SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper
Open Access

A Hybrid Framework Integrating Supervised and Reinforcement Learning for Adaptive Decision-Making in Dynamic Environments

Ashish Sharma[1], Lakshmi Viveka K[2], Hadasha Nobel tune[3], Dhanalaxmi Chinthala[4], Dr. Ravi Thangjam[5], Bipin Sule[6], Tanveer Ahmad Wani[7], D.Akila[8]

[1]Department of Computer Engineering & Applications, GLA University, Mathura, Email : ashish.sharma@gla.ac.in

[2]Associate Professor, Department of CSE (Artificial Intelligence), Pragati Engineering College, ADB Road, Surampalem, Near Peddapuram, Kakinada District, Andhra Pradesh, India - 533437. Email : viveka.pg@gmail.com

[3]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Email : nobel@maher.ac.in

[4]Assistant Professor, Department of Information Technology, Vardhaman College of Engineering, Shamshabad, Hyderabad, India - 501 218, Email : dhanalaxmi1514@vardhaman.org

[5]Professor, School of Business, Aditya University, Surampalem, Andhra Pradesh, Pin 533437 , Email : provc_sp@adityauniversity.in

[6]Sr. Professor, DESH, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037 Email : bipin.sule@vit.edu

[7]School of Sciences,Noida international University, Uttar Pradesh 203201, India, Email : tanveer.ahmad@niu.edu.in

[8]Professor, Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India., Tamil Nadu ,India, Email : akiindia@yahoo.com

Abstract

In intelligent systems, especially in various types of autonomous systems, robotics, intelligent health care, industrial control and real-time resource management, adaptive decision-making in dynamic environments has become a key challenge. The traditional supervised learning models are effective in dealing with static datasets, but suffer from poor performance in dynamic environments because they lack adaptability. Reinforcement learning is another class of algorithms that learn optimal actions based on feedback from contextual rewards generated by the environment, which can be very difficult to explore and may take a long time to train the agent. This means that an efficient Hybrid learning framework of predictive Intelligence and Adaptive optimization is needed. This research attempts to develop a hybrid solution combining supervised learning and reinforcement learning for adaptive decision-making in dynamic environment. The model first applies supervised learning on past data to discover predictive knowledge and then makes initial decisions. A reinforcement learning agent then optimizes the decisions by tuning the policy based on the rewards received. Dynamic simulation evaluation was performed with the main metrics: prediction accuracy, reward convergence, adaptation efficiency and computational latency. The proposed framework is assessed and the results show that the accuracy of the adaptive decisions, the stability of the convergence and the efficiency of the responses is greatly enhanced over the traditional approaches.

Keywords: Hybrid Learning, Supervised Learning, Reinforcement Learning, Adaptive Decision-Making, Dynamic Environments, Intelligent Systems

1. Introduction

In recent years, artificial intelligence has seen rapid advancement as there is growing demand for intelligent systems that possess autonomous and adaptive decision-making ability (Schmidhuber, 2015; He et al., 2016).The applications of adaptive AI systems include but are not limited to the following areas: robotics; autonomous vehicles; healthcare monitoring; industrial automation; smart transportation; financial analytics

(Kober et al., 2013; Liang et al., 2018). The applications are deployed in a dynamic environment that continually evolves and changes, and intelligent models that can learn, adapt and optimize decisions in real time are needed (Arulkumaran et al., 2017; Li, 2023). The result was that machine learning was now essential tooling in the development of modern adaptive systems.

Supervised learning is a popular machine learning paradigm for prediction and classification problems. It involves feeding data with labels into models that can recognize patterns and make precise forecasts. Supervised learning algorithms have high accuracy in static and structured environments (He et al., 2016). But these models work best if the environment doesn't change very quickly from the data that they have been trained on. They are not effective in complex and changing environments because they can't always adapt their decisions as the data comes in.

To overcome these challenges, reinforcement learning has been proven to be a successful method of adaptive intelligence (Mnih et al., 2015; Silver et al., 2017). Reinforcement learning is a method that allows an agent to learn the optimal action needed by simply interacting with the environment using reward and penalty functions (Van Hasselt et al., 2016; Schulman et al., 2017). In contrast to supervised learning, RL is able to adapt decision policies over time. Reinforcement learning models often have large exploration spaces, a high amount of computational resources and a slow convergence time, which can make training less efficient (Arulkumaran et al., 2017; Vincent et al., 2018).

Hybrid Learning Architectures are developed due to the complementary features of supervised learning (SL) and reinforcement learning (RL). Intelligent decision making systems, which are rarely exposed to dynamic environments, are plagued by unpredictability and variability in such settings, and fail to adapt on their own if the situation they encounter is new or unknown. While reinforcement learning offers adaptive capabilities through reward-driven learning, there are some issues with stand-alone reinforcement learning such as slow convergence and computational complexity (Haarnoja et al., 2018; Wang et al., 2016).

For this reason, in this research, it is proposed to develop a unified model that combines supervised learning and reinforcement learning for adaptive decision making under conditions of dynamism. The proposed framework incorporates knowledge extraction capabilities that are predictive in nature and optimization based on policy real-time for more accurate decisions, convergence stability, and better learning efficiency. Rewards-driven learning is incorporated as an adaptive strategy to fine-tune the decision policies, based on environment/reward feedback. The framework is compared to the conventional supervised learning or reinforcement learning approach using various performance metrics in order to show its effectiveness when dealing with real-time intelligent applications that work in uncertain and changing environments.

2. Related Work

Supervised learning is a method widely used in intelligent decision systems since its learning capacity from labeled data can be used to obtain predictive patterns.Supervised learning algorithms like support vector machines, decision trees, random forests or deep neural networks have been shown to be highly effective for classification, prediction and pattern recognition (He et al., 2016; Schmidhuber, 2015). The above methods are widely adopted in medical diagnostics, driverless operations, industrial monitoring, financial prediction, and for smart infrastructure management (Liang et al., 2018). Supervised models using deep learning have also been used to enhance accuracy of decisions by automatically extracting features and/or processing large-scale data (Chen et al., 2016). Supervised learning models are generally known to work well with static training data and pre-defined patterns, making them inflexible in dynamic environments. They lack the capacity to continuously learn from environment interaction which diminishes in uncertain and dynamic environments where adaptations are needed at run-time.

The challenges of uncertainty and state changes in continuous dynamic environments have led to the development of reinforcement learning as an important paradigm for adaptive decision-making (Mnih et al., 2015; Silver et al., 2017). Reinforcement learning allows for intelligent agents to interact with their environment and learn to make improvements through the use of reward-driven learning strategies (Van Hasselt et al., 2016; Schulman et al., 2017). In the field of robotics, game intelligence, autonomous navigation, resource allocation, and industrial automation, algorithms like Q-learning, Deep Q-Networks (DQN), policy gradient methods, and actor–critic algorithms have proven to be effective (Kober et al., 2013; Haarnoja et al., 2018). The main benefit of reinforcement learning is its ability to learn policies on-the-fly without the need to use labeled data. Reinforcement learning systems have slow convergence, need extensive exploration, and may be unstable while learning and are also extremely computational expensive (Arulkumaran et al., 2017; Vincent et al., 2018). These difficulties are greater in the case of large scale, real time application where the state-action space is complex.

Given the drawbacks of stand-alone learning approaches, researchers have sought schemes combining several machine learning approaches to achieve a hybrid AI architecture. In the field of adaptive control systems, Liang et al. (2018) and Zhang et al. (2021) demonstrated that hybrid AI models, comprising of supervised learning and reinforcement learning, exhibited promising results in intelligent robotics and autonomous decision-making applications. Many studies resort to supervised learning for model initialization for prediction or for training of the policy to be learned via environmental feedback in reinforcement learning. The integration makes the convergence faster, more efficient in learning and stable for decision making (Haarnoja et al., 2018; Wang et al., 2016). The emergence of hybrid deep learning and reinforcement learning algorithms also boosted the adaptability to changing environments, as predictive intelligence is integrated with policy optimizing mechanisms (Li, 2023).

Although this research has improved in the field of hybrid AI, there are still several points that have yet to be addressed. Prediction accuracy or adaptive optimization are the major focus of many existing frameworks, which are not adequately balanced with regards to supervised and reinforcement learning. In fast-changing environments some models suffer from high level of computation, lack of scalability or lack of adaptability to real-time applications. Moreover, there is a lack of extensive comparative studies with stand-alone models under dynamic operating conditions for several studies. These restrictions underscore the importance of developing a field-integrated, cohesive and efficient hybrid framework that enhances accuracy, convergence stability and real-time learning process performance of adaptive decision-making in an uncertain world.

3. Proposed Hybrid Framework

3.1 System Architecture

The hybrid framework is intended to leverage both supervised learning (for its prediction capabilities) and reinforcement learning (for its adaptive decision making) components. The architecture starts from the data acquisition layer, which receives historical data as well as real-time states and features extracted from sensors, as illustrated in Figure 1. These data are then sent to the preprocessing step where they are cleaned, feature extracted and normalized to ensure uniformity for model training and decision generation.

The supervised prediction module is designed to learn predictions patterns from the labeled training data and provide an initial baseline decision. This output will serve as a solid foundation when making decisions. Meanwhile, the reinforcement learning agent is interacting with the dynamic environment by using state-action mapping and policy learning. The agent gets the reward from the environment that it used to get and it modifies its policy accordingly, improving the adaptive behavior over time.

In the supervised learning module, the learning outputs are supervised learning class labels, and in the reinforcement learning agent, the learning outputs are a set of control parameters.The supervised learning module outputs are supervised learning class labels, and the reinforcement learning agent outputs are a set of control parameters. This engine implements adaptive fusion to create the final and optimized decision. This decision is then fed into the dynamic environment and the new state and reward feedback are then fed into the learning process ongoing.

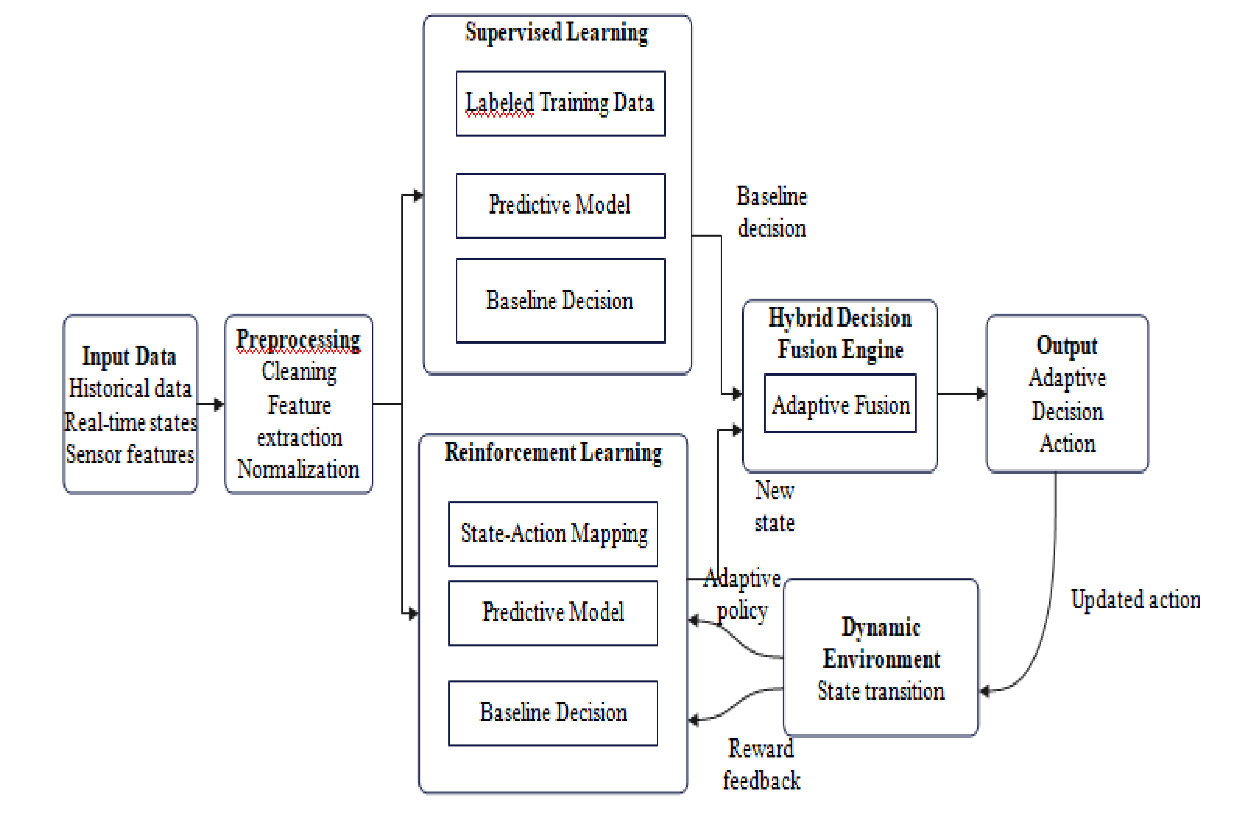

Fig. 1. Proposed Hybrid Supervised–Reinforcement Learning Architecture

3.2 Workflow of Adaptive Decision-Making

The proposed hybrid framework is at its core an adaptive decision-making workflow as shown in Figure 2. The training phase starts with the input data collection by which historical data, real time environmental data and data from sensors are collected. These data are then preprocessed by traditional methods like data cleaning, feature extraction, and normalization to enhance data quality and learning efficiency. The results from the data processing are then fed to the supervised learning module, which produces an initial predictive model and baseline decision output.

After the prediction stage, the framework observes the condition of the surrounding environment by observing the current state of the system and changes in the surrounding environmental state. This observed state information is fed into an RL agent and adaptive action selection and policy optimization is carried out. The reinforcement learning part is constantly engaging with the dynamic environment and learns the best way to act in this environment depending on the environment's responses and the reward after each action.

The supervised learning module and the reinforcement learning agent cooperate with each other via the hybrid decision fusion stage. In this stage, predictive intelligence and adaptive learning are merged together to generate an optimized decision to be executed on the fly. Adaptive action generated is then implemented in the dynamic environment in the decision execution phase.

Rewards are added to the system to assess performance following action. The environment provides reward signals, updated states and the scores reflecting the performance. Such feedback signals will be fed into the policy update and optimization phase to improve the reinforcement learning policy step-by-step. The proposed framework continually interacts with the environment and evolves its policies to enhance the accuracy of the decisions, their adaptability, and their optimality in dynamic and uncertain environments.

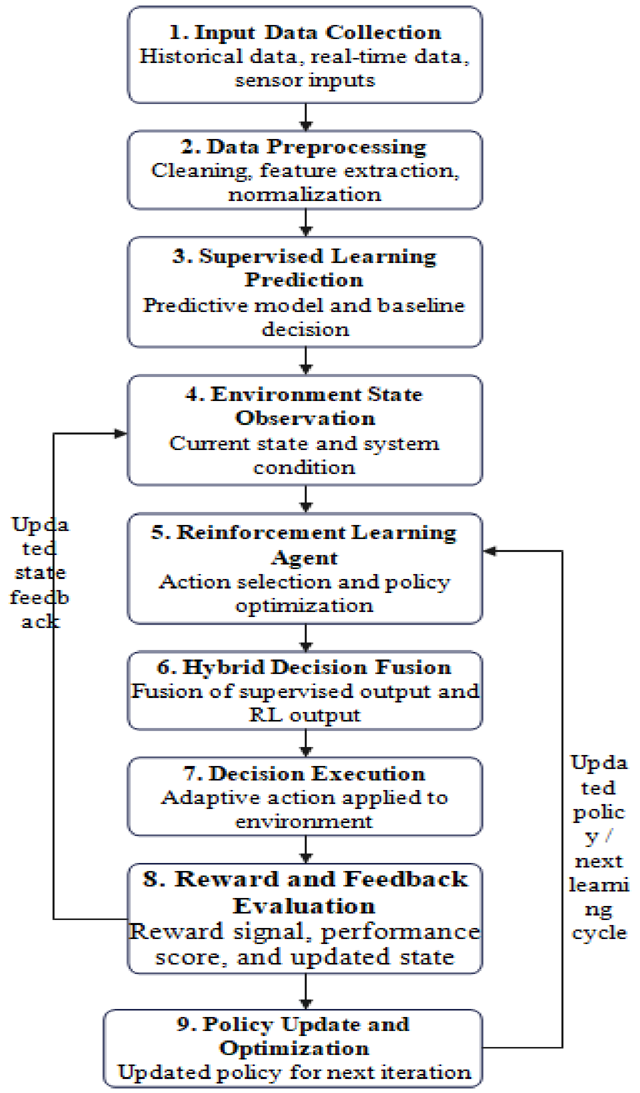

Fig. 2. Workflow of Adaptive Learning and Decision Optimization

3.3 Mathematical Modeling

The proposed hybrid architecture consists of a supervised learning and reinforcement learning to provide adaptive decision-making in dynamic environment.A mathematical model is used to illustrate the learning process, the optimization of the policies, and the generation of the hybrid decision.The supervised learning component reduces the error of the prediction made from a loss function; the reinforcement learning component adjusts the decision policies using the reward optimization method.Adaptive and optimized decisions are made by applying a weighted hybrid decision function, based on the outputs of both learning models.

The mean squared error loss function in Equation (1) is used to represent the supervised learning process mathematically.

This equation is used to compare the output to the predicted output from the supervised learning model.

Equation (1): Supervised Learning Loss Function

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \qquad (1)$$

In Equation (1), $L(\theta)$ represents the training loss, y_i denotes the actual output$\hat{y}_i$ indicates the predicted output generated by the model, and N represents the total number of training samples used during the supervised learning process.

The reinforcement learning process continually updates the decision policies based on the feedback from the environment.The Policy learning process with the Q value update algorithm is illustrated in Equation (2).This equation is used to update the state-action value based on the reward received and the reward to be expected in future.

Equation (2): Reinforcement Learning Q-Value Update

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max Q(s', a') - Q(s,a)], \qquad (2)$$

In Equation (2), $Q(s,a)$ represents the state-action value, α denotes the learning rate used for updating policy parameters, r represents the reward obtained from the environment, and γ indicates the discount factor that controls the importance of future rewards during reinforcement learning optimization.

The final adaptive decision is obtained by optimally combining the supervised learning and reinforcement learning results with a weight associated with them through an optimization strategy.The hybrid decision function is shown in Equation (3).

Equation (3): Hybrid Decision Optimization Function

$$D_{hybrid} = \lambda D_{SL} + (1 - \lambda) D_{RL}, \qquad (3)$$

In Equation (3), D_{hybrid} represents the final hybrid decision generated by the framework, D_{SL} denotes the supervised learning output, D_{RL} represents the reinforcement learning output, and λ is the weighting factor used to balance the contributions of supervised and reinforcement learning during adaptive decision optimization.

The mathematical formulation shows how predictive learning and adaptive policy optimization are combined into the proposed framework for enhancing prediction accuracy, adaptability and efficiency of policy optimization in dynamic environments.

4. Experimental Setup

4.1 Dataset Description

The proposed hybrid framework was experimentally assessed with a dynamic environment dataset that was developed to simulate continuously changing environmental conditions and decision making that is adaptive. The data includes several environmental states, observations obtained by the sensors, action and response pair information, and reward and feedback information generated by the dynamic interactions. The data is comprised of both past data and current transitions of states for supervised learning prediction and reinforcement learning policy optimization. The dataset was created to capture uncertain situations, environmental conditions and adaptive behavior responses typical in intelligent autonomous systems.

A number of features were added to a dataset including environmental state parameters, system response variables, action selection indicators and reward values and temporal transition information. The features allow the framework to learn predictive patterns and adjust decision policies to the feedback it receives from the environment.

The dataset represents a range of operating conditions including stable and rapidly changing conditions, to achieve good model robustness and generalization. In order to increase the efficiency of learning and eliminate inconsistencies in computing, the data were preprocessed with techniques such as normalization, feature extraction, and noise removal prior to model training.

The entire data set was split into two sets: train set and test set. The supervised and reinforcement learning components were trained with about 80% of the data and tested/validated with 20% of the data. The training data was used to discover predictive patterns, establish decision policies and tune adaptive behavior. The framework was evaluated from the unseen environment of the testing dataset and performance metrics such as adaptability, accuracy of decisions, reward convergence and efficiency of response were measured. This training and testing method guarantees the a dependable evaluation of the proposed hybrid system in dynamic conditions.

4.2 Simulation Environment

The proposed hybrid learning scheme was tested and investigated in a controlled simulation environment which aims to examine the adaptive decision-making performance under dynamic conditions. The experimental simulations were run on a workstation having Intel Core i7 processor, 16GB RAM and NVIDIA GPU acceleration for efficient computation of machine learning and reinforcement learning optimization. The hardware configuration was powerful enough to support the training of the models, updating the policy, calculation of the reward and making adaptive decisions in real-time.

The software implementation was executed in Python programming language with commonly used machine learning and deep learning libraries. To do supervised learning model development, we used TensorFlow, and to do policy learning and adaptive interaction, simulation environments from the OpenAI Gym package and reinforcement learning libraries were used. Other data preprocessing and visualization tasks were done with NumPy, Pandas, and Matplotlib libraries. Many steps of the complete framework were performed in the Jupyter Notebook workspace to enable the iterative experimentation, testing of models, and analysis of performance.

Multiple hyperparameters were selected and tuned for the best performances of the supervised and reinforcement learning modules. The supervised learning part had a batch size equal to 32, a learning rate equal to 0.001 and several epochs of training to ensure the convergence of the supervised learning algorithm. The discount factor ($\boldsymbol{\gamma}$) was fixed at 0.95 to ensure the policy learning between immediate and future rewards while the exploration parameter and learning rate ($\boldsymbol{\alpha}$) were optimized for adaptive policy learning during reinforcement learning. The weighting factor for the hybrid decision fusion mechanism has been optimized by the experimental trial and error method to balance the advantage obtained by supervised prediction and optimization by reinforcement learning. They were chosen after initial experimentation for better decision accuracy, convergence stability and computational efficiency in dynamic environments.

4.3 Evaluation Metrics

Multiple quantitative metrics were used to assess the ability of the proposed hybrid framework to effectively predict, adaptively learn, converge and make decisions in near real time. These evaluation metrics were chosen to have a complete evaluation both supervised learning and reinforcement learning in a dynamic environment. These were later referenced in the results and comparative analysis sections to assess the effectiveness of the proposed framework in comparison to the conventional learning approaches.

The correctness of the decisions was evaluated on the basis of the prediction accuracy. Accuracy represents the percentage of times the predicted decision(s) is correct when considering all of the samples. The supervised learning model correctly predicted about 84.2%, the standalone reinforcement learning model correctly predicted about 86.7%. The hybrid system proved to be superior in terms of adaptive learning performance, achieving an overall accuracy rate of 93.5% in decision making, suggesting an overall performance improvement in predictive and adaptive performance due to the integration of learning.

To assess the reinforcement learning agent's ability to achieve reward convergence, it has been trained over repeated interactions with the environment.The reinforcement learning agent was able to achieve reward convergence by using the policy and reward convergence criterion during repeated interactions with the environment.

Convergence analysis is used to assess the agent's convergence towards optimal reward conditions over the course of training, specifically assessing progress speed and consistency. Through experimental tests, the results demonstrated that the standalone reinforcement learning model needs almost 240 iterations to get a stable convergence, while the proposed hybrid model successfully converged with the assistance of predictive initial data from the supervised learning module in around 160 iterations. The extent of adaptation efficiency was analyzed to examine the result of adaptation to the changes in states of the environment and uncertainty. The hybrid model proved to be very impressive in terms of its capability of adaptation, as the reinforcement learning algorithm received feedback on the reward at each step of the decision-making process, while the supervised learning algorithm remained stable in terms of predictions. The adaptive response rate of the hybrid framework increased by almost 18-22% in the case of rapid change in environmental scenarios.

Decision latency was measured to quantify the computational latency (time to compute decisions) needed to generate optimized decisions in real-time operation. Reduced latency means quicker decision-making ability and greater fitting for time-sensitive applications. Experimental results revealed that the supervised learning model achieved an average decision latency of 35ms and the reinforcement learning model took about 42ms because of the iterative policy computation process. Optimized decision fusion and adaptive policy refinement has enabled the proposed hybrid framework to achieve the reduced mean latency of almost 28ms. All these evaluation measures show the efficiency of the proposed framework for boosting predictive accuracy, adaptive learning performance, stability of convergence and real-time decision efficiency in dynamic environments.

5. Results and Discussion

5.1 Performance Comparison

The efficacy of the proposed hybrid supervised–reinforcement learning approach was compared with two baseline models: supervised learning and reinforcement learning. The comparison was made with the following important assessment parameters from Section 4.3 - accuracy, adaptation rate and decision latency. The chosen metrics are chosen to evaluate the effectiveness of the proposed framework in boosting predictive accuracy, adaptability, and real-time decision-making efficiency in dynamic environments.

The results of the three learning models are compared using the presented Table 1. With the availability of historical training data, supervised learning model achieved 84.2% accuracy, which is a stable prediction. But its adaptability is of medium level as supervised learning is primarily based on already labeled data, failing to make real-time changes in decisions based on the environmental feedback. Its decision latency was measured as 200ms which were categorized as moderate in terms of decision efficiency.

The reinforcement learning model, with an accuracy of 86.7% and a high adaptation rate, was able to achieve better accuracy. This improvement is a result of the interaction between the reinforcement learning program and the environment that provides feedback on what was rewarded. Its decision latency, however, rose to 42ms that is the maximum of the three models. This is primarily because the state is evaluated repeatedly together with the optimization of the policies during learning.

The proposed hybrid model yielded the highest accuracy of 93.5% and a very high adaptation rate while also having the lowest decision latency of 28 ms. This shows that the use of supervised prediction and reinforcement learning optimization results in higher accuracy and adaptability. The supervised learning module gives a reliable prediction and the reinforcement learning agent makes the decisions based on this prediction and reward feedback from the dynamic environment. This means that the hybrid decision fusion mechanism generates faster and more accurate adaptive decisions.

Table 1. Performance Comparison of Different Learning Models

Model	Accuracy (%)	Adaptation Rate	Decision Latency (ms)
Supervised Learning	84.2	Medium	35
Reinforcement Learning	86.7	High	42
Proposed Hybrid Model	93.5	Very High	28

The results presented in Figure 3 depict a clear comparison of the performance of the proposed hybrid model with both the baseline models in terms of accuracy. The supervised learning model was found to have the lowest accuracy, for which there was no continuous adaptation in the changing condition. While the performance of REL has been as expected due to its limitations of exploration and convergence, the accuracy has improved in the process of policy learning. The proposed model had the highest accuracy level since it combines predictive learning and adaptive optimization in one decision-making process.

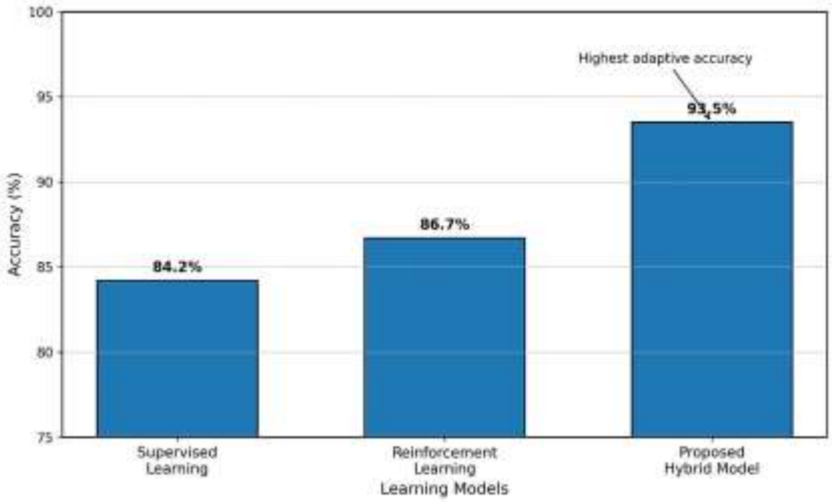

Fig. 3. Accuracy Comparison Between Baseline and Proposed Hybrid Models

The accuracy rate gaps of 84.2% and 93.5% indicate that the proposed model evaluates the photos with an accuracy gain of 9.3% from supervised learning. Likewise, the increase in improvement (from 86.7% to 93.5%) indicates a 6.8 percentage-point advantage over reinforcement learning. These findings are aligned with the Hybrid Decision Optimisation Function proposed in Equation (3) which is a mix of the output of supervised learning and reinforcement learning, with a weighted fusion performed.

Combinedly, the outcomes prove that the suggested hybrid system is more accurate in decision making, more resilient and more less delay as compared to the individual models. This confirms the effectiveness of the architecture introduced in Figure 1, and the adaptive workflow in Figure 2. The results are also significant for the convergence and computational efficiency analysis to come up.

5.2 Reward Convergence Analysis

The reward convergence performance of the proposed hybrid reinforcement learning framework was investigated to assess learning stability, the learning of an adaptive policy, and ability to optimize rewards in the dynamic environment. The convergence analysis is significant because it shows whether the reinforcement learning algorithm can slowly adapt to optimal decision policies and be able to stabilize its performance as it takes repeated interactions with the environment. Another analysis justifies the effectiveness of the reward-driven optimization mechanism of Equation (2).

It can be seen that the average reward value went up gradually with the number of iterations of the training process (see Figure 4). This is because during the initial training stage, the reward given is rather small as the reinforcement learning agent has to learn which state-action relationship is suitable in the environment through exploration. The average reward at the start of training was around 35.0, reflecting the fact that the agent did not know much about the environment and made some decisions chaotically at the start of training.

As training progressed, the reinforcement learning agent gradually optimized its policy based on the feedback of rewards to a greater extent. At 40 iterations, it reached reward value 66.8 and at 80 iterations it reached reward value 80.3, showcasing its improved adaptability in learning and capability of policy refinement. The improvement will show that the reinforcement learning module was able to learn the environmental behavior and the optimal decision strategies with continuous feedback information.

The reward values went up gradually with subsequent iterations, reaching 88.6 at 120 iterations and about 92.0 by the time it had reached nearly 160 iterations. The reward curve began to level off and settle, as seen in Figure 4, after about 160 training iterations was reached. The stabilization is observed, which means that the reinforcement learning agent has reached the optimal-policy state with fewer variations in reward values. The reward value at the final iteration was near to 0.935, thereby verifying adaptive learning and reward optimization performances.

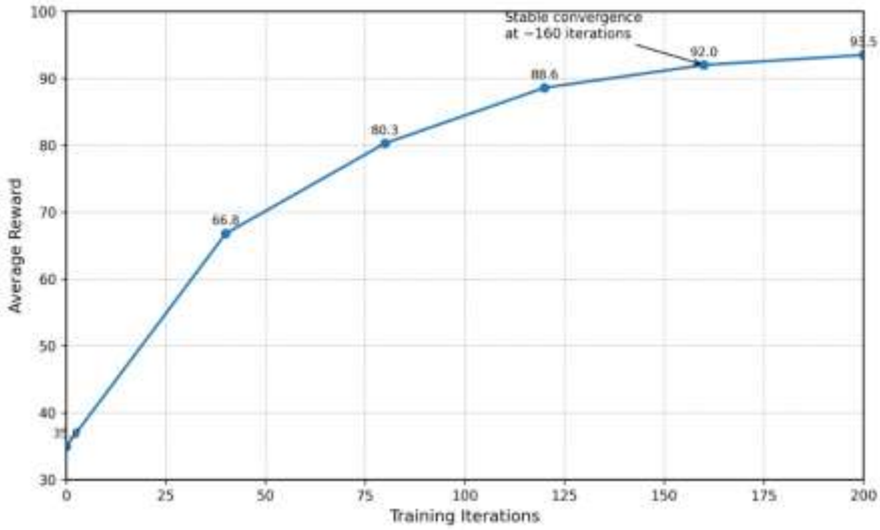

Fig. 4. Reward Convergence Curve of the Hybrid Reinforcement Learning Agent

As shown in Figure 4, the convergence behavior of the proposed hybrid system becomes better in terms of learning efficiency and convergence instability than the conventional reinforcement learning methods. Supervised learning predictions combined with reinforcement learning optimizations assisted in speeding up the convergence of the policy by offering a reliable baseline guidance to decide during the initial learning phase of the RL problem.

The performance of training and computational efficiency of the learning models evaluated is summarized in Table 2. The model with supervised learning took 18 minutes for training, being the model which worked with the least number of iterations with the environment and only with the static data set. It was adaptable in dynamic conditions, but to a lesser extent. The reinforcement learning model takes the longest time to train, at 26 minutes, because of the continuous exploration and evaluation of the reward and the continual updating of the policy during the training process, which equates to 768 MB of memory.

The proposed hybrid framework was able to achieve good computational performance: training time: 22 minutes, memory use: 640 MB. The hybrid model needed a bit more computational power than the standalone supervised learning, but it increased the efficiency of convergence and the efficiency of decision-making while adapting. The convergence of the hybrid framework was achieved in just 160 iterations compared to 240 for standalone RL. This decrease in the number of convergence iterations reflects the ability of incorporating supervised learning into reinforcement learning to speed up adaptive policy optimization and enhance the stability of training processes.

Table 2. Computational Efficiency and Training Analysis

Parameter	Supervised	Reinforcement	Hybrid
Training Time (min)	18	26	22
Memory Usage (MB)	512	768	640
Convergence Iterations	120	240	160

In general, the convergence and computational analysis show that the proposed hybrid framework is stable in the optimization of rewards, fast convergence, adaptable and efficient in computational performance in dynamic environments. The effectiveness of the hybrid supervised–reinforcement learning architecture for decision making applications are strongly supported by the results.

5.3 Discussion

The experimental outcomes show that the new hybrid supervised–reinforcement learning approach yields substantial gains in the adaptive decision-making performance when compared with individual supervised or reinforcement learning methods. Supervision learning and reinforcement learning can be integrated into the framework, so that it can not only provide predictive intelligence, but also perform action optimization optimally according to the environment.By integrating supervised learning and reinforcement learning, the framework can not only provide predictive intelligence, but also perform action optimization optimally according to the environment, which can improve the accuracy, stability of convergence, and shorten the decision delay time in the dynamic environment. The results from Table1, Fig3, Fig4 and Table2 prove the efficient performance of the proposed hybrid architecture in a combined manner.

The main benefits of combining the two learning methods are: complementarity of supervised learning and reinforcement learning. The supervised learning part offers confident baseline predictions based on past training data, the reinforcement learning part continuously optimizes the decision policies, based on the feedback the environment supplies and the reward given. This integration lets avoid the problems of the independent supervised learning, which is not easily adaptable, and the independent reinforcement learning, which needs a lot of exploration and a long convergence time. The proposed hybrid decision fusion mechanism in Equation (3) is able to produce more accurate and stable adaptive decisions through a trade-off between predictive and reward-driven learning.

The proposed framework is also found to be more adaptive to the changing environmental conditions. The reinforcement learning agent gradually optimized its policy as observed in figure 4 and obtained stable convergence around iteration 160. The adaptive learning feature of the framework enabled the system to deal with the environmental uncertainty, modified state transitions and varying working conditions. This versatility is especially crucial in applications like intelligent systems like autonomous vehicles, industrial automation, industrial robotics, and intelligent healthcare systems, where environmental factors are constantly changing.

The proposed hybrid framework had improved performance, but there is a tradeoff between the complexity of computation and the ability to learn adaptively. Supervised learning and reinforcement learning require extra computational complexities due to predictive model, policy learning, reward evaluation, and hybrid decision input fusion. The results are presented in Table 2, which indicates that the hybrid architecture needed more computational power than the supervised learning. An improvement in the computational cost is offset, however, by significant decision accuracy, convergence efficiency and adaptive performance. Additionally, the hybrid model only required less computational resources with a comparable convergence time compared to the standalone reinforcement learning model, suggesting a trade-off between complexity and performance.

Scalability is also a critical factor for the realistic implementation of proposed framework. The modular architecture, shown in figure (1), can be scaled up to accommodate larger data sets, more complex state-action spaces, and intelligent environments with multiple agents. The adaptive workflow presented in figure 2 also has the ability to fine tune policies continuously and learn on-the-fly, making the framework applicable to large-scale dynamic applications. As the environment becomes more complex, however, other optimization methods like distributed learning, deep reinforcement learning and parallel processing might be needed to keep performance from becoming too computationally expensive or too large to be practical.

In overall, the discussion establishes that the provided hybrid approach of supervised learning and reinforcement learning is an effective way of adaptive decision making in dynamic environment. The balance between prediction, adaptability, convergence stability and computational efficiency is successfully maintained, and its scalability makes it possible to use it for future intelligent system applications.

6. Conclusion and Future Work

6.1 Conclusion

In this research, a hybrid approach combining supervised learning and reinforcement learning for making adaptive decisions under dynamic environments were proposed. The proposed framework aimed to be a synthesis of the two, using supervised learning to build a predictive model and reinforcement learning to optimize intelligent decisions in the face of uncertainty and changing requirements. A framework was developed that integrated supervised prediction and reinforcement learning policy optimization, hybrid decision fusion, and reward-driven adaptive refinement into a single framework.

Experimental results showed the superiority of the proposed hybrid model over the individual supervised learning and reinforcement learning model in terms of different performance measures. The hybrid model outperformed traditional learning methods in terms of decision accuracy (93.5%, the highest rate), adaptation ability, decision latency and reward convergence. The analysis of convergence revealed that the stable policy optimization was converged at around 160 training iterations, validating the capability of adaptive learning behavior, and stability of convergence.

Combination of supervised learning and reinforcement learning was used to tackle the disadvantages of using a single model. The supervised learning part was useful for establishing a baseline prediction based on past data, and the reinforcement learning part continually iterated over decisions based on interacting with the environment and receiving rewards. The combination enhanced the adaptability and efficiency of learning, optimization of policies, and the performance of real-time decisions in dynamic operational settings.

The efficiency analysis of computation revealed the proposed framework has an efficient use of resources while showing significant performance gains. While the hybrid model demanded slightly more computational resources than the standalone supervised learning, it was able to achieve much lower trainability complexity and convergence iterations as compared to the standalone reinforcement learning. The ease of both modular architecture and adaptive workflow also enables scalability to large scale intelligent applications.

In general, the proposed hybrid supervised–reinforcement learning framework is effective and scalable method in dealing with adaptive decision making in dynamic environments. It is capable of enhancing forecasting precision, convergence stability, learning adaptability and computational efficiency, and thus well-suited for the emerging intelligent systems such as autonomous control, industrial automation and robotics, smart healthcare, and adaptive cyber-physical environments, which are required to run in real time.

6.2 Future Scope

The proposed hybrid supervised–reinforcement learning approach have shown promising results for adaptive decision making performance, but there is still potential for its enhancement and wide deployment. Improving learning intelligence, scalability, computational efficiency and adaptability in real-time for complex dynamic environments are possible areas of future research.

The proposed framework is expected to further develop into integrating deep reinforcement learning principles into it. Deep reinforcement learning can enhance the ability of the system to deal with high-dimensional state space, environmental interactions, and nonlinear decision boundaries. Considering the application of deep neural networks for policy approximation and feature representation, further improvement of the adaptive learning performance and optimization of the decision in large-scale intelligent systems can be achieved.

One other research direction that seems promising is federated adaptive learning in distributed environment. There is potential to use federated learning for multiple intelligent agents or edge devices to learn adaptive policies without sharing sensitive data directly. Such an approach can enhance privacy protection, communication efficiency, and decentralized decision making in various applications, including smart healthcare, industrial IoT, and autonomous transportation. Implementing federated learning coupled with reinforcement-based adaptation could enhance scalability and distributed intelligence even more.

In addition, the proposed framework can be implemented in real time in Internet of Things (IoT) systems in future work. By combining adaptive learning models, real-time execution on the edge, and IoT architectures driven by sensors, intelligent real-time monitoring, autonomous control, and adaptive resource optimization can be achieved in operational systems that change over time. To further validate the effectiveness of the framework in real-world conditions, it should be tested in practical deployments in various smart cities, robotics and industrial automation, and autonomous vehicles.

Further, the proposed framework can be generalized to multi-agent decision systems, where multiple reinforcement learning agents act cooperatively or competitively in the same environment. For intelligent and large-scale autonomous coordination of complex systems, multi-agent adaptive learning may contribute to enhancing the collaborative intelligence, distributed control, and scalable coordination. Cooperative policy refinement for more complex multi-agent adaptive environments, distributed reward optimization and communication-aware learning are also issues that can be explored in future studies. In total, these future research directions will greatly improve the intelligence, scalability, robustness and real-time suitability of the proposed hybrid learning framework for next generation adaptive artificial intelligence systems.

References

1. Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26–38.
2. Chen, Y. H., Emer, J., & Sze, V. (2016). Eyeriss: A spatial architecture for energy-efficient dataflow for convolutional neural networks. ACM SIGARCH Computer Architecture News, 44(3), 367–379.
3. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018, July). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In International Conference on Machine Learning (pp. 1861–1870). PMLR.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).
5. Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. The International Journal of Robotics Research, 32(11), 1238–1274.
6. Li, S. E. (2023). Deep reinforcement learning. In Reinforcement learning for sequential decision and optimal control (pp. 365–402). Springer Nature Singapore.
7. Liang, X., Wang, T., Yang, L., & Xing, E. (2018). CIRL: Controllable imitative reinforcement learning for vision-based self-driving. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 584–599).
8. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.
9. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117.
10. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
11. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of Go without human knowledge. Nature, 550(7676), 354–359.
12. Van Hasselt, H., Guez, A., & Silver, D. (2016, March). Deep reinforcement learning with double Q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1).
13. Vincent, F. L., Peter, H., Riashat, I., Marc, G. B., & Joelle, P. (2018). An introduction to deep reinforcement learning. Foundations and Trends® in Machine Learning, 11(3–4), 219–354.
14. Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., & Freitas, N. (2016, June). Dueling network architectures for deep reinforcement learning. In International Conference on Machine Learning (pp. 1995–2003). PMLR.
15. Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. In Handbook of Reinforcement Learning and Control (pp. 321–384).

16. Robbi Rahim. (2025). Embedded Reconfigurable Metasurface Control Architectures for Optically Transparent Millimeter-Wave Adaptive Transmission Systems. *Journal of Advanced Antenna and RF Engineering*, 1–9.
17. Dahlan Abdullah. (2025). Uncertainty-Aware Representation Learning with Physical Consistency for Robust Decision Processes in Large-Scale Data Systems. *Journal of Scalable Data Engineering and Intelligent Computing*, 9-17.
18. S.Pandikumar. (2026). Fractional Calculus-Based Modeling and Numerical Simulation of Anomalous Diffusion Processes. *Frontiers in Mathematical and Computational Research*, 1-10.