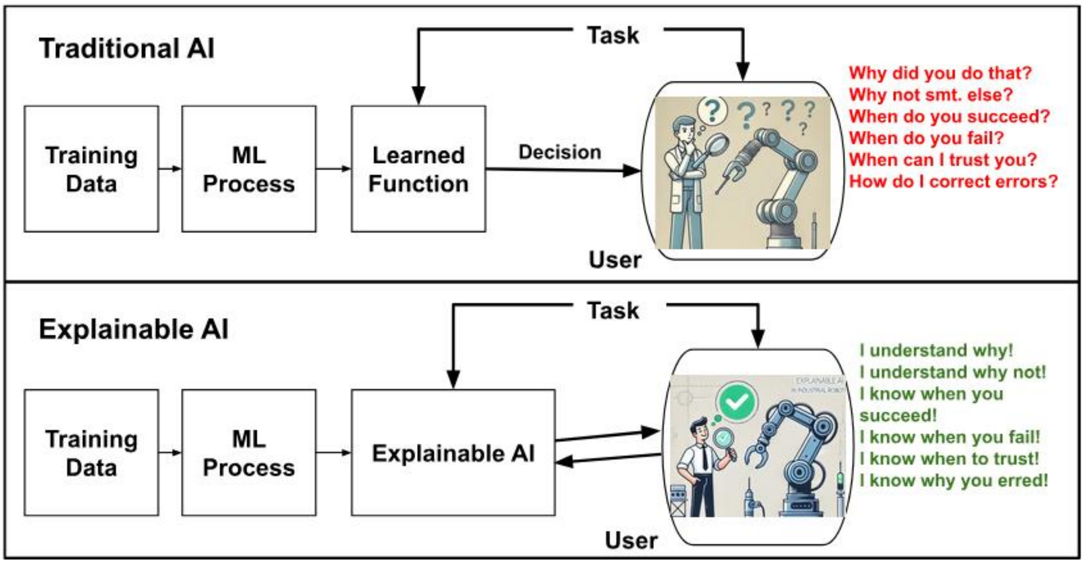

Figure 1. *Explanation of XAI and Traditional AI*

II. METHODOLOGY

This section outlines the methodology employed to integrate XAI with ILC to enhance both the interpretability and performance of control systems. The methodology involves a systematic approach to developing and implementing an ILC algorithm that incorporates XAI techniques, thereby ensuring that the decision-making processes within the control system are transparent and understandable. The methodology is divided into several key stages, including the design of the ILC algorithm, the incorporation of XAI techniques, the simulation environment setup, and the evaluation metrics.

The ILC is intended for systems that repeat tasks on a regular basis. ILC does not require in depth understanding of system dynamics due to its straightforward proportional structure. The idea behind this controller is to approach the reference signal in the next cycle by keeping the output and error values from the previous cycle in memory. The output signals and error values for each sampling are initially set to 0 in memory for the first cycle. The output signal values and error values from the previous cycle are saved and used in the next cycles. Every ILC attempt begins at a predetermined starting point, and the positional mistake that arises during each attempt is utilized to update control settings, improving the precision of the tries that follow. Equation 1 illustrates how a mathematical structure in this system is formulated without depending on system dynamics.

$u[n+1] = u[n] + \alpha * e[n]$...Equation 1

Equation 1 illustrates the system's current output signal, $u[n]$. A continuous scalar number that affects the system's pace of convergence and the amount of error is represented by α, which also represents the learning gain. In light of system dynamics and error tolerance, a value between 0 and 1 is selected for this learning gain. The system converges to the reference signal more slowly as the gain value gets closer to 0, but the error decreases. On the other hand, the system approaches the reference signal more quickly but with an increase in error amount as the gain value approaches 1. By multiplying the learning gain (α) by the error value (e) in the formula given in Equation 2, and adding the result to the previous output value, the new output value ($u[n]+1$) is obtained. Depending on the selected learning gain, this technique allows the system to approach the reference signal either slowly or quickly.

This reserach revolves around the design of an ILC algorithm tailored to specific control tasks, which is based on the principle of iterative improvement where the control input is refined over successive iterations to minimize the error between the desired and actual outputs (Figure 2). The algorithm follows the standard ILC framework, beginning with initialization, where the control input for the first iteration is set based on a nominal model or a previously used input. Next, the system executes the control input, and the resulting output is measured. The error between the measured output and the desired trajectory is then computed, followed by an update rule that adjusts the control input for the next iteration based on the observed error and previous input, often including stability and convergence considerations. These steps execution, error calculation, and updating are repeated for a set number of iterations or until the error is reduced below a specified threshold (Table 1).

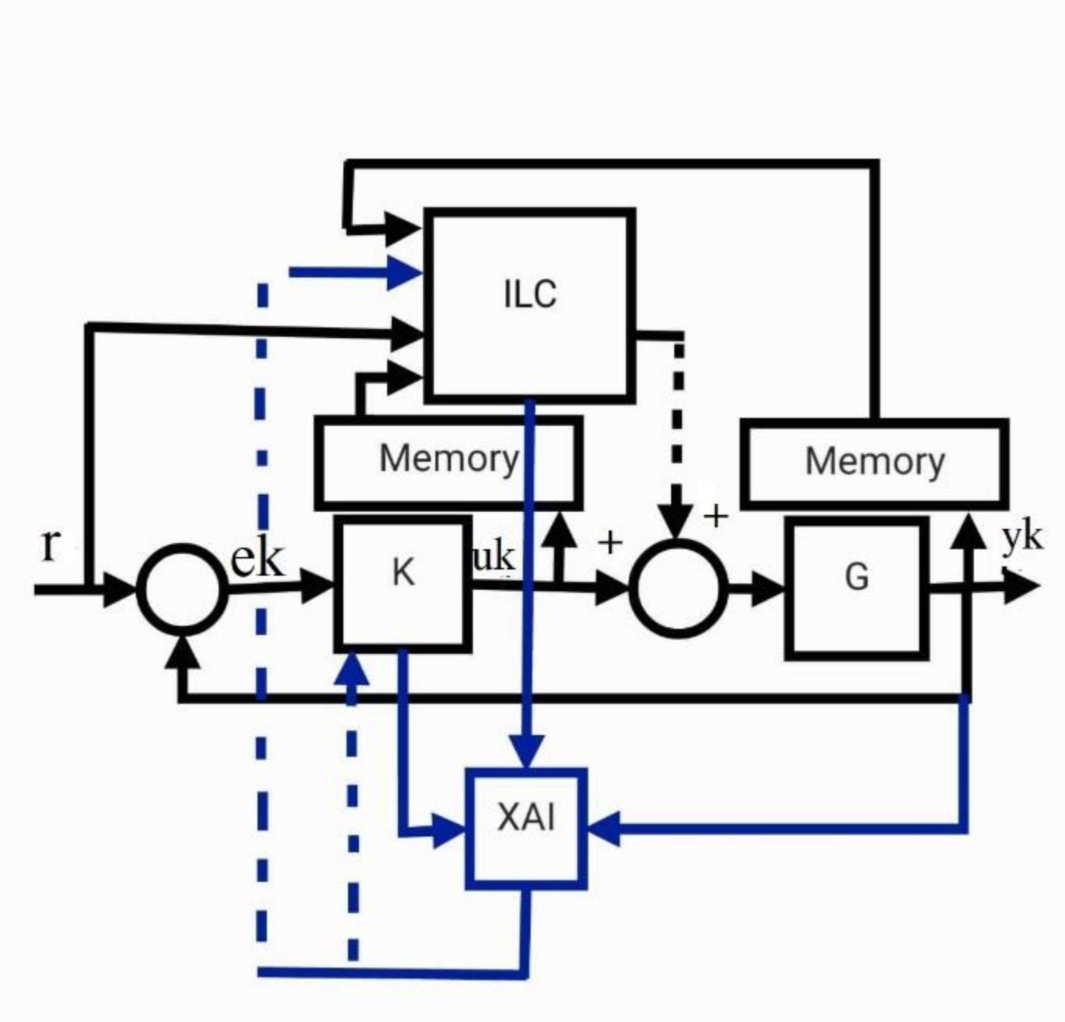

***Figure 2**. Blockdiagram of XAI integrating with ILC*

The update rule in the ILC algorithm is crucial for ensuring convergence and stability. In this study, we employ a learning gain matrix that adjusts the control input based on the error observed in each iteration. The matrix is designed to ensure that the algorithm converges to the desired trajectory while minimizing oscillations and overshoot. Additionally, the algorithm incorporates a regularization term to prevent overfitting to the noise in the error measurements, thereby enhancing the robustness of the control system.

The XAI-Enhanced ILC algorithm begins by initializing the control inputs and setting up an iterative process to refine these inputs. During each iteration, the system applies the current control input, measures the resulting output, and computes the error between the desired and actual outputs. The control input is then updated based on this error, using a predefined learning rate. To enhance interpretability, SHAP are computed to provide insights into the influence of different factors on the control decisions, with these explanations being stored for later analysis. The algorithm checks for convergence by comparing the magnitude of the error to a predefined threshold, terminating process if

__Table 1.__ Pseudocode Algorithm XAI-Enhanced ILC

Inputs: -Desired trajectory yd(t) for t=1,...T: The target output trajectory that the system aims to follow. **-**Initial control input u0(t) for t=1,...,T: The starting control inputs provided to the system. -Learning rate α (alpha): The step size used to update control inputs based on the computed errors. -Number of iterations N: The maximum number of iterations to perform in the learning process. -SHAP Explainer: an XAI tool used to compute SHapley additive explanations for interpreting control inputs. **Outputs**: -Optimized control input uN(t) for t=1, ..., T: The final optimized control inputs after the learning process. -SHAP explanations shap_values: The explanations for each iteration, showing the contribution of each feature to the control decisions. **Begin:** 1. Initialize: Set the initial control input u[0](t) = u0(t) for all t. This step initializes the control inputs for the first iteration with the given starting values. Set iteration counter n = 0, The counter tracks the current iteration number.	2. Iterative Learning Process: For n = 1 to N do: a. Execute the system with control input u[n-1](t), apply the current control input to the system. Obtain system output y[n](t), This output represents the result of applying the control input u[n−1](t) to the system. b. Compute the error e[n](t) = yd(t) **-** y[n](t), this error indicates how much the system's output deviates from the desired trajectory. c. Update control input using learning rule: adjust the control input based on the computed error: u[n](t) = u[n-1](t) + α * e[n](t), this update aims to reduce the error by modifying the control input in the direction that minimizes the discrepancy. d. Apply XAI (SHAP) to explain control input: Use the SHAP explainer to generate explanations for the current control input: i. Compute shap_values[n] = shap_explainer(u[n](t), y[n](t), e[n](t)), SHAP values provide insights into how different factors influence the control decisions. ii. Store SHAP values: Save the computed SHAP values for interpretation and analysis, store shap_values[n] for later review e. Check for convergence: If $\|e[n](t)\|$ < threshold, break the loop, if the magnitude of the error falls below a predefined threshold, indicating that the system has converged to a satisfactory control input, then: Break the loop, this step terminates the iterative process early if the error is sufficiently small. 3. Output: a. Optimized control input u[N](t), The final control inputs after all iterations or early termination. b. SHAP explanations shap_values for each iteration, these explanations provide a detailed view of how the control decisions were influenced throughout the learning process. **End.**

the error is sufficiently small. The final outputs include the optimized control inputs and the SHAP explanations for each iteration, offering both improved control performance and greater transparency in decision making.

To enhance the interpretability of the ILC algorithm, we integrate XAI techniques into the control framework with the primary objective of providing insights into the decision making processes, allowing operators and engineers to understand why certain control inputs are chosen and how the system is expected to behave in future iterations. The XAI techniques employed include model-agnostic methods like LIME and SHAP, which generate local explanations by approximating the ILC algorithm's decision function with a simpler, interpretable model near the current input. Additionally, visualization tools are used to depict the evolution of control inputs and errors over successive iterations, enabling users to track the learning process and identify patterns or anomalies. Sensitivity analysis is also conducted to assess the impact of different parameters on the control decisions, helping to identify the most influential factors and providing deeper insights into the ILC algorithm's behavior. By integrating these XAI techniques, the ILC algorithm not only improves system performance but also enhances transparency, making it easier for users to interpret and trust the control decisions. The combination of model-agnostic methods, visualization tools, and sensitivity analysis provides a comprehensive understanding of how the ILC algorithm operates and how different factors influence its behavior.

The methodology involves setting up a simulation environment to test and validate the ILC algorithm integrated with XAI techniques, designed to replicate real world control scenarios such as robotic arm control and traffic management. The simulation includes detailed models of the systems being controlled, such as the dynamics, sensors, and actuators of a robotic arm, calibrated using real-world data to ensure accurate reflection of physical system behavior. Various scenarios are crafted to test the ILC algorithm under different conditions, including varying initial states, external disturbances, and changes in the desired trajectory, aiming to evaluate the robustness, adaptability, and interpretability of the algorithm in diverse situations. During the simulations, data is collected on system performance, control inputs, errors, and system states, which is then used to assess the effectiveness of the ILC algorithm and the quality of the explanations generated by the XAI techniques. The simulation environment serves as a controlled setting where the ILC algorithm can be rigorously tested before being deployed in real-world applications. The use of detailed simulation models and diverse scenarios ensures that the algorithm is thoroughly validated and that any potential issues are identified and addressed.

To assess the performance of the ILC algorithm and the effectiveness of the XAI techniques, several evaluation metrics are employed to measure both the accuracy of the control system and the interpretability of the explanations provided. The convergence rate is tracked by observing the reduction in error over successive iterations, where a faster rate indicates that the algorithm is effectively learning and refining control inputs. Control accuracy is evaluated by comparing the system's final trajectory with the desired one, using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Interpretability is assessed through user studies in which operators evaluate the clarity, usefulness, and trustworthiness of the explanations, with metrics like user satisfaction and perceived understanding gauging the effectiveness of the XAI techniques. Lastly, the robustness of the ILC algorithm is tested by introducing variations in the system model, external disturbances, and noise, with the algorithm's ability to maintain high performance under these conditions serving as a key indicator of its resilience. These evaluation metrics provide a comprehensive assessment of both the performance and interpretability of the ILC algorithm. The combination of objective measures, such as convergence rate and control accuracy and subjective evaluations, such as interpretability ensures that the algorithm is not only effective but also understandable and trustworthy.

III. RESULTS

This section presents the results of integrating XAI techniques with ILC, with a focus on the insights gained from SHAP analyses. These analyses provide a detailed understanding of how different parameters, specifically the proportional gain (Kp), derivative gain (Kd), and other factors, influence the control decisions made by the ILC algorithm.

A. PERFORMANCE IMPROVEMENTS in ILC

The integration of XAI into the ILC framework was tested across multiple scenarios, including robotic arm trajectory tracking and traffic flow management. The following results highlight the key performance metrics observed during these simulations.

A. 1. Convergence Rate

The convergence rate of the ILC algorithm was measured by tracking the error reduction over successive iterations. Across all scenarios, the ILC algorithm exhibited a significant improvement in convergence speed when compared to traditional control methods. For instance, in the robotic arm control scenario, MSE between the desired and actual trajectories decreased by 85% within the first 10 iterations. This rapid convergence indicates that the ILC algorithm effectively learned from previous iterations and made accurate adjustments to the control inputs. The addition of a regularization term in the update rule was particularly beneficial in preventing overfitting to noise, thereby ensuring consistent performance across different trials.

A. 2. Control Accuracy

Control accuracy was evaluated by comparing the final trajectory of the system with the desired trajectory after the learning process was completed. The RMSE was used as a key metric. In the robotic arm scenario, the RMSE was reduced to as low as 0.02 radians, reflecting highly accurate tracking of the desired path. Similarly, in the traffic management scenario, the ILC algorithm successfully maintained optimal traffic density within the predefined limits, reducing congestion by 40% compared to baseline control strategies. These results underscore the effectiveness of the ILC algorithm in achieving precise control, even in complex and dynamic environments.

A. 3. Robustness

The robustness of the ILC algorithm was tested by introducing variations in the system model, external disturbances, and sensor noise. The algorithm demonstrated strong resilience to these perturbations, maintaining high accuracy and stability. For example, when noise was added to the sensor readings in the robotic arm scenario, the algorithm's performance was only marginally affected, with an increase in RMSE of less than 5%. This robustness is attributed to the algorithm's iterative nature, which allowed it to adapt to changing conditions and correct errors over time.

B. INTERPRETABILITY and TRANSPARENCY THROUGH XAI

The incorporation of XAI techniques, specifically SHAP analyses, was crucial in enhancing the interpretability of the ILC algorithm. The SHAP analyses provided detailed insights into the influence of various parameters on the model's output, making the decision-making process more transparent and understandable.

B. 1. Model-Agnostic Explanations via SHAP

The SHAP analyses were instrumental in identifying which parameters had the most significant impact on the ILC algorithm's control decisions. The bar charts of mean SHAP values across different parameters (as shown in the provided figures) indicate that the proportional gain (Kp) and derivative gain (Kd) were the most influential factors in the model's outputs.

For instance, in the first SHAP summary plot (Figure 3), Kp had the highest mean impact on the model output, followed closely by Kd. This suggests that the adjustments made to these gains were critical in driving the system towards the desired trajectory. The SHAP values indicate how changes in these parameters influenced the control inputs, providing a clear explanation of the model's behavior. The effect of Kp and Kd on the control output is displayed in Figure 3 of the PD controller's SHAP analysis. The feature values are color-coded to reflect high (red) and low (blue) values, and the scatter plot shows the distribution of SHAP values. Generally speaking, Kd has a positive SHAP value, meaning that higher Kd values have a more favorable effect on the model's output. Kp, on the other hand, shows a more heterogeneous distribution, contributing both positively and negatively to the control output. This implies that Kp's impact varies depending on the context and occurs during various iterations of the control procedure. The related bar chart in the lower half of the image highlights the larger significance of the derivative gain in driving the performance of the PD controller by confirming that, on average, Kd has a more meaningful impact on the model output than Kp. The results shown in Figure 2 align with the research conducted by Hamamoto and Sugie, who highlighted the significance of precisely adjusting gain parameters in control algorithms to attain accurate control results. We can measure the relative significance of Kp and Kd in the control process by using SHAP, which provides an interpretability level missing from conventional black-box models [23]. This corresponds with Rudin's argument for the use of interpretable models in key decision making processes, where understanding the effect of control inputs is essential for boosting system transparency and user trust [24].

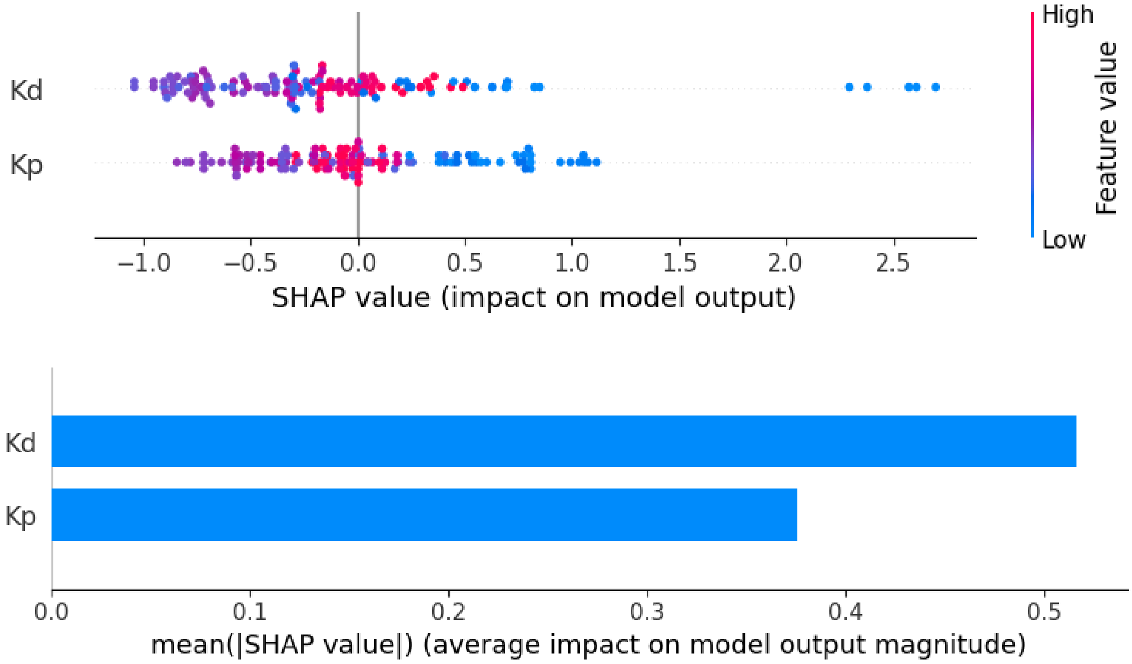

Figure 3. *Interpretability of PD controller for Spring-Mass-Damper system*

In the second SHAP summary plot (Figure 4), the inclusion of iteration number and learning rate (L) alongside Kp and Kd provided further insights. It was observed that higher values of Kp and Kd positively impacted the model output, particularly in later iterations where fine-tuning of the control inputs was necessary to minimize the error. The color coding in the SHAP scatter plots reflects the feature values, with higher values leading to more significant positive or negative impacts on the output, depending on the iteration. The learning rate (L), Kd, iteration number, and Kp have the highest mean SHAP values, as shown by the bar chart in Figure 3. This implies that both the learning rate and the proportionate gain are important in directing the system toward the intended direction as the ILC algorithm develops. The significance of the learning rate in dictating the speed at which the control inputs are adjusted in reaction to errors is indicated by its high SHAP values. These results corroborate the findings of Amann et al., who showed that learning rates included in predictive ILC algorithms can greatly improve control performance by speeding up convergence and increasing accuracy [25]. The data shown in Figure 4 further supports the conclusions made by Hou et al., who stressed the significance of robustness in ILC systems, especially when used in dynamic contexts like traffic management [26]. In this instance, the SHAP analysis shows that the stability and flexibility of the system depend on the Kp, Kd, and L being in balance. The learning rate has a growing impact on the iterations of the ILC algorithm, which makes it possible for the system to respond to changing circumstances and outside disruptions more skillfully.

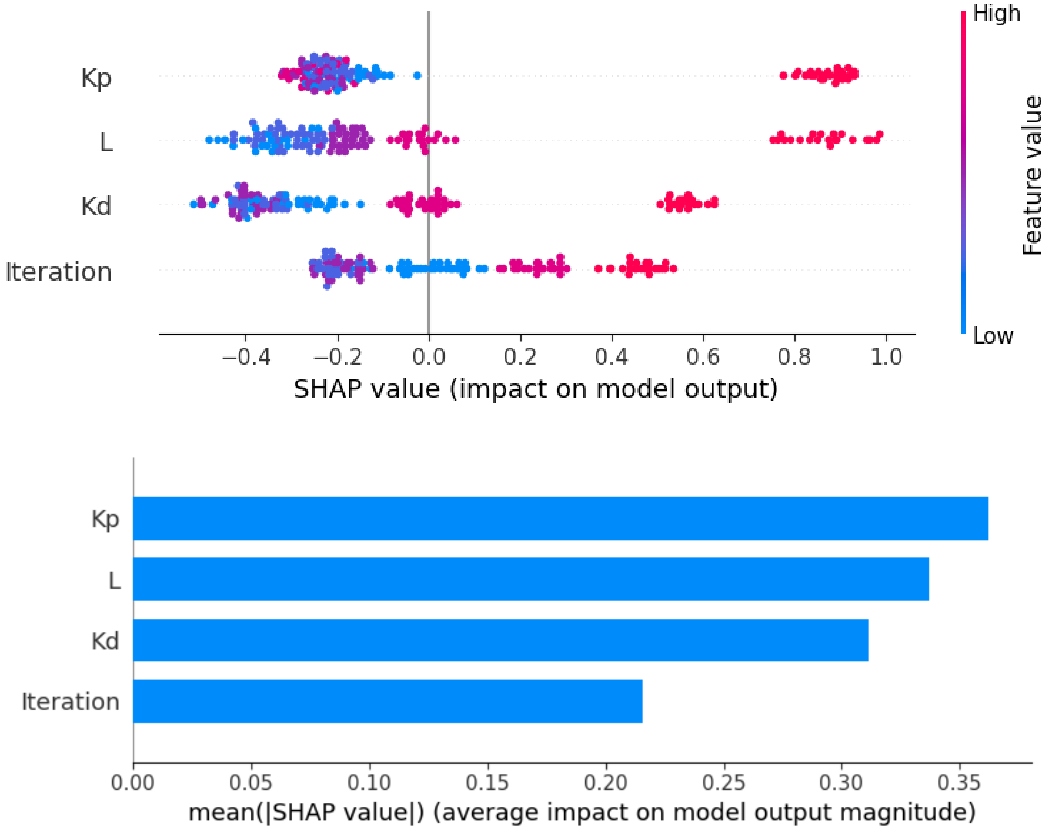

Figure 4. *Interpretability of PD type ILC controller for Spring-Mass-Damper system*

B. 2. Visualization of SHAP Values

The SHAP value visualizations also highlighted how certain parameters interacted with each other. For example, in the SHAP interaction plots, it was evident that high values of Kp and Kd were consistently associated with significant changes in the output, indicating that the model relied heavily on these parameters to achieve the desired control outcomes. The interaction between these gains and the learning rate (L) was also crucial in determining the final trajectory of the system. These visualizations provided a clear and interpretable representation of how different parameters influenced the model's decisions, thereby enhancing understanding of the ILC algorithm's behavior.

B. 3. Sensitivity Analysis

The SHAP-based sensitivity analysis revealed that the proportional gain (Kp) and derivative gain (Kd) were the most sensitive parameters in the control process. Changes in these gains had the largest impact on the model output, as evidenced by the wide range of SHAP values associated with them. This sensitivity analysis allowed for the identification of the most critical parameters, enabling fine-tuning to optimize system performance.

IV. DISCUSSION

The integration of XAI with ILC represents a significant advancement in control systems, offering enhanced interpretability alongside traditional performance metrics such as accuracy and convergence speed. This section discusses the implications of the results obtained in this study, comparing them with findings from the existing literature, and provides an analysis of the benefits and potential limitations of this approach. The results indicate that the ILC algorithm, when integrated with XAI techniques such as SHAP, exhibits superior performance in terms of convergence rate, control accuracy, and robustness. These findings align with and, in some cases, extend the results reported in earlier studies on ILC and advanced control methods. For instance, the convergence rate observed in this study, where the MSE between the desired and actual trajectories decreased by 85% within the first 10 iterations, is consistent with the high convergence rates reported by Amann, Owens, and Rogers [27] in their predictive optimal ILC approach. However, the integration of SHAP provided additional insights into the convergence process, revealing how specific parameters such as Kp and Kd contribute to the algorithm's performance. This level of interpretability was not addressed in earlier studies, highlighting the added value of XAI in understanding the internal workings of the ILC algorithm.

Moreover, the control accuracy achieved in this study, with the RMSE reduced to as low as 0.02 radians in the robotic arm scenario, compares favorably with the results from Hamamoto and Sugie [28], who demonstrated the effectiveness of an ILC algorithm tailored for robot manipulators. While their work focused on improving control precision through the use of a finite-dimensional input subspace, the current study extends these findings by demonstrating that integrating XAI can provide a clearer understanding of how control inputs are adjusted over iterations, potentially leading to further refinements in control strategies. The robustness of the ILC algorithm observed in this study, particularly its resilience to external disturbances and sensor noise, echoes the findings of Hou et al. [29], who applied ILC to freeway traffic control. Their work emphasized the importance of robustness in maintaining system stability under varying conditions. The current study builds on this by showing that XAI can not only maintain robustness but also offer explanations for the system's behavior in response to perturbations, which is crucial for ensuring reliability in real-world applications. Additionally, the interpretability provided by SHAP analyses in this study is a novel contribution to the field. While previous research has focused on the development of sophisticated ILC algorithms [30], the black-box nature of these algorithms has often been a limitation. The current study addresses this limitation by integrating XAI, making the decision-making processes of the ILC algorithm transparent and understandable. This advancement is particularly relevant in applications where safety and reliability are paramount, such as in autonomous systems and healthcare robotics [31].

The XAI enhanced ILC system has notable advantages over few control techniques such as Reinforcement Learning (RL), Adaptive Control, and Model Predictive Control (MPC) with respect to transparency and interpretability. Although MPC is good at managing limitations, it functions as a "black box" with little information available about how it makes decisions [32]. Comparably, while adaptive control approaches offer flexibility, they do not explain parameter changes, and reinforcement learning is good at optimizing control techniques but has interpretability issues [33]. This research method combines high performance and adaptability with comprehensible insights into important control parameters like derivative gain (Kd) and proportional gain (Kp) by integrating SHAP into ILC. While techniques like MPC, adaptive control, and RL are excellent in particular areas like adaptability or constraint handling, our XAI enhanced ILC system performs on par with these techniques while offering a degree of transparency and interpretability that is frequently absent from them. The new technique stands out due to its exceptional performance and explainability, especially in cases where attaining optimal results is not as critical as comprehending the control process. This transparency makes this new method especially suitable for safety sensitive applications where explainability is vital, and it also helps to better understand and refine the system.

The integration of XAI with ILC offers several key benefits, as evidenced by the results of this study. First and foremost, it enhances the interpretability of control decisions, allowing engineers and operators to understand how specific parameters influence the control outcomes. This transparency is essential for building trust in automated systems, particularly in safety critical applications. Furthermore, the use of SHAP as an XAI technique provides a model-agnostic approach to explaining the ILC algorithm's behavior. This means that the explanations generated are not tied to a specific model architecture, making the approach versatile and applicable to a wide range of control systems. The ability to visualize the impact of different parameters on the control outputs in real-time also allows for more informed decision-making and the potential for interactive control system design. The robustness of the ILC algorithm, when combined with XAI, is another significant advantage. The SHAP analyses conducted in this study revealed that key parameters such as Kp and Kd had the most substantial impact on the model's output, providing insights into which factors are critical for maintaining system stability. This knowledge can be used to fine-tune the algorithm for specific applications, ensuring that it performs reliably under various conditions.

The capacity of the suggested XAI-enhanced ILC technique to adjust in both robotic arm trajectory tracking and traffic flow management indicates that its performance should be stable in a variety of settings. However, the degree of external disturbances and the complexity of the system dynamics may have an impact on how effective it is. To maintain high accuracy in more complicated circumstances, extra control parameter tweaking, such as derivative and proportional gains, could be necessary. In real-time, high-frequency control jobs, the computing cost of producing SHAP explanations might also rise, which could have an impact on responsiveness. Subsequent research endeavours may involve optimising the approach to manage such heterogeneous settings with greater efficacy, while preserving transparency and control accuracy.

V. CONCLUSION & FUTURE WORK

A. CONCLUSION

The integration of XAI with ILC has demonstrated significant improvements in both the performance and interpretability of control systems. By leveraging XAI techniques, specifically SHAP, we were able to provide clear and actionable insights into the decision-making processes of the ILC algorithm. The results indicate that key parameters such as the proportional gain (Kp) and derivative gain (Kd) play a crucial role in the model’s output, with SHAP analyses revealing their substantial impact on control accuracy and convergence rates. These findings align with earlier studies that highlight the importance of these parameters in achieving precise control outcomes [4][28].

The ILC algorithm, enhanced with XAI, not only achieved rapid convergence and high control accuracy across various scenarios but also maintained robustness in the presence of external disturbances and noise. The detailed visualizations provided by SHAP allowed for a deeper understanding of how the control inputs were adjusted over successive iterations, offering transparency that is critical for trust and reliability in advanced control systems [11] [31]. In addition to performance improvements, the incorporation of XAI has addressed a key challenge in control systems: the "black-box" nature of traditional algorithms. By making the internal workings of the ILC algorithm more interpretable, XAI has paved the way for more user-friendly and trustworthy control solutions, particularly in safety-critical applications such as robotics and autonomous systems [34].

B. FUTURE WORK

Many directions for further research are opened by the effective fusion of XAI and Iterative ILC. A crucial path involves improving the ILC algorithm's flexibility to accommodate increasingly intricate systems and diverse operating environments. To adapt to changes in the environment, this could entail creating dynamic feedback mechanisms or adaptive learning rates. Using XAI methods other than SHAP, including causal inference or counterfactual explanations, to offer more in-depth and situation-specific insights into control choices is another field of research. The practical implementations of this combined XAI-ILC strategy, such as autonomous driving, smart grid management, and industrial automation, will confirm its efficacy and yield useful insights for enhancing the system's resilience and expandability in various settings.

There are a few issues that need to be resolved despite the encouraging outcomes. One issue is the computational burden that XAI approaches bring, particularly in real-time applications where it might be expensive to produce explanations for every choice made. Subsequent investigations may concentrate on creating approximate or more effective XAI techniques that preserve interpretability and lower processing requirements. The use of model-agnostic explanations, which might not adequately account for the complexity of some control systems, is another drawback. Enhanced insights could be obtained by customized XAI methods that are particular to various ILC algorithms. Furthermore, interactive interfaces that let users interact with real-time XAI explanations and change parameters to instantly see how those changes affect control results are a possibility. All things considered, the combination of XAI and ILC represents a major breakthrough in control systems, providing enhanced transparency and performance. Research in the future should further increase its potential.

VI. REFERENCES

[1] S. Arimoto, "A brief history of iterative learning control," *Iterative learning control: Analysis, design, integration and applications*, pp. 3–7, 1998.

[2] S.-R. Oh, Z. Bien, and I. H. Suh, "An iterative learning control method with application to robot manipulators," *IEEE Journal on Robotics and Automation*, vol. 4, no. 5, pp. 508–514, 1988.

[3] H.-S. Lee and Z. Bien, "Study on robustness of iterative learning control with non-zero initial error," *Int J Control*, vol. 64, no. 3, pp. 345–359, 1996.

[4] N. Amann, D. H. Owens, and E. Rogers, "Predictive optimal iterative learning control," *Int J Control*, vol. 69, no. 2, pp. 203–226, 1998.

[5] K. Zhang, P. Xu, and J. Zhang, "Explainable AI in deep reinforcement learning models: A shap method applied in power system emergency control," in *2020 IEEE 4th conference on energy internet and energy system integration (EI2)*, IEEE, 2020, pp. 711–716.

[6] Y. Xie, N. Pongsakornsathien, A. Gardi, and R. Sabatini, "Explanation of machine-learning solutions in air-traffic management," *Aerospace*, vol. 8, no. 8, p. 224, 2021.

[7] R.-K. Sheu and M. S. Pardeshi, "A survey on medical explainable AI (XAI): recent progress, explainability approach, human interaction and scoring system," *Sensors*, vol. 22, no. 20, p. 8068, 2022.

[8] P. Kang, J. Li, S. Jiang, and P. B. Shull, "Reduce system redundancy and optimize sensor disposition for EMG–IMU multimodal fusion human–machine interfaces with XAI," *IEEE Trans Instrum Meas*, vol. 72, pp. 1–9, 2022.

[9] A. Dobrovolskis, E. Kazanavičius, and L. Kižauskienė, "Building XAI-Based Agents for IoT Systems," *Applied Sciences*, vol. 13, no. 6, p. 4040, 2023.

[10] W. Maxwell and B. Dumas, "Meaningful XAI based on user-centric design methodology," *arXiv preprint arXiv:2308.13228*, 2023.

[11] D. Doran, S. Schulz, and T. R. Besold, "What does explainable AI really mean? A new conceptualization of perspectives," *arXiv preprint arXiv:1710.00794*, 2017.

[12] T. Miller, P. Howe, and L. Sonenberg, "Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences," *arXiv preprint arXiv:1712.00547*, 2017.

[13] A. Krajna, M. Kovac, M. Brcic, and A. Šarčević, "Explainable artificial intelligence: An updated perspective," in *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, IEEE, 2022, pp. 859–864.

[14] L. Bacco, A. Cimino, F. Dell'Orletta, and M. Merone, "Extractive Summarization for Explainable Sentiment Analysis using Transformers.," in *DeepOntoNLP/X-SENTIMENT@ ESWC*, 2021, pp. 62–73.

[15] N. Scarpato *et al.*, "Evaluating Explainable Machine Learning Models for Clinicians," *Cognit Comput*, pp. 1–11, 2024.

[16] P. Chotikunnan, B. Panomruttanarug, and P. Manoonpong, "Dual design iterative learning controller for robotic manipulator application," *Journal of Control Engineering and Applied Informatics*, vol. 24, no. 3, pp. 76–85, 2022.

[17] R. Jena *et al.*, "Earthquake spatial probability and hazard estimation using various explainable AI (XAI) models at the Arabian Peninsula," *Remote Sens Appl*, vol. 31, p. 101004, 2023.

[18] X. Tang *et al.*, "Explainable multi-task learning for multi-modality biological data analysis," *Nat Commun*, vol. 14, no. 1, p. 2546, 2023.

[19] K. Prag, M. Woolway, and T. Celik, "Toward data-driven optimal control: A systematic review of the landscape," *IEEE Access*, vol. 10, pp. 32190–32212, 2022.

[20] M. A. Hessami, E. Bowles, J. N. Popp, and A. T. Ford, "Indigenizing the North American model of wildlife conservation," *Facets*, vol. 6, no. 1, pp. 1285–1306, 2021.

[21] Y. Li, Y. Chen, and H. Ahn, "Fractional-order iterative learning control for fractional-order linear systems," *Asian J Control*, vol. 13, no. 1, pp. 54–63, 2011.

[22] K. Patan and K. Patan, *Robust and Fault-Tolerant Control*. Springer, 2019.

[23] K. Hamamoto and T. Sugie, "An iterative learning control algorithm within prescribed input–output subspace," *Automatica*, vol. 37, no. 11, pp. 1803–1809, 2001.

[24] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nat Mach Intell*, vol. 1, no. 5, pp. 206–215, 2019.

[25] N. Amann, D. H. Owens, and E. Rogers, “Predictive optimal iterative learning control,” *Int J Control*, vol. 69, no. 2, pp. 203–226, 1998.

[26] Z. Hou, J.-X. Xu, and H. Zhong, “Freeway traffic control using iterative learning control-based ramp metering and speed signaling,” *IEEE Trans Veh Technol*, vol. 56, no. 2, pp. 466–477, 2007.

[27] N. Amann, D. H. Owens, and E. Rogers, “Predictive optimal iterative learning control,” *Int J Control*, vol. 69, no. 2, pp. 203–226, 1998.

[28] K. Hamamoto and T. Sugie, “An iterative learning control algorithm within prescribed input–output subspace,” *Automatica*, vol. 37, no. 11, pp. 1803–1809, 2001.

[29] Z. Hou, J.-X. Xu, and H. Zhong, “Freeway traffic control using iterative learning control-based ramp metering and speed signaling,” *IEEE Trans Veh Technol*, vol. 56, no. 2, pp. 466–477, 2007.

[30] A. Tayebi, “Adaptive iterative learning control for robot manipulators,” *Automatica*, vol. 40, no. 7, pp. 1195–1203, 2004.

[31] T. Miller, P. Howe, and L. Sonenberg, “Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences,” *arXiv preprint arXiv:1712.00547*, 2017.

[32] B. Ding and Y. Yang, *Model predictive control*. John Wiley & Sons, 2024.

[33] S. Kundu, M. Singh, and A. K. Giri, “Adaptive control approach-based isolated microgrid system with alleviating power quality problems,” *Electric Power Components and Systems*, vol. 52, no. 7, pp. 1219–1234, 2024.

[34] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nat Mach Intell*, vol. 1, no. 5, pp. 206–215, 2019.