Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends

Vasileios P. Rekkas [1,*] **, Sotirios Sotiroudis** [1,*] **, Panagiotis Sarigiannidis** [2] **, Shaohua Wan** [3] **, George K. Karagiannidis** [4] **and Sotirios K. Goudos** [1,*]

1   ELEDIA@AUTH, School of Physics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece
2   Department of Informatics and Telecommunications Engineering, University of Western Macedonia, 501 00 Kozani, Greece; psarigiannidis@uowm.gr
3   School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China; shwanhust@gmail.com
4   School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; geokarag@auth.gr
*   Correspondence: vrekkas@physics.auth.gr (V.P.R.); ssoti@physics.auth.gr (S.S.); sgoudo@physics.auth.gr (S.K.G.)

Abstract: Artificial Intelligence (AI) and especially Machine Learning (ML) can play a very important role in realizing and optimizing 6G network applications. In this paper, we present a brief summary of ML methods, as well as an up-to-date review of ML approaches in 6G wireless communication systems. These methods include supervised, unsupervised and reinforcement techniques. Additionally, we discuss open issues in the field of ML for 6G networks and wireless communications in general, as well as some potential future trends to motivate further research into this area.

Keywords: 6G; wireless communications; artificial intelligence; machine learning

1. Introduction

Wireless communication systems have experienced substantial revolutionary progress over the past years. With the rapid progress of 3GPP 5G phase 2 standardization, the commercial deployment of 5G applications being deployed all over the world cannot fully meet the challenges brought by the rapid increase of traffic and the real-time requirement of services [1]. In this behalf, industry and academia are already working towards realizing the sixth generation (6G) communication systems. ML, as part of AI, involves teaching the machines to perform tasks independently based on making data-driven decisions. ML can accurately estimate various parameters and support interactive decision-making. In [2], the deployment of ML techniques as potential solutions upcoming 6G wireless communications challenges is being discussed. The application of ML techniques in 6G wireless communication systems has been the subject that attracts interest in recent years. In this paper, we extend our earlier work [3].

The remainder of the paper is as follows. Section 2 briefly discusses the 6G network requirements and challenges. In Section 3, we present some basic ML algorithms. In Section 4, we present some of the emerging new 6G applications and services and the role of ML. Finally, Sections 5 and 6 discuss some open issues and future trends in the application of ML algorithms in 6G and wireless communications, whereas Section 7 concludes this review paper with some remarks.

2. 6G Network Requirements and Challenges

The global mobile traffic volume is anticipated to reach 5016 exabytes per month (Eb/mo) in 2030, while in 2010 it was 7.462 EB/mo in 2010 [4] and so 5G will not be able to address the traffic load. 6G will try to address the shortcomings of 5G by trying creating smart radio environments through Intelligent Reflecting Surfaces (IRS) and adjusting the

communication in higher frequency bands (THz and mm-wave) [5]. IRS emerges as a key technology in future 6G networks. IRS receives a signal from the base station (BS), and reflects the signal with induced phase changes, which are adjusted by a controller. The reflected signal can be added coherently with the signal from the BS to either boost or attenuate the overall signal at the receiver. IRS may not amplify the signal power without power but has minimal power requirement for the operation of the controller and reconfiguration of the elements to have full control over the reflection signal.

IRS is energy and cost efficient, by inducing smart radio environments, and is free from self-interference, so can be used as other related wireless technologies such as, conventional relaying, backscatter communication (BackCom), and mMIMO relaying. IRS can be a solution for energy and spectral-efficient issue in 6G systems [6]. IRS will play a crucial role in 6G communication networks, similar to that of massive MIMO in 5G networks. Thus, IRS can be used to help achieve massive MIMO 2.0 in 6G networks [7].

6G networks will enhance and expand 5G applications and will meet the following requirements [8,9]:

- Achieve higher data rate per user/device (10–100 times greater than 5G);
- Support wider coverage;
- Support larger number of connected devices;
- Integrate low latency communications;
- Reduce the energy consumption;
- Support massive Internet of Things (IoT) and integrate virtual reality (VR) and augmented reality (AR) into one extended reality (XR);
- Generate large amounts of data through the Internet of Everything (IoE);
- Suppor distributed massive MIMO;
- Support high and reliable connectivity;
- Support real-time dynamic analysis and self-awareness;
- Support trust and security mechanisms for safer integration.

Application and feature description of 5G and 6G networks [9–12] are presented in Table 1.

Table 1. Comparison of 5G and 6G networks.

Technology	5G	6G
Applications	Enhanced Mobile Broadband Communications (eMBB), Ultrareliable Low Latency Communications (URLLC), Massive Machine Type Communications (mMTC)	Holographic-Type Communication (HTC), Tactile Internet, Intelligent Transport and Logistics, Intelligent and automated machines, Virtual Reality (VR), Augmented Reality (AR), Extended reality (XR)
Peak data rate	10 Gbps	1 Tbps
Frequency	3–300 GHz	1000 GHz
Latency	10 ms	<1 ms
Mobility support	Up to 500 km/h	Up to 1000 km/h
Spectral efficiency	30 bps/Hz	100 bps/Hz
Reliability	99.9999%	99.99999%

3. Machine Learning

Machine Learning (ML) models are computational systems that are able to learn the features of a system that cannot be represented by using a conventional mathematical model approach. These models are commonly used in tasks such as regression, classification, and any interaction between an intelligent agent and an environment. After the model is trained on the given training data-set, it can be effectively applied to unknown data and

perform any decision based on the training data. ML is usually classified into three major categories [13]: supervised, unsupervised, and reinforcement learning.

3.1. Supervised Learning

Supervised learning algorithms are trained using a labeled data-set. In supervised approach, both the input data and the desired output data to be predicted, are known to the system. In supervised learning it is essential to have enough data, in order to be effectively applied in any application [14]. Supervised learning is mostly used for classification and regression problems and some typical supervised algorithms are logistic regression, Artificial Neural Networks (ANN), k-Nearest Neighbor (kNN) [15], naive Bayes, random forest and decision tree [16].

- ANNs: ANNs are inspired by nature and try to imitate biological neural networks, and so are able to learn from complicated data. In wireless communication systems, ANNs can be used to learn the structure of the network and predict user's behavior to solve different problems such as spectrum and resource allocation, cell association etc. [17]. Recently deep learning has extended the ANN applicability and capabilities with Deep Neural Networks (DNN) [18]. Moreover, there are ANN types like the Autoencoders that are applied for unsupervised learning or other ANN structures that are used for reinforcement learning.

- K-Nearest Neighbor: KNN is a classification and regression algorithm based on the distance between different feature values. The classification of an unknown data sample is determined based on the class of K nearest neighbors. If the majority of the nearest neighbors belongs to a certain class, then the sample is assigned into this class. The advantages of the algorithm are many: it is insensitive to outliers, easy to realize and suitable for multiclass classifications. The big disadvantage of the approach is that, for large input dataset, is very time-consuming [16].

- Naive Bayes: it is a simple probabilistic classification model based on the Bayes theorem. The Bayes theorem provides a model of the conditional probability of a result Y with the given inpu/ condition X. The Naïve Bayes classifiers can effectively handle a large number of independent continuous or categorical features. This is due to the ability to transform a high-dimensional density estimation task into a one dimensional kernel density estimation task, assuming the features are independent with one another [19].

- Decision tree: This model imitates trees in natures. Each node of the decision tree represents a feature of a data, each branch the conjunction of features that are needed for the classification, and each leaf node represents a specific class. The model tries to maximize the information gain of each variable split. After the model is trained by the known labeled dataset, the classification of the unlabeled sample can be achieved by comparing the feature value with the trained nodes of the decision tree. The basic advantages of the approach, include simple implementation, and high classification accuracy. However, it suffers from including many-level data variables because information gains are biased towards multi-level features [16].

- Random Forest: A random forest usually consists of multiple decision trees. The method randomly selects a subset of features to be the base of constructing each decision tree. Each decision tree classifies any new dataset and the unknown data samples are categorized into a specific class, based on the majority of the decision trees [16]. The algorithm only examines part of the attributes for an attribute of the best split and so low correlation between trees is essential to avoid the domination of a few strong attributes [19]. Figure 1 depicts an example of a Random Forest model.

- Convolutional Neural Networks (CNN): These models are made up of neurons that can self-optimize through unsupervised learning. They are mostly used for pattern recognition, especially in classification applications for image recognition. CNN consists of three layers: the convolutional layer, the pooling layer, and the fully connected layer. When these layers are stacked together, the complete CNN architecture

is formed [17]. CNNs can be used for both supervised or unsupervised learning depending on the task in which it is used.

- Recurrent Neural Network (RNN): A RNN is an ANN type that uses sequential data or time series data. Some common applications of RNNs include ordinal or temporal problems, like as language translation, natural language processing, speech recognition, and image captioning. An artificial Recurrent Neural Network type is Long Short Term Memory (LSTM), which have been introduced in order to overcome the vanishing gradient problems, which are observed when training traditional RNNs. LSTM networks can be applied for classification, processing and making predictions based on time series data. As with CNNs, RNNs can be applied for both supervised or unsupervised learning.
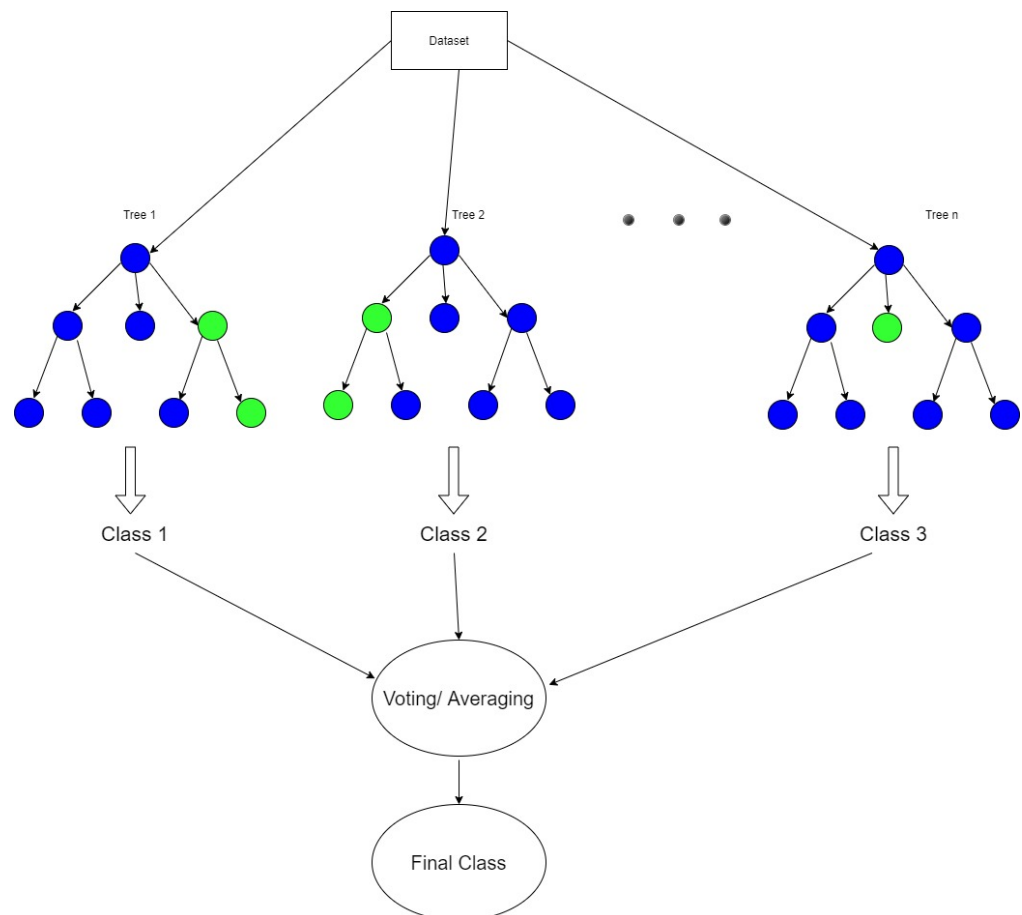

Figure 1. Random Forest model.

In our study, many different algorithms were applied, but all of them were based and inspired from the previously mentioned supervised algorithms. The advantages and limitations of the most common supervised ML methods that were introduced [20–24], are analyzed in Table 2:

Table 2. Advantages and limitations of supervised ML methods.

ML Approach	Advantages	Limitations
ANN	• High fault tolerance • Distributed memory • Parallel processing capability • Robust to noise	• Hardware dependence • Reduced trust • Structure through trial and error
knn	• One hyperparameter (k) • Non-parametric • No training step • Easy to implement in multi-class problems	• Computationally expensive • Sensitive to noise • Curse of dimensionality • Needs homogenous features
Naive Bayes	• Fast and can be used in real-time • Insensitive to irrelevant features • Performs well with high dimensional data • Scalable with large datasets	• Not so accurate • Zero-frequency problem • Assumes independent features
Decision tree	• Does not require normalization or scaling of data • Missing values in data do not affect process • Simple implementation • High accuracy	• Several level-data variables • High complexity • Instable for data variation
Random Forest	• Accurate and robust • Insesitive to overfitting • Offers feature importance	• Low correlation between trees • High complexity
CNN	• Automatically detects important features • Weight sharing • Minimizes computation	• Lacks ability to be spatially invariant from input data • Slow training procedure
RNN	• Can process inputs of any length • Model size does not increase with larger input • Minimizes computation	• Computationally expensive • Cannot process long sequences for certain activation functions

3.2. Unsupervised Learning

Unsupervised learning algorithms are given a set of unlabeled data to correctly predict the output, which is the basic difference with the supervised learning approach. These algorithms are mostly used for clustering and aggregation problems, but can also achieve great results for regression problems. Some typical unsupervised algorithms include K-means, Self-Organizing Maps (SOMs), Hidden Markov Model (HMM), Auto Encoders (AEs), Principal Component Analysis (PCA), Restricted Boltzmann Machine (RBM), fuzzy C-means etc. Furthermore, unsupervised ML have been applied to enhance the performance of Deep Learning (DL) algorithms such as Convolutional Neural Networks (CNNs) and Long short-term memory (LSTM) algorithms [16].

- K-means: It is a widely used method to classify unlabeled raw input data into different clusters. K-means algorithm assigns each new data point to a cluster, based on its distance from the nearest associated centroid. The centroids are updated based on the previously assigned data point and the procedure is repeated until there is no alteration in the input data points and the centroids. K represents the number of desired clusters and can greatly impact the performance of the algorithm [16].
- Self-Organizing Map (SOM): This approach is mostly used for data clustering and dimensionality reduction. The model has one input layer and a map layer, with each layer containing many neurons and a different weight vector is assigned to each neuron. During the training process, SOM builds the map by using an unsupervised competitive learning approach. The winning neuron from this competition determines the cluster in which any new input vector is classified [16]. Figure 2 displays the architecture of a traditional Self Organizing Map model.
- Autoencoders: learning circuits that copy inputs into outputs, aiming to have the least possible deviation. The have great results on both classification and regression problems. Autoencoders are stacked approaches and are trained unsupervised bottom-

up, followed by a supervised learning method. In this way the top layer is trained based on known input, and so fine-tuning the whole architecture.
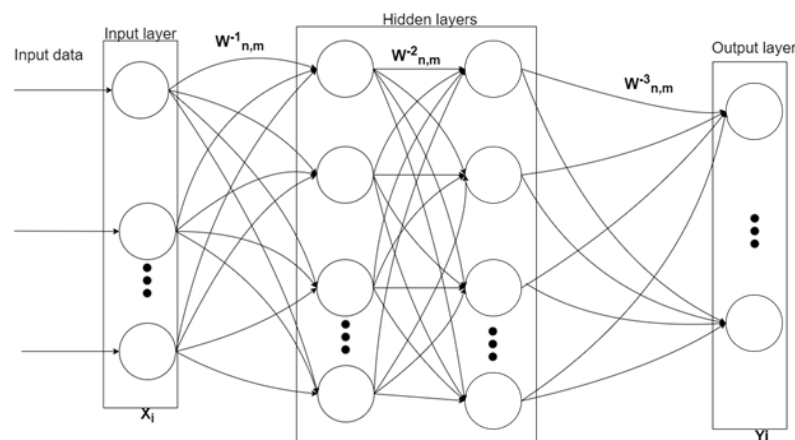

Figure 2. Self Organizimg Map Model.

In our study, many different algorithms were applied for unsupervised ML, but all of them were based and inspired from the previously mentioned algorithms. The advantages and limitations of the most common unsupervised ML methods that were introduced [16,25–28], are analyzed in Table 3:

Table 3. Advantages and limitations of unsupervised ML methods.

ML Approach	Advantages	Limitations
k-means	• Easy to implement • Suitable for large datasets • Adapts easily to new examples	• Manual choice of k • k greatly impacts performance • Can cluster outliers • Scales with dimensionality
SOM	• Easily understood • Capable of clustering large and complex datasets	• Needs large number of data • Nearby data points must behave similarly • Finds different similarities among sample vectors
Auto-encoders	• Denoising training • Dimensionality reduction • Able to learn non-linear feature representations	• Computationally expensive • High complexity • Prone to overfitting
PCA	• Fast • Dimensionality reduction • Reduces overfitting	• Independent variables are less interpretable • Needs data standarization beforehand • Incapable of learning non linear feature representations
HMM	• Can handle inputs of variable lengths • Can combine into libraries • Can learn from raw input data	• Does not take into account the sequence of states leading into any given state • Dependency between appliances cannot be represented
RBM	• Can encode any distribution • Computationally efficient	• Difficult training procedure • Needs weight adjustment

3.3. Reinforcement Learning

Reinforcement Learning (RL) is based on the principles of behaviourist psychology and the model learns the same way as a child learns to perform a new task. RL is realized on the basis of a feedback performance indicator (reward) conceived from the model's environment. The model pursues the ideal performance of the output by maximizing the indicator of the reward. RL is a hybrid of supervised and unsupervised learning, because (indirect) supervision is required for the model to understand and learn the ideal

system's performance, while there is no available training dataset paired with the desired output [15]. Basically, RL is a trial and error procedure where an agent interacts with the environment and based on whether the action tried was good or bad, gets feedback in terms of reward or penalty. RL tries to learn the best policy that would enable the agent to make an optimal decision at any given state of the environment.Figure 3 displays an example of RF. RL algorithms can be categorized to value-based (e.g., Q-learning, SARSA) and policy-based algorithms (e.g., Policy Gradient (PG), Proximal Policy Optimization (PPO) and Actor-Critic (A2C) [29].
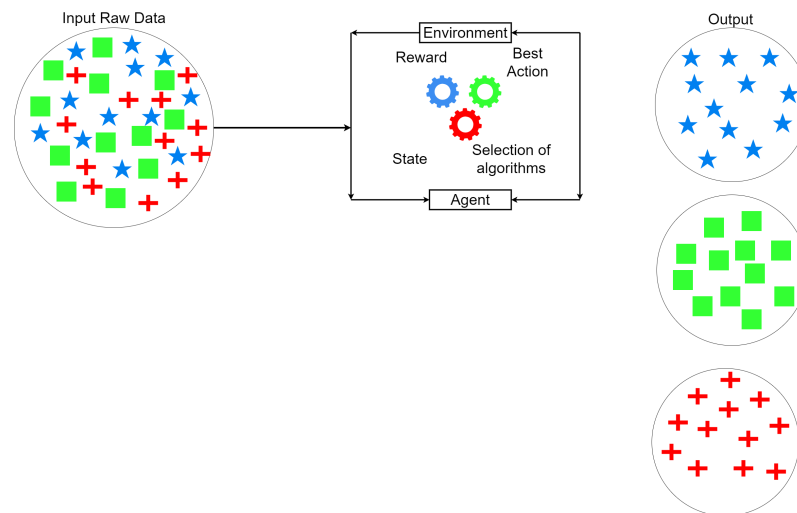
Figure 3. Example of reinforcement learning.

- Q-learning: Q-learning is the most common used RL algorithm. It is an off Policy technique and uses a greedy approach to learn the needed Q-value. The algorithm learns the Q-value given to the agent in a certain state, based on a specific action. The approach creates the Q-table, where the number of rows represent the number of states, and the number of columns represent the number of actions. The Q-value is the reward of the action at a certain state. Once the Q-values are learned the agent can make quick decisions under a current state by taking the action that has the largest Q-value from the table [30].
- SARSA: It is an on-policy algorithm which uses each time the action performed by the current policy of the model, in order to learn the Q-values [19].
- Policy Gradient (PG): The approach uses a random network, and a frame of the agent is applied to produce a random output action. This output is sent back to the agent and then the agent produces the next frame and the procedure is repeated until a good solution is reached. During the training of the model, the network's output is being sampled in order to avoid repeating loops pf the action. The sampling allows the agent to randomly explore the environment and find the better solution [17].
- Actor Critic: The actor-critic model learns a policy (actor) and value function (critic). Actor-critic learning is always on-policy because the critic needs to learn correct the Temporal Difference (TD) errors from the 'actor' or the policy [19].
- Deep reinforcement learning. In recent years, deep learning has significantly advanced the field of RL, with the use of deep learning algorithms within RL giving rise to the field of "deep reinforcement learning". Deep learning enables RL to operate in high-dimensional state and action spaces and can now be used for complex decision-making problems [31,32].

Some advantages and limitations of the most common RL algorithms [33–36], are listed below in Table 4:

<p align="center">**Table 4.** Advantages and limitations of RL methods.</p>

ML Approach	Advantages	Limitations
Q-learning	• Learns directly the optimal policy • Less computation cost • Relatively fast • Efficient for offline learning	• Use of biased samples • High per-sample variance • Computationally expensive • Not very efficient for online learning
SARSA	• Fast • Efficient for online learning datasets	• Learns a near-optimal policy while exploring • Not very efficient for offline learning
Policy Gradient	• Capable of finding best stochastic policy • Effective for high dimensionallity datasets	• Slow convergence • High variance
Actor Critic	• Reduces variance with respect to pure policy methods • More sample efficient than other RL methods • Guaranteed convergence	• Must be stochastic • Estimators need high variance

4. Beyond 5G/6G Applications and Machine Learning

6G will be able to support enhanced Mobile Broadband Communications (eMBB), Ultrareliable Low Latency Communications (URLLC) and massive Machine Type Communications (mMTC), but with enhanced capabilities compared to 5G networks. Furthermore, will be able to support application such as Virtual Reality (VR) Augmented Reality (AR) and ultimately Extended Reality (XR). Based on the problem different ML algorithms are applied as analyzed below.

4.1. Supervised Learning

4.1.1. Optimization Problems

Coverage, power and capacity optimization are critical challenges in future 6G networks services [16]. In [37], Random Forest and knn algorithms are proposed to predict and optimize the Path Loss (PL). The results show a higher accuracy and reduced Mean Squared Error (MSE) compared with conventional approaches. The authors in [38] propose a novel approach, namely GRL, to address the problems of joint user association and power allocation. In the proposed model, for optimization purposes, the learning process is split into two parts, the generalization-representation learning (GRL) part, and the specialization-representation learning (SRL) part The authors assume a function that can represent the connection between the network's parameters and the optimal resource allocation, and problems are addressed by optimizing this selected function. In this approach, the data-driven (supervised learning) and model-driven (unsupervised learning) training methods are combined to accurately predict the optimal function and the results are satisfactory.

In [39], a supervised ANN-based algorithm, named MLP-DBA, is proposed to predict the dynamic bandwidth allocation (DBA). The authors, aim to achieve bandwidth allocation close to optimal conventional approaches. The simulation results indicate that the proposed model can adaptively allocate the bandwidth, while improving the latency performance over the conventional DBA schemes. In [40], a DNN algorithm is proposed to predict the user's requirements in high dynamic (UAV) network. The results show better performance than the conventional Q-learning based algorithms that were mostly used. In [41], an RNN algorithm is proposed for intelligent load balancing. The proposed intelligent load balancer, named APRIL, can effectively load forecast information to maximize server utilization. Results show that the proposed forecasting model performs by between 5.88 and 92.6 better than the alternatives. The deviation in the performance is because the user's role greatly impacts the performance of the model.

In [42], machine learning-based Cooperative Spectrum Sensing schemes (CSSs) have been proposed. In the proposed approaches, some nodes send the received signal power from the users to the Fusion Center (FC), where some artificial neural networks (ANNs)

and SVM approach are used to determine whether the channels are idle or not. ANN is used to recognize the transmit power while SVM is used to find the best decision boundary, acting as a classifier. The results show that proposed approaches can offer great results in terms of accuracy and performance. In [43], the authors compare different supervised ML algorithms to predict data rate (ANN, SVM, random forest). Results show that random forest approach can achieve the lowest prediction error. The error is minimized in the uplink transmission direction (in downlink it is more significant). In [44], a supervised cooperative data rate prediction approach is introduced. This cooperative model reduces average prediction error by 30%.

In [45], combination of 2 well-known beamforming schemes (maximum ratio transmission and zero-forcing) is used in a K-user Multiple Input Single Output (MISO) channel. The proposed approach is based on a DNN in which the input nodes take channel vector with transmit power and the output returns the combining factors from transmitter's beamforming. The model achieves a sum rate of 99% when compared with conventional approaches.

A K-means clustering model for users in Thz MIMO-NOMA systems is proposed in [46]. Based on whether the user belong to Small Cell Base Stations (SBSs) coverage or Macro Base Station (MBSs) coverage, they are separated into different cluster. The great path spreading path loss and molecular absorption loss are two important challenges in THz systems. So an efficient clustering scheme can both reduce interference and improve the channel quality, resulting in higher throughput and Signal-to-interference-plus-noise ratio (SINR).For the user's clustering an enhanced K-means approach is proposed in the same paper. The channel's correlation parameters of different cluster are examined and the one that maximizes the metric is used to address the issue of fluctuation of clustering centers. The simulation results show the efficiency of the proposed schemes.

In [47], a machine learning based predictive DBA algorithm is proposed for the contention of upstream bandwidth and bottleneck latency in Passive Optical Networks (PONs). The proposed algorithm using an ANN at the Central Office (CO) to learn the uplink latency and estimate the bandwidth demand of every units. Using this approach, the CO can allocate the required bandwidth to forthcoming packet bursts without the need to have them wait until the following transmission cycle. The simulation results show that the model is able to achieve a >90% accuracy in predicting the Optical Network's status leading to the improvement of the accuracy of estimating the bandwidth demands of the optical units. Table 5 holds a brief summary of the supervised ML models in Beyond 5G(B5G)/6G optimization problems.

4.1.2. Fault/Anomaly Management

In [48], the authors propose an extended SVM, which is called support Tucker machine, to detect any fault/outlier detection in IoT systems. The model improved the accuracy and efficiency of anomaly detection and was able to retain the structure of the big sensor data.

Table 5. Supervised ML models in B5G/6G optimization problems.

Paper	ML Approach	Application Problem	Description
[37]	Random Forest, Knn	Path loss	Prediction and optimization of PL in mm-wave systems
[38]	Novel semi-supervised method	Power allocation, joint user association	Prediction of optimal function for network's parameters and power allocation
[39]	ANN	Dynamic Bandwidth Allocation (DBA)	Allocation of bandwitdth and improvement of latency performance

Table 5. *Cont.*

Paper	ML Approach	Application Problem	Description
[40]	DNN	User's requirements	Prediction of user's requirements in high dynamic UAV networks
[41]	DNN& RNN	Trafic load, power allocation, Load Balancing	Prediction of trafic load and optimization of power allocation
[42]	ANN, SVM	Cooperative Sensing Schemes (CSSs)	Prediction of trasmit power and boundary decision classifier
[43,44]	ANN, Random Forest, SVM	Data rate	Accuraty prediction of data rate with lowest possible prediction error
[45]	DNN	Beamforming Schemes in MISO channels	Prediction of trasmit power and trasmitter's beamforming
[46]	k-means	Clustering in MIMO-NOMA systems	Efficient clustering with higher throughput and lower SNR
[39]	ANN	DBA in Pasive Optical Networks (PONs)	Bandwidth and uplink latency prediction

4.1.3. Channel Estimation/Allocation

Estimation of future radio communication channels is rather challenging, due to their growing complexity [16]. In [49], data-driven supervised DNN estimators are used to predict channels, with results showing that using this approach the authors can predict more accurate channels compared to conventional channel estimation algorithms. The authors in [50] propose a supervised deep neural network (DNN) approach for adaptive bit allocation with imperfect Channel State Information (CSI) in heterogeneous networks. The accurate CSI estimation in heterogeneous networks can greatly impact the system's performance. Furthermore, the reduction of feedback overhead is an important challenge in heterogeneous networks. Even though many different quantization techniques have been used to address this issue, the system's performance cannot increase linearly with the number of bits increasing exponentially. The bits need to be distributed to the cells and then they are further allocated to each channel optimally. This conventional approach is time-consuming and so in order to enable direct allocation for the entire network, the proposed method is used. Using the supervised DNN the optimized number of bits can be directly obtained for a different number of bits and scenarios, leading to complexity reduction. Simulations show that the proposed method achieves a closer to optimal performance than the conventional approaches.

4.1.4. Beam Selection

The authors in [51] propose a combined supervised ML approach for beams selection in mm-wave communications. The beam selection problem was addressed as a multi-class problem, using two supervised learning algorithms (kNN and Support Vector Classifier-SVC) to address the issue, with simulation results showing that the proposed ML schemes can retain 90% of the sum rate with optimal beam selection. In [52], a supervised SVM for beam selection is proposed, aiming to achieve high sum-rate at lower computational complexity. The results verified that the proposed ML approach can achieve higher Average Sum Rate (ASR) with substantially lower computational complexity than conventional approaches. In [53], the authors propose a DNN model for beam selection in mm-wave systems, to reduce space required for the initial beam. The results show that the proposed beam selection reduces the beam overhead by up 79.3%. In [54], a DNN for optimal downlink beam in mm-wave networks is proposed, to enhance prediction accuracy and data rate. The simulation results show superior performance and robustness of the proposed model. The conventional approaches mostly rely on the sub 6GHz information, especially in the low signal-to-noise ratio (SNR) regions. In [55], a novel deep learning solution

based on a RNN, namely the Gated Recurrent Unit (GRU) is proposed for beam selection. The model can predict the serving base station and beam for each drone based on their prior trajectories and locations, extending their coverage. Simulation results show that the proposed scheme can achieve more than 90% accuracy for beam prediction.

4.1.5. Caching/Computing

In [56], the authors use an ANN-based approach to address the issue of code caching, with results showing the effectiveness of the model, In [57], a supervised DNN is proposed to address the issue of caching in IoT systems, with results being close to the optimal of conventional ones.

4.1.6. Security

In [58], the authors use decision tree algorithms to boost trust management using eXplainable Artificial Intelligence (XAI) for intrusion detection. Simple decision tree algorithms are applied to split the sub-choices for the intrusion detection system (IDS), which resemble a human approach to decision-making.Results show that the accuracy of the proposed approach is comparable with state-of-the-art algorithms. The authors in [59] used a supervised-based LSTM algorithm for intrusion detection model. They applied 6 different optimizer to investigate the performance of the model and the results show that LSTM model with Nadam optimizer can achieve an accuracy of 97.5%, which outperforms conventional approaches. In [60], the authors propose a supervised CNN-based method to classify and detect malware traffic, with classification accuracy of up to 99.4%.

4.1.7. MIMO

In [61], the authors propose a combination of ML-estimators, using CNN with Autoregressive Network (ARN)) for predicting Channel State Information (CSI) and RNN for channel prediction in massive MIMO systems with channel aging property. Results show that proposed model can improve the prediction accuracy and user's throughput gains for both low and high mobility scenarios. In [62], the issue of channel mapping in space and frequency domain in massive MIMO is addressed, by using a novel supervised deep learning approach, reducing overhead in both the training and feedback aspects.

4.1.8. UAV

In [63], a supervised deep learning approach is proposed for UAV systems. The proposed model uses a Clustering-based Two-layered (CBTL) algorithm for addressing this joint caching and trajectory prediction issue. Then, a DL approach of a CNN is used to enhanced make fast decisions online. This approach aims to maximize the network's throughput by jointly optimizing cache and trajectory. Simulation results show the effectiveness of the proposed approach in terms of accuracy. In [64] an ANN-based algorithm is proposed, to detect GPS spoofing signals in UAV systems. The results show high detection accuracy of spoofing signals and can reduce possible false alarms in the UAV system. In [65], the authors propose a SVM-based supervised approach for detecting jamming, spoofing and intrusion attacks in UAV systems. The proposed model shows high accuracy in detecting any attacks, reassuring safer UAV systems against cyber security attacks. The authors in [66] proposed a supervised ANN approach combined with an evolutionary algorithm, to predict the Received Signal Strength (RSS) in a UAV system. Moreover, in [67] an ensemble approach is selected, which exhibits satisfactory results in terms of performance and accuracy. Table 6 reports some supervised ML models used for B5G/6G problems.

Table 6. Supervised ML models in B5G/6G problems.

Paper	ML Approach	Application Problem	Description
[48]	Support Tucker Machine	Fault detection	Accurately predicts faults/outliers, while retaining structure of big sensor data in IoT systems
[49]	DNN	Channel estimation	Effectively predicts channels and CSI
[50]	Deep DNN	Adaptive bit allocation	Accurately predicts system's CSI in heterogeneous networks, reducing feedback overhead
[51]	knn & SVC	Beam selection	Addresses beam selection in mm-wave communication systems as multi-class problem
[52]	SVM	Beam selection sum-rate	Achieves higher Average Sum Rate (ASR) with substantially lower computational complexity
[53]	DNN	Beam selection	Optimal beam selection to reduce space for initial beam, reducing beam overhead
[54]	DNN	Downlink beam prediction	Accurately predicts downlink beam in mm-wave systems, enhancing data rate
[55]	GRU	Beam prediction	Predicts BS and beam for each drone, extending their coverage leading to optimal beam prediction
[56]	ANN	Caching	Effectively addresses challenge of code caching
[57]	DNN	Caching	Optimizes caching in IoT systems
[58]	Decision Tree	Security	Boosts trust management using XAI for intrusion detection
[59]	LSTM	Security	Boosts accuracy using Nadam optimizer for intrusion detection
[60]	CNN	Security	Boosts accuracy for classification and detection of malware traffic
[61]	CNN & ARN, RNN	Channel Estimation	Accurately predicts CSI in massive MIMO systems with channel aging property
[62]	Deep supervised mapping model	Channel Mapping	Addresses channel mapping in space and frequency domain for massive MIMO systems, reducing training and feedback overhead
[63]	CBTL, Deep CNN	Trajectory prediction	Maximizes network's throughput by jointly optimizing cache and trajectory, then DCNN makes fast decisions online
[64]	ANN	Security	Detects GPS spoofing signals in UAV systems, reducing possible false alarms
[65]	SVM	Cyber Security	Detects jamming, spoofing and intrusion attacks in UAV systems
[66,67]	Ensemble, ANN	RSS prediction	Accurately predicts RSS in UAV systems

4.2. Unsupervised Learning

4.2.1. Optimization Problems

Coverage, power and capacity optimization are critical challenges in future 6G networks services [16]. In [68,69], an unsupervised K-means algorithm is used to address the user selection and optimization of power allocation challenges in NOMA systems. Results show that the proposed model achieves great results in terms of accuracy and optimization. In [70], two Power Control (PC) algorithms, which are trained both using supervised and unsupervised learning, were proposed for Device-to-Device (D2D) scenarios. The comparison of the hybrid algorithms with conventional PC methods, show satisfactory results in terms of computational complexity, throughput, energy efficiency, resource allocation and

power control optimization. This work is categorized in unsupervised ML, because for the approach the supervised decision tree occurs from the unsupervised Q-learning method, so for the final hybrid approach the most significant impact factor is the performance of the unsupervised model that defines the supervised phase of the model and so the final performance of the approach.

Conventional approaches in modulation recognition of the received signals include several procedures such as preprocessing, classification and feature extraction. The authors in [71,72] addressed the challenge of modulation recognition, by investigating the performance of different deep learning algorithms such as CNN, LSTM etc, by using unsupervised learning paradigms for optimization purposes. The comparison results suggest that LSTM can achieve better performance than other DL based approaches.

CNN and LSTM are categorized as supervised learning methods, but they can be used in an unsupervised learning approach with satisfactory results. CNN is mostly supervised ML approach, but can be also used in an unsupervised way depending on the problem at hand [73]. The authors in [74] propose an automatic unsupervised cell event detection and classification method, which expands convolutional Long Short-Term Memory (LSTM) neural networks. The LSTM network could be trained in an unsupervised manner, by using a branched structure where one branch learns the regular appearance and movements of objects and the second learns the stochastic events, which occur rarely and without warning in a cell video sequence. Furthermore, the authors in [75] investigated anomaly detection in an unsupervised framework and introduce long short-term memory (LSTM) neural network-based algorithms with significant performance gains. The authors in [76] propose a new architecture for extracting features from images in an unsupervised manner, which is based on CNN. The model, namely Unsupervised Convolutional Siamese Network (UCSN), is trained to embed a set of images in a vector space, in a way that the local distance structure in the image space is preserved.The results indicate that the UCSN produces representations that are suitable for classification purposes. So LSTM and CNN are mainly used as supervised ML approaches, they can also be used in an unsupervised manner and as an unsupervised learning paradigm.

4.2.2. Fault Management

Fault management includes detection, identification and mitigation of any abnormal status of networks. Fault management in future 6G network needs to be effective, due to their heterogeneous, complex and dynamic nature. The authors in [77] compared five different unsupervised learning approaches (including K-means clustering, Fuzzy C-means clustering, Local Outlier Factor- LOF, Local Outlier Probabilities- LoOP and Kohonen's Self Organizing Maps-SOM) for fault detection in 6G networks. The results show that SOM-based approach outperforms Fuzzy C-means and K-means in detecting and predicting faults/abnormalities in 6G networks.

In [78], an extension of the conventional K-Means clustering algorithm, named K-Aware K-means, is used for fault detection in 6G network systems. In this extended version of K-means, the model uses an unsupervised learning phase to acquire a temporary expert knowledge of what the smallest cluster of the current data is like and then labels them as outliers, while updating the temporary knowledge. In this way, the model self-optimizes the K value ($K \leq 1$). and achieves a prediction accuracy of 99.7%. The authors in [79] propose an unsupervised learning approach with a SOM algorithm as the centerpiece for both fault recognition and recovery, achieving great accuracy results.

4.2.3. Channel Estimation

Estimation of future 6G radio communication channels is rather challenging, due to their growing complexity [16]. State-of-the-art unsupervised learning approaches (DL unsupervised model, CNN and RNN) have been used for channel detection in molecular communication [80,81]. A DL-based detector called DetNet was proposed in [82] and is able to achieve similar accuracy as conventional algorithms with much lower computation time.

The unsupervised DL-based detectors suggested in [81] can also outperform conventional detectors. Especially, the LSTM-based detector shows an outstanding performance for molecular communication use-cases, when dealing with inter-symbol interference [80].

4.2.4. User Mobility Estimation

Predicting user's position, movement and trajectory can improve resource allocation and reduce signal overhead in 6G networks [16]. The authors in [83] used a discrete-time Markov chain based approach to predict the next cell a user is most likely to move into. Results show that the solution can accurately predict both the movement and trajectory of the users. Furthermore, in [84] the authors used HMM algorithm to predict user's location. The model addresses the mobile network as a state-transition graph. The efficiency and accuracy results of the approach were satisfactory. Two unsupervised algorithms for user equipment (UE) association are proposed in [85] in heterogeneous networks at RF and THz frequencies. The simulation results show that proposed algorithms can outperform conventional approaches in both data rate and balancing traffic load.

4.2.5. Security

AI/ML technologies can also be considered in applications of authentication and access control to detect different kinds of attacks, such as jamming and malware attacks, Denial of Service (DoS) or Distributed DoS (DDoS) attacks. In IoT devices, it is important to address authentication and access control without leaking privacy-sensitive information such as localization. In [86], the authors use non-parametric Bayesian methods for IoT authentication, access control, malware detection, with satisfactory results. The authors in [87] propose a DRL based approach that detects various attacking possibilities through unsupervised learning to address the security issue, with result showing a 6 percent extra gain in accuracy. The authors in [88] propose an unsupervised Gausian Mixture Model (GMM) approach for Physical Layer security, enhancing the performance of the model, whereas the authors in [89] used an unsupervised approach combining CNN and Stacked Encoders (SAE) for intrusion detection, achieving a precision of 98.44% black.

4.2.6. UAV Networks

Future 6G networks will support high transmission data rates and wireless broadcast. Unmanned Aerial Vehicle (UAV)-assisted communication networks will be widely used towards achieving these challenges [90]. In UAV-NOMA systems, an UAV often acts as a flying BS to boost the capacity of an existing terrestrial network. In [91], a K-means clustering algorithm is used to spatially cluster correlated users and then a reinforcement Q-learning algorithm is used to place the UAV as BS in a 3-D manner. The authors in [92] proposed MLP and LSTM algorithms techniques to predict the optimal UAV location and optimize user throughout and system performance. The proposed model accurately predicts UAV position and enhances user throughput and system performance.

4.2.7. MIMO

With multiple antennas at the transmitter and receiver, Multiple Input Multiple Output (MIMO) has been widely adopted in wireless systems. The authors in [93] propose an unsupervised fast beamforming DNN design method for maximization of sum-rate in a MIMO single base station system. The proposed approach can preserve the performance, while improving considerably the computational speed, thus achieving results close to optimal.

4.2.8. Visible Light Communications

Effective Radio Frequency (RF) communications systems in indoor use-cases emerge as an important challenge in 6G networks. Visible Light Communications (VLC) as a potential technology, can offer various solutions to this issue. VLC is based on the principle of modulating Light Emitted by Diodes (LEDs), without affecting the human eye, giving an

opportunity to exploit the existing illumination infrastructure for wireless communication. VLC technology is expected to offer very high data-rate short-range communications, needed for 6G Networks [90]. 6G is expected to support transmission rates 100–1000 times higher than those for 5G, so there will be growing frequency and bandwidth demands.VLC can employ high transmission rates and use unlicensed bands. So, it is a promising technique to replace conventional wireless local area networks for indoor communications in 6G networks [94].

Optical Wireless Communications (OWCs) will be widely used in 6G networks and among them, VLC is the most promising frequency spectrum because of the technology advancement and extensive using of light-emitting diodes (LEDs). VLC-based communications do not emit electromagnetic (EM) radiation and have minor interference with other potential EM interference source. Furthermore, VLC has significant advantages in terms of communication security and privacy [95].

VLC can also be widely exploited in Vehicle to everything (V2X) applications and especially in n Vehicle to Vehicle (V2V) applications [90]. In [94], some clustering unsupervised ML techniques (K-means and clustering algorithm perception decision-CAPD)) have been proposed to reduce non linearity in VLC systems. In 2017, CAPD was applied in a multi band VLC system, with the results showing an improvement in the Q-factor by 1.6–2.5 dB. Furthermore, in 2018 a K-means-based pre-distorter was proposed, leading to a 50% improvement of performance [94]. The data for the unsupervised ML models used in 6G problems are listed in Table 7.

Table 7. Unsupervised ML models in 6G problems.

Paper	ML Approach	Application Problem	Description
[68,69]	K-means	Power allocation	Addresses user selection issue and achieves power optimization in NOMA systems
[70]	PC algorithms	D2D systems optimization	Optimizes power control, computational complexity, throughput and resource allocation
[71,72]	CNN, LSTM	Modulation recognition	Achieves better performance results in modulation recognition
[77]	k-means clustering, Fuzzy C-means clustering, LOF, LoOP, Kohonen's SOM	Fault management	Effectively predicts and detects faults/outliers
[78]	k-aware k-means	Fault detection	Self optimizes k-value and accurately detects anomalies
[79]	SOM	Fault detection	Effective fault recognition and recovery
[82]	DetNet	Channel estimation	Effectively estimates channel with much lower computation time
[80,81]	LSTM	Channel estimation	Deals with inter-symbol interference for molecular communication cases
[83]	Discrete Markov chain model	User's mobility estimation	Predicts next cell a user is most likely to move into, predicting movement and trajectory
[84]	HMM	User's location	Addresses the mobile network as a state-transition graph, accuratyle predicting user's location
[85]	Unsupervised UE association algorithms	Data rate, traffic load	Optimizes data rate and traffic load in RF and THz frequency systems
[86]	Non-parametric Bayesian approach	Security	Access control, malware detection in IoT systems
[87]	Unsupervised trained DRL	Security	Detects attack possibilities in 6G systems

Table 7. *Cont.*

Paper	ML Approach	Application Problem	Description
[88]	Unsupervised GMM	Security	Enhances physical layer security
[89]	Unsupervised CNN,SAE	Intrusion detection	Accurately detects intrusion
[91]	k-means clustering & Q-learning	Clustering	Spatially clusters correlated users and places the UAV in 3-D manner
[92]	MLP, LSTM	UAV location	Predicts optimal UAV location and optimizes user throughout and system performance
[93]	DNN	Sum rate	Maximization of sum-rate in a MIMO single base station system, while improving considerably the computational speed
[94]	k-means, CAPD	VLC	Reduces non-linearity in VLC and multi-band VLC systems. CAPD was also applied as pre-distorter, with great performance improvement

4.3. Reinforcement Learning

4.3.1. Optimization Problems

In [96], the authors propose a multi-agent deep reinforcement learning-based model, named Neighbor-Agent Actor Critic (NAAC), for spectrum allocation in 6G network D2D scenarios. This model uses information from user's neighbors for centralized training and utilizes any cooperation between the users to optimize system's performance. The simulation results show that the proposed approach can improve the sum rate of D2D links and have good convergence.

In [97], a deep Q-learning based approach is proposed, namely a Generative Adversarial Network-powered Deep Distributional Q Network (GAN-DDQN) for spectrum allocation per network slice. Simulation results show enhanced performance accuracy compared with conventional deep Q-learning algorithms. In [98], the authors propose a reinforcement Q-learning-based algorithm, for resource allocation. The model minimizes the outage probability of information by assigning the channel resources. The results demonstrate the superior performance and effectiveness of the proposed scheme while satisfying the average power constraint at the energy harvesting node.

In [99], the authors propose a Q-learning based algorithm for channel selection, scanning the order of the channel and so reducing the overhead and possible delays. The proposed approach achieves higher detection probability and accuracy, and reduction of scanning overhead and access delay when compared with state-of-the-art algorithm, resulting to enhanced spectrum sharing. In [100], a deep Q network based algorithm is proposed for cooperative communications in 6G networks. The model aims to select optimal relay from different nodes without needing a network model. Results show that the proposed algorithms can achieve better performance probability, and reduced energy consumption with lower convergence time than existing approaches.

In [101], a deep RL-based algorithm is developed for dynamic power allocation. Each transmitter exploits its neighbors to collect CSI and QoS information and then adapt its needed transmit power. Random variations and delays in the CSI are addressed using deep Q-learning based approach. The proposed algorithm is shown to achieve near-optimal power allocation results based on delayed CSI measurements and is excellent for scenarios where the CSI is significant.

In [102], novel reinforcement learning-based transmission approaches, named Reinforcement Learning Channel-aware Transmission (RL-CAT) and Reinforcement Learning pCAT (RL-pCAT), for data rate optimization are proposed. The proposed models significantly outperform conventional probabilistic approaches and achieve data rate improvements of up to 181 in uplink and up to 270 in downlink transmission direction.

In [103], a DRL-based approach for joint mode selection and resource management is proposed. Each user equipment (UE) can operate either in cloud RAN (C-RAN) mode or D2D mode. The network controller makes intelligent decisions on UE communications and aims to minimize system's power consumption. The proposed approach is compared with other different models to show its effectiveness. In [104], the authors propose a DRL based model to maximize downlink SNR in Intelligent Reflecting Surface (IRS) communications. Simulations results show that the system can, not only achieve almost the upper bound of received SNR, but also reduce the time consumption.

In [105], a DRL actor-critic based model is used for resource allocation optimization and to solve the joint network control challenge in IoT systems. The actor-critic based algorithms reduce the data rate assigned to each IoT network and IoT devices. The algorithm also chooses whether transmission will be in space or terrestrial network. The proposed model outperforms conventional approaches with different network parameters and metrics.

In [106], a Single-Agent Q-learning (SAQ-learning) algorithm is proposed for resource allocation using historical experience with satisfactory result. In the same paper, a Bayesian Learning Automated (BLA) Multi-Agent Q-learning (MAQ-learning) algorithm is proposed for task offloading decision. The effectiveness of the proposed algorithm is confirmed from the comparison with the results of conventional algorithms in various network scenarios.

4.3.2. Caching/Computing

In [107], a DRL MDP-based algorithm is proposed to enhance caching and computing capabilites in cache-aided MEC networks. This approach lead to resource allocation optimization with low complexity and thus is able to achieve quasi-optimal performance under various system setups, and significantly outperform the conventional methods. In [108], the authors propose a deep actor-critic reinforcement learning based model for caching (centralized and decentralized). For centralized edge caching, the model aims at the maximization of cache hit rate, where both the cache hit rate and transmission delay are addressed as performance metrics that need optimization. Results show that the proposed approach outperforms previously applied conventional approaches, such as least frequently used (LFU), least recently used (LRU, etc. In [109], a Multi-Agent Multi-Armed bandit (MAMAB) approach is proposed for caching in 6G networks. The proposed model learns online the caching strategy in various environments (stationary and non-stationary), whereas conventional approaches first estimate the users preference and need and then tries to optimize the caching. Results show great accuracy and performance results of the proposed algorithm. Table 8 reports the RL models used in 6G for optimization and caching problems.

4.3.3. Channel Estimation/Allocation

In [110], the authors propose a RL-based algorithm (based on auction theory model) for channel allocation. Each user try to converge to the optimal allocation while achieving an optimal regret order O (log T), where T is the length of time horizon. The algorithm is based on a Carrier Sensing Multiple access (CSMA) implementation. Simulations show that the algorithm performs very well on realistic LTE and 5G channels and has great potential for B5G systems. In [111], a Markov decision process (MDP)-based algorithm for channel allocation is proposed. The model allocates channels in densely deployed WLANs, leading to enhancement of throughput. The proposed method can achieve more efficient channel allocation or realizes the optimal channel allocation and reducing the number of changes in the systems performance, when compared with state-of-the-art approaches.

4.3.4. Energy Consumption/Harvesting

In [112], author propose a Q-learning and a deep Q-learning algorithms for cooperative networks to user devices and the Small Base Station (SBS) due to different complexities. Results show greater energy saving performance of these approaches over existing methods. In [113], a DRL approach is proposed for optimizing energy consumption in 6G

networks. This model takes mobility into account and accelerates block verification. The reward function considers the total consumed energy for transmission and caching. In this paper, also, a security study is conducted, with the model providing security and privacy protection, while maintaining low-energy consumption. The proposed algorithms achieves 86% of successful content caching requests against 76% of a conventional greedy algorithm and 5% of a random content caching approach.

In [114], the authors propose two DRL-based algorithms for energy harvesting: one hybrid-decision-based actor–critic learning (Hybrid-AC) algorithm and one multi-device hybrid-AC (MD-Hybrid-AC) algorithm for dynamic computation offloading scenarios. Hybrid-AC applies an improvement in the actor–critic architecture. In this approach, the actor outputs offloading ratio and local computation capacity and the critic evaluates these continuous outputs with discrete server selection. MD-Hybrid-AC applies centralized training with decentralized execution in the scenarios. The model constructs a centralized critic for output server selections, and considers the continuous action policies of all devices for actor. Simulation results show that the proposed algorithms have a significant performance improvement compared with conventional and can maintain good balance between time and energy consumption.

In [65], a Deep Q-Network (DQN) based algorithm for energy consumption is proposed. Furthermore, the authors develop a RL algorithm for minimization of prediction error, in order to address a battery's energy prediction challenge. Finally, a two-layer RL network approach is developed to solve the joint access control and battery prediction issue. In this approach the first RL layer deals with the battery's energy prediction and the second, depending on the output of the first layer, produces the access policy of the system. Simulation results show that the three proposed RL algorithms can achieve better performances compared with existing approaches in terms of optimizing energy consumption, sum rate and minimizing the prediction loss.

In [115], a multi-agent DRL-based framework was proposed for power control and maximization of throughput in energy-harvesting super IoT systems. Furthermore, a DNN based for distributed online power control is developed to study the policies in the system. Simulation results show the efficiency of the proposed power control policies, outperforming conventional optimal approaches like Markov decision process, and also achieving throughput close to optimal.

4.3.5. Handover

In [116], the authors propose an offline RL algorithm to optimize Handover decisions. The model is able to decrease excess Handover up to 70% by studying the prolonged user's connectivity. This model can also achieve higher than conventional Handover reduction approaches. In [117], a DRL framework is proposed for handover optimizing and timing in mm-wave systems. The model uses camera images for predicting future data rate of mm-wave links and ensuring that proactive Handover is performed before the presence of obstacles leads to decreasing system's data rate. The proposed approach achieves better performance results than conventional model and is also able to predict the degradations of date rate 500 ms before the occur. In [118], a distributed RL model for Handover optimization in mm-wave systems is proposed, with results showing reduction in signal overhead.

4.3.6. V2V

In [119], a DRL algorithm is adopted to map the correlation between observation and optimal resource allocation in V2V systems. The proposed model satisfies the latency constraints on V2V links and is able to minimize any interference in the V2V system. In [120], a RL-based approach for sum rate optimization in V2V systems is being introduced. The model is a reinforcement distributed Resource Allocation (RA) algorithm, modeled as a multi-agent system. Furthermore, a double deep Q-learning algorithm is applied to jointly train the agents and maximize the sum-rate. Simulation results show that the proposed

RL-based algorithms achieve close to optimal performances, while ensuring limited latency and accurate packet delivery in the V2V link.

4.3.7. UAV

In [121], the authors propose a two-stage DRL algorithm for joint content placement and trajectory design. The two stages of the proposed scheme include offline content placement and online user tracking. In the first stage, the authors maximize users hit rate while constraining cache capacity. In the second stage, a Double Deep Q-Network (DDQN) is developed for online tracking mobile users, while maintaining energy constrains. Simulation results show that the proposed algorithm can easily adapt to dynamic conditions, predict trajectory and provide enhanced achievable throughput.

Table 8. RL models in 6G optimization and caching problems.

Paper	ML Approach	Application Problem	Description
[96]	NAAC	Spectrum allocation	Uses information from user's neighbors and improves sum rate in D2D uses cases
[97]	GAN-DDQN	Spectrum allocation per network slice	Uses a deep Q-learning based approach to optimize spectrum allocation
[98]	Q-learning	Resource allocation	Miimizes outage probability of information by assigning the channel resources, while satisfying average power constraint at the energy harvesting node
[99]	Q-learning	Channel selection	Achieves higher detection probability and accuracy, reduces scanning overhead and access delay
[100]	Deep Q-learning	Energy consumption	In cooperative communications, model selects optimal relay from different nodes without needing network model, reducing consumption with lower convergence time
[101]	Deep Q-learning	Dynamic power allocation	Each transmitter exploits its neighbors to collect CSI and QoS information and then adapt its needed transmit power
[102]	RL-CAT, RL-pCAT	Data rate	Achives data rate improvements both in uplink and downlink directions
[103]	DRL-based approach	Resource management	Makes intelligent decisions on UE communications and minimizes system's power consumption
[105]	Actor-critic	Resource allocation, joint user control	Reduces data rate assigned to each IoT network and IoT device,chooses whether transmission will be in space or terrestrial network.
[106]	SAQ-learning, BLA-MAQ algorithm	Resource allocation, offloading decision	Uses historical experience and achieves optimal accuracy in various network scenarios
[107]	MDP-based algorithm	Caching	Leads to resource allocation optimization with low complexity in cache-aided MEC networks
[108]	Deep actor-critic	Caching	Used for centralized and decentralized uses cases, optimizing cache hit rate and transmission delay
[109]	MAMAB	Caching	Learns online the caching strategy in both stationary and non-starionary environments

4.3.8. Security

In [122], a DRL is proposed to maximize throughput, and security metrics against jamming attacks, in 6G network. Simulation results show that the proposed approach is robust against jamming and can achieve throughput enhancement, compared with conventional policies. In [123], the authors use a Markov model to deal with several ad-

vanced jamming attacks. When dealing with attacks such as swept jamming and dynamic jamming, the authors model a multi-agent reinforcement learning (MARL) algorithm for effective defense. The simulation results show that the algorithm can effectively avoid these advanced jamming attacks, thanks to collaboratively sharing the spectrum to its agents. In [104], a novel DRL-based algorithm is proposed to ensure secure beamforming approach against eavesdroppers in dynamic IRS-aided environments. The model uses post-decision state (PDS) and prioritized experience replay (PER) approaches to boost the learning efficiency and secrecy performance of the system. The proposed novel approach can significantly improve the system secrecy rate and QoS (thus optimal beamforming is required) in IRS-aided secure communication systems.

4.3.9. Visible Light Comunication

In [124], the authors propose a DQN based multi-agent multi-user algorithm for hybrid networks for power allocation. These networks are composed of radio frequency (RF) and visible light communication (VLC) access points (APs). The users are capable of multi-hopping, which can link RF and VLC systems in terms of bandwidth requirements. In the proposed DQN algorithm, each AP is considered an agent and so the transmit power needed for users is optimized by an online power allocation strategy. Simulation results demonstrate faster median convergence time training (90 shorter than typical Q-Learning based algorithm) and convergence rate is 96.1% (whereas conventional QL-based algorithm's convergence rate in 72.3%). In [125], a multi-agent Q-learning algorithms is proposed for power allocation strategy in RF/VLC systems. In these systems, in order to ensure QoS satisfaction, the transmit power at the Aps needs to be optimized. Simulation results demonstrate the effectiveness of the proposed Q-learning based strategy in terms of accuracy and performance.

4.3.10. Fault/Anomaly Management

In [126], a deep Q-learning approach is proposed for fault detection and diagnosis in 6G networks. Simulation results show that the algorithm can use less features and achieve higher accuracy, up to 96.7%

Table 9 holds a brief summary of the RL models used in various 6G problems.

Table 9. RL models in 6G various problems.

Paper	ML Approach	Application Problem	Description
[110]	RL-based on auction model	Channel allocation	Based on a carrier sensing multiple access (CSMA) implementation, performs well for LTE scenarios
[111]	MDP	Channel allocation	Allocates channels in densely deployed WLANs, leading to throughput enhancement
[112]	Q-learning, Deep Q-learning	Energy consumption	Used in cooperative networks on user devices and SBS, respectively, achieving great energy saving results
[113]	DRL	Energy consumption, security	Accelerates block verification, where the reward function considers energy for trasnmission and caching, while providing privacy protection
[114]	Hybrid-AC, MD-Hybrid-AC	Dynamic computation offloading	Theactor outputs offloading ratio and local computation capacity and the critic evaluates these continuous outputs with discrete server selection
[65]	DQN, two-layered Rl algorithm	Energy consumption, joint access control	Minimizes prediction error and predict a battery's energy consumption, while producing access policy
[115]	Multi-agent RL, DNN	Power control	Maximizes throughput in energy-harvesting super IoT systems, while studying PC policies

Table 9. *Cont.*

Paper	ML Approach	Application Problem	Description
[116]	Offline RL model	Handover	Optimizes handover decisions by studying prolonged user's connectivity
[117]	DRL	Handover, timing	Uses camera images for predicting future data rate of mm-wave links and ensuring that proactive Handover is performed
[118]	Distributed RL model	Handover	Minimizes Handover inn mm-wave systems, reducing signal overhead
[119]	Novel RL model	Resource allocation	Satisfies latency constraints on V2V links and is able to minimize any interference in the V2V system
[120]	Distributed Resource allocation model, Double DQN	Resource allocation, Sum rate	Modeled as multi-agent system, while jointly training agents maximizing sum-rate and reducing latency
[121]	Two-stage DRL	Jointcontent placement, trajectory design	Maximize users hit rate while constraining cache capacity, tracks online mobile users while maintaining energy constrains
[122]	DRL	Security	Robust against jamming attacks, achieves throughput enhancement
[123]	MARL	Securiry	Deals with advanced jamming attacks (swept jamming and dynamic jamming) by collaboratively sharing the spectrum to its agents
[104]	Novel RL model	Security	Uses PDS and PER approaches to boost learning efficiency and secrecy performance in dynamic IRS-aided environments
[124]	Multi-agent multi-user DQN	Power allocation	Capable of multi-hoping in VLC/RF systems and addressing each AP as agent, boosting convergence rate and achieving optimal power allocation
[125]	Multi-agent Q-learning	Power allocation	Optimizes transmit power at Aps, ensuring QoS satisfaction and optimal power allocation in VLC/RF systems
[126]	Deep Q-learning	Fault detection	Detects diagnoses faults, using less features and achieving great accuracy

5. Open Issues

ML application can offer new research directions and solutions in wireless communication systems and also support the realization of 6G wireless communication networks and services. Although significant research has emerged on the field of ML in wireless communication systems, there are still many challenges and open issues to be resolved:

- Time Convergence: A careful investigation of the relatively long convergence time of ML methods, as well as the factors that influence the convergence, is needed. Optimizing the time convergence is critical, as long ML time convergence can undermine the performance in highly dynamic wireless networks [127].
- Resource allocation: AI-enabled networks also impact e-health applications. For instance, advancing outside-of-clinic operations by using wearable sensor requires harmonizing network resource allocation across several technologies, and ML can be helpful for such harmonization [127].
- QoS and QoE: A network encompassing a large and diverse set of users will have very dynamic operation, as users may have very different QoS and QoE requirements. For example, users require high throughput and low delay in video stream applications, in the expense of security, but when it comes to payment software, the users demand high security, even in the expense of throughput. In this direction, a design of a

cross-layer, action based ML protocol for different applications is a critical issue, as to meet various requirement while balancing network resources [128].

- UAVs as an Intelligent Service(UaaIS): UaaIS employs UAVs to intelligently provide fundamental services in terms of wireless communication, edge computing, and edge caching, using advanced ML techniques. Due to the scarce resources, it is urgent to perform energy-efficient ML model training and inference for UaaIS, a rather challenging open issue in the field. For example, when a UAV acts as an edge intelligence trainer, energy-efficient training strategies for all participants should be designed, and especially for the UAVs with relatively limited energy [129].

- CSI Acquisition in IRS: The acquisition of timely and accurate CSI plays a crucial role in IRS-enhanced wireless systems and especially in MIMO-IRS and MISO-IRS networks. Obtaining CSI in IRS-enhanced wireless networks is a non-trivial task, that requires a non-negligible training overhead. Additionally, in IRS-assisted NOMA networks, users in each cluster have to share the CSI with each other. Due to the passive characteristic of IRS, CSI acquisition and exchanging are non-trivial tasks. A challenging issue is the employment of ML and DL approaches for exploiting CSI in cases beyond linear correlations [130].

6. Future Trends

6.1. Model Agnostic Meta Learning (MAML)

Meta-learning is an exciting research direction in the field of ML. Model Agnostic Meta Learning (MAML) is a gradient-based meta-learning algorithm that is able to learn a sensitive initialization to perform fast adaptation. Compared to other meta learning methods, MAML has much less complexity. MAML does not depend on any specific model, and only requires the use of gradient descent algorithm to update the parameters. So MAML can be applied to multiple learning problems, such as regression, classification and reinforcement learning, etc. [131,132]. MAML is a field of ML that needs to be further investigated and developed. To this end, few studies are exploring potential solutions. For example, in [133] a MAML- based method is proposed o solve the challenge of associated large number of samples in a wireless channel environment, in order to train a deep neural network (DNN) with good results in terms of Normalized Mean Squared Error (NMSE). Furthermore, the authors in [134] propose a new decoder, namely Model Independent Neural Decoder (MIND) based on a MAML methodology achieving satisfactory parameter initialization in the meta-training stage and accuracy results. The authors in [135] use state-of-the-art meta-learning schemes,namely MAML, FOMAML, REPTILE, and CAVIA, for IoT scenarios using offline and online meta learning approach. The results show the advantage of meta-learning in both offline and online cases as compared to conventional ML approaches. It is an interesting and ongoing direction to developing ML methods that can be utilized in 6G networks in future work.

6.2. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) is a novel class of deep generative models in which training is a minimax zero-sum game between two networks: a Generator (G) and a Discriminator(D) [136]. These networks compete in a unified training process where the generator uses its neural network to produce samples and the discriminator tries to classify these samples as real or fake [137]. The game is played until Nash equilibrium using a gradient-based optimization technique (Simultaneous Gradient Descent), i.e., G can generate images like sampled from the true distribution, and D cannot differentiate between the two sets of images [136]. GANs has gained a lot of attention recently for different applications and seem to be a potential solution to various challenges. For example, the authors in [138] employ a GAN approach to pre-train a deep-RL framework to provide resource allocation for ultra reliable low latency communication (URLLC) in the downlink of a 6G wireless network, with results showing near-optimal performance within the rate-reliability-latency region, depending on the network and service requirements.

Furthermore, the authors in [139] proposea GAN based joint trajectory and power optimization (GAN-JTP) algorithm for a UAV trajectory prediction and power optimization, with results being close to optimal with high convergence speed. In the context of a complex 6G network system, the development of GANs seems crucial for the upcoming challenges.

7. Conclusions

In this review, we focused on the various enhanced capabilities that 6G has to offer, but also to the solutions that ML has to offer to the emerging 6G wireless communication challenges. We have summarized the state of-the-art 6G applications and the deployment of ML algorithms in various fields and applications. The most important ML were explained in detail, focusing on their advantages in dealing with upcoming 6G wireless communications challenges and enhancement of different systems. The interest in exploiting ML in 6G wireless communications challenges will sky rocket in the upcoming years, as 6G networks will soon be realized and the various challenges in the networks can be effectively addressed using ML approaches and models. Finally, we outlined out a handful of open problems and directions worth future research efforts.

Author Contributions: Conceptualization: V.P.R.; methodology, V.P.R. and S.S.; validation, P.S.; data curation, P.S. and S.W.; writing—original draft preparation, V.P.R. and S.S.; formal analysis, V.P.R. and S.W.; writing—review and editing, S.W. and S.S.; visualization, S.S. and P.S.; investigation, S.K.G., G.K.K.; supervision, S.K.G. and G.K.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, R.W.; Nie, J.; Garg, S.; Xiong, Z.; Zhang, Y.; Hossain, M.S. Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems. *IEEE Internet Things J.* **2020**, *8*, 5374–5385. [CrossRef]
2. Piran, M.J.; Suh, D.Y. Learning-driven wireless communications, towards 6G. In Proceedings of the 2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE), London, UK, 22–29 August 2019; pp. 219–224.
3. Rekkas, V.P.; Sotiroudis, S.; Sarigiannidis, P.; Karagiannidis, G.K.; Goudos, S.K. Unsupervised Machine Learning in 6G Networks-State-of-the-art and Future Trends. In Proceedings of the 2021 10th International Conference on Modern Circuits and Systems Technologies (MOCAST), Thessaloniki, Greece, 5–7 July 2021; pp. 1–4.
4. Akhtar, M.W.; Hassan, S.A.; Ghaffar, R.; Jung, H.; Garg, S.; Hossain, M.S. The shift to 6G communications: vision and requirements. *Hum. Centric Comput. Inf. Sci.* **2020**, *10*, 1–27. [CrossRef]
5. Matthaiou, M.; Yurduseven, O.; Ngo, H.Q.; Morales-Jimenez, D.; Cotton, S.L.; Fusco, V.F. The road to 6G: Ten physical layer challenges for communications engineers. *IEEE Commun. Mag.* **2021**, *59*, 64–69. [CrossRef]
6. Basharat, S.; Hassan, S.A.; Pervaiz, H.; Mahmood, A.; Ding, Z.; Gidlund, M. Reconfigurable Intelligent Surfaces: Potentials, Applications, and Challenges for 6G Wireless Networks. *IEEE Wirel. Commun.* **2021**, 1–8. [CrossRef]
7. Zhao, J. A survey of intelligent reflecting surfaces (IRSs): Towards 6G wireless communication networks. *arXiv* **2019**, arXiv:1907.04789.

8. Ji, B.; Han, Y.; Liu, S.; Tao, F.; Zhang, G.; Fu, Z.; Li, C. Several key technologies for 6G: challenges and opportunities. *IEEE Commun. Stand. Mag.* **2021**, *5*, 44–51. [CrossRef]

9. Yaklaf, S.K.A.; Tarmissi, K.S.; Shashoa, N.A.A. 6G Mobile Communications Systems: Requirements, Specifications, Challenges, Applications, and Technologies. In Proceedings of the 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, Tripoli, Libya, 25–27 May 2021; pp. 679–683.

10. Jiang, W.; Han, B.; Habibi, M.A.; Schotten, H.D. The road towards 6G: A comprehensive survey. *IEEE Open J. Commun. Soc.* **2021**, *2*, 334–366. [CrossRef]

11. Malik, U.M.; Javed, M.A.; Zeadally, S.; ul Islam, S. Energy efficient fog computing for 6G enabled massive IoT: Recent trends and future opportunities. *IEEE Internet Things J.* **2021**. [CrossRef]

12. Vinesh, R.; Ancy, C.A. Understanding the Future Communication: 5G to 6G. *Int. Res. J. Adv. Sci. Hub* **2021**, *3*, 17–23.

13. Kaur, J.; Khan, M.A.; Iftikhar, M.; Imran, M.; Haq, Q.E.U. Machine learning techniques for 5G and beyond. *IEEE Access* **2021**, *9*, 23472–23488. [CrossRef]

14. Chen, M.; Challita, U.; Saad, W.; Yin, C.; Debbah, M. Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3039–3071. [CrossRef]

15. Nawaz, S.J.; Sharma, S.K.; Wyne, S.; Patwary, M.N.; Asaduzzaman, M. Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future. *IEEE Access* **2019**, *7*, 46317–46350. [CrossRef]

16. Zhang, S.; Zhu, D. Towards artificial intelligence enabled 6G: State of the art, challenges, and opportunities. *Comput. Netw.* **2020**, *183*, 107556. [CrossRef]

17. Dahrouj, H.; Alghamdi, R.; Alwazani, H.; Bahanshal, S.; Ahmad, A.A.; Faisal, A.; Shalabi, R.; Alhadrami, R.; Subasi, A.; Alnory, M.; et al. An Overview of Machine Learning-Based Techniques for Solving Optimization Problems in Communications and Signal Processing. *IEEE Access* **2021**, *9*, 74908–74938. [CrossRef]

18. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, CA, USA, 2016.

19. Zhou, I.; Makhdoom, I.; Shariati, N.; Raza, M.A.; Keshavarz, R.; Lipman, J.; Abolhasan, M.; Jamalipour, A. Internet of Things 2.0: Concepts, Applications, and Future Directions. *IEEE Access* **2021**, *9*, 70961–71012. [CrossRef]

20. Zou, J.; Han, Y.; So, S.S. Overview of artificial neural networks. *Artif. Neural Netw.* **2008**, *458*, 14–22.

21. Nugrahaeni, R.A.; Mutijarsa, K. Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification. In Proceedings of the 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang, Indonesia, 5–6 August 2016; pp. 163–168.

22. Al-Aidaroos, K.M.; Bakar, A.A.; Othman, Z. Naive Bayes variants in classification learning. In Proceedings of the 2010 International Conference on Information Retrieval & Knowledge Management (CAMP), Shah Alam, Malaysia, 17–18 March 2010, pp. 276–281.

23. Rokach, L.; Maimon, O. Decision trees. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2005; pp. 165–192.

24. Wang, J.; Yang, Y.; Mao, J.; Huang, Z.; Huang, C.; Xu, W. Cnn-rnn: A unified framework for multi-label image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2285–2294.

25. Celebi, M.E.; Kingravi, H.A.; Vela, P.A. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Syst. Appl.* **2013**, *40*, 200–210. [CrossRef]

26. Charte, D.; Charte, F.; García, S.; del Jesus, M.J.; Herrera, F. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Inf. Fusion* **2018**, *44*, 78–96. [CrossRef]

27. Degirmenci, A. Introduction to hidden markov models. *Harv. Univ.* **2014**, 1–5. [CrossRef]

28. De la Rosa, E.; Yu, W. Data-driven fuzzy modeling using restricted Boltzmann machines and probability theory. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 2316–2326. [CrossRef]

29. Mollel, M.S.; Abubakar, A.I.; Ozturk, M.; Kaijage, S.F.; Kisangiri, M.; Hussain, S.; Imran, M.A.; Abbasi, Q.H. A survey of machine learning applications to handover management in 5G and beyond. *IEEE Access* **2021**, *9*, 45770–45802. [CrossRef]

30. Mohammed, S.; Anokye, S.; Guolin, S. Machine learning based unmanned aerial vehicle enabled fog-radio aerial vehicle enabled fog-radio access network and edge computing. *ZTE Commun.* **2020**, *17*, 33–45.

31. Taha, A.; Zhang, Y.; Mismar, F.B.; Alkhateeb, A. Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation. In Proceedings of the 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 26–29 May 2020; pp. 1–5.

32. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [CrossRef]

33. Manju, S.; Punithavalli, M. An analysis of Q-learning algorithms with strategies of reward function. *Int. J. Comput. Sci. Eng.* **2011**, *3*, 814–820.

34. Arabnejad, H.; Pahl, C.; Jamshidi, P.; Estrada, G. A comparison of reinforcement learning techniques for fuzzy cloud auto-scaling. In Proceedings of the 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), Madrid, Spain, 14–17 May 2017, pp. 64–73.

35. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [CrossRef] [PubMed]

36. Konda, V.R.; Tsitsiklis, J.N. Actor-critic algorithms. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2000; Volume 42, pp. 1008–1014.

37. Yang, G.; Zhang, Y.; He, Z.;Wen, J.; Ji, Z.; Li, Y. Machine-learning-based prediction methods for path loss and delay spread in air-to-ground millimetre-wave channels. *IET Microwaves Antennas Propag.* **2019**, *13*, 1113–1121. [CrossRef]

38. Zhang, X.; Zhang, Z.; Yang, L. Joint User Association and Power Allocation in Heterogeneous Ultra Dense Network via Semi-Supervised Representation Learning. *arXiv* **2021**, arXiv:2103.15367.

39. Ruan, L.; Dias, M.P.I.; Wong, E. Machine learning-based bandwidth prediction for low-latency H2M applications. *IEEE Internet Things J.* **2019**, *6*, 3743–3752. [CrossRef]

40. Chen, M.; Saad, W.; Yin, C. Liquid state machine learning for resource and cache management in LTE-U unmanned aerial vehicle (UAV) networks. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 1504–1517. [CrossRef]

41. Nadig, D.; Ramamurthy, B.; Bockelman, B.; Swanson, D. APRIL: An Application-Aware, Predictive and Intelligent Load Balancing Solution for Data-Intensive Science. In Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications, Paris, France, 29 April–2 May 2019; pp. 1909–1917.

42. Kim, J.; Choi, J.P. Sensing coverage-based cooperative spectrum detection in cognitive radio networks. *IEEE Sens. J.* **2019**, *19*, 5325–5332. [CrossRef]

43. Sliwa, B.; Adam, R.; Wietfeld, C. Client-Based Intelligence for Resource Efficient Vehicular Big Data Transfer in Future 6G Network. *arXiv* **2021**, arXiv:2102.08624.

44. Sliwa, B.; Falkenberg, R.; Wietfeld, C. Towards cooperative data rate prediction for future mobile and vehicular 6G networks. In Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT), Virtual, 17–20 March 2020; pp. 1–5.

45. Kwon, H.J.; Lee, J.H.; Choi, W. Machine Learning-Based Beamforming in K-User MISO Interference Channels. *IEEE Access* **2021**, *9*, 28066–28075. [CrossRef]

46. Zhang, H.; Zhang, H.; Liu, W.; Long, K.; Dong, J.; Leung, V.C. Energy efficient user clustering, hybrid precoding and power optimization in terahertz MIMO-NOMA systems. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2074–2085. [CrossRef]

47. Ruan, L.; Dias, I.; Wong, E. Machine intelligence in supervising bandwidth allocation for low-latency communications. In Proceedings of the 2019 IEEE 20th International Conference on High Performance Switching and Routing (HPSR), Xi'an, China, 26–29 May 2019; pp. 1–6.

48. Deng, X.; Jiang, P.; Peng, X.; Mi, C. An intelligent outlier detection method with one class support tucker machine and genetic algorithm toward big sensor data in internet of things. *IEEE Trans. Ind. Electron.* **2018**, *66*, 4672–4683. [CrossRef]

49. Yang, Y.; Gao, F.; Ma, X.; Zhang, S. Deep learning-based channel estimation for doubly selective fading channels. *IEEE Access* **2019**, *7*, 36579–36589. [CrossRef]

50. Beyazıt, E.A.; Özbek, B.; Le Ruyet, D. Deep learning based adaptive bit allocation for heterogeneous interference channels. *Phys. Commun.* **2021**, *47*, 101364. [CrossRef]

51. Antón-Haro, C.; Mestre, X. Learning and data-driven beam selection for mmWave communications: An angle of arrival-based approach. *IEEE Access* **2019**, *7*, 20404–20415. [CrossRef]

52. Yang, Y.; Gao, Z.; Ma, Y.; Cao, B.; He, D. Machine learning enabling analog beam selection for concurrent transmissions in millimeter-wave V2V communications. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9185–9189. [CrossRef]

53. Sim, M.S.; Lim, Y.G.; Park, S.H.; Dai, L.; Chae, C.B. Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation. *IEEE Access* **2020**, *8*, 51634–51646. [CrossRef]

54. Gao, F.; Lin, B.; Bian, C.; Zhou, T.; Qian, J.; Wang, H. FusionNet: Enhanced beam prediction for mmWave communications using sub-6GHz channel and a few pilots. *IEEE Trans. Commun.* **2021**. [CrossRef]

55. Abuzainab, N.; Alrabeiah, M.; Alkhateeb, A.; Sagduyu, Y.E. Deep Learning for THz Drones with Flying Intelligent Surfaces: Beam and Handoff Prediction. *arXiv* **2021**, arXiv:2102.11222.

56. Zhang, Z.; Hua, M.; Li, C.; Huang, Y.; Yang, L. Placement Delivery Array Design via Attention-Based Deep Neural Network. *arXiv* **2018**, arXiv:1805.00599.

57. Wei, Y.; Yu, F.R.; Song, M.; Han, Z. Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor-critic deep reinforcement learning. *IEEE Internet Things J.* **2018**, *6*, 2061–2073. [CrossRef]

58. Mahbooba, B.; Timilsina, M.; Sahal, R.; Serrano, M. Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model. *Complexity* **2021**, *2021*, 11. [CrossRef]

59. Kim, J.; Kim, H. An effective intrusion detection classifier using long short-term memory with gradient descent optimization. In Proceedings of the 2017 International Conference on Platform Technology and Service (PlatCon), Busan, Korea, 13–15 February 2017; pp. 1–6.

60. Wang,W.; Zhu, M.; Wang, J.; Zeng, X.; Yang, Z. End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, 22–24 July 2017; pp. 43–48.

61. Yuan, J.; Ngo, H.Q.; Matthaiou, M. Machine learning-based channel prediction in massive MIMO with channel aging. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 2960–2973. [CrossRef]

62. Alrabeiah, M.; Alkhateeb, A. Deep learning for TDD and FDD massive MIMO: Mapping channels in space and frequency. In Proceedings of the 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 3–6 November 2019; pp. 1465–1470.

63. Wu, H.; Lyu, F.; Zhou, C.; Chen, J.; Wang, L.; Shen, X. Optimal UAV caching and trajectory in aerial-assisted vehicular networks: A learning-based approach. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2783–2797. [CrossRef]

64. Manesh, M.R.; Kenney, J.; Hu,W.C.; Devabhaktuni, V.K.; Kaabouch, N. Detection of GPS spoofing attacks on unmanned aerial systems. In Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2019; pp. 1–6.

65. Chu, M.; Li, H.; Liao, X.; Cui, S. Reinforcement learning-based multiaccess control and battery prediction with energy harvesting in IoT systems. *IEEE Internet Things J.* **2018**, *6*, 2009–2020. [CrossRef]

66. Goudos, S.K.; Tsoulos, G.V.; Athanasiadou, G.; Batistatos, M.C.; Zarbouti, D.; Psannis, K.E. Artificial neural network optimal modeling and optimization of UAV measurements for mobile communications using the L-SHADE algorithm. *IEEE Trans. Antennas Propag.* **2019**, *67*, 4022–4031. [CrossRef]

67. Goudos, S.K.; Athanasiadou, G. Application of an ensemble method to UAV power modeling for cellular communications. *IEEE Antennas Wirel. Propag. Lett.* **2019**, *18*, 2340–2344. [CrossRef]

68. Cui, J.; Ding, Z.; Fan, P.; Al-Dhahir, N. Unsupervised machine learning-based user clustering in millimeter-wave-NOMA systems. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 7425–7440. [CrossRef]

69. Ren, J.; Wang, Z.; Xu, M.; Fang, F.; Ding, Z. An EM-based user clustering method in non-orthogonal multiple access. *IEEE Trans. Commun.* **2019**, *67*, 8422–8434. [CrossRef]

70. Fan, Z.; Gu, X.; Nie, S.; Chen, M. D2D power control based on supervised and unsupervised learning. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 558–563.

71. Rajendran, S.; Meert, W.; Giustiniano, D.; Lenders, V.; Pollin, S. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. Cogn. Commun. Netw.* **2018**, *4*, 433–445. [CrossRef]

72. West, N.E.; O'Shea, T. Deep architectures for modulation recognition. In Proceedings of the2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; pp. 1–6.

73. Guérin, J.; Gibaru, O.; Thiery, S.; Nyiri, E. CNN features are also great at unsupervised classification. *arXiv* **2017**, arXiv:1707.01700.

74. Phan, H.T.H.; Kumar, A.; Feng, D.; Fulham, M.; Kim, J. An unsupervised long short-term memory neural network for event detection in cell videos. *arXiv* **2017**, arXiv:1709.02081.

75. Ergen, T.; Kozat, S.S. Unsupervised anomaly detection with LSTM neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3127–3141. [CrossRef]

76. Trosten, D.J.; Sharma, P. Unsupervised feature extraction—A cnn-based approach. In Proceedings of the Scandinavian Conference on Image Analysis, Norrköping, Sweden, 11–13 June 2019; pp. 197–208.

77. Hashmi, U.S.; Darbandi, A.; Imran, A. Enabling proactive self-healing by data mining network failure logs. In Proceedings of the 2017 International Conference on Computing, Networking and Communications (ICNC), Silicon Valley, CA, USA, 26–29 January 2017; pp. 511–517.

78. Mohamed, A.; Ruan, H.; Abdelwahab, M.H.H.; Dorneanu, B.; Xiao, P.; Arellano-Garcia, H.; Gao, Y.; Tafazolli, R. An Inter-disciplinary Modelling Approach in Industrial 5G/6G and Machine Learning Era. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6.

79. Gómez-Andrades, A.; Munoz, P.; Serrano, I.; Barco, R. Automatic root cause analysis for LTE networks based on unsupervised techniques. *IEEE Trans. Veh. Technol.* **2015**, *65*, 2369–2386. [CrossRef]

80. Liu, L.; Song, D.; Geng, Z.; Zheng, Z. A Real-Time Fault Early Warning Method for a High-Speed EMU Axle Box Bearing. *Sensors* **2020**, *20*, 823. [CrossRef]

81. Farsad, N.; Goldsmith, A. Detection algorithms for communication systems using deep learning. *arXiv* **2017**, arXiv:1705.08044.

82. Samuel, N.; Diskin, T.; Wiesel, A. Deep MIMO detection. In Proceedings of the 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, 3–6 July 2017; pp. 1–5.

83. Mohamed, A.; Onireti, O.; Hoseinitabatabaei, S.A.; Imran, M.; Imran, A.; Tafazolli, R. Mobility prediction for handover management in cellular networks with control/data separation. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 3939–3944.

84. Si, H.;Wang, Y.; Yuan, J.; Shan, X. Mobility prediction in cellular network using hidden markov model. In Proceedings of the 2010 7th IEEE Consumer Communications and Networking Conference,, Las Vegas, NV, USA, 9–12 January 2010; pp. 1–5.

85. Hassan, N.; Hossan, M.T.; Tabassum, H. User Association in Coexisting RF and TeraHertz Networks in 6G. In Proceedings of the 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 30 August–2 September 2020; pp. 1–5.

86. Xiao, L.;Wan, X.; Lu, X.; Zhang, Y.;Wu, D. IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? *IEEE Signal Process. Mag.* **2018**, *35*, 41–49. [CrossRef]

87. Chen, Y.; Zhang, Y.; Maharjan, S.; Alam, M.; Wu, T. Deep learning for secure mobile edge computing in cyber-physical transportation systems. *IEEE Netw.* **2019**, *33*, 36–41. [CrossRef]

88. Sattiraju, R.; Weinand, A.; Schotten, H.D. AI-assisted PHY technologies for 6G and beyond wireless networks. *arXiv* **2019**, arXiv:1908.09523.

89. Yu, Y.; Long, J.; Cai, Z. Network intrusion detection through stacking dilated convolutional autoencoders. *Secur. Commun. Netw.* **2017**, *2017*, 1–10. [CrossRef]

90. Maraqa, O.; Rajasekaran, A.S.; Al-Ahmadi, S.; Yanikomeroglu, H.; Sait, S.M. A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2192–2235. [CrossRef]

91. Liu, Y.; Qin, Z.; Cai, Y.; Gao, Y.; Li, G.Y.; Nallanathan, A. UAV communications based on non-orthogonal multiple access. *IEEE Wirel. Commun.* **2019**, *26*, 52–57. [CrossRef]

92. Munaye, Y.Y.; Lin, H.P.; Adege, A.B.; Tarekegn, G.B. UAV positioning for throughput maximization using deep learning approaches. *Sensors* **2019**, *19*, 2775. [CrossRef]

93. Huang, H.; Xia,W.; Xiong, J.; Yang, J.; Zheng, G.; Zhu, X. Unsupervised learning-based fast beamforming design for downlink MIMO. *IEEE Access* **2018**, *7*, 7599–7605. [CrossRef]

94. Chi, N.; Zhou, Y.; Wei, Y.; Hu, F. Visible light communication in 6G: Advances, challenges, and prospects. *IEEE Veh. Technol. Mag.* **2020**, *15*, 93–102. [CrossRef]

95. Shahraki, A.; Abbasi, M.; Piran, M.; Chen, M.; Cui, S. A comprehensive survey on 6g networks: Applications, core services, enabling technologies, and future challenges. *arXiv* **2021**, arXiv:2101.12475.

96. Li, Z.; Guo, C.; Xuan, Y. A multi-agent deep reinforcement learning based spectrum allocation framework for D2D communications. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.

97. Hua, Y.; Li, R.; Zhao, Z.; Chen, X.; Zhang, H. GAN-powered deep distributional reinforcement learning for resource management in network slicing. *IEEE J. Sel. Areas Commun.* **2019**, *38*, 334–349. [CrossRef]

98. Kang, J.M. Reinforcement learning based adaptive resource allocation for wireless powered communication systems. *IEEE Commun. Lett.* **2020**, *24*, 1752–1756. [CrossRef]

99. Ning, W.; Huang, X.; Yang, K.; Wu, F.; Leng, S. Reinforcement learning enabled cooperative spectrum sensing in cognitive radio networks. *J. Commun. Netw.* **2020**, *22*, 12–22. [CrossRef]

100. Su, Y.; Lu, X.; Zhao, Y.; Huang, L.; Du, X. Cooperative communications with relay selection based on deep reinforcement learning in wireless sensor networks. *IEEE Sens. J.* **2019**, *19*, 9561–9569. [CrossRef]

101. Nasir, Y.S.; Guo, D. Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2239–2250. [CrossRef]

102. Sliwa, B.;Wietfeld, C. A reinforcement learning approach for efficient opportunistic vehicle-to-cloud data transfer. In Proceedings of the 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea, 25–28 May 2020; pp. 1–8.

103. Sun, Y.; Peng, M.; Mao, S. Deep reinforcement learning-based mode selection and resource management for green fog radio access networks. *IEEE Internet Things J.* **2018**, *6*, 1960–1971. [CrossRef]

104. Feng, K.; Wang, Q.; Li, X.; Wen, C.K. Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 745–749. [CrossRef]

105. Shah, H.A.; Zhao, L.; Kim, I.M. Joint Network Control and Resource Allocation for Space-Terrestrial Integrated Network Through Hierarchal Deep Actor-Critic Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2021**, *70*, 4943–4954. [CrossRef]

106. Yang, Z.; Liu, Y.; Chen, Y. Distributed reinforcement learning for NOMA-enabled mobile edge computing. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6.

107. Yang, Z.; Liu, Y.; Chen, Y.; Tyson, G. Deep reinforcement learning in cache-aided MEC networks. In Proceedings of the ICC 2019-2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; pp. 1–6.

108. Zhong, C.; Gursoy, M.C.; Velipasalar, S. Deep reinforcement learning-based edge caching in wireless networks. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 48–61. [CrossRef]

109. Xu, X.; Tao, M.; Shen, C. Collaborative multi-agent multi-armed bandit learning for small-cell caching. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 2570–2585. [CrossRef]

110. Zafaruddin, S.M.; Bistritz, I.; Leshem, A.; Niyato, D. Multiagent Autonomous Learning for Distributed Channel Allocation in Wireless Networks. In Proceedings of the 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, 2–5 July 2019; pp. 1–5.

111. Nakashima, K.; Kamiya, S.; Ohtsu, K.; Yamamoto, K.; Nishio, T.; Morikura, M. Deep reinforcement learning-based channel allocation for wireless lans with graph convolutional networks. *IEEE Access* **2020**, *8*, 31823–31834. [CrossRef]

112. Tang, J.; Tang, H.; Zhang, X.; Cumanan, K.; Chen, G.; Wong, K.K.; Chambers, J.A. Energy minimization in D2D-assisted cache-enabled Internet of Things: A deep reinforcement learning approach. *IEEE Trans. Ind. Inform.* **2019**, *16*, 5412–5423. [CrossRef]

113. Dai, Y.; Xu, D.; Zhang, K.; Maharjan, S.; Zhang, Y. Deep reinforcement learning and permissioned blockchain for content caching in vehicular edge computing and networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4312–4324. [CrossRef]

114. Zhang, J.; Du, J.; Shen, Y.;Wang, J. Dynamic computation offloading with energy harvesting devices: A hybrid-decision-based deep reinforcement learning approach. *IEEE Internet Things J.* **2020**, *7*, 9303–9317. [CrossRef]

115. Sharma, M.K.; Zappone, A.; Assaad, M.; Debbah, M.; Vassilaras, S. Distributed power control for large energy harvesting networks: A multi-agent deep reinforcement learning approach. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *5*, 1140–1154. [CrossRef]

116. Mollel, M.S.; Kaijage, S.F.; Michael, K. *Deep Reinforcement Learning Based Handover Management for Millimeter Wave Communication*; The Nelson Mandela African Institution of Science and Technology (NM-AIST): Arusha, Tanzania, 2021; Volume 9.

117. Koda, Y.; Nakashima, K.; Yamamoto, K.; Nishio, T.; Morikura, M. Handover management for mmwave networks with proactive performance prediction using camera images and deep reinforcement learning. *IEEE Trans. Cogn. Commun. Netw.* **2019**, *6*, 802–816. [CrossRef]

118. Sana, M.; De Domenico, A.; Strinati, E.C.; Clemente, A. Multi-agent deep reinforcement learning for distributed handover management in dense mmWave networks. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 8976–8980.

119. Ye, H.; Li, G.Y.; Juang, B.H.F. Deep reinforcement learning based resource allocation for V2V communications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3163–3173. [CrossRef]

120. Vu, H.V.; Liu, Z.; Nguyen, D.H.; Morawski, R.; Le-Ngoc, T. Multi-agent reinforcement learning for joint channel assignment and power allocation in platoon-based C-V2X systems. *arXiv* **2020**, arXiv:2011.04555.

121. Wu, C.; Shi, S.; Gu, S.; Zhang, L.; Gu, X. Deep reinforcement learning-based content placement and trajectory design in urban cache-enabled UAV networks. *Wirel. Commun. Mob. Comput.* **2020**, *2020*, 1–11. [CrossRef]

122. Yazdinejad, A.; Parizi, R.M.; Dehghantanha, A.; Choo, K.K.R. Blockchain-enabled authentication handover with efficient privacy protection in SDN-based 5G networks. *IEEE Trans. Netw. Sci. Eng.* **2019**, *8*, 1120–1132. [CrossRef]

123. Wang, X.; Xu, Y.; Chen, J.; Li, C.; Liu, X.; Liu, D.; Xu, Y. Mean field reinforcement learning based anti-jamming communications for ultra-dense internet of things in 6G. In Proceedings of the 2020 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 21–23 October 2020; pp. 195–200.

124. Ciftler, B.S.; Abdallah, M.; Alwarafy, A.; Hamdi, M. DQN-Based Multi-User Power Allocation for Hybrid RF/VLC Networks. In Proceedings of the ICC 2021-IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6.

125. Kong, J.; Wu, Z.Y.; Ismail, M.; Serpedin, E.; Qaraqe, K.A. Q-learning based two-timescale power allocation for multi-homing hybrid RF/VLC networks. *IEEE Wirel. Commun. Lett.* **2019**, *9*, 443–447. [CrossRef]

126. Zhang, P.;Wu, M.; Zhu, X. Research on Network Fault Detection and Diagnosis Based on Deep Q Learning. In Proceedings of the International Conference on Wireless and Satellite Systems, Nanjing, China, 17–18 September 2020; pp. 533–545.

127. Elsayed, M.; Erol-Kantarci, M. AI-enabled future wireless networks: Challenges, opportunities, and open issues. *IEEE Veh. Technol. Mag.* **2019**, *14*, 70–77. [CrossRef]

128. Tang, F.; Mao, B.; Kawamoto, Y.; Kato, N. Survey on Machine Learning for Intelligent End-to-End Communication towards 6G: From Network Access, Routing to Traffic Control and Streaming Adaption. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1578–1598. [CrossRef]

129. Dong, C.; Shen, Y.; Qu, Y.; Wang, K.; Zheng, J.; Wu, Q.; Wu, F. UAVs as an Intelligent Service: Boosting Edge Intelligence for Air-Ground Integrated Networks. *IEEE Netw.* **2021**, *35*, 167–175. [CrossRef]

130. Liu, Y.; Liu, X.; Mu, X.; Hou, T.; Xu, J.; Di Renzo, M.; Al-Dhahir, N. Reconfigurable intelligent surfaces: Principles and opportunities. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1546–1577. [CrossRef]

131. Finn, C.; Levine, S. Meta-learning and universality: Deep representations and gradient descent can approximate any learning algorithm. *arXiv* **2017**, arXiv:1710.11622.

132. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.

133. Zeng, J.; Sun, J.; Gui, G.; Adebisi, B.; Ohtsuki, T.; Gacanin, H.; Sari, H. Downlink CSI Feedback Algorithm with Deep Transfer Learning for FDD Massive MIMO Systems. *IEEE Trans. Cogn. Commun. Netw.* **2021**; doi:10.1109/TCCN.2021.3084409. [CrossRef]

134. Jiang, Y.; Kim, H.; Asnani, H.; Kannan, S. Mind: Model independent neural decoder. In Proceedings of the 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, 2–5 July 2019; pp. 1–5.

135. Park, S.; Jang, H.; Simeone, O.; Kang, J. Learning to demodulate from few pilots via offline and online meta-learning. *IEEE Trans. Signal Process.* **2020**, *69*, 226–239. [CrossRef]

136. Saxena, D.; Cao, J. Generative Adversarial Networks (GANs) Challenges, Solutions, and Future Directions. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–42. [CrossRef]

137. Alqahtani, H.; Kavakli-Thorne, M.; Kumar, G. Applications of generative adversarial networks (gans): An updated review. *Arch. Comput. Methods Eng.* **2021**, *28*, 525–552. [CrossRef]

138. Kasgari, A.T.Z.; Saad, W.; Mozaffari, M.; Poor, H.V. Experienced deep reinforcement learning with generative adversarial networks (GANs) for model-free ultra reliable low latency communication. *IEEE Trans. Commun.* **2020**, *69*, 884–899. [CrossRef]

139. Li, Z.; Liao, X.; Shi, J.; Xue, X.; Li, L.; Xiao, P. MD-GAN Based UAV Trajectory and Power Optimization for Cognitive Covert Communications. *IEEE Internet Things J.* **2021**. [CrossRef]