Volume: 5, Issue: 3
Page: 174-189
2021

International Journal of
Science and Business
Journal homepage: ijsab.com/ijsb

IJSAB International

Reinforcement Learning and Modeling Techniques: A Review

Hindreen Rashid Abdulqadir & Adnan Mohsin Abdulazeez

ABSTRACT

The Reinforcement learning (RL) algorithms solve a wide range of problems we faced. The topic of RL has achieved a new, complete standard of public opinion. High difficulty in large-scale real-world implementations is the effective use of large data sets previously obtained in augmented learning algorithms. Q-learning (QL), by learning a conservative Q function that allows a policy to be below the predicted value of the Q function, is introduced by us, which aims to circumvent these restrictions. We revealed technical reinforcement learning in this study. In principle, we demonstrate that QL creates a lower relation to current policy importance and that this can be correlated with guarantees of political learning theoretical change. In reality, QL strengthens the benchmark objective with a simple, standardized Q value which, in addition to existing Q-learning and essential applications, is quickly applied. The findings indicate that all algorithms are needed to learn how to play successfully. In comparison, all dual Q-learning variables have a significantly higher score compared with Q-learning, and the incremental reward function shows no improved effects than the normal reward function. We present an attack mechanism that uses the portability of competing tests to execute policy incentives and to prove their usefulness and consequences by means of a pilot study of a play learning scenario.

Keywords: *Machine learning, Reinforcement learning, Modelling – Technique, Q- learning.*

IJSB
Literature Review
Accepted 11 February 2021
Published 16 February 2021
DOI: 10.5281/zenodo.4542638

About Author (s)

Hindreen Rashid Abdulqadir (corresponding author), Information Technology Department, Akre Technical College of Informatics, Duhok Polytechnic University, Duhok Kurdistan Region, Iraq. Email: Hindreen.rashid@dpu.edu.krd

Professor Adnan Mohsin Abdulazeez, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq. E-mail: adnan.mohsin@dpu.edu.krd

1. INTRODUCTION

Machine learning Applications of machine learning instruments to problems of physical interest are often criticized at the expense of transparency for generating sensitivity (Zeebaree et al., 2019; Maulud & Abdulazeez, 2020) . To answer this issue, we investigate a method of data preparation to classify combinations of variables that can discriminate against signals from the context, aided by physical intuition (Chang et al., 2018). First of all, supervised learning is useful for predicting or classifying a specific outcome of interest. We have three types of machine learning (Jiang et al., 2020; Powell et al., 2020). Secondly, Unsupervised ML deals with unlabeled data that includes the excoriates, but not the results, defined as data. To recognize trends and correlations between data points, this technique is used (Wang & Filipi, 2020). As a first step in the analysis, dimensional reduction methods can be applied. The following sections explain many commonly used algorithms in unsupervised ML (Talevi et al., 2020; Zeebaree et al., 2017). Third, reinforcement learning focuses on paper.

Reinforcement learning is the study of how to utilize historical data to improve the future exploitation of a dynamic structure. How would that happen to regular computer learning? Where the dynamics are unclear, the survey focuses on reinforcement learning as maximum regulation (Recht, 2018). Reinforcement learning algorithm to date, the main formalism used in RL has been implemented, and some problems in RL have been briefly noted. In the next step, we will differentiate between different groups of (Arulkumaran et al., 2017; Zhang and Han., 2018; Hein et al., 2017). In RL, four fundamental components exist agent, environment, incentive, and behavior. RL's aim is for the agent to maximize the incentive in response to a changing situation by taking a sequence of actions. Q-learning (Pan & Wu, 2018; Adeen et al., 2020; Debnath et al., 2018). An animal engages with its surroundings by model-based learning through biological learning and aims to optimize its perceived integrated reward strategies(Hein et al., 2018; Cazé et al., 2018). Q-learning a method of time differential learning to maximize the political position of each nation with respect to trade parameters. By using the neural network, Q-Learning is implemented by doing approximation. Created to teach stock market behavior and strategy to traders, (Brim, 2020; Mahmood & Abdulazeez, 2018). Q-Learning Algorithm Q-Learning is an algorithm for the learning of time-diversity (TD), a further case of enhanced learning and one of the most important enhanced learning algorithms(Guo et al., 2004; Hasselt, et al., 2016; Sadiq et al., 2020).

This paper offers a thorough analysis of the new and most efficient methods made by researchers over the past three years to decision-making trees in various areas of machine learning. The particulars of each methodology are also outlined, such as the use of algorithms/approaches, datasets and results. In comparison, we have illustrated the most widely used techniques and the highest precision methods achieved. The arrangement of the remaining document shall be as follows. Section 2 includes a Reinforcement Learning algorithm that lists its forms, benefits, and drawbacks; Section 3 gives a related work on Reinforcement Learning Algorithm; Section 4 comparison and discussion on the Reinforcement Learning, and the last section conclude the research work.

2. BACKGROUND THEORY

2.1 MACHINE LEARNING (ML)

It is the area in which different computer algorithms are studied that is progressing steadily. Within machine learning, Classified are supervised schooling, unattended learning, semi-supervised learning, and enhanced learning. Learning supervised is a machine learning task that has a part to play in the specified training data (Sulaiman et al., 2019). Data is not classified as unsupervised learning; we have unmarked information. Semi-supervised

learning is a fusion of knowledge that is graded and not labelled. In order to improve the programmer's learning, the agent gathers from the contact with the world to take measures to optimize rewards. Figure 1 reflects the classification of the machine learning system(Brifcani & Brifcani, 2010; Thomas & Gupta, 2020; Maulud & Abdulazeez, 2020). Many powerful applications of gene expression analysis have been proved by machine learning and data mining. ML is used in all fields of computational work where algorithms are designed and performance is improved (Abdulqader et al., 2020). The standard machine learning strategies are typically disappointing to perform well for cancer classification (Ahmed & Brifcani, 2019). On the theory of probability and statistics, machine learning is based (Li, 2018; Moerland et al., 2020; Zebari et al., 2020).

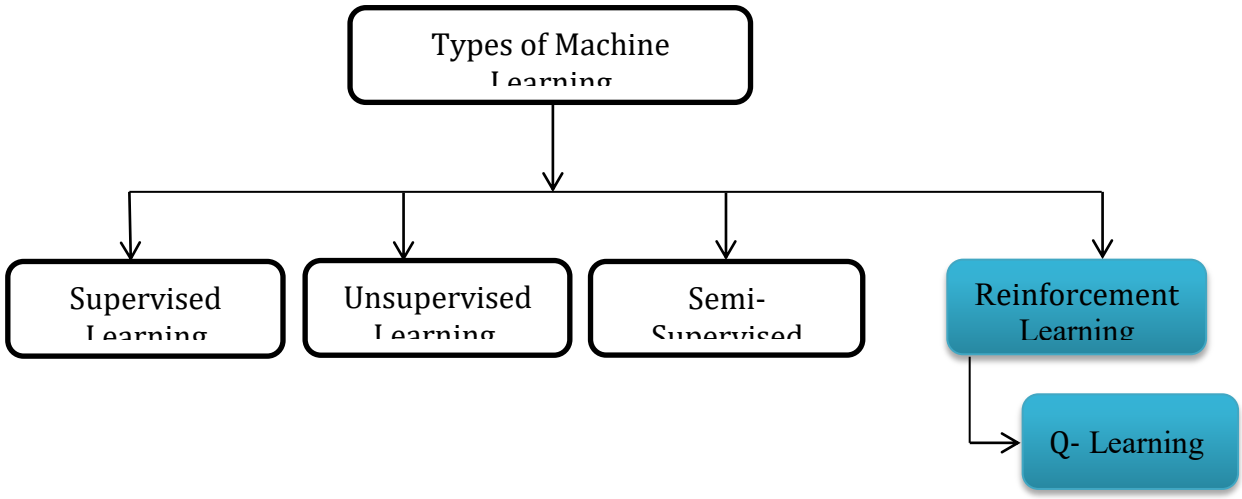

Figure1 : Algorithms of Machine Learning Classification (Thomas & Gupta, 2020).

2.2 LEARNING TO IMPROVE BUILDING CONTROLS

Reinforcement learning is a significant subfield of machine learning that deals with sequential decision-making. In terms of the forms of input that the agent/algorithm can receive after making a decision/prediction, as seen in Figure 1, the three classes of machine learning problems vary from one another. The agent would automatically know how reliable his forecast is relative to the ground reality given for supervised learning by the label results. And this knowledge will be used to upgrade and improve the indicator (Salih Hassan et al., 2018). No input is provided for unsupervised learning as the dataset is unmarked. Strengthening learning sits at the heart of the two examples that generate delayed input (Wang & Hong, 2020; Lopez-Martin et al., 2020).

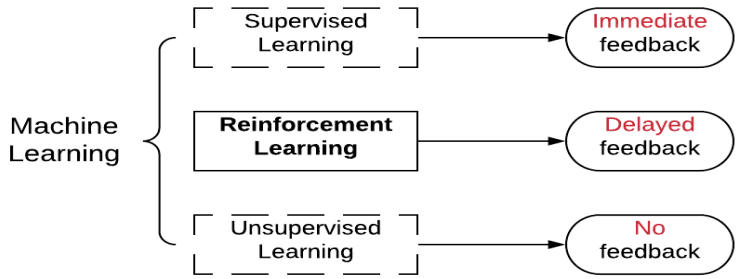

Figure 2: Three kinds of problems with machine learning (Z. Wang & Hong, 2020).

2.3 REINFORCEMENT LEARNING (RL)

Reinforcement is one of the ways that the program agent takes steps to optimize awards in machine learning, where reinforcement is gained through interaction with the world. As a Markov decision system, the universe is formulated (Kintsakis et al., 2019). There is no supply of reinforcement learning input/output variables. The software agent will acquire the

data, the actual environmental condition, and then the performance behavior is determined by the program agent. The scalar feedback signals, the state change values and the state of the environment that are altered by the behavior of the software agent are transmitted. Strengthening learning advises the automated agent to compensate its corresponding condition after the action has been picked. The software agent is not advised which activity is going to be better in terms of long-term interest. The program agent needs to gather information for optimal state function, actions, transition, and rewards. The algorithms used for reinforcement learning are reviewed in this section (Thomas & Gupta, 2020; Valladares et al., 2019) . Whose agents benefit from the optimal set of actions, that is, by their interaction with the environment, the optimal strategy (see Figure. 3). We're briefly analyzing it here. The interested reader for a detailed introduction (Vázquez-Canteli & Nagy, 2019; Kasgari et al., 2020).

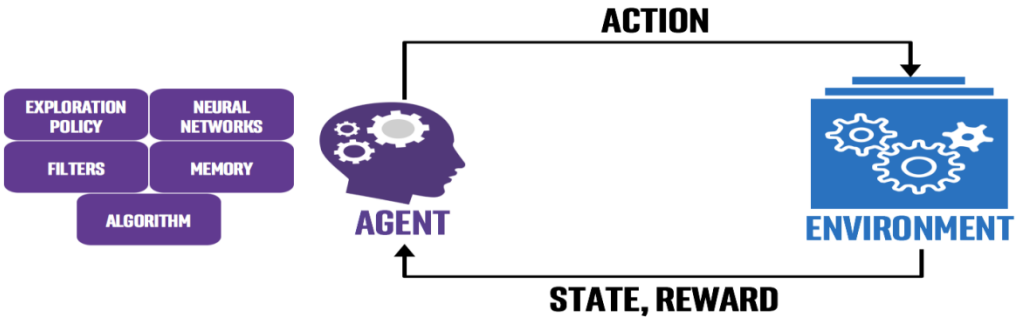

Figure 3: Reinforcement learning (Suerich & Young, 2020).

2.4 APPLICATIONS OF REINFORCEMENT LEARNING

Reinforcement learning can overcome a number of issues. Any of these areas are gambling, robots and several more. In the optimization of chemical reactions, reinforcement education may also be utilized (Shastha et al., 2019; Akalin & Loutfi, 2020) .

Figure 4: Applications of Reinforcement Learning (Shastha et al., 2019).

2.5 INTEGRATING MODEL-FREE AND MODEL-BASED METHODS

We now explain how advantages can be accomplished Worlds by the convergence of learning and preparation in one end-to-end training phase, in order to obtain an effective and performing algorithm both in processing time and inefficiency. Figure 5, displays a Venn diagram of the numerous variations. One direct solution is to use the tree when the model is open. Search approaches using both merit networks and policy networks. The main property of the model is to provide an algorithm that generalizes well when the model is inaccessible and the agent only has access to a small range of trajectories. One alternative is to construct a

specification that will be used to produce additional samples of an algorithm for model-free strengthening (Francois-Lavet et al., 2018).

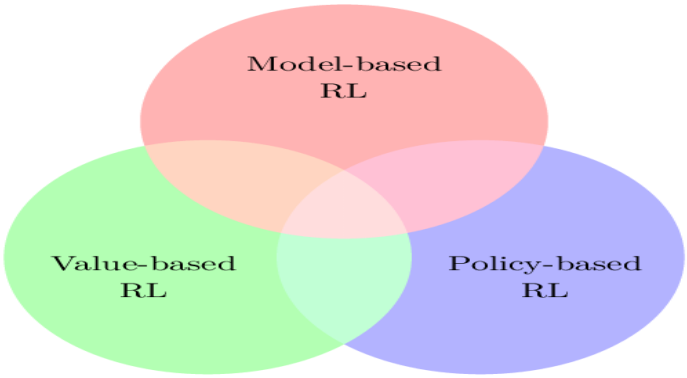

Figure 5: Venn diagram in the space of possible RL algorithms(Francois-Lavet et al., 2018).

2.6 BENEFITS OF REINFORCEMENT LEARNING

First, in truth, supervised learning imitates who supplied the data for the algorithm. Increasing learning will create totally different strategies, unlike improving learning. Second, strengthened learning algorithms are better instruments for discovering answers without bigotry or partiality. Third, successful time improving curriculum takes place in actual time. This implies that it will generate performance while also developing other algorithms. Fourth, for sequences of behavior aim-oriented learning may be used. The most widely employed input-output method is guided learning. Finally, reinforcement learning needs no reconstruction and it immediately adapts to different flight conditions. Increased learning doesn't, unlike supervised learning algorithms (Francois-Lavet et al., 2018; Renaudo et al., 2015).

2.7 REINFORCEMENT LEARNING MODELING: STATE, ACTION, REWARD

The learner, also known as the Agent, communicates with his or her surroundings and decides to adapt his or her behavior to the world in accordance with his or her present circumstance and the environmental support he or she gathers. (Figure.2). For example, in order to make routing decisions, a router communicates with its neighboring nodes.(Chen et al., 2020; Zhao et al., 2019). In such a case, the agent is a router, the field is the vicinity of the router, and the next neighbor nodes to send data packets are chosen. The reward functions are the basis of the RL algorithms. The function of rewards returned to the agent by the environment is to provide insight into the learning algorithm on the effects of the recent action taken (Soni et al., 2019; Xie et al., 2020; Wang et al., 2020). Here, as a reward feature reveals in an immediate way what is good (or bad), a benefit feature indicates what is good in the long term (or bad)(Mammeri, 2019; Mammeri, 2019) .

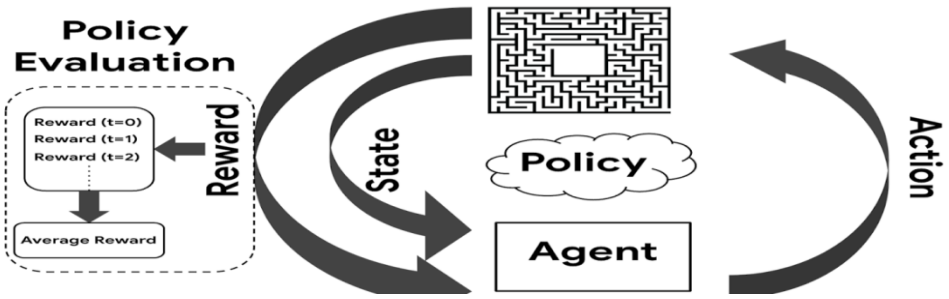

Figure 6: Reinforcement learning Modeling. State, Action, Reward (Suerich & Young, 2020; Yin & Wang, 2020) .

2.8 THE REINFORCEMENT METHOD OF LEARNING

The theory of improving learning evolved from the term "optimal control" that originated at the end of the 1950sWith the controller's configuration to reduce the approximation of the device's output over time, the problem was formulated. The notion of Markov (MDP), or finite MDPs, was developed as a fundamental principle of RL to express optimal control problems(Rocchetta et al., 2019). To maximize the numerical delayed incentive signal, the learner or RL agent knows how to map conditions to actions. It does not require a "teacher" to tell him how to take any action, but rather to make choices by trying trial and error and knowing the action (Pervaiz & Jäntti, 2020). A benefit from the environment the agent is communicating in (Perera et al., 2020; Soni et al., 2019). No RL, the third form of computer education, is neither regulated nor unregulated education. Instead of collecting signals for good behavior, directed learning generates signals from a reward of behavior without knowing whether the action was right or not. RL, In a sense, the nature of machine learning is hidden. RL lets an entity understand actions spontaneously in the sense of artificial intelligence, which cannot be carried out through controlled or unattended learning (Han et al., 2019; Pan & Wu, 2018)

2.9 Q-LEARNING

It's a kind of model-free schooling for reinforcing. It can also be remembered as a dynamic approach to asynchronous programming (DP). In the Markovian region, Q-learning allows agents to learn how to exemplify domain maps by realizing the results of their behavior that no longer allow them to create domain maps. A Q- learning algorithm is a form of learning time gap that forecasts a sum dependent on the future values of the signal. The agent is in the state of st at each point, chooses an action at, and travels to the next state st+1 thus receiving the rt reward. The aim is to optimize the cumulative reward intake and Q-learning through the use of experience. (s_t, a_t, r_t, s_t+1) To understand the significance of the role of state-action, Q. (s, a). Quantity Q(s, a) is a measure of an agent's potential probability of a total sum of discounted rewards and the program is implemented subsequently. The Q-learning law is updated in general: As seen in Equation (1),

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t)+\alpha[r_t +\gamma \max Q(s_t+1, a)-Q(s_t, a_t) \quad (1)$$

Where the

s_t= Time State
a_t = Action Time
r_t = Reward Time
γ= discount factor

This equation has two constants that influence it. The value between 0 and 1 is the constant alpha that denotes the learning rate and affects the degree to which the Q-values are updated during the process. The value between 0 and 1 is also the constant Δ, called the discount factor, and how many future gains can be accomplished with Q worth changes. A short-term Discount Factor of near 0 implies that a discount factor of close to 1 would in the short term allow the agent to more accurately offset the benefits earned in the distant future. The Q-learning algorithm uses a state-action lookup table in order to store and use data. The Q-value, which represents the consistency of the decision, is associated with each pair of State-Actions The Q values are based on the benefit gained by the selection of activity plus the overall possible reward predicted. We use another method, which combines the Q-learning with a thoroughly specified estimated function since our status space is too big to hold all Q values in a table (Schilperoort et al., 2018; Stember & Shalu, 2020).

We can learn estimations to solve sequential decision problems for the optimum value of each action, described as the estimated amount of potential incentives for this action, and then

follow the optimal strategy. Pursuant to this procedure.π, the real worth of a state s, as shown in equation (2).

$Q\pi(s, a) \equiv E [R1 + \gamma R2 + \ldots | S_0 = s, A_0 = a, \pi]$ (2)

If $\gamma \in [0, 1]$is a discount vector that varies immediately and subsequently from the value of the rewards. The highest value is indicated.

Then $Q*(s, a) = \max\pi\ Q\pi(s, a)$. By selecting the highest action value in - state, the optimum strategy is easily derived from optimal values. Using Q-learning, a form of time difference in learning, estimates of optimal action values may be learned. Any of the interesting topics are too significant for all-action principles to be learned independently in all nations. We should learn the parameterized value function instead. $Q(s, a; \theta_t)$. Standard Update of
Q –(Raghu et al., 2017). Learning parameters after in-state St behavior and observation of Instant compensation Rt+1 and the resulting state of S_t+1. As shown in equation (3).

$\theta_t+1 = \theta_t+\alpha(Y^{Q}{}_t - Q(S_t,A_t; \theta_t))\nabla\ \theta_t Q(S_t,A_t; \theta_t)$. (3)

Where a is a scalar phase size and the target YQt is defined as seen in the equation. (4).
$Y^{Q}{}_t \equiv R_t+1 + \gamma\ \max a\ Q(St+1, a; \theta_t)$. (4)
This update is reminiscent of stochastic gradient regression, upgrading The present value is $Q(S_t,A_t; \theta_t)$ to the goal value $Y^{Q}{}_t$. (Schilperoort et al., 2018; Brim, 2020;).

Algorithm 1: Q-learning algorithm. Scan algorithm description in section 2.

```
---------------------------------------------------------------------------
initialize s and Q
repeat
if explore() then
a ← randomMove()
else
a ← arg maxa Q(s, a)
end if
s∗ ← newState(s, a)
r ← reward(s, a, s∗)
Qtarget(s, a) ← r + γ maxa Q(s∗, a)
update(Q(s, a),Qtarget(s, a))
until end
---------------------------------------------------------------------------
```

$Qtarget(s_t, a_t) \leftarrow r_t + \gamma\ \max a\ Q(s_t+1, a)$

3. RELATED WORK

Reinforcement Learning has been utilized in many different implementations of Modeling Techniques, in this section; we will study some of its utilizations and how it has performed under different implementations and approaches. Han et al. (2020) Proposed The creation of a hierarchical granular computing model for long-term IP generation, where probabilistic simulation leads to a long number estimation horizon, and hierarchically expands the result to an interval-rated format, by using angular results. In order to build long-term IPs for the steel by-product gas system a method based on HGrC where a multi-input one-output relationship is developed. Granules were created based on the data. Experimental findings indicate that the proposed solution is preferable to the most used methods of the framework

and that the performance of the learning structure is higher than other improvement methods, guy. Saleh et al. (2020) Argued that the paper makes the following contributions: (a) Creates a new hierarchical control strategy, VHRL; (b) Shows the usefulness of a number of dialog models; Trains open-domain VHRL and self-play dialog models, demonstrates changes to state-of-the-art dialog architectures for both human and automated measurements; and (c) Provides and contrasts metrics; Repeatable, and more enjoyable, constructive, user-dependent features to guide conversations that are less toxic and less toxic. It is the review. It can lead to an unacceptable, biased, or aggressive generation of text goals and training on regular films or online datasets. We have seen that our proposal The VHRL strategy is more effective for long-term optimization, theoretically solving these problems. Increased metrics such as toxicity are incentives for conversation and increase the consistency of conversation. Zhang et al. (2020) Explained Multi-Agent Enhancing Learning (MARL) was commonly used in many applications for its workable implementation and performance. In order to obtain an ideal common behavior or equilibrium in any mode of action, automatic learning which can be categorized into a separate apprentice category under the Marl shall be used. Repeat observational evidence reveals that LA-OCA converges. Ample exploration has been given for efficient collective action. LA-OCA has an impressive approach in terms of stochastic games for both games. Of the three tasks with a 100% success rate and the other algorithms in terms of learning level. Simulation findings indicate that for all three cooperative workloads and execute additional algorithms as regards learners pace, LA-OCA has a purely optimally organized approach to. Chen et al. (2021) Explained The current Midfield Behavior Scheme (MFTRPO), a UAV management system that uses the medium-field approach for constructing a Hamilton Jacobi-Bellman/Fokker-Planck-Kolmogorov equation that offers ideal solutions and solves practical problems in applying the faith area by improving trust regions and incorporating the capabilities of neural networks Interplaying multiple objectives with multiple UAVs contributes to a wide modern room which makes it difficult to achieve practical applications on a large scale. For the intermediate game, we model the UAV control problem to simplify complex interactions (MFG). The effects of the simulation have shown that MFTRPO beats the two narrowly applied basic methods of coverage, equity, and energy efficiency significantly and consistently. Aind et al. (2020) Introduced Our new algorithm, Q-Bully, can be used to track cyber-bullying on various social networking and online gaming sites utilizing natural language reinforcement learning techniques. The application of reinforcement learning and duration studies to feed the input and messages of bullies and victims have also been integrated. Description learning agent's improvement. Description we equate our model to other models based on F1 (0.86 16K annotated datasets) and on F2 ratings. Our model is related to the other models. Might claim our platform is crushing other cutting-edge ones. Quite dynamic and populous structures of datasets. Conditions meant to mislead the traditional identity scheme Conditions.

Mukherjee et al. (2020) Explained the concept of a model-free power control device with an undefined network where the system has been introduced to disrupt ambient and triggered oscillations. In presence of external disruption inputs, build continuous WACs based on RLs, especially with two types of recurrent oscillations found in the device. Our proposed RL algorithm restores the maximum possible input. React completely in view of all of these interruptions On-line calculation, data, and unrest model-free application. Simulation of multiple effects Situations demonstrates the various complexities of the system. At home, we expect that the proposed definition will be applied to potential research. Reduced-size learning through structured energy system models with a structure may be used to save learning time. Yang and Wang (2020) developed the incentive function for RL to prevent fake dissemination problems induced by the clustering sampling of Gibbs. We add to the RL case

the desired subject modeling techniques and use the policy quest with the Gibbs EM parameter estimation algorithm. We model the clustering purpose as an MDP in conventional theme models instead of a generative method and suggest TAM with RL for fine-grain themes. The ranking for F1 and the proposed uniform shared information-F1 are used for clustering and theme creation steps respectively. Our testing found that TAM would surpass the leading-edge models to obtain an average binary cluster gain of 25.7%, which is the F1 performance.

Du & Li (2020) explained the Deep Neural Network (DNN) and Model-Free Enhancing Learning Technologies are suggested for smart multi Micro Grid (MMG) Storage Capacity Methods, indeed. There have been several tested micro grids Linked with the major supply chain and the delivery system's buying force to maintain local use. In order to increase power sales benefit and reduce PAR. The simulation results suggest that due to its automated extraction capacity, the DNN regression model is quite exact. The numerical efficiency of the traditional model approach often exudes high generalization. Da Silva et al. (2020) Introduced a modern RA method combined with the potential to test uncertainty and spectral effectiveness for Q Learning. The solution suggested is to use MTC devices Select time slots and move power to improve efficiency. A minimum of extra sophistication is needed for the solution since only one simple equation must be used. 3 K numeric values are introduced and processed when the system stops the transmitting capability from being undesirable. We suggest a clustered approach for hierarchical assigning of RA slots on MTC computers focused on non-orthogonal Multiple Access (NOMA) and Q-Learn. Numerical findings indicate that, relative to recent work, the new approach would dramatically increase the network performance. Zhang et al., (2020) explained the Q-learning has been claimed to be a kind of learning process. In a number of charging applications, compliance learning (RL) is used. Oh, sir, scenarios. The Q-learning methodology suggested raises the calculation of conventional methods of artificial intelligence, such as the Recurring Neural Network (RNN), and the Artificial Network (ANN). Results indicate that PHEV loads can be accurate predictions utilizing the under-three Q-learning approach. Similar conditions (smart, uncoordinated, and coordinated). In the future, the proposed forecasting strategies could be assisted by a broad power grid with more complicated PHEV load specifications. The Q-learning technic, as can be seen in the simulation results of the method, is capable of estimating tons of PHEVs more specifically than those of the RNN and ANN techniques (i.e. smart, smart). Loading) showed more than that in the Q-learning process. 50% more than traditional technology ANN and RNN. As has been demonstrated, procedures can maximize precision by utilizing a larger number of Iterations using the ANN method. (MSE) in comparison with the RNN process, in the ANN technique, PHEV filling measurement. Lavaei et al. (2020) Introduced This strategy would refer to Markov's continuous policies. The purpose of this research project is to obtain Algorithms for MDPs that can ultimately calculate the MDPs Finite State MDBs, eliminating the need to specifically specify the state space of the MDPs. Of the remainder of this article, this paper is structured as follows. We include context descriptions and notation for the principle of LUCC in the next portion. Next, we present the key topic in the sense of "Then". These findings are especially significant for computer vision. Finally, we illustrate the effectiveness of our proposed findings by implementing our own methodologies for MDPs for several physical and state parameters in the final portion. The purpose of this strategy is to optimize the likelihood that this method satisfies a simultaneous formula.

Lei et al. (2020) presented templates, implementations, and complexities in AioT structures. A description of how new RL / DRL approaches are being used. first, implement an AIoT tutorial, then gave recommendations for real-world implementations of IoT. Next is a survey of related works on AIoT which is added to the group of writers. As stated above, future

researches are expected on the topic. The combining of the Internet of Things (IoT) and successful sampling result on the Internet of Independent Things (IoT) (AIoT). Sensors gather various kinds of data in real-time from several sources and those data are processed to turn into an intelligent signal to render necessary behavior. For freedom, problems and concerns that are available to the future to meet are deliberately decided. Chang et al. (2019) Explained the New methods of building campuses are explored in this article. It is to measure whether there is a reciprocal interaction between architecture variables and urban efficiency variables, including energy demands, solar harvesting capacity and sky view factor. The key objective of this research is to help urban planning to create an energy effective and visually competent urban setting . What this study is doing is introducing a new technique applying a generative architecture strategy by applying reinforcement learning algorithms to a multivariate adaptive regression splines system to define associations between design parameters and urban results. The solar potential can be preserved by covering the region of one sun with the field of two sun. In order to preserve our optimum energy balance, the recommended threshold for sky view factor is 54.17%. It enables one to extract planning techniques and recommendations to design a sustainable campus. Lillicrap et al. (2019) Argued that We apply the concepts behind the performance of Deep Q-Learning to the continuous action domain. We propose an actor-critic algorithm that includes continuous action spaces. Our device will overcome the same physics problems using the same algorithm with the same network design and hyper-parameters to leverage the strength of the learning technique. Our novel algorithm may identify solutions that need fewer measures of experience than DQN, offering more stable results even in high uncertainty Atari games. What's more interesting is that nearly all the issues of our scenario were addressed by the 2.5 million encounters. We have a strategy that will reduce overhead and raise sales. These findings showed that consistent learning can be attained without any modifications between environments. Anderson et al. (2018) The reinforcement learning (RL) methodology was launched. To get the awareness of how antivirus models work upon violating machine learning on a constant basis. The firmware identification is an essential component in the protection community before enabling entry to the device. Our motivation for studying the anti-software has two goals: to develop an effective and secure software "machine learning", and to target the intent of machine learning software to create samples Here are several new steps for achieving Self Control by applying reinforcement learning (RL) to automation. Besides, training sample evasive ransomware samples has a big gap. An Open-air gymnasium has been launched for researchers to review artificial intelligence, malware samples, and their artificial neural networks agent. It would be important for our analysis in the future.

Table 1: Summary of Literature Review Related of Reinforcement Learning Algorithm.

Use reference	Datasets	Objective	Result and accuracy
(Z. Han et al., 2020)	----------	In order to construct long-term IPs for the steel industry by-product gas grid, a method based on HGrC has been proposed where there is a relationship of multi-input-single-output In view of the details; the probability of it has been developed in granules.	Experimental results suggest that the proposed approach is similar to more commonly used methods and that the accuracy of the learning structure is greater than that of other reinforcement learning structures.
Saleh et al. (2020)	movie dialogs	Objectives and instructions on routine films or on-line datasets This could be linked to an inappropriate, biased or violent age. Well. Yeah. Address, please. These issues are likely to be fixed.	We have shown that our plan for a VHRL approach is the most effective for long-term optimization. Conversation advantages and increased quality of conversation Improved metrics such as toxicity.
(Zhang et al., 2020a)	(Multiagent reinforceme	Repeat observational evidence reveals that LA-OCA converges. Enhanced	Simulation results demonstrate that LA-OCA is 100% effective across all three cooperative

	nt learning)MARL Algorithms	collective action exploration gave. LA-OCA provides an ideal solution for all sports in the case of Stochastic games. Of the three works, the success rate is 100% and the speed of learning beats other algorithms.	workflows with pure and competitive strategy, which outperforms other learning speed algorithms.
(Chen et al., 2021)	Unmanned Aerial Vehicle (UAV) networks, aerial Base Stations (UAV-BSs)	The interaction between multi-targets and multiple UAVs makes it difficult for massive, practical applications to build a comprehensive state-of-the-art space. Ode guided the midfield game to simplify dynamic relationships in UAV control issues (MFG).	Simulation results indicate that MFTRPO exceeds two widely employed basic approaches, coverage, fairness, and energy usage, to a large degree and reliably.
(Aind et al., 2020)	social media websites.	We also involve the usage of motivation training and an educational study in which the bullies and the victims' remarks and messages were fed. An expanded learning agent overview.	We equate our model to other models based on F1 (0.86 16K annotated datasets) and on F2 ratings. Our model is related to the other models. Can say that our technology overtakes other cutting-edge systems. Quite complicated and very crowded database versions. Words that are purposely fooled to trick the traditional detection of the device.
(Mukherjee et al., 2020)	online measureme nts of the states	Optimal feedback is restored by our proposed RL algorithm. Absolute response in view of all these disruptions Model-free use of online measures of the economy, inputs, And with the unrest.	Simulation of multiple consequences The scenarios explains the various complexities of the architecture. We hope to apply the suggested concept to possible experiments in our home. Reduced-dimensional learning can be used to save learning time by organized simulations of power systems with a framework.
Yang & Wang, (2020)	Tweetset and Google News	In conventional theme models, we model the clustering target as an MDP activity rather than a generational strategy.	Score metrics for F1 and F1. The proposed uniform shared information-F1 is used for clustering and theme generation evaluation purposes, respectively. Ours, ours Experiments have shown that TAM will surpass the condition of the-Art-models-F1 score for binary clustering, explicitly achieving an improvement of 25.7 per cent on average.
Du & Li (2020)	Historical market price evidence and sharing of electricity Data.	To increase the benefit from selling power and to minimize PAR.	Simulation findings demonstrate that, because of its automated extraction capabilities, the DNN regression model is very precise. It also exudes the high generalization of numerical efficiency of the traditional model-based system.
da Silva et al., (2020)		For the method, a minimum of additional complexity is appropriate on the device side since only one equation is required. Implement and store 3 K numerical values, mitigating an insufficient transmission capability for the device.	We suggest a distributed system of allocating RA slots hierarchically to MTC devices based on NOMA and Q-Learn. Compared to recent studies, several results indicate that the latest approach dramatically increased network efficiency.
(Zhang et al., 2020b)	Time series	Proposed forecasting methods may be used in future large-scale power grids and could be implemented. Criteria: The criteria.	This is shown in the simulation results that the Q-learning methodology estimates the PHEVs' load more accurately than that of the ANN and RNN techniques. According to the details in Table IV, in the worst situation, PHEV is the one most susceptible (i.e. smart, smart). Q-learning has proved to be more efficient than logistic regression. General advances are relative to traditional ANN and RNN technologies. Techniques-well. Contrary to the popular assumption, using further successive iterations of an ANN would improve accuracy. (MSE) of ANN technique relative to the one-armed RNN technique (MSE), is able to produce more modified performance.

(Akalin & Loutfi, 2020)	Fuzzification of emotions	This paper attempts to allow researchers involved in utilizing and implement methods of reinforcement learning as a starting point In this basic area of study.	The paper focuses especially on studies that involve social physical robots and real-world interactions between human robots and users. We often define current RL methods, depending on the nature of the award structures, in addition to a sample.
(Babaeizadeh et al., 2017)	Environment	The study of the numerical dimensions of RL algorithms can be called a consistent subject in the future for RL. GA3C scales are far higher than the DNN.	We allow other researchers to further explore this field, review deep-RL algorithms in-depth, and test new algorithms, including strategies for the combined use of CPU/GPU computing resources.
(Rahimi et al., 2018)	Graphical Explanation	A complex world composed of static artifacts, a static target location, a dynamic panel position, and a dynamic agent position has implemented all sides of the algorithms.	The four algorithms are evaluated in a simulated setting and their output is contrasted with graphs from test reports. The comparison reveals that the reinforcement algorithm recently used outperforms the current algorithms in a complex set in the robot box moving the problem.
(Arabnejad et al., 2017)	OpenStack platform	The main question is how and how to add/remove services to satisfy commitments on the negotiated service standard. SLAs are two crucial factors of dynamical regulating architecture that minimize application expense and guarantee service-level arrangements. We compare in this article two complex learning methods that use a fluid logic system to learn and change the flowing scaling laws.	The experimental findings indicate that FSL and FQL work in the correct number of virtual machines to maximize enforcement with and answer times for SLA.

4. DISCUSSION

This part is a discussion about Machine Learning Q-learning selection reinforcement algorithms. Strengthening learning (RL) algorithms can successfully solve a wide range of problems that we have encountered. The issue of RL has reached a new, completed level of public opinion. Today, machine learning is an important part of solving the problem of many data sets, such as gaming. The best way to find the accuracy of the data set is to combine two supervised algorithms because all of the above methods using the combination will have the highest percentage. All algorithms, on the other hand, can be used. Furthermore, all dual Q-Learning variables have a slightly higher output than Q-Learning, and there are no better outcomes than the normal reward function in the incremental reward function pilot analysis of the game-learning scenario proposes an attack mechanism that exploits the portability of adversarial tests in order to carry out policy rewards and to demonstrate their efficacy and effects. With respect to (Han et al., 2019; Schilperoort et al., 2018 ; Stember & Shalu, 2020) allowed substantial discovery for optimal cooperative action. In this game, LA-OCA has a basic strategy. This research utilizes Hamilton-Jacobi-Bellman/Fokker-Plan Q-Bully algorithm to track the phenomena called cyber-bullying. Results show that LA-OCA has an ideal joint model based on LA-OCA. The MFTRPO is significantly higher than customary steps. Two commonly recognized words for identity are much and straightforwardly overshadowed by MTRPO. This is perceived to be a significant advancement in weather and environment forecasting techniques. The findings of these studies demonstrated that a hierarchical resource distribution system focused on Non-Orthogonal Multiple Access (NOMA) and Q-Learning is suggested. The approach would have a great effect on the throughput of the network. One Q-Learning method that is being developed is a methodology utilized in robotics learning: Reinforcement Learning. Q-learning system aims to boost the estimation capability of traditional AI models (RNN). In conclusion, the load provided by PHEVs can be calculated more reliably than that produced by an ANN and RNN model. Increasing the number of neural network iterations would increase the performance of the algorithm. The

MSE expense estimation of the ANN methodology relates to the load calculation of the PHEV on the national energy.

5. CONCLUSION

In this study, various reinforcement learning strategies and their approaches were studied. Learning enhancement and its complex algorithms play a key role in this research. We looked at different aspects of the reward function operate Difference learning algorithms Q-learning. The regular incentive function is balanced by constant values and the incentive feature accrued increases award independent of the growth stage. Play all the bonus roles of the game well. The emphasis of this initial study is on a series of studies on enhancement learning protection. In future work, new countermeasures should be explored to minimize the impact of such electronic, physical, and critical network attacks, and our results also suggest that all Q-Learning variables deliver considerably improved performance over Q-Learning. Algorithms have been able to perform well in the first steps to reach standards that have proven impossible for humans. In this model, the behavior of RL agents can be further established. Analytical management of the topic of determining limits and the relationship of model parameters, such as network layout and exploration mechanisms, with the weakness of regulation induction will provide insight and guidance on developing robust reinforcement architectures.

References

Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). *Machine Learning Supervised Algorithms of Gene Selection: A Review*. *62*(03), 13.

Adeen, I. M. N., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). *Systematic Review of Unsupervised Genomic Clustering Algorithms Techniques for High Dimensional Datasets*. *62*(03), 21.

Ahmed, O., & Brifcani, A. (2019). Gene Expression Classification Based on Deep Learning. *2019 4th Scientific International Conference Najaf (SICN)*, 145–149. https://doi.org/10.1109/SICN47020.2019.9019357

Aind, A. T., Ramnaney, A., & Sethia, D. (2020). Q-Bully: A Reinforcement Learning based Cyberbullying Detection Framework. *2020 International Conference for Emerging Technology (INCET)*, 1–6. https://doi.org/10.1109/INCET49848.2020.9154092

Akalin, N., & Loutfi, A. (2020). Reinforcement Learning Approaches in Social Robotics. *ArXiv:2009.09689 [Cs]*. http://arxiv.org/abs/2009.09689

Anderson, H. S., Kharkar, A., Filar, B., Evans, D., & Roth, P. (2018). Learning to Evade Static PE Machine Learning Malware Models via Reinforcement Learning. *ArXiv:1801.08917 [Cs]*. http://arxiv.org/abs/1801.08917

Arabnejad, H., Pahl, C., Jamshidi, P., & Estrada, G. (2017). A Comparison of Reinforcement Learning Techniques for Fuzzy Cloud Auto-Scaling. *ArXiv:1705.07114 [Cs]*. http://arxiv.org/abs/1705.07114

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). A Brief Survey of Deep Reinforcement Learning. *IEEE Signal Processing Magazine*, *34*(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

Babaeizadeh, M., Frosio, I., Tyree, S., Clemons, J., & Kautz, J. (2017). Reinforcement Learning through Asynchronous Advantage Actor-Critic on a GPU. *ArXiv:1611.06256 [Cs]*. http://arxiv.org/abs/1611.06256

Brifcani, A. M. A., & Brifcani, W. M. A. (2010). Stego-Based-Crypto Technique for High Security Applications. *International Journal of Computer Theory and Engineering*, 835–841. https://doi.org/10.7763/IJCTE.2010.V2.249

Brim, A. (2020). Deep Reinforcement Learning Pairs Trading with a Double Deep Q-Network. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 0222–0227. https://doi.org/10.1109/CCWC47524.2020.9031159

Cazé, R., Khamassi, M., Aubin, L., & Girard, B. (2018). Hippocampal replays under the scrutiny of reinforcement learning models. *Journal of Neurophysiology*, *120*(6), 2877–2896. https://doi.org/10.1152/jn.00145.2018

Chang, Soowon, Saha, N., Castro-Lacouture, D., & Yang, P. P.-J. (2019). Multivariate relationships between campus design parameters and energy performance using reinforcement learning and parametric modeling. *Applied Energy*, *249*, 253–264. https://doi.org/10.1016/j.apenergy.2019.04.109

Chang, Spencer, Cohen, T., & Ostdiek, B. (2018). What is the machine learning? *Physical Review D*, *97*(5), 056009. https://doi.org/10.1103/PhysRevD.97.056009

Chen, C., Cui, M., Li, F. F., Yin, S., & Wang, X. (2020). Model-Free Emergency Frequency Control Based on Reinforcement Learning. *IEEE Transactions on Industrial Informatics*, 1–1. https://doi.org/10.1109/TII.2020.3001095

Chen, D., Qi, Q., Zhuang, Z., Wang, J., Liao, J., & Han, Z. (2021). Mean Field Deep Reinforcement Learning for Fair and Efficient UAV Control. *IEEE Internet of Things Journal*, *8*(2), 813–828. https://doi.org/10.1109/JIOT.2020.3008299

Chen et al. - 2020—Model-Free Emergency Frequency Control Based on Re.pdf. (n.d.).

Chen, J., Gong, Z., Wang, W., Liu, W., Yang, M., & Wang, C. (2020). TAM: Targeted Analysis Model With Reinforcement Learning on Short Texts. *IEEE Transactions on Neural Networks and Learning Systems*, 1–10. https://doi.org/10.1109/TNNLS.2020.3009247

China Electric Power Research Institute, Zhang, D., Han, X., Taiyuan University of Technology, Deng, C., & China Electric Power Research Institute. (2018). Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE Journal of Power and Energy Systems*, *4*(3), 362–370. https://doi.org/10.17775/CSEEJPES.2018.00520

da Silva, M. V., Souza, R. D., Alves, H., & Abrao, T. (2020). A NOMA-Based *Q* -Learning Random Access Method for Machine Type Communications. *IEEE Wireless Communications Letters*, *9*(10), 1720–1724. https://doi.org/10.1109/LWC.2020.3002691

Debnath, S., Sukhatme, G., & Liu, L. (2018). Accelerating Goal-Directed Reinforcement Learning by Model Characterization. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1–9. https://doi.org/10.1109/IROS.2018.8593728

Du, Y., & Li, F. (2020). Intelligent Multi-Microgrid Energy Management Based on Deep Neural Network and Model-Free Reinforcement Learning. *IEEE Transactions on Smart Grid*, *11*(2), 1066–1076. https://doi.org/10.1109/TSG.2019.2930299

Francois-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An Introduction to Deep Reinforcement Learning. *Foundations and Trends® in Machine Learning*, *11*(3–4), 219–354. https://doi.org/10.1561/2200000071

Guo, M., Liu, Y., & Malec, J. (2004). A New Q-Learning Algorithm Based on the Metropolis Criterion. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *34*(5), 2140–2143. https://doi.org/10.1109/TSMCB.2004.832154

Han, M., May, R., Zhang, X., Wang, X., Pan, S., Yan, D., Jin, Y., & Xu, L. (2019). A review of reinforcement learning methodologies for controlling occupant comfort in buildings. *Sustainable Cities and Society*, *51*, 101748. https://doi.org/10.1016/j.scs.2019.101748

Han, M., Zhang, X., Xu, L., May, R., Pan, S., Wu, J., & Fleyeh, H. (n.d.-a). *A review of reinforcement learning methodologies on control systems for building energy*. 26.

Han, M., Zhang, X., Xu, L., May, R., Pan, S., Wu, J., & Fleyeh, H. (n.d.-b). *A review of reinforcement learning methodologies on control systems for building energy*. 26.

Han, Z., Pedrycz, W., Zhao, J., & Wang, W. (2020). Hierarchical Granular Computing-Based Model and Its Reinforcement Structural Learning for Construction of Long-Term Prediction Intervals. *IEEE Transactions on Cybernetics*, 1–11. https://doi.org/10.1109/TCYB.2020.2964011

Hein, D., Hentschel, A., Runkler, T., & Udluft, S. (2017). Particle Swarm Optimization for Generating Interpretable Fuzzy Reinforcement Learning Policies. *Engineering Applications of Artificial Intelligence*, *65*, 87–98. https://doi.org/10.1016/j.engappai.2017.07.005

Hein, D., Udluft, S., & Runkler, T. A. (2018). Interpretable Policies for Reinforcement Learning by Genetic Programming. *ArXiv:1712.04170 [Cs]*. http://arxiv.org/abs/1712.04170

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, *51*(5), 675–687. https://doi.org/10.1016/j.beth.2020.05.002

Kasgari, A. T. Z., Saad, W., Mozaffari, M., & Poor, H. V. (2020). Experienced Deep Reinforcement Learning with Generative Adversarial Networks (GANs) for Model-Free Ultra Reliable Low Latency Communication. *IEEE Transactions on Communications*, 1–1. https://doi.org/10.1109/TCOMM.2020.3031930

Kintsakis, A. M., Psomopoulos, F. E., & Mitkas, P. A. (2019). Reinforcement Learning based scheduling in a workflow management system. *Engineering Applications of Artificial Intelligence*, *81*, 94–106. https://doi.org/10.1016/j.engappai.2019.02.013

Kumar Shastha, T., Kyrarini, M., & Gräser, A. (2019). Application of Reinforcement Learning to a Robotic Drinking Assistant. *Robotics*, *9*(1), 1. https://doi.org/10.3390/robotics9010001

Lavaei, A., Somenzi, F., Soudjani, S., Trivedi, A., & Zamani, M. (2020). Formal Controller Synthesis for Continuous-Space MDPs via Model-Free Reinforcement Learning. *2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS)*, 98–107. https://doi.org/10.1109/ICCPS48487.2020.00017

Lei, L., Tan, Y., Zheng, K., Liu, S., Zhang, K., & Shen, X. (2020). Deep Reinforcement Learning for Autonomous Internet of Things: Model, Applications and Challenges. *IEEE Communications Surveys & Tutorials*, *22*(3), 1722–1760. https://doi.org/10.1109/COMST.2020.2988367

Li, Y. (2018). Deep Reinforcement Learning: An Overview. *ArXiv:1701.07274 [Cs]*. http://arxiv.org/abs/1701.07274

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2019). Continuous control with deep reinforcement learning. *ArXiv:1509.02971 [Cs, Stat]*. http://arxiv.org/abs/1509.02971

Lopez-Martin, M., Carro, B., & Sanchez-Esguevillas, A. (2020). Application of deep reinforcement learning to intrusion detection for supervised problems. *Expert Systems with Applications*, *141*, 112963. https://doi.org/10.1016/j.eswa.2019.112963

Mahmood, M. R., & Abdulazeez, A. M. (2018). A Comparative Study of a New Hand Recognition Model Based on Line of Features and Other Techniques. In F. Saeed, N. Gazem, S. Patnaik, A. S. Saed Balaid, & F. Mohammed (Eds.), *Recent Trends in Information and Communication Technology* (Vol. 5, pp. 420–432). Springer International Publishing. https://doi.org/10.1007/978-3-319-59427-9_45

Mammeri, Z. (2019). Reinforcement Learning Based Routing in Networks: Review and Classification of Approaches. *IEEE Access*, *7*, 55916–55950. https://doi.org/10.1109/ACCESS.2019.2913776

Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(4), 140–147. https://doi.org/10.38094/jastt1457

Moerland, T. M., Broekens, J., & Jonker, C. M. (2020). Model-based Reinforcement Learning: A Survey. *ArXiv:2006.16712 [Cs, Stat]*. http://arxiv.org/abs/2006.16712

Mukherjee, S., Bai, H., & Chakrabortty, A. (2020). Reinforcement Learning Control of Power Systems with Unknown Network Model under Ambient and Forced Oscillations. *2020 IEEE Conference on Control Technology and Applications (CCTA)*, 346–351. https://doi.org/10.1109/CCTA41146.2020.9206271

Perera, A. T. D., Wickramasinghe, P. U., Nik, V. M., & Scartezzini, J.-L. (2020). Introducing reinforcement learning to the energy system design process. *Applied Energy*, *262*, 114580. https://doi.org/10.1016/j.apenergy.2020.114580

Powell, B. K. M., Machalek, D., & Quah, T. (2020). Real-time optimization using reinforcement learning. *Computers & Chemical Engineering*, *143*, 107077. https://doi.org/10.1016/j.compchemeng.2020.107077

Raghu, A., Komorowski, M., Celi, L. A., Szolovits, P., & Ghassemi, M. (2017). Continuous State-Space Models for Optimal Sepsis Treatment—A Deep Reinforcement Learning Approach. *ArXiv:1705.08422 [Cs]*. http://arxiv.org/abs/1705.08422

Rahimi, M., Gibb, S., Shen, Y., & La, H. M. (2018). A Comparison of Various Approaches to Reinforcement Learning Algorithms for Multi-robot Box Pushing. *ArXiv:1809.08337 [Cs]*. http://arxiv.org/abs/1809.08337

Recht, B. (2018). A Tour of Reinforcement Learning: The View from Continuous Control. *ArXiv:1806.09460 [Cs, Math, Stat]*. http://arxiv.org/abs/1806.09460

Renaudo, E., Girard, B., Chatila, R., & Khamassi, M. (2015). Respective Advantages and Disadvantages of Model-based and Model-free Reinforcement Learning in a Robotics Neuro-inspired Cognitive Architecture. *Procedia Computer Science*, *71*, 178–184. https://doi.org/10.1016/j.procs.2015.12.194

Rocchetta, R., Bellani, L., Compare, M., Zio, E., & Patelli, E. (2019). A reinforcement learning framework for optimal operation and maintenance of power grids. *Applied Energy*, *241*, 291–301. https://doi.org/10.1016/j.apenergy.2019.03.027

Sadiq, S. S., Abdulazeez, A. M., & Haron, H. (2020). Solving multi-objective master production schedule problem using memetic algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(2), 938. https://doi.org/10.11591/ijeecs.v18.i2.pp938-945

Saleh, A., Jaques, N., Ghandeharioun, A., Shen, J., & Picard, R. (2020a). Hierarchical Reinforcement Learning for Open-Domain Dialog. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(05), 8741–8748. https://doi.org/10.1609/aaai.v34i05.6400

Saleh, A., Jaques, N., Ghandeharioun, A., Shen, J., & Picard, R. (2020b). Hierarchical Reinforcement Learning for Open-Domain Dialog. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(05), 8741–8748. https://doi.org/10.1609/aaai.v34i05.6400

Salih Hassan, O. M., Mohsin Abdulazeez, A., & Tiryaki, V. M. (2018). Gait-Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis. *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 173–178. https://doi.org/10.1109/ICOASE.2018.8548909

Schilperoort, J., Mak, I., Drugan, M. M., & Wiering, M. A. (2018). Learning to Play Pac-Xon with Q-Learning and Two Double Q-Learning Variants. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1151–1158. https://doi.org/10.1109/SSCI.2018.8628782

Soni, R., Guan, J., Avinash, G., & Saripalli, V. R. (2019). HMC: A Hybrid Reinforcement Learning Based Model Compression for Healthcare Applications. *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, 146–151. https://doi.org/10.1109/COASE.2019.8843047

Stember, J. N., & Shalu, H. (2020). Reinforcement learning using Deep Q Networks and Q learning accurately localizes brain tumors on MRI with very small training sets. *ArXiv:2010.10763 [Cs]*. http://arxiv.org/abs/2010.10763

Suerich, D., & Young, T. (2020). Reinforcement Learning for Efficient Scheduling in Complex Semiconductor Equipment. *2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–3. https://doi.org/10.1109/ASMC49169.2020.9185293

Sulaiman, D. M., Abdulazeez, A. M., Haron, H., & Sadiq, S. S. (2019). Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization. *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 82–87. https://doi.org/10.1109/ICOASE.2019.8723749

Talevi, A., Morales, J. F., Hather, G., Podichetty, J. T., Kim, S., Bloomingdale, P. C., Kim, S., Burton, J., Brown, J. D., Winterstein, A. G., Schmidt, S., White, J. K., & Conrado, D. J. (2020). Machine Learning in Drug Discovery and Development Part 1: A Primer. *CPT: Pharmacometrics & Systems Pharmacology*, *9*(3), 129–142. https://doi.org/10.1002/psp4.12491

Thomas, R. N., & Gupta, R. (2020). A Survey on Machine Learning Approaches and Its Techniques: *2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)*, 1–6. https://doi.org/10.1109/SCEECS48394.2020.190

Valladares, W., Galindo, M., Gutiérrez, J., Wu, W.-C., Liao, K.-K., Liao, J.-C., Lu, K.-C., & Wang, C.-C. (2019). Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm. *Building and Environment*, *155*, 105–117. https://doi.org/10.1016/j.buildenv.2019.03.038

van Hasselt, H., Guez, A., & Silver, D. (n.d.). *Deep Reinforcement Learning with Double Q-Learning*. 7.

Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied Energy*, *235*, 1072–1089. https://doi.org/10.1016/j.apenergy.2018.11.002

Wang, J., Liu, Y., & Li, B. (2020). Reinforcement Learning with Perturbed Rewards. *ArXiv:1810.01032 [Cs, Stat]*. http://arxiv.org/abs/1810.01032

Wang, Z., & Hong, T. (2020). Reinforcement learning for building controls: The opportunities and challenges. *Applied Energy*, *269*, 115036. https://doi.org/10.1016/j.apenergy.2020.115036

Xie, H., Xu, X., Li, Y., Hong, W., & Shi, J. (2020). Model Predictive Control Guided Reinforcement Learning Control Scheme. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN48605.2020.9207398

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, *1*(2), 56–70. https://doi.org/10.38094/jastt1224

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019). Machine learning and Region Growing for Breast Cancer Segmentation. *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 88–93. https://doi.org/10.1109/ICOASE.2019.8723832

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zeebaree, S. R. M. (2017). *Combination of K-means clustering with Genetic Algorithm: A review*. *12*(24), 8.

Zhang, Z., Wang, D., & Gao, J. (2020a). Learning Automata-Based Multiagent Reinforcement Learning for Optimization of Cooperative Tasks. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14. https://doi.org/10.1109/TNNLS.2020.3025711

Zhang, Z., Wang, D., & Gao, J. (2020b). Learning Automata-Based Multiagent Reinforcement Learning for Optimization of Cooperative Tasks. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14. https://doi.org/10.1109/TNNLS.2020.3025711

Zhao, T., Xie, K., & Eskenazi, M. (2019). Rethinking Action Spaces for Reinforcement Learning in End-to-end Dialog Agents with Latent Variable Models. *ArXiv:1902.08858 [Cs]*. http://arxiv.org/abs/1902.08858

Cite this article:

Abdulqadir, H. R. & Abdulazeez, A. M. (2021). Reinforcement Learning and Modeling Techniques: A Review. *International Journal of Science and Business, 5*(3), 174-189. doi: https://doi.org/10.5281/zenodo.4542638

Retrieved from http://ijsab.com/wp-content/uploads/696.pdf

Published by