International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-020-01178-4

ORIGINAL ARTICLE

A survey of 5G network systems: challenges and machine learning approaches

Hasna Fourati[1] · Rihab Maaloul[1] · Lamia Chaari[1]

Received: 11 December 2019 / Accepted: 1 August 2020

Abstract
5G cellular networks are expected to be the key infrastructure to deliver the emerging services. These services bring new requirements and challenges that obstruct the desired goal of forthcoming networks. Mobile operators are rethinking their network design to provide more flexible, dynamic, cost-effective and intelligent solutions. This paper starts with describing the background of the 5G wireless networks then we give a deep insight into a set of 5G challenges and research opportunities for machine learning (ML) techniques to manage these challenges. The first part of the paper is devoted to overview the fifth-generation of cellular networks, explaining its requirements as well as its key technologies, their challenges and its forthcoming architecture. The second part is devoted to present a basic overview of ML techniques that are nowadays applied to cellular networks. The last part discusses the most important related works which propose ML solutions in order to overcome 5G challenges.

Keywords 5G cellular network · 5G services · 5G key technologies · 5G architectures · 5G challenges · ML solutions · Intelligence · SON

1 Introduction

1.1 Problem statement

Due to the potential increase in mobile traffic and the rapid expansion of communication infrastructures, 4G can no longer meet the actual needs of users. In order to improve Quality of Services (QoS) and reach user satisfaction, a new generation "5G" is emerging and it is perceived to fulfil several requirements such as:

- Support a huge number of connected devices 10–100 times higher than 4G [1–3].
- Support a mobile data volume per area 1000 times higher than 4G [1, 3].
- Provide a data rate 10–100 × higher than 4G [1–4].
- 5 times reduced End-to-End (E2E) latency, reaching 5 ms [1].
- Provide nearly 100% availability [1].
- Provide 100% geographical coverage [1, 5, 6].
- Support the coexistence of different Radio Access Network (RAN) technologies [4].
- Improved security and privacy [4].
- Energy consumption 10 times less than 4G [2, 3].
- Support real-time processing and transmission.
- The network management operation expenses 5 times less than 4G.
- Smooth current wireless technologies integration.
- Network will be more flexible, intelligent, dynamic and open service [4].
- Cost-efficient in terms of CAPital and OPerational EXpenditures (CAPEX and OPEX) [4].
- Battery life for devices 10 times longer than 4G [1, 7]
- Provide Millimeter-wave (mm-wave) spectrum per Hz 10–100 × less than 4G spectrum below 3 GHz [8].
- Small cells should be 10–100 × cheaper than macro cells and power-efficient higher than macro cells [8].

✉ Hasna Fourati
hasna.fourati21@gmail.com

Rihab Maaloul
rihab.maaloul.abid@gmail.com

Lamia Chaari
lamiachaari1@gmail.com

[1] Laboratory of Technology and Smart Systems (LT2S), Digital Research Center of SFAX (CRNS), University of Sfax, Sfax, Tunisia

In order to address these requirements, 5G must involve technologies that can transmit and receive large volumes of data fast with lower or equal cost as the current mobile technology, while integrating and managing heterogeneous network elements in a seamless way. Therefore, several key technologies adopted by 4G will be integrated into 5G with certain improvement. Among them we mention: ultra-dense deployment of small cells (UDSC) [8, 9], dense multiple-radio access technology (Multi-RAT) in heterogeneous networks (HetNets) [10, 11], massive multi-input multi-output (Massive-MIMO) [12, 13], beamforming, orthogonal frequency-division multiplexing (OFDM) amelioration [2, 7], machine-to-machine communication (M2M) [14], device-to-device communication (D2D) [2, 15], cognitive radio (CR) [16, 17], cloud network (CN) [18], network virtualization (NV) [19], network function virtualization (NFV) and software defined networking (SDN) [20, 21].

In addition, other new technologies are integrated in 5G network, the most common ones are: millimeter-wave signals (Mm-wave) [22], full duplex (FD) wireless communication [23, 24], mobile edge computing (MEC) [25, 26], and network slicing (NS) [27].

The integration of these technologies in 5G is introducing several challenges. The most important ones are: large data volume management caused by the UDSC and the interference between macro and small cells, Radio Access Technology (RAT) selection, massive MIMO challenges, synchronism and orthogonality of 5G applications, efficient proximity detection and interference in D2D, control and data plane separation, NFV and SDN security, control and confidentiality in SDN, high path loss, antenna array, high atmospheric attenuation, high penetration loss, Self-Interference in FD communication.

Consequently, it is being necessary to automate and manage the increasingly more complex 5G network in order to tackle the facing challenges and to meet the 5G requirements.

Therefore, new management approaches must be deployed with a view to improve more intelligence in 5G.

In fact, these approaches should be more accurate and practical to operate the network resources and large data generated in 5G and to meet the needs of both users and operators [4, 28]. Artificial Intelligence (AI) has the potential to deliver and operate new services with different and various requirements on top of the future network architecture. Recently, AI has followed by ML which is one of the most promising AI techniques. ML helps systems to learn, optimize and analyzing large data volume that humans are not able to use systematically. It gives the ability to not only prevent problems but also to predict them accurately from arising in real time [4, 29, 30].

1.2 Motivation and contributions

Over the past few years, preliminary interest and discussions about the migration towards 5G cellular networks that are expected to support extremely high data rates and requiring intelligent paradigms to design and manage the networks. In this respect, the employment of ML tools into 5G systems have received significant attention from international projects [31, 32] and research bodies.

This paper covers not only the research on 5G background and ML techniques but also a basic overview of the challenges related to the adoption of advanced technologies in 5G and explains how ML solutions can overcome these challenges.

There are extensive surveys that cover approaches related to the application of AI techniques in various technologies of cellular networks. In order to compare our survey to others, Table 1 summarizes relevant surveys of five past years, as [4, 33–37]. Others focus on understanding 5G background, its architectures, technologies, services and applications without addressing AI technology and especially ML technique, such as [1, 6, 10, 11, 38–45].

We summarize in Table 2 the current surveys that focusing particularly on ML techniques for 5G challenges, such as [46–51]. All of them list the existing key enablers 5G technologies and discuss the problems of their integrations in 5G such as [4, 48, 50–52].

Authors in [53] give special focus on 5G applications, especially healthcare application, their problems and how 5G can serve it using AI and ML solutions. In addition, [47] enumerates the 5G applications and services highlighting their challenges and their ML solutions without considering architectures of 5G. The article [50] lists 5G services and gives an insight into 5G architectures. The contributions of this survey are as follows:

- We mention some key enablers technologies toward 5G and we study their positive contribution to improve the quality of service and to meet the 5G requirements, then, we identify the challenges of the integration on the network operation. Indeed, we will see how big an impact they have.
- We highlight a comprehensive view on 5G background, especially the 5G benefits and impacts on modern applications, the different models of 5G architecture and their challenges to improve quality of service, user experience and management expectation.
- We provide an extensive overview of ML techniques, their definitions, the ML categories classification and how their algorithms perform.
- We provide a basic overview of current ML solutions applied to face 5G challenges.

Table 1 Major related works in cellular networks

Cellular network			5G					
With artificial intelligence			Without artificial intelligence			With artificial intelligence		
Ref+Y	Pros	Cons	Ref+Y	Pros	Cons	Ref+Y	Pros	Cons
[33] 2015	Illustration of Typical AI techniques in accordance with various HetNets problems. AI-based techniques for resource management applications	Ignorance of networking applications (clustering, routing, BS switching, etc.), mobility and localization applications Limited insights to ML techniques	[10] 2014	Guideline for technology development toward the 5G Significant study for achieving 1000-times capacity increase	Ignorance of key concepts like Self Organizing Network (SON), SDN, slicing and virtualization	[51] 2016	Interdisciplinary research of big data and channel modeling Illustrate the advantage of combining stochastic and deterministic models	Neglect issues related to trusted database structure
[54] 2015	Overview of surveys, providing a vision on how AI can simplify network management	Very limited issues in resource management and inter-cell interference coordination	[45] 2014	Illustrate the usefulness of small cell, D2D communication, SDN, M2M Points out the posed challenges when adopting these technologies	Ignorance of autonomous and SON network concepts	[48] 2016	Present the potential of applying ML to 5G network applications including massive MIMO and smart grids Show research gaps between ML and 5G that remain unexplored	Overlook deep learning techniques
[37] 2016	Vision of how the building RAN management procedures based on AI concepts Focus on the time domain traffic pattern	No identification of spatio-temporal traffic complementarities among neighbor cells	[1] 2015	Overview for the envisioned flexible 5G network architecture Description of 5G key enablers in accordance with requirements	Ignorance of slicing concept	[55] 2017	A tutorial on neural networks for wireless networks Various neural network such as recurrent, spiking, and deep neural networks	Overlook applications related to significant technologies such as D2D communication, cognitive radio
[56] 2016	Overview on the vehicular cloud and operational components Focus on the resources allocation for VMs in vehicular clouds	Limited related works of AI in provisioning and allocation of resources	[41] 2015	Generic review of 5G architecture and technologies Framework that incorporates networking ingredients (D2D, cloud computing, Internet of Things, etc.)	Overlook SDN, SON and slicing concepts	[57] 2017	Discuss the potential of AI and ML on the networking ecosystem as enabler for cognitive SON Significant description of SON and its generic categories	Overlook deep learning techniques
[58] 2017	Framework for real-time decision informatics for internet of things systems	Limited to internet of things	[11] 2015	Generic review of 5G technologies, and the impact their adoption	Focus only on radio resource management	[50] 2017	Highlight opportunities and challenges of incorporating AI into 5G architectures	Neglect mobile big data prospect

Table 1 (continued)

Cellular network			5G					
With artificial intelligence			Without artificial intelligence			With artificial intelligence		
Ref+Y	Pros	Cons	Ref+Y	Pros	Cons	Ref+Y	Pros	Cons
[59] 2017	A survey on the progress of deep learning for network traffic control	Very generic consideration of network traffic control systems	[43] 2016	Analysis on the impact of pilot contamination in massive MIMO systems Helpful investigation in the standardization and the ongoing research on mm-wave and massive MIMO	Lack the review of the hybrid beamforming design techniques in the presence of pilot contamination and the channel estimation errors	[60] 2017	A survey of future mobile networks analysis and optimization	Overlook deep learning and big data
[4] 2017	A survey of ML applications in self evolution networks Important applications emerge in the computer network domain	Neglect the posed optimization challenges with the emergence of 5G key technologies	[6] 2016	Exhaustive review of wireless evolution toward 5G Point out relevant research problems that remain unsolved	Neglect issues related to mobile data traffic, i.e. mobile networking big data prospect	[4] 2017	Holistic view of the applicability, challenges and limitations of ML techniques in SON Significant classification and comparison	Overlook deep learning and big data
[35] 2017	Insights on the need for intelligence functionalities for network automation	Limited examples of AI mechanisms	[42] 2016	Design guidelines and key considerations for 5G radio access network architecture	Overlook important technologies: massive MIMO, Ultra dense network and D2D	[53] 2017	Overview of how 5G-enabled healthcare revolution can fix the deficiencies of the current healthcare system Overview of how ML algorithms can be integrated with healthcare systems to detect the anomalies	Discuss the technological trends only in terms of healthcare
[34] 2018	Reveals the potential of employing ML and big data analytics to realize self-sustaining and proactive wireless networks	Limited related works of AI and ML techniques applications	[61] 2016	Survey of promising technologies for 5G networks based on standardization status	Only three main technologies SDN, internet of things, cloud	[52] 2018	Comprehensive survey on the utilization of AI integrating ML, data analytics and natural language process techniques for efficient wireless networks	Limited literature evidence on how ML can assist in meeting the specific and practical 5G requirements

Table 1 (continued)

Cellular network			5G					
With artificial intelligence			Without artificial intelligence			With artificial intelligence		
Ref+Y	Pros	Cons	Ref+Y	Pros	Cons	Ref+Y	Pros	Cons
[62] 2018	Application of ML techniques in different networking area including traffic prediction, routing and classification, congestion control, resource and fault management, QoS and QoE management, and network security	Missing study in terms of complexity and implementation feasibility	[38] 2017	Introduction to the potential physical technologies adopted in 5G MIMO, mm-wave, small cell	Limited physical and software technologies	[63] 2018	Key technologies used in 5G to meet requirements: ultra-dense deployment, Mm-wave, Massive MIMO	Only three technologies of 5G: not discuss on FD, MEC, NFV, D2D, M2M, beamforming, network slicing technologies
[64] 2018	Identify the development and requirement of ML techniques based on the mobile big data process	Missing challenges related to scalability, of a learning model	[44] 2018	A Survey on hybrid beamforming techniques in 5G: hybrid beamforming architectures, Mm-wave, Massive MIMO, and HetNets/small cells	Interference management, handover and mobility management and antenna selection are not addressed	[49] 2018	Overview relevant solutions of the 5G and internet of things technologies Discuss the need for intelligence in internet of things-based 5G	Slight undertake of the ML methods
[65] 2018	Overview the applications of data analysis in mobile big data context	Slight presentation for the corresponding challenges	[39] 2018	State-of-the-art of 5G internet of things, key enabling technologies, requirements 5G enabled internet of things, 5G-internet of things architecture	Consider limited physical technologies	[46] 2019	Review various deep learning models Focus on upper layers	Neglect MAC and physical layer ML approaches
[36] 2018	Analysis of self-organized network management, with an end-to-end perspective of the network	Ignorance of mobility-awareness	[40] 2019	Extensive survey of achievements and research findings for 5G wireless communication	Limited to ultra dense network, Mm-wave and CoMP cooperation	[47]2019	Supervised learning based qos assurance architecture for 5G networks	Limited related work and approaches

Table 1 (continued)

Cellular network			5G					
With artificial intelligence			Without artificial intelligence			With artificial intelligence		
Ref+Y	Pros	Cons	Ref+Y	Pros	Cons	Ref+Y	Pros	Cons
						Our contribution	All 5G technologies and applications Most 5G challenges Physical and service based 5G architecture More than three ML categories Various works related to ML solutions applied in 5G network	

Table 2 Major related works in ML applied in 5G challenges

Ref+Y	5G Key enablers technologies		5G Background					ML Background				ML for 5G
	Key technologies	Key technologies Challenges	Services	Applications	Applications challenges	Architecture	Quality of expectation	SL	UL	RL	DL and Others ML	
[51] 2016	✓	✓										✓
[66] 2016	✓	✓										✓
[48] 2017	✓	✓						✓	✓	✓		✓
[50] 2017	✓	✓	✓			✓		✓	✓	✓		✓
[4] 2017	✓	✓						✓	✓	✓	✓	✓
[53] 2017	✓		✓	✓	✓						✓	✓
[52]2018	✓	✓									✓	✓
[63] 2018	✓	✓						✓	✓	✓	✓	✓
[49] 2018	✓			✓	✓							✓
[46] 2019	✓							✓	✓	✓	✓	✓
[47] 2019	✓	✓	✓	✓	✓		✓	✓				✓
Our contribution	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

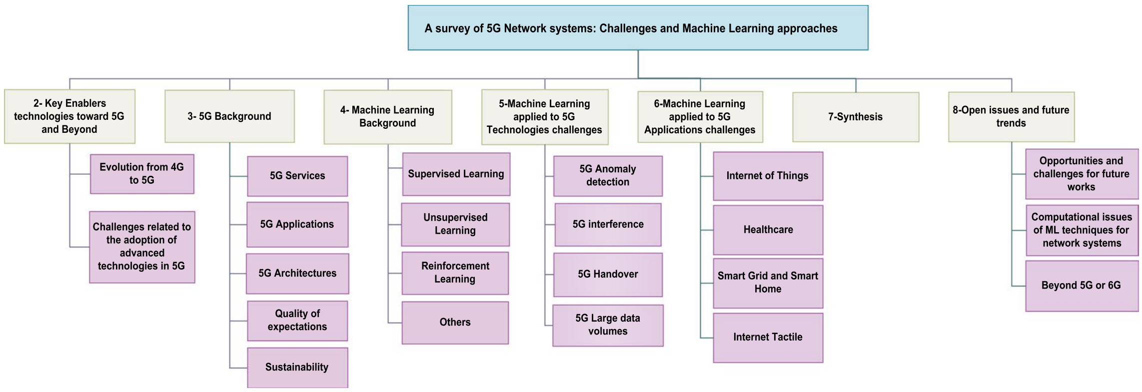

Fig. 1 Paper organization

1.3 Scope and paper organization

The objective of this survey is to give a definite guide for 5G network and ML methods researchers and understand how the existing works addressed 5G problems and developed solutions in order to tackle the 5G challenges, what are ML methods that must be applied and how it can solve the problems. This paper is organized as follows. Section 2 discusses the key enablers technologies toward 5G, beyond and their 5G challenges. Section 3 provides a tutorial on the 5G context, its applications, architectures, quality of expectations and sustainability. Section 4 provides a tutorial on the Machine Learning context. Sections 5 and 6 examine the benefits of ML solutions to tackle the 5G challenges proposed in some current works. Section 7 synthesizes the global vision of proposed challenges and ML solutions. Section 8 discusses some open issues and future trends. Figure 1 includes this organization.

1.4 List of acronyms

The following Table 3 presents the list of acronyms used in our work.

2 Key enablers technologies toward 5G and beyond

2.1 Evolution from 4G to 5G

The 4G network has already embodied several key technologies in order to meet their requirements such as the deployment of small cells, RAT, MIMO, OFDM, M2M and D2D communication, cloud computing, NFV and SDN. Critical improvements should take place in all these technologies to guarantee the high QoS in 5G networks.

- Ultra-dense deployment of small cells (UDSC): Over the years and with the presence of new technologies, the size of the cell has decreased more and more [63]. The dense deployment of the small cell was started with 4G and it is expected to be more dense with 5G to achieve higher throughput and capacity [7]. The deployment of small cells, e.g., microcells, picocells and femtocells provides an increase of the network capacity and better spectrum exploitation. Besides, it off-loads the macro cells and minimizes energy consumption. Consequently, this deployment minimizes the number of users belonging to the off-loading macro cells [2, 8, 9, 67, 68].
- Dense multiple-radio access technology (Multi-RAT) in heterogeneous networks (HetNet) the networks continue to be more and more heterogeneous as we move towards 5G. These networks can connect several devices with several operating systems and protocols [8]. In HetNet, different types of nodes perform with different transmission powers and data processing capability and it supports different RATs (multi-RAT) [10]. Besides, the user can connect to several Base Stations (BS) or Access Point (AP) simultaneously which uses the same RAT or different RATs [9, 11].
- Massive multi-input multi-output (Massive-MIMO): It can improve signal strength and cell edge performance and increase cell throughput compared to traditional 4G systems [2]. MIMO is a multiplexing technique. It is used in wireless and mobile networks that allow data transfer at longer range and higher speed. It consists of multiplying the signals to transmit the same information. It improves the transmission rate and the signal-to-noise ratio quality by using a large number of transmit and receive antennas [8, 63]. This concept was defined by the 3GPP in LTE-Advanced (LTE-A) [15]. Due to the massive growth in the number of devices, the 5G infrastructure needs a massive MIMO technique to serve large devices without congestion. Massive-

Table 3 List of acronyms

Symbol	Description
ML	Machine learning
QoS	Quality of service
E2E	End-to-end
RAN	Radio access network
CAPEX	CAPital EXpenditures
OPEX	OPerational EXpenditures
Mm-wave	Millimeter-wave
UDSC	Ultra-dense deployment of small cells
Multi-RAT	Multiple-radio access technology
HetNets	Heterogeneous networks
Massive-MIMO	Massive multi-input multi-output
OFDM	Orthogonal frequency-division multiplexing
M2M	Machine-to-machine
D2D	Device-to-device
CR	Cognitive radio
CN	Cloud network
NV	Network virtualization
NFV	Network function virtualization
SDN	Software defined networking
FD	Full duplex
MEC	Mobile edge computing
NS	Network slicing
RAT	Radio access technology
AI	Artificial intelligence
SON	Self organizing networks
BS	Base stations
AP	Access point
LTE-A	Long term evolution-advanced
ISI	Inter-symbol-interference
OFDMA	Orthogonal frequency division multiple access
LTE	Long term evolution
FBMC	Filter bank multi-carrier
GFDM	Generalized frequency division multiplexing
UFMC	Universal filtered multicarrier
CRN	Cognitive radio network
ISPs	Internet service providers
InPs	Infrastructure providers
SPs	Service providers
VNs	Virtual networks
ONF	Open networking foundation
HD	Half-duplex
NF	Network functions
SINR	Signal-to-interference-plus-noise ratio
CSI	Channel state information
FD-MIMO	Full-dimension MIMO
VM	Virtual machine
VNFs	Virtual network functions
IDS	Intrusion detection system
SI	Self-interference
eMBB	enhanced Mobile Broadband
mMTC	massive Machine-Type Communications

Table 3 (continued)

Symbol	Description
URLLC	Ultra-reliable and low-latency communications
IoT	Internet of things
IoV	Internet of vehicles
POCT	Point-of-care testing
EHR	Electronic health record
SG	Smart grid
H2H	Human-to-human
MTC	Machine type communications
MAC	Medium access control
MBS	Macro cell base stations
SBS	Small Cell base stations
M2M GW	M2M GatWay
LAN	Local area network
C-RAN	Centralized-radio access network
BBU	BaseBand unit
RRH	Remote radio heads
FFT	Fast Fourier Transform
AUSF	Authentication server function
AMF	Access and mobility management function
DN	Data network
NEF	Network exposure function
NRF	Network repository function
NSSF	Network slice selection function
PCF	Policy control function
SMF	Session management function
UDM	Unified data management
UDR	Unified data repository
UPF	User plane function
AF	Application function
UE	User equipment
NWDAF	Network data analytics function
CHF	CHarging function
IP	Internet protocol
QoE	Quality of experience
HTTP	Hypertext transfer protocol
HAS	HTTP adaptive streaming
SL	Supervised learning
UL	Unsupervised learning
RL	Reinforcement learning
K-NN	K-nearest neighbors
NB	Naive Bayesian
SVM	Support vector machines
ANN	Artificial neural network
GLM	Generalized linear models
BDT	Binary decision tree
RNN	Recurrent NN
DT	Decision tree
PCA	Principle component analysis
ICA	Independent component analysis
MDP	Markov decision process
QL	Q-learning

Table 3 (continued)

Symbol	Description
TL	Transfer learning
DL	Deep learning
MC	Markov chains
HMM	Hidden Markov models
GA	Genetic algorithm
5G-PPP	5G infrastructure public–private partnership
KPI	Key performance indicator
ASD	Anomaly symptom detection
NAD	Network anomaly detection
DBN	Deep belief network
LSTM	Long short term memory
SAE	Stacked AutoEncoders
CDR	Call detail record
DOHM	Detecting outage by hidden Markov model
LOF	Local outilier factor
M-LOF	Modified local outilier factor
CTMC	Continuous time Markov chain
SIR	Signal-to-interference ratio
H-ELM	Hierarchical extreme learning machine
APPC	Affinity propagation power control
VACR	Victim-aware channel rearrangement
DDRM	Data-driven resource management
RSRP	Reference signal received power
HMS	HeNB management system
FBS	Femto base station
FUE	Femto user equipment
MUE	Macro user equipment
RBs	Resource blocks
DHO	Data-driven handover optimization
HOM	HandOver margin
TTT	Time-to-trigger
RLF	Radio link failures
LECC	LEarning-based cooperative caching
MPC	Model predictive control
5GCS	5G cognitive system
CART DT	CART decision tree
NN	Neural network
BL	Bayesian learning
ZSM	Zero-touch network and service management
EB	ExaByte
B5G	Beyond 5G
MBRLLC	Mobile broadband reliable low latency communication
HCS	Human-centric services
MPS	Multi-purpose 3CLS and energy services
LISs	Large intelligent surfaces
RF	Radio frequency
VLC	Visible light communications
V2X	Vehicle-to-everything
XR	eXtended reality
AR/MR/VR	Augmented, virtual and mixed reality ecosystem
CRAS	Connected robotics and autonomous systems

Table 3 (continued)

Symbol	Description
WBCI	Wireless brain–computer interactions
DLT	Blockchain and distributed ledger technologies
DRL	Deep reinforcement learning

MIMO (or large-scale antenna systems, very large MIMO or Hyper MIMO or Full Dimension MIMO), is an emerging technology, its principle is to equip the BS with a higher number of antennas than the active users' number per time-frequency signalling resource [12, 13, 69]. This technology can strengthen the new generation 5G, improve the energy-efficiency in the order of 100 times and increase the capacity to 10 times. Further, it minimizes latency in the air interface and simplifies transceiver design [8, 13, 63, 68].

- Beamforming: It is a signal processing method that assigns the transmitted signal to the desired direction without receiving an undesirable noise signal. It concentrates the power in the desired direction with a large gain. Using the multiple antennas at the transmitter, beamforming can produce directional antenna beam patterns [2, 70].
- Orthogonal frequency-division multiplexing (OFDM) amelioration: It allows to divide a mainstream into subframes, send them in subcarriers simultaneously and modulate them in different frequencies. This operation improves the spectral efficiency and helps combat multipath and Inter-Symbol-Interference (ISI) [2, 7]. OFDM and orthogonal frequency division multiple access (OFDMA) is used in long term evolution (LTE) Advanced 4G. The integration of OFDM can result in inefficient power consumption, synchronization of signals and exacerbation of interferences. To avoid these problems, the following amendments are proposed [71–73]:
  - Filter bank multi-carrier (FBMC): It uses high quality filters to prevent both ingress and egress noise. In addition, it provides a spectral efficiency higher than OFDM and it robustly works with asynchronous communication between a transmitter and a receiver [74, 75].
  - Generalized frequency division multiplexing (GFDM): It is a filtering OFDM enhancement. Each subcarrier band is filtered to reduce the overlapping among subcarriers and minimize the equalization and synchronization issues [76, 77].
  - Universal filtered multicarrier (UFMC): Signal filtering is performed on groups of adjacent subcarriers to minimize of the sidelobe levels and intercarrier interference [74, 78].
- Machine-to-machine (M2M) communication: It is a technique that connects a large number of smart devices or machines, such as smart metering, sensors and smart grid equipment in wide coverage areas with little or no human intervention [6, 14]. It meets several requirements in 5G such as: The support of a huge number of devices with a low rate and very low latency [69]. M2M undergoes changes in architecture, protocols, security and standards to be compatible with 5G [6].
- Device-to-device (D2D) communication: It allows direct exchange between mobile users without the use of a base station [7, 39]. D2D communication is considered as a type of M2M communication such as D2D users are close and they exchange small quantities of data [79]. It is expected to be a part of LTE-A in 3GPP Release 12 and it will be used in 5G [80]. M2M ensures a reduction in latency and energy consumption, handles local traffic efficiently and improves the spectral efficiency and cellular coverage [2, 7, 15]. Compared to 4G, the communication range between users in 5G will be reduced, which will increase the frequency of interaction between users [23].
- Cognitive radio (CR): It is a software defined radio technologies that tend to improve the spectrum efficiency and to gain access to new shared spectrum bands. Indeed, in the cognitive radio network (CRN), users belonging to a secondary system (unlicensed) can share frequency bands by users of primary system (licensed). This technology allows a higher traffic load and a lower delay in 5G [6, 16, 17]. The CRN is an autonomous network which can learn and adapt the system parameters interacting with its environment [81, 82].
- Cloud network (CN): NIST has defined cloud computing as follows: cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [18]. Cloud computing is considered a solution that solves the calculation problem in smart devices that are unable to handle a large amount of data. The cloud can facilitate the complicated management of ultra-dense networks in 5G. It controls the elements of the network, collects the statistics and configures the necessary parameters in a centralized way. The cloud also facilitates the

network scalability by dynamically adapting cells and users with low OPEX [16, 41, 83].

- Network virtualization (NV): It is a technology that abstracts away the physical hardware functions and offers virtualized resources for high-level applications. It uncouples the traditional internet service providers (ISPs) roles into infrastructure providers (InPs) and service providers (SPs). InP provides the physical infrastructure management and SP creates virtual networks (VNs) using resources collected from multiple InPs and provides E2E services. In 5G HetNet, NV can integrate different RANs and manage also the 5G networks and resources in order to provide users with a flexible QoE. Moreover, virtualization techniques enable network resources to be shared between cells in order to reduce investment and operating costs and facilitate the inclusion of new services. NV is a key process for the management of ultra-dense networks, it provides wide coverage, monitoring and management of mobility and a high data rate to small cell users. In cloud computing, network virtualization can pool computing resources from servers, then, it dynamically assigning virtual resources to applications on-demand [16, 19, 84–86].
- Software defined networking (SDN): It is a promising technology for network virtualization. It is a novel networking approach that allows operators to easily handle all the network elements. It decouples the control plane (the network control traffic) from the underlying data plane (the user specific traffic) and logically centralizes the control of network resources in a controller unit [20, 87, 88]. SDN minimizes the complexity of the 5G network by minimizing network construction costs and reinforcing the intelligence of the network [21]. SDN facilitates network management, simplifies the addition of new services in 5G.
- Network function virtualization (NFV): NFV is complementary to SDN networking technology. It is a technology that separates software implementation of network functions from the infrastructure and underlying hardware. Network resources can be shared and migrated on demand, allowing operators to choose network resources at different locations. Furthermore, the network functions execution can be done on the cloud infrastructure in data centres [8]. Generally, NFV and SDN are two potential technologies that are keys to accomplish 5G networks. Actually, SDN and NFV are developed as two separated technologies. However, in order to realize intelligent and flexible services, SDN and NFV are combined in the same architecture. In this respect, the Open Networking Foundation (ONF) [89] has defined the envisioned scenario of cooperation between SDN and NFV, using a common interface to simplify their integration on the top of the network infrastructure. NFV and SDN were recently proposed to make the future network more flexible, open and service-oriented in terms of control and management, as well as NFV and SDN use data centres to provide load balancing and better resources allocation in cloud [16, 90]. In addition, NFV contributes to reducing the operational and capital expenditures in 5G RAN by the use of virtual machines and SDN. Furthermore, NFV and SDN improve the implementation of D2D communication technology in 5G architecture. To support IoT and M2M communications in 5G network, NFV can effectively manage the large number of connection and provide a dynamically expandable infrastructure to manage spatio-temporal load variations. Furthermore, the SDN controller can use ML algorithms to perform an intelligent global network control [21, 91].

In addition, research has proposed other technologies that reinforce the proposed enhancement of 4G technologies in order to achieve 5G requirements in a more efficient and intelligent way.

- Millimeter-wave signals (Mm-wave): Taking into account the rapid increase in data traffic, a new technology mm-wave is introduced in 5G to improve cell capacity and enhance the capabilities of new technologies adopted in 5G [7]. Its band is between 30 and 300 GHz, as well as the available bandwidths are much wider than the current cellular networks, and they are suitable for 5G communication systems [22, 68].
- Full duplex (FD) wireless communication: In an FD communication, the transmission and reception are operated in the same frequency at the same time which doubles the spectral efficiency. Thus, it is viewed as one of the promising technology in 5G [23]. Thanks to the evolution of antenna and digital baseband technologies and innovative techniques of a complete cancellation of interference, the antennas will be able to decode the signal received at the time of sending the signal in the same frequency band. For example, MIMO technology is an important factor in minimizing self-interference in the space domain, allowing a successful FD relay despite interference [6, 24]. FD communication is a promising technology for 5G wireless systems as it can double the spectral-efficiency at the physical layer [11, 67, 92]. In addition, FD communication provides much higher data rates than conventional half-duplex (HD) communication systems by minimizing congestion and significant E2E delay [7, 11, 93]
- Mobile edge computing (MEC): Cloud computing faces the challenges of high latency and heavy burden on the

backhaul links.[1] Therefore, MEC is a promising solution applied in 5G which deploy cloud servers in base stations. Generally, it pushes the network function (control and storage) and mobile computing to the network edges (gateways, base stations and end user devices). In addition, the computation-intensive tasks of mobile applications will be offloaded to the MEC servers for cloud execution. On the other hand, the duplication transmission of the same popular content requests causes high content requests concentration. Therefore, MEC integrates content caching technique to avoid frequent duplication of the same content. In effect, the same contents will be cached at BSs using MEC server. The computation offloading and caching content can enhance a computation capability close to mobile devices, further providing a highly efficient network operation, energy saving, high throughput, flexible service delivery and ultra-low latency and improving scalability, automation and user experience [25, 26, 94, 95].
- Network slicing (NS): It is mainly based on NFV, SDN and network softwarization technologies to provide a programmability, flexibility, modularity and powerful virtualization capability in order to separate the physical network to multiple (logical) virtual networks. i.e network slices. These slices are E2E logical networks which perform on a common underlying (physical or virtual) network. NS is composed of network functions (NFs) and infrastructure resources. In 5G network, NS can support different RANs to minimize network construction costs [27, 96].

2.2 Challenges related to the adoption of advanced technologies in 5G

The adoption of the above mentioned technologies faces new crucial challenges. We identify here these challenges with their associated requirements.

- UDSC challenges: The UDSC causes a huge collection of management data, resulting in network configuration and maintenance complexity [4]. In addition, the interference between already existing macros cells and small (pico and femto) cells is a big challenge in 5G. The densification made by the HetNets exacerbates the interference between small and macro cells. Signals from macro cell users can go through the small cells and disrupt the signals of these users. Therefore, the ratio signal-to-interference-plus-noise ratio (SINR) will be lower because of interference, which affects network performance and user experience [6, 83]. In addition, insufficient control of the small cell causes insufficient control in its form and location [2]. Furthermore, the UDSC poses a cost problem for operators such as the cost of installing new resources, the additional costs of electricity and backhaul [83].
- RAT selection challenges: In 5G HetNet, most devices today use multiple technologies in a multi-RAT environment (such as 3G, 4G/LTE, Wi-Fi, Bluetooth, and potential 5G technologies). By addressing the needs of users, an optimal and intelligent selection of RATs must be carried out in order to optimize the QoS [4, 8, 97, 98]. RAT selection can be influenced by the selection of the optimal bandwidth among several available bandwidths, the spectrum to be used [8]. Moreover, the huge number of devices and multiple RATs, operating at different frequencies and using different protocols, can influence the effective RAT selection and the rapid transfer between RATs. In fact, it generates high interference and a huge number of unnecessary signals. In addition, the optimized partitioning of capacity between RATs can be degraded because of the high number of RATs. The optimal selection of the user is a crucial challenge for the 5G, it depends on the capacity of available resources and BS and also the choice of RAT for other users [7].
- Massive-MIMO challenges: MIMO creates the following challenges [13, 63] :
  - Pilot contamination: The orthogonal uplink transmission of the pilot sequence helps users to estimate the channel in the same cell. The same sequence can be reused by another user in another cell in order to consume resources. This reuse will cause pilot contamination which is interference caused by the interfering orthogonal uplink pilot sequences of different cells [8, 13]. This problem degrades the performance of the channel detection and estimation in a cell because of the interference causes in another cell which contaminates Channel State Information (CSI). Moreover, pilot contamination appears with the increase of the antenna number of BSs, which causes a problem in the design of massive-MIMO [2, 48, 63].
  - Architectural design: Massive-MIMO has a wide number of antennas that are powered by multiple low power amplifiers and each antenna should be integrated into its own amplifier. The best mapping between the antennas, the cost of resources installation and the mutual couplings are challenging tasks [8, 63].
  - Full dimension MIMO: In the current cellular network, MIMO and beamforming techniques use 2D

[1] A backhaul network is an intermediary that enables the data transmission and reception between core networks, or macro base station and small base stations. It can be a wired or wireless link.

directional antennas with a horizontal plane to control the beam pattern radiation. The BSs with a linear horizontal plane can only accommodate a restricted number of antennas and it uses only azimuth angle dimension. To overcome this limitation, it is possible to add a vertical plane with elevation angle dimension exploitation. This amelioration is called Full-dimension MIMO (FD-MIMO) or 3D MIMO, it is based on 3D directional antennas which transmit the signals with certain pattern radiation in both the azimuth and elevation planes. Massive-MIMO is based on 3D directional antennas, it can provide a high capacity, energy capacity and spectral efficiency and less inter-cell interference to 5G network [2, 70, 99]. FD-MIMO presents problems of implementation and complexity in channel estimation for a large number of channels and an interference problem in both azimuth and elevation beamforming [8, 63].

  – Channel designing: Modelisation of the antenna correlations impact and coupling for massive networks [6, 8, 63].

- OFDM challenges: The configuration and implementation of new types of multiple access such as FBMC and GFDM [6].
- D2D challenges: Efficient proximity detection, network integration, native support, secure and respectful data transfer, efficient network coding scheme to improve throughput, self-interference minimization due to the use of FD transmission, multi-mode selection (simultaneous use of both types of communication: D2D and communication with a BS) [2, 16, 100, 101]. In the future, manufacturers tend to implement hybrid architectures to solve the problems of M2M and D2D networks [4].
- CRN challenges: Interference can be caused between primary and secondary users [16]. In CRN security, the risk of attacks is possible due to frequency sharing [23].
- SDN and NFV challenges: NFV is used by 5G network service providers to decouple network functions from their specific hardware components and build a more flexible 5G network. However, it faces challenges which have a profound impact on performance and prevent its implementation.

  – In NFV security, the network functions implementation in a virtualized environment can increase the number of attacks which substantially degrade network security. Thus, it is necessary to exploit a hypervisor which avoids any unauthorized access or data leakage. In addition, it is necessary to provide secure execution of processes such as data communication and virtual machines (VM) migration [88]. In addition, virtual network functions (VNFs) are network functions implemented by vendors in the software component, they exist in NFV architecture. The VNF can be an intrusion detection system (IDS) or firewall, routers or switch. The implementation of VNFs in 5G RAN facilitates network management and minimizes CAPEX and OPEX, reduces congestion and power consumption through dynamic allocation of infrastructure resources and traffic balancing [4, 21, 102]. Security becomes a fundamental feature of the design and operation in the 5G network. In fact, 5G must manage and provide solutions to have high-level security [28]. Moreover, the IDSs implemented in current cellular become obsolete and it will be unable to effectively detect new types of intrusion. Consequently, IDS become a new challenge in 5G mobile technology and to solve it, several approaches based on ML algorithms have been proposed as an powerful solution.

  – In computing performance, VNF is a software instance in NFV based on some number of VMs that execute different processes for a network function [87]. VNF poses three principles of computing challenges:

    ⋆ NFV needs to instantiate VNFs in the right places at the appropriate times, allocate, control and interconnect their hardware resources to chain services.

    ⋆ The implementation of a set of techniques in VNF software can ensure high performance, such as the multithreading technique that implements a VNF file with high demand, the software instances that have independent memory structures to avoid the operating-system deadlocks, processor affinity techniques must be used to take advantage of cache memories, a technique of direct access to input/output interfaces that reduces latency and increases data throughput. These techniques require an automated resource allocation approach in a server pool [88].

    ⋆ NFV provides scheduled and easy maintenance. In fact, network operators will be able to determine, correlate and recover errors, monitor the use of computing, storage and network resources during the life cycle of a VNF. Thus, VNF will be able to dynamically create and migrate. In this case, keeping track of where a given VNF is running will be more complex and VNF can behave erratically even which influences the detection of the problem even if the underlying infrastructure works well [87].

  - NFV presents other challenges in inter-slice isolation, customizable intra-slice resource allocation, control signalling and boot stamping problems, resource discovery, pricing-based allocation, mobility management, fairness and revenue/price optimization [11].

  In the 5G network, SDN faces the challenges of control and confidentiality, it must provide indiscriminate services and seamless handoff between service providers. It must support the heterogeneity of devices and technologies and control messages with performance and survivability, it should ensure a balance between performance, security and flexibility, and maintain standardization of the control interfaces and information of the controlling network-big data development [103, 104]. In addition, the attacks on the control plane, congestions, bottlenecks and lack of scalability and resilience of the control plane are main risks that affect the security of the control plane and centralization [28]. Furthermore, SDN meets several other challenges as the definition of the SDN controller functions, the interfaces and protocols standardization of the controller and forwarding hardware [105].
- Mm-wave challenges: There are four known challenges in mm-wave band 5G system:

  - High path loss: The increase of the transmission frequencies mm-wave causes the increase of omnidirectional path loss [106]. It is higher than the microwave bands below 3 GHz [2].
  - Antenna arrays: To achieve mm-wave communication, the antenna arrays technique is proposed as a promising research direction in 5G mm-wave network. Due to unfavourable propagation characteristics, large antenna arrays with beamforming are used in mm-wave communications to overcome propagation losses. In addition, narrow beams are integrated to reduce the spectrum overlap and improve QoS. However, the pole sway when using narrow beams increases sensitivity to movement. When beams are blocked, antenna arrays must be adapted rapidly which will cause a high processing complexity [6, 23].
  - High atmospheric attenuation: Oxygen and water vapour absorb mm-wave energy with the attenuation of the received signal. Oxygen absorbs large electromagnetic energy at about 60 GHz with attenuation of about 15 dB/km and water vapour absorbs greater electromagnetic energy at 164–200 GHz with even higher attenuation. Consequently, the attenuation at the high frequency band limits signal propagation [2, 8, 107].
  - high penetration loss: the penetration into solid materials causes attenuation of mm-wave signals with very high losses. The attenuation values are highly dependent on the material [2, 107].

  Diffraction, propagation, Doppler, scattering, refraction and reflection are other mm-wave challenges [6].
- FD communication challenges: The key challenge in Full-duplex communication is self-interference (SI) mitigation. SI minimizes the gain of the FD transmission, some amount of SI rests in the system even with the use of various mechanisms of SI cancellation. Moreover, the interference between the BSs is high because of the high power signals influenced by the path loss and shadowing, in addition to the interference between the spatially close users [11, 16]. Cross-layer resource, interference and power allocation management, synchronization and time adjustment during the FD transmission establishment, dynamic mode selection are other challenges posed by FD communication [11].
- E2E connectivity challenges: In future, the network will be more dense and heterogeneous and the user decision of cell connection will be more complex. The challenge arises when analyzing mobile connections in the RAN taking into account the network complexity and the diversity of applications to process [4].

3 5G background

This section describes the 5G wireless networks architectures with emerging technologies and applications.

3.1 5G services

Mainly, 5G is expected to support three general types of services: enhanced Mobile Broadband (eMBB), massive Machine-Type Communications (mMTC) and ultra-reliable and low-latency communications (URLLC) based on different QoS requirements of 5G. It aims to improve 5G applications demands using 5G technologies [108–111].

- enhanced Mobile Broadband (eMBB): It aims to provide 20 Gbit/s in the downlink and 10 Gbit/s in the uplink data rates and supply low latency communications. Mm-wave and Massive-MIMO are two key technologies used to support this higher data rate and ultra dense network provides the data rates [1, 112].
- massive Machine-Type Communications (mMTC): It provides scalable connectivity solutions for the massive number of devices using MEC as a technology solution.
- Ultra-reliable and low-latency communications (URLLC): It provides ultra-reliability communication

with packet error rates equal to $\leq 10^{-5}$ and low E2E latencies in order to support 5G applications such as healthcare and smart city. It is based on NFV, NS and MEC technologies.

3.2 5G applications

The response to 5G requirements, for instance low latency, fast data transmission, high throughput, is expected to serve a large range of applications. Therefore, we cite some applications which are improved by 5G.

- Internet of things (IoT) IoT allows interconnection between the internet and objects offering the possibility to share information across multiple platforms. In fact, it establishes an interconnection between intelligent sensing and devices to share information and provide a common picture of operation for allowing innovative applications. This interconnection is realized with seamless large scale sensing, data analytics and representation of information using cutting edge ubiquitous sensing and cloud computing [113]. Varieties of devices, connected homes, smart grids and smart transportation systems are connected simultaneously under the IoT. To achieve the interconnection of this large number of connected objects, IoT can use high bandwidth 5G wireless networks, as well, the capacity improvement and high speed of 5G to provide reliable and fast connectivity in the future IoT. In addition, 5G can provide interaction between IoT devices and the smart environment through smart sensors connected. To reinforce M2M technology in 5G, IoT can transform the current internet of human-centred interactions to an M2M platform equipped with 5G [6, 39, 114].
  - IoT challenges: Sensor automation, acquisition, modelling and reasoning, discovery and context sharing, the higher density of devices, data rate compatibility, interoperability and heterogeneity of devices, the complexity of the IoT implementation, the security and flexibility of IoT management 5G network, the effectiveness of 5G data networking [6, 39, 49, 115–117].
- M2M communication: The generation, processing, transfer and exchange of automated data between intelligent machines are M2M communication characteristics with human control. M2M in 5G system aims to provide E2E latency less than 5 ms and more reliability in the system e.g., packet loss rate 10^{-9}. Moreover, the 5G network can offer advantages for M2M communication, it can introduce a new mm-wave spectrum that facilitates the proliferation of the device, improve cell capacity and enhances M2M communication. Besides, it can offer a new waveform to minimize sporadic traffic and solves the problem of critical time [6, 118].
  - M2M challenges: Massive access, security and privacy [45]
- Internet of vehicles (IoV): IoV allows communication between vehicles or communication between vehicles and sensors, roads and humans to collect information on their surrounds [64]. The evolution of IoT automatically leads to the evolution of IoV [119]. Thus, it presents robust traffic management and reduced collision probabilities [5]. Under the IoT, Internet of Vehicles (IoV) has been evolved and especially with 5G. Intelligent communication between vehicles can be ensured thanks to the high bandwidth, generalized availability and low latency provided by 5G. Besides, 5G can also offer a small cell-specific environment, cloud capabilities, content-oriented networking [6].
  - IoV challenges: Ensuring a good method of communication between the drivers and the IoV system [64, 120]
- Healthcare: with the evolution of 5G and the improvement in detection and security technologies, will be lightly monitored by opening up healthcare solutions. The concept is to send the patient's body data to the cloud or health centres. Thereafter, a prediction of the disease will be quickly provided to patients by relevant and urgent medical services. Moreover, the multiple physiological signals can be registered over a long time period to in order to understand the disease pathophysiology. Thus, high bandwidth and real-time sending in 5G ensure reliable data change [6, 121]. 5G technologies can reduce overall medical costs such as Point-Of-Care Testing (POCT) that saves money. Also, patients can take advantage of health technologies, digital platforms or remote monitoring devices rather than going to large medical institutions [122]. Furthermore, 5G aims to provide effective, comfortable personalization, sustainable, smart service, low cost to patients.
  - Healthcare challenges: Lack of a data-driven culture, healthcare disparities, problems with healthcare information systems, resource constraints, ageing populations, the long-term chronic care burden, lack of universal access, challenges with Electronic Health Record (EHR) [53].
- Smart grid (SG): In order to improve efficiency, reliability and security, SG is proposed as an intelligent power grid infrastructure. In fact, it helps to gently integrate renewable and alternative energy sources. Based on the optimizing of demand and energy consumption, it uses intelligent communication infrastructure and high power converters. Besides, it counts on intelligent sensing and metering technologies in order to improve efficiency and reliability [123]. Decentralized energy

distribution and improved energy consumption analysis are Smart Grid's goals. 5G delivers these goals to improve efficiency and deliver economic benefits. It collects energy data, monitors power lines, protects and manages requests and responses from remote sensors, and then adjusts power distribution. The large bandwidth and low latency of 5G solve the problems related to SG response and demand [6].

- Smart homes and smart societies: Smart homes offer human–machine interaction and use the information analysis in life problems [124]. Smart society is the concept of interconnected all digital and electrical services/devices in home, office or magazine in an intelligent way to improve the QoE. It extends the scope of smart homes into the social environment to social problems solving [124]. The 5G offers to these two applications a highly directional beamforming antenna, with a long battery life with low probability of failure. It provides high throughput in large coverage areas with low installation costs. When using a network simultaneously, higher capacity is accumulated [6, 125].
- Internet tactile: It is a communication infrastructure that aims to ensure high availability, efficient reliability and high level of security. Internet tactile can use protocols which provide a high QoE in a virtual environment with the minimal response time. It addresses in real-time gaming, industrial automation, transportation systems, health and education. These areas are very critical, it must have sufficient capacity to enable the communication between a massive number of devices in a simultaneous and autonomous way, and ensure very low latency.
  - Internet tactile challenges: The finite speed of light affects the facilitation of the required real-time experience [126], transmission rate, latency requirements for visual, auditory, and haptic modalities, instability of haptic communication control due to communication induced artefacts which causes the time-varying delays and packet losses issues, ensures stringent reliability for the transport of haptic sessions, E2E latency optimization, available resources sharing between haptic applications and other human-to-human (H2H) or machine type communications (MTC) applications, safety and privacy improvement [127, 128].

  Generally, 5G supports the Internet Tactile to meet the requirements and to cope with the challenges discussed above, it uses the underlying 5G-driven communication architecture, composed of the RAN and Core network [126]. In fact, the implementation of 5G RAN in internet tactile ensures :

  - An efficient support of multiple RATs.
  - Through multiple radio protocols and physical layer, efficient packet delivery is provided.
  - The use of novel medium access control (MAC) techniques to provide an optimal resolution of air-interface conflicts.
  - Application-aware QoS provisioning is dynamic.
  - Access to edge-cloud and security

3.3 5G architectures

To achieve the envisioned goals of 5G, significant architectural improvements should be made and defined [23].

3.3.1 5G physical architectures

The deployment of the advanced technologies: dense cells, D2D, CR, cloud, SDN, mm-wave, multi-RAT, massive-MIMO, lead to more complicated architectures with multiple tiers. Each tier has different size, transmit power, backhaul connections, and different RATs [10, 11]. Figure 2 presents the 5G general physical architecture.

- Multi-tier architecture: Tiers are presented as macro cell underlaid with small cells such as femtocell, picocell and microcell, relays and D2D communication, which in its turn lead to a heterogeneous network (HetNet) [6, 16, 83, 84, 129] In HetNet, cellular service is delivered by the use of a mixture of macro base stations (MBSs) underlaid with small base stations (SBSs). The MBSs and the SBSs are dynamically operated over different spatial and time scales. SBSs are deployed for indoor coverage (home, office, and subway) serving different small cells. This architecture can include a mobile small cells in the form of train or bus, communication in a building using an SBS, use to mobile small cell located inside a vehicle and large antennas located outside to communicate with MBS [130]. This architecture is based on hierarchical cell structures:
  - Small cell: Macro cells use macro cell base stations (MBS) and small cells manage the data transmission of their users by using small cell base stations (SBS). Indeed, SBS operates under the direction of MBS and its cost of installation is cheaper than MBS. In addition, Small cells offload traffic from MBS to SBSs and reducing both OPEX and CAPEX. Small cells use the high-frequency mm-waves to provide high throughput. But, the frequency used by macro cell remains stronger than the frequency of small cells, it can be reused by small cells to avoid interference and increase the spectral efficiency. In fact, MBS integrates

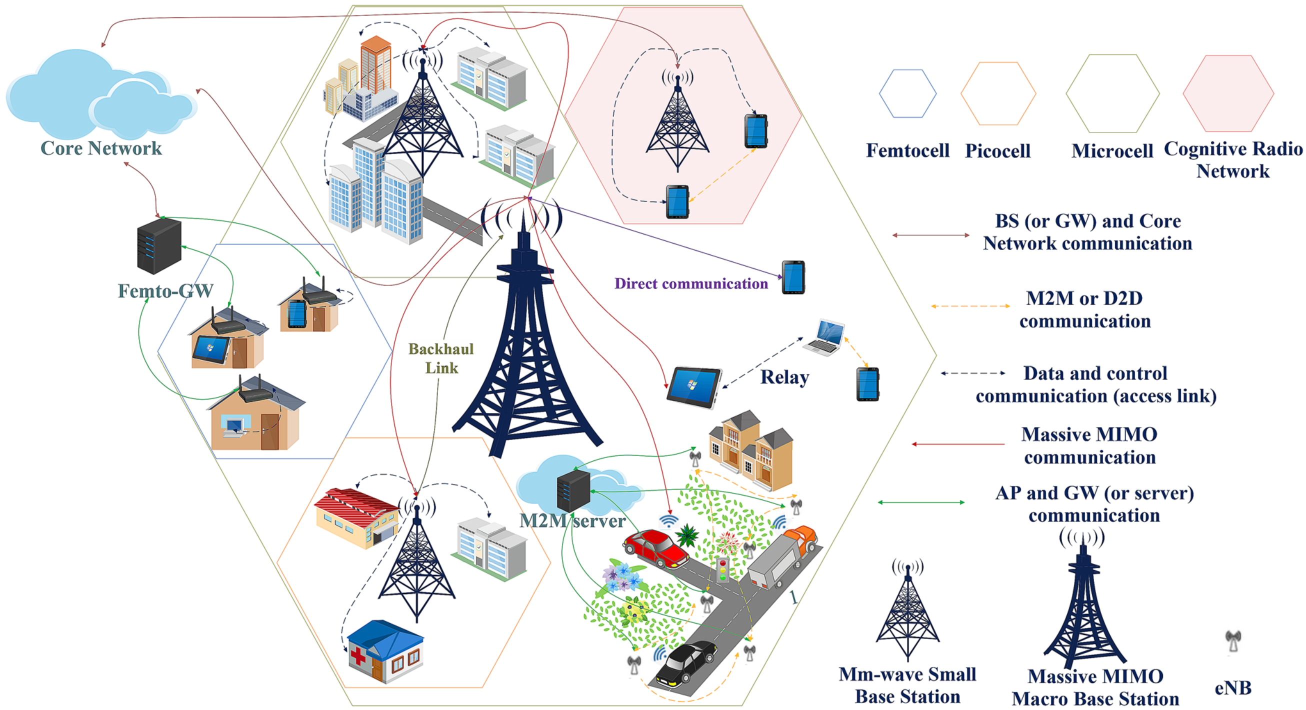

Fig. 2 General 5G physical architecture [16, 40, 130, 137, 157]

massive-MIMO technology equipped by a multiple numbers of antennas [1, 6, 16, 83, 84, 131]. 5G is an ultra-dense network with small coverage area i.e., SBS are very close. Transfer management ensures the minimization of frequent transfers, it consists of division of tasks between the MBS and SBS, the MBS is responsible for the handover control while the SBS are responsible for the transmission of the data of the users [132, 133]. Generally, small cells are divided into femtocells, picocell and microcells. Indeed, the femtocell is the smallest, it manages a few users. Microcell is the largest, it covers 1–2 km and it can be used in the large events. Picocell is in the middle, it can cover 250 m and manage 100 users [134].

– CRN: It is used to enhance the multi-tier architecture. It uses MBS and SBS and it works on several heterogeneous channels. CRN is composed of two types of architectures [135]:

  – Non-cooperative CRN: It uses a multi-RAT system with unoccupied channels. There are two types of channels in non-cooperative architecture: cognitive channels used for users near or far from MBS and cognitive channels used for QoS.
  – Cooperative CRN: SBS can be used to analyze the activities of a macro cell. Besides, SBS elaborates on temporarily unoccupied frequency bands or spectrum holes and it uses a dynamic pricing model. This type of architecture uses a licensed channel that serves users.

  The application of CRN at the small cell avoids the selection of identical channels for different cells, which minimizes interference between cells. In addition, the spectrum holes help to improve network capacity and bandwidth utilization. In contrast, energy efficiency in the CRN network is presented as a big challenge, In fact, all the energy is consumed in the circuits and cooling systems and beams in the air.

– D2D communication: Without using an MBS, users can communicate using D2D communication by sharing network frequency resources. The social-conscious D2D architecture is a D2D communication architecture for social networks. It consists of components: bridge, community, centrality, links. D2D communication allows instant communication, peer to peer file sharing, and local video streaming [23]. D2D communication in multi-tier architecture presents several challenges such as the interference caused by using of the same DL channel, the optimal allocation of resources especially channels and bandwidth, real-time processing and minimizing latency [100, 101].
– M2M communication: ETSI proposed an M2M communication included in 5G architecture [136], it is composed of intelligent sensors in M2M devices

(mobile, vehicle, train, smart metering), M2M communication domain that includes M2M GateWay (M2M GW) such as roadside units or cellular base stations e.g., eNBs, and M2M server/application domain. In fact, each device is connected to local area networks (LANs) in order to transmit data to M2M GW. Then, M2M GW will establish a communication link between M2M devices and the application servers using wired and wireless communication networks in order to forward data collected from devices to M2M server [137, 138].

– Mm-wave network: During a communication inside a cell, mm-wave works successfully but during communication with the outside, various external factors limit the mm-wave bands' operation in the applications of mobile communication systems. As MIMO, beamforming scheme can be applied at the mm-wave frequency bands to overcome the severe path loss and to provide user mobility in the access links in outdoor scenarios [139, 140]. Mm-wave wireless backhaul links are used in the multi-tier architecture such as the transmission links can be composed of two parts: the access links which use a highly directional and flexible beam to ensure high speed and flexible mobility of the users and a second part: backhaul links that support high data rate transmissions between a BS and an AP where BS and AP locations are generally fixed [140]. In the Hybrid mm-wave Network, the system operates on two phases: the 5G mm-wave and legacy 4G network, it provides low latency and cost-effective solutions thanks to a dual-mode modem that allows user switching between the two networks. The data will be transmitted by the mm-wave spectrum and control information will be transmitted using the traditional 4G network. In Standalone mm-wave Network, the system operates except with 5G mm-wave and it is characterized by the deployment of the same mm-wave spectrum for both backhaul and wireless access links. Using of mm-wave BS grids provide good QoS with low latency at the cell edges and horizontal and vertical beams under the use of large beamforming gains. In addition, the use of narrow beams reduces spectrum overlap and improves QoS between BS grids and the large number of users[6].
– Massive-MIMO: Massive-MIMO implementation in BS improves the performance network in multi-tier architecture [6]. A high penetration loss through the walls is performed during communications between inside users and outside base station users. Thus, this will result in a reduction in spectral efficiency, data throughput, and energy efficiency in wireless communications. By using the multiplexing technique defined by MIMO technology, the high penetration loss will decrease [41].

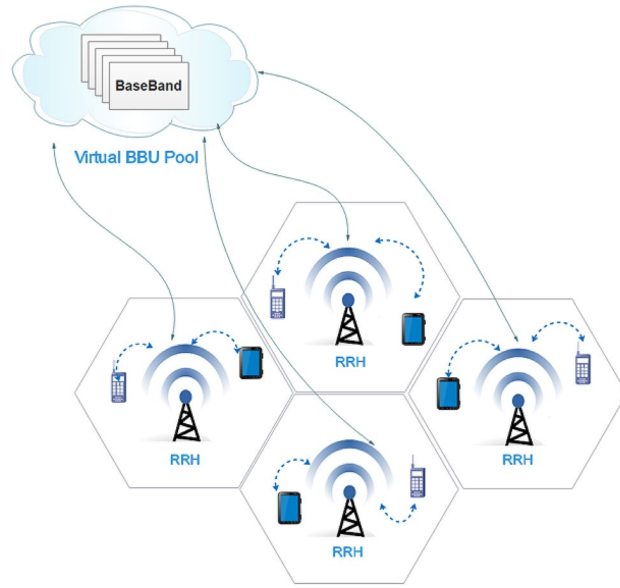

Fig. 3 C-RAN architecture

- Centralized-radio access network or cloud-radio access network architecture (C-RAN): Virtualization and centralization are the C-RAN architecture foundation which consists of several MBS. Most functions of MBS are run in the cloud. Each MBS consists of BaseBand Unit (BBU) and remote radio heads (RRH). BBU is placed in the cloud and RRH stays in MBS [141]. A virtualized baseband unit (BBU) pool enables the centralization and sharing of the Baseband processing among sites. Using VN mechanism, the BaseBand Units (BBUs) operations can be centralized and their functionalities can be virtualized in central servers. Figure 3 depicts C-RAN architecture, it can pool BBUs into centralized BBU Pool from multiple base stations in order to achieve a statistical multiplexing gain by unloading the configurable computing resources of the base stations to the central servers [142]. In this area, C-RAN can ensure several 5G requirements, while bringing an important saving for mobile operators in terms of CAPEX and OPEX. Indeed, C-RAN helps to improve network capacity, power-efficiency and load balancing, it reduces latency, it ensures integration of different services and it provides simple network management. In addition, it provides dynamic service provisioning without costly networking devices. Moreover, centralization and virtualization can reduce the number of processing units required in a small cell dense environment [6, 16, 40, 143]. The use of MIMO technology in C-RAN minimize energy consumption and ensure a global optimal utilization of system resources [11, 105]. In general,

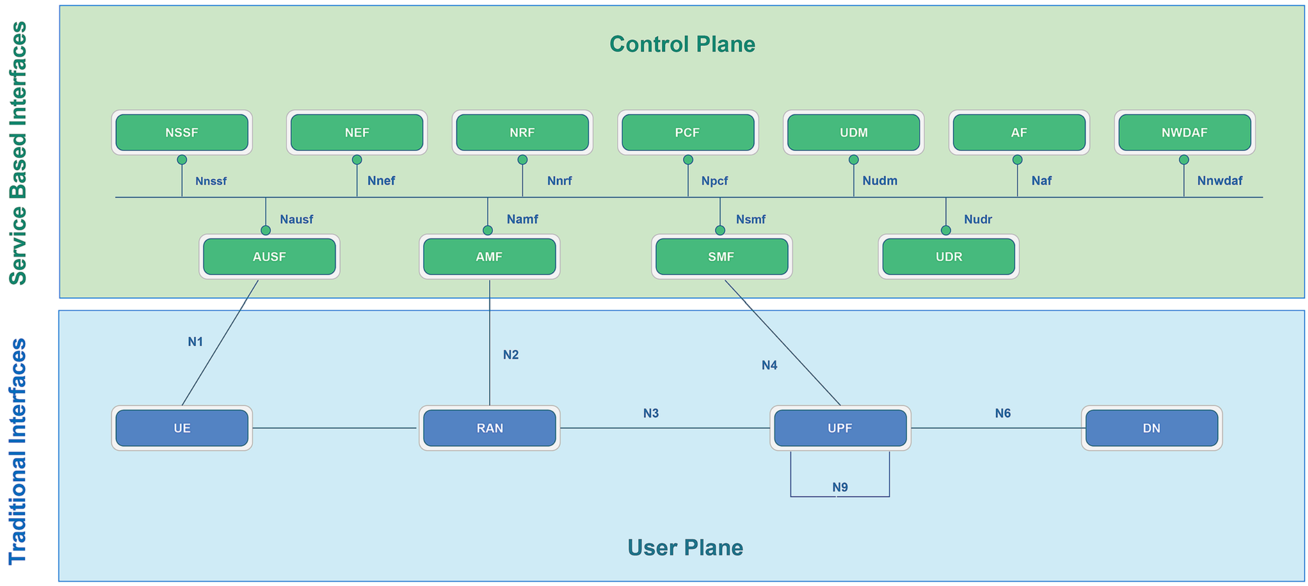

Fig. 4 5G service based architecture

C-RAN architecture is composed of two layers: control and data. The data layer consists of heterogeneous physical resources, it performs channel decoding and the Fast Fourier Transform (FFT). On the other hand, the control layer manages resources such as real-time communication, security, network management and baseband processing. This architecture has several drawbacks in the 5G network: the high processing time caused by the distance between the data layer and the control layer and the raw exchange of the baseband samples between the two layers. In order to solve these inconveniences, an additional layer is added, Software defined services layer. This architecture provides traffic management, cell configuration, interference control, assignment of functional components to video streaming services and physical elements. We mention some challenges of C-RAN [144, 145]: high computational complexity, the necessity to an efficient and fast transfer between the RRH and BBU, the need for an element that provides a transfer and data processing in real-time, an efficient power resource allocation is necessary to minimize power consumption and meeting demands of wireless users over a long operational period, a safe execution by the cloud provider, unsecured C-RAN poses a management challenge in RAN networks.

3.3.2 Service based 5G architecture

By applying the SDN, the problem of configuration and maintenance of multiple servers and routers in a dense 5G environment is solved. This solution consists of separating the control plane from the data plane (or user plane) by using the software components. This introduces speed and flexibility in the 5G network. The increase in the capacity of the resources of the data plane does not pose any overload in terms of control. RANs use SDN to provide intelligence, self-configuration and control plane optimization [6]. Generally, 5G architecture is based on softwarization. Furthermore, it is characterized by network functions (NF) elements that facilitate the operation of the 5G network to meet user requirements and quality of service. According to 3GPP release 15 [146], the defined NFs are as follow: authentication server function (AUSF), Access and mobility management function (AMF), data network (DN), e.g., operator services, internet access or 3rd party services, network exposure function (NEF), network repository function (NRF), network slice selection function (NSSF), policy control function (PCF), session management function (SMF), unified data management (UDM), unified data repository (UDR), user plane function (UPF), application function (AF), user equipment (UE), RAN, network data analytics function (NWDAF), charging function (CHF). In fact, service based architecture, shown in Fig. 4, is separated by two plans: control and user plans. Control plane includes service based interfaces which are composed of: AMF, SMF, UDR, NEF, UDM, AUSF, AF, PCF, NSSF, NWDAF.

- Mobility management function (AMF): It terminates the RAN CP interface (N2) and NAS ciphering and integrity protection (N1). In addition, it ensures registration, connection, reachability and mobility management, it provides a UE and SMF communication, an access authentication, access authorization and notifies the mobility of the user.

- Session management function (SMF): It manages the session such as establishment, modify and release including UPF tunnel maintenance and internet protocol (IP) address allocation of UE. Moreover, it configures traffic steering at UPF to achieve the proper destination. Besides, it selects and controls the function of UP.
- Unified data repository (UDR): It stores subscription data by the UDM, policy data by the PCF and structured data for exposure.
- Network exposure function (NEF): It stores information received by NFs as structured data using in the UDR, then, this information can be reposed or exposed to others NF and AFs which can be used for analysis. Moreover, it helps to manage network behaviour by developing policy rules for the control plane. In addition, it ensures secure NF capabilities and events and provides secure AF information to the network.
- Unified data management (UDM): It generates the authentication credentials to allow access based on subscription and it stores the information in UDR. In addition, it identifies users and manages SMSs.
- Authentication server function (AUSF): It supports the authentication of the UE functionality.
- Application function (AF): It provides services to support: the application influence on traffic routing and access to NEF. In fact, the operator permits some AF to interact with other network functions. AFs, that are not approved by the operator, must contact NEF to interact with the NFs corresponding to them.
- Policy control function (PCF): It provides a unified policy framework to network behaviour management. It supports control plane function by providing policy rules.
- Network slice selection function (NSSF): It selects the network slice instances based on provided UE information to serve UE and it chooses AMF set which serves the UE.
- Network data analytics function (NWDAF): It provides to NSSF and PCF the information related to slice-specific congestion. In the other hand, user plane includes UE, RAN, DN and UPF which provides intra-inter-RAT mobility. In fact, UPF provides flexible connectivity to the user in external networks and enforces the operator policy in user plane during packet routing and forwarding [146–148].

3.4 Quality of expectations

The 5G network must provide high QoS, QoE, reliability and strong security to meet the massive demands of users.

1. Improve the quality of service (QoS): It is necessary to take up existing challenges, such as high quality multimedia traffic in different wireless channels, limited wireless network resources, technical challenges presented by the high quality of applications despite the benefits presented by mm-wave to ensure QoS, technology limitations and existing multimedia transmission protocols, blocking calls and handoff failures for multi-class traffic. Based on existing research, solutions are proposed to meet requirements such as [6]:
   - Solutions based on the multimedia scheduling system which is compatible with QoS and used to achieve a compromise between complexity and performance and to perform accurate propagation analysis and adapts countermeasure techniques in order to satisfy QoS [149]. These solutions can be based on ML which provides automation to the scheduling system such as [150, 151].
   - Approaches allow users to evaluate the proposed QoS. The evaluation is done by calculating the bandwidth, the error rate, the signal strength, etc [152].
2. Improve user experience (QoE): The official definition of the QoE is given by the recommendation P.10/G.110 of the International Telecommunication Union. It is " The degree of the delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the users' personality and current state" [153]. Another definition is presented by the European Network on QoE in Multimedia Systems and Services, Qualinet (COST Action IC 1003): degree of the delight of the user of a service. In the context of communication services, it is influenced by content, network, device, application, user expectations and goals, and context of use [154]. In addition, ITU-T Recommendation G.1011 has published a reference guide on methodologies for evaluating the quality of experience [155]. QoE has features such as interactivity, product awareness, ability to serve goals, and contextual adaptation. The satisfaction of the user depends on the QoS enhancements [156]. [6] presents the relationship between QoS and QoE. Research has considered neural networks as a correlation method between QoE values and QoS parameters [158]. However, the rate of packet loss, delay and response time are considered inefficient and traditional QoS metrics because of the exponential growth of Internet video traffic when using 5G. In fact, some QoS parameters are used to indirectly estimate QoE. In addition, the techniques used for QoS become insufficient. The perceptual video quality of the users is subjective and diversified, which implies complexity in the relation between QoE

and QoS [156]. Due to interference problems in mobile small cell, the user will repeat the reload of the pages, which will influence later on the QoE [158]. To improve QoE in internet video traffic, researchers have proposed HTTP adaptive streaming (HAS) which can minimize the interruptions of the video playback and massive bandwidth utilisation. Despite that, HAS client can still suffer from video playback blocking which will reduce the QoE. [159] proposes a network-based framework based on ML techniques to enhance QoE and minimize this blockage. To improve QoS and QoE in a 5G environment, network management needs to be advanced, automated and uses a new cloud, big data and SDN technologies based on automation and intelligence.

3. Improve quality management by Self Organizing Network (SON) paradigm: The densification and heterogeneity in 5G architecture increase the complexity of network maintenance and configuration. Thus, the 5G network needs the automation and the intelligence to solve this issue. SON is the solution, it is among a potential and effective paradigm that will lead to 5G performance enhancement. In fact, it provides the intelligence and autonomous adaptivity inside the network to simplify the work of operators, to facilitate the coordination, optimization and configuration procedures and to minimize the overall complexity CAPEX and OPEX in 5G network. Besides, it improves the user experience, the quality of service, the capacity of small cells and it optimizes the automatic interference control. Furthermore, ML is applied in SON functionalities which provides the intelligence to the network in order to learn relevant parameters and provide a network with high performance [6, 82, 160].

3.5 Sustainability

The high number of base stations and cells causes an increase in energy exploitation. Thus, it will be necessary to implement new sustainable technologies to save energy. In order to minimize the exploitation of energy by the BS, a solution proposed by [161] presents BS that saves energy with techniques in standby mode. The 5G network has self-organizing backhaul links that minimize interference and energy consumption. Ml techniques can be used to intelligently match users to corresponding small cells using backhaul connection which serves different 5G requirements [162–164]. The complex operations performed by the network must be scalable. To achieve this objective, the energy-efficiency of these operations must be sustainable. In addition, reducing the energy consumption rate minimizes the energy spent on packet transmission. Under the smart antennas algorithms, the diffusion algorithms help to minimize costs, redundancy and energy consumption.

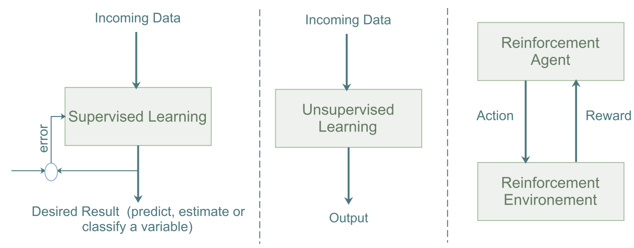

Fig. 5 Supervised, unsupervised and reinforcement learning [165, 171]

The C-RAN architecture can minimize overhead costs and energy drains, also the combination of the two HetNet and cloud architectures gives a new Heterogeneous CRAN (H-CRAN) architecture that presents an efficient allocation of energy resources [6].

4 Machine learning background

Machine learning is a collection of methods that allows computers to be able to learn, automate and optimize using large data sets that can not be used systematically by humans. Without AI and ML, network operators will not be able to efficiently delivering 5G services with its various requirements. Recently, promising approaches of ML have been proposed to enable SON the key feature of 5G [36]. Indeed, the promise is that ML approaches have the potential to automatically learn the system experience, predict future scenarios and to adapt to fluctuating environments [4, 48]. As in 5G systems, there are multiple parameters and complex multi-variable scenarios, the pattern of learning will take different categories for each kind of problem and within each stage of the learning model. ML approaches are classified into three main categories Supervised Learning (SL), Unsupervised Learning (UL) and Reinforcement Learning (RL) and other secondary categories, such as Markov models, Heuristics, Controllers, which are applied in accordance with the 5G applications. Each ML category can be divided into several sub-classes dealing with very specific algorithms that will be applied in cellular networks [4]. Figure 5 illustrates the main categories of ML: SL, UL and RL.

4.1 Supervised learning

Supervised learning (SL) requires a supervisor who tells the system the desired result based on incoming data. The use of supervised algorithms allows to predict, to estimate or to classify a variable. The idea is to train a learning model that try to generate a general rule mapping inputs to outputs with samples of the problem for which the solution is known. Then, the model is used to find optimal solutions from new samples [165]. SL techniques are based on the assumption that we have the mapping from a dataset of instances to their corresponding labels. Based on the continuity/discreteness of the output data, SL can be divided into two learning tasks: regression and classification. For example, Bayes' theory, K-nearest neighbors (K-NN), Naive Bayesian (NB) and support vector machines (SVMs) use classification learning agent. Artificial neural network (ANN), Generalized linear models (GLM) use regression learning agent [36].

- Regression: It estimates the relationship between variables and predicts a value of one or more continuous-valued targets. Regression is a very simple ML that predicts accurate results with minimum errors. There are two types of regression: linear when the regression function is linear and polynomial regression which is more reliable when the curve is built on a large number of observations that are distributed in non-linear curve or a series of humps [48]. Regression analysis is used to predict a *continuous* quantity output for a given example (e.g., predict radio parameter, future failures, define scheduling policies to model energy, etc.). Regression algorithms that are used in wireless networks mainly include: Support Vector regression and the Gaussian process for regression [62].
- Classification: Is the agent of predicting a *discrete* class label output for an input in order to group data together based on predetermined criteria. The classification problem in a network can be presented as a problem of predicting a type of attack. Algorithms in this category include SVM, K-NN, NB, Decision Tree (DT) and others.

4.2 Unsupervised learning

Unlike SL, unsupervised learning (UL) does not require a supervisor, therefore the expected result is unknown in advance (i.e., there is no output vector). The ultimate goal of UL is to find efficient inferences from labelled data samples to describe a hidden feature or structure of data. It is widely used to detect patterns and relationships in the dataset. UL is useful for identifying an anomaly, recognizing patterns or minimizing the dimensionality of the data. Clustering, dimensionality reduction, anomaly detection, latent variable models are among the well-known models of UL [4, 36].

- Clustering: It allows to gather elements that have similar features. K-means clustering is the most common technique in anticipation of networks [60]. This technique makes it possible to partition *n* observations in *m* clusters as each observation belongs to the nearest cluster [48]. Algorithms in clustering category include K-means, hierarchical and fuzzy-c-means algorithms.
- Dimensionality reduction: Methods of this category are capable of reducing both the computational complexity and the storage requirements. Minimizing the size of the original data presents a powerful regression analysis. Moreover, a data set with large dimension undergoes analysis challenges. Dimensionality reduction is mainly based on principal component analysis (PCA) [166] and independent component analysis (ICA) [167]. These techniques are applied to solve various issues such as localization and tracking accuracy in an indoor location, and recovering smart meter data.
- Density estimation: It enables to encode density information. The algorithm directly models the scenes without detecting all pedestrians and vehicles and then estimates the density according to the response function modelled [168].

4.3 Reinforcement learning

As UL, the system will have to learn the expected result by itself. Unlike SL and UL, reinforcement learning (RL) neither tries to find category and reconstruct the system model as SL does, nor it tries to find hidden structures as UL does. Instead, RL tries to make decision and find the optimal solution given the constraints imposed by the inputs. RL comprises single or multiple agents that are decision-makers. The learner agent interacts with the environment, getting feedback in terms of reward and penalties. The agent tries to maximize the rewards within closed-loop fashion. The reward is given if the selection decision of the system is good, otherwise, the system has a penalty. RL aims to map situations *S* into actions *A* so as to reward a good behavior. RL models the environment as Markov decision process (MDP) and is useful for solving problems that may have multiple optimal solutions [4, 36, 169]. Q-learning, double Q-learning and prioritized experience replay are the most popular models in RL.

- Q-learning (QL): QL algorithm is one of the most popular algorithms of RL, it overestimates the actions under certain conditions. This overestimation is not known whether it harms the performance or not. It uses unrealized values to maximize the estimated action values[170].
- Double Q-learning: It aims to minimize the overestimations of QL by decomposing the max operation in the target into action selection and action evaluation [46, 170].

4.4 Others

There exist numerous other ML techniques such as Markov models, heuristic algorithms, controllers, transfer learning (TL) and deep learning (DL).

- Markov models: They are used in systems that randomly change and it can be presented such as stochastic models. The property of Markov models is used generally in the statistic. Generally, the models applied in cellular networks are Markov chains (MC) and hidden Markov models (HMM) [4].
- Heuristic algorithms: They are used to present the best decision and result solution, these algorithms used simple rules and guideline. Sometimes, these algorithms are used to make a solution to a specific problem that it has not

a solution or consumes an high cost solution calculation. There are two popular Heuristic algorithms: heuristic and genetic algorithms (GAs) [4].

- Controllers: Their simplicity and their flexible implementation help to apply the basic SON tasks in cellular networks. Feedback controllers and fuzzy logic controllers are two most commonly used in a cellular application [4].
- Transfer learning (TL): The learning mechanism can be empowered through the transfer of knowledge from a related task that has been already pre-trained. It is useful for applications that have reduced amount of data [4]. For instance, TL has been successfully applied in caching in order to improve the QoE [172].
- Deep learning (DL): An important member of the ML categories. The term *deep* refers to having multiple layers in the network. DL algorithms can be further classified into deep SL, deep UL as well as deep RL. Based on the extraction of data features, it allows algorithms to classify, predict and make a well-known decision. Besides, it is composed of well-known multilayer perceptron and such previous layer's result is used as an input. The increasing number of layers helps to rapidly obtain the desired result [102, 173].

5 Machine learning applied to 5G technologies challenges

We split this section according to the above mentioned challenges of 5G. Then, we report in each sub-section, the proposed ML solutions and their domains of applications.

5.1 Machine learning approaches applied to 5G anomaly detection challenges

5.1.1 Machine learning approaches applied to 5G intrusion detection system challenge

To prevent against malicious attacks, the researchers proposed solutions based on cybersecurity defence systems , in particular the well-known IDS which are widely used to defend cyber threats from wireless communications by anticipating and removing vulnerabilities. The heterogeneity and the huge unstructured data of 5G, make IDS procedures obsolete and not efficient enough to detect potential cyber-attacks in real-time [174]. The European 5G Infrastructure Public–Private Partnership (5G-PPP) consortium has identified a relevant set of key performance indicators (KPI) [175] serving as a critical driver when analyzing and inspecting network flows during the detection process. For instance, the authors in [102] have proposed a network anomaly detection system embedded into the 5G architecture which rapidly and reliably identifies cyber threats in 5G mobile networks thanks to DL techniques. In general, AI, especially ML techniques, are introduced to traditional IDS in order to intelligently enhance intrusion detection. Based on ML classification techniques, IDS will able to classify the abnormal traffics in optimal way with self-learning ability [176], and predict with high accuracy the network evolution and traffic behaviour. The authors of [102] succeed in providing an accurate detection of anomalies with low response time and especially with the enormous amount of data expected in 5G. This article focuses on both modules virtualized anomaly symptom detection (ASD) modules which are located in the RAN and network anomaly detection (NAD) module which is related to high-level management and orchestration NFV plane in 5G infrastructure. The two modules are based on DL models deep belief networks (DBN) and Long short-term memory (LSTM). In general, DL is an efficient solution that ables to identify the network flows in order to detect the anomaly symptoms. The proposed system begins with ASD by detecting symptom from the arrival feature vector using a supervised or semi-supervised learning model with DBN or stacked autoencoders (SAE). Indeed, two criteria are imposed on the choice of DBN and SAE: the same structure shared by both models and it can be used in both supervised and UL. The supervised learning methods are trained when labelled set are available, in this work, the training labelled set is anomalous or normal traffic classes, it will classify the traffics according to these two classes. In semi-supervised learning methods, only the normal class is available, when the methods cannot characterize the traffic as normal, it will be marked as anomalous [177]. Then, the collected symptom packets by NAD will be the inputs of LSTM model, this model is trained in a supervised way which recognizes temporal patterns of typical attacks e.g., sorts them by their timestamps, assembles a time sequence of symptoms. Then, NAD will decide the symptom in any category will be classified (i.e., anomalous or normal). The authors are interested in time execution. Using DL, the initial classification of anomalies will be quick in order to offer accuracy for lower time response. The selected ML model ensures an efficient execution on a GPU, it can be developed in a constant operations' number for a given feature vector size, besides, when using GPU, the increase in input batch size, extracted from the feature vector, will decrease the execution time and increase the spending time when transferring the batch to the memory of GPU. One GPU is used in this work, thus, the batch size is limited by the GPU's memory. Although DL models usually aim to optimize memory usage, they replicate certain structures in order to improve parallelism with neglect the greater optimization of memory usage. In the other hand, the selected model provides good accuracy in the classification of anomalies, it has the same memory requirements regardless of the number of samples used in training. In [178], the authors proposed an extension of the architecture of [102]: 5G-oriented cyber defense architecture, helping to determine quickly cyber threats in 5G mobile networks. To analyze the traffic, DL was used by extracting the features from the network flows. The extension is to use new virtualized resources to detect anomaly symptoms when network traffic increases. In this respect, researchers in [178] focused on ASD model and prediction on time evaluation. ASD uses a supervised or semi-supervised learning model with DBN or SAE to detect symptom anomaly. Moreover, the authors use a NetFlow protocol that transports the flows to the flow collector where flow extraction is performed. The article also uses an existing labelled datasets CTU which is a set of input data. It contributes to the creation of features vector and it is used in the test of this article's proposal. In addition, CTU is a publicly available dataset suitable to be used in botnets attack detection. It is filled by real botnets attacks without simulations, ground-truth labels for training and evaluating and different types of botnets. It has thirteen scenarios illustrated in [179]. Like in [102], Maimó et al. [178] aim to quickly adapt to new data while using minimal computation resources. This article compares Tensorflow and cuBLAS libraries in CPU and GPU performance in three neural architecture, such as Tensorflow has a higher CPU performance than cuBLAS, contrariwise with GPU performance. Besides, GPU performance is the same in the three models.

Another approach used the same technology as in [178] is presented in [176], the authors proposed a new intelligent IDS based on Software Defined 5G architecture. SDN technology can integrate security function modules in the new platform and with centralization of their management and control. Besides, the proposed system can intelligently learn the rules of a large data amount and detect intrusions using ML algorithms: Random Forest and k-means ++ with Adaboost. To intelligently learn the rules and detect intrusions, the flows arrived at the proposed system will pass through three main layers. From the beginning, the flows are collected by open flow-controlled entities in forwarding layer and they are uploaded to the control layer with blocking malicious flows. Then, the suspicious flows will be identified by the Management and control layer and the anomalies will be detected preliminarily using uploaded flow information from the forwarding layer. Finally, the intelligent centre, in data and intelligence layer, will analyze the features selection of and classify flows through ML algorithms. The paper finishes by the optimality of the system that can achieve a higher accuracy reached 97.3% than other solutions and without sacrification of high time complexity. To detect the intrusion, Random Forest is used to select the optimal subset of flow features then K-means is used to classify the flow into different classes of attacks based on selected features. The use of K-means provides a fine-grained classification, in another way, it caused a high processing time compared with other systems. The use of Random Forest helps to provide high precision of predicted intrusion, a high number of correctly predicted intrusions and a good trade-off between precision and number of correctly predicted intrusions than other ML solution, a less percentage of normal traffic which is misclassified than other ML solutions. Among the advantages of using these two ML solutions, the less overhead of misclassification than other ML solutions.

Due to the enormous number of devices, intrusion detection, such as botnets detection, is presented as a challenge that still needs to be resolved, especially with the emergence of 5G and the increased number of devices powered by ultra-wide broadband. Like [176], the authors in [180] proposed a new system that uses multiple ML algorithms applied in SDN and NFV technologies to previously detect intrusion. Dimensionality Reduction techniques help to avoid the Curse of Dimensionality problems i.e., minimize the higher dimensionality by selecting or extracting the features considered by the model in order to minimize the resources consumption and prevent performance deterioration. Principal component analysis helps to reduce the dimensionality of the data. Besides, the authors use Class Sorter ML solution to analyze the flows and offer flexibility in detecting the known attacks. This ML solution presents a high precision compared to other ML solutions.

5.1.2 Machine learning approaches applied to 5G anomaly detection based on big data analytics challenge

big data analytic is a powerful tool that can cope with the explosive data of large scale 5G networks. The typical big data analytics schemes in the design of wireless networks are presented in Fig. 6 [181]. In fact, it offers E2E visibility to collect, manage and analyze a large scale data in real-time, it coordinates intelligently the network functions and entities and it ensures a minimum energy consumption with a good performance evaluation [182]. It captures the traffic using Call Detail Records (CDR) as its capacity has reached a terabyte per day [183]. The use of Big data analytics enhances the end to end visibility in 5G and build a proactive network

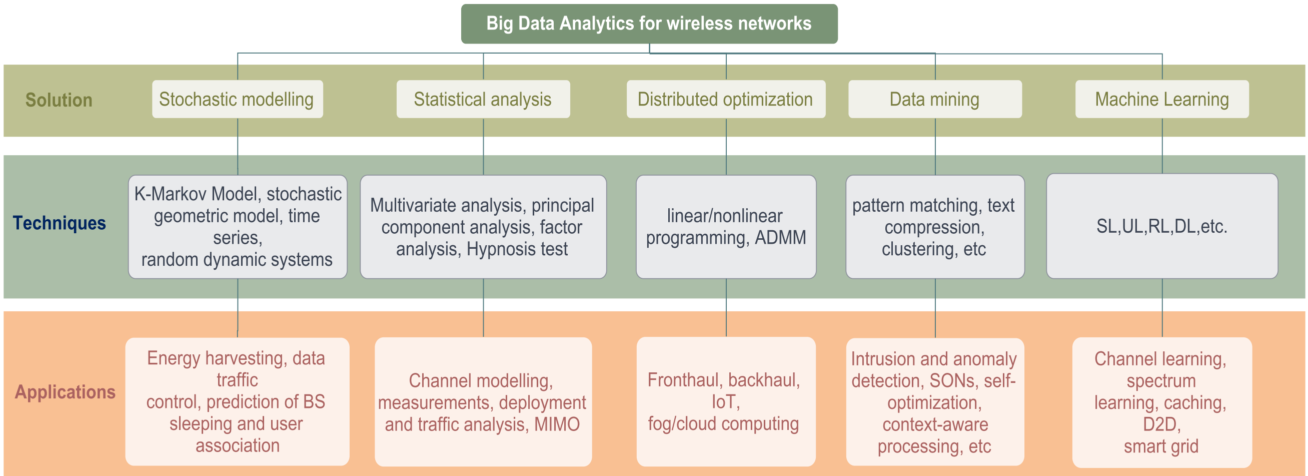

Fig. 6 Solutions/techniques applied in Big data analytics wireless networks [181]

by extracting intelligence through the application of ML techniques [160].

In [184], the authors use CDR to analyze anomalous behaviour of mobile network big data and unsupervised clustering algorithms: K-means clustering and hierarchical clustering to detect and classify anomalies into anomaly-free and anomalous data. Then, they used the neural-network-based prediction model that forecasts future traffic pattern, it evaluates the performance of data anomaly-free by analysing the mean square error in prediction after passing anomalous and anomaly-free data. K-means is less complex and has less large dataset than hierarchical clustering. In contrast, hierarchical clustering offers, in general, a better performance than K-means. In this study, the two techniques present the same performance. Using Neural Network, the difference error of the prediction, using training data with and without anomaly, decreases in term of epoch. Besides, less mean square error has resulted in anomaly-free data than with anomalous data, it must provide an ML solution that search on the minimization of the difference error between predicted and real anomalous data.

5.1.3 Machine learning approaches applied to 5G mobile edge caching challenge

mobile edge caching is a technique used by mobile edge computing (MEC), which minimizes latency and computing overhead by caching the content in 5G edge node, such as BSs and UEs. In order to make an accurate caching strategy, a high volume of data should be collected and treated by edge network. In general, the edge node in MEC system should select carefully the data, transmission power, time and canal in traffic offloading. Unfortunately, the collected data can be exposed to attack. In order to protect against the smart attackers, most existing solutions, based on accurate knowledge parameters that can be difficultly obtained by the edge node, are vulnerable and insecure. The attack parameters are difficult to be estimated because of their significant change over time, the limited memory, energy and computational resources [95, 181]. To secure mobile edge caching system as well as the collected data, the edge node uses RL techniques to learn the current state, in order to decide their actions such as the security complexity and defence levels. Indeed, the node learns the status of the other nodes and observes the attacks characters to intelligently decide its actions. The strategy will be optimal after enough interactions with attackers. In order to avoid jammer signal and the interference from other radios, the authors of [185] provide a solution based on RL techniques, especially, QL technique that helps to learn current state and optimally select a sub-band policy in wideband autonomous cognitive radios. This solution allows the radios to simultaneously operate in the same wide spectrum band avoiding each other and the jammer signal. Compared to random sub-band selection policy, the proposed solution offers a substantial improvement. Although in interference case between radios, the QL offers a flexible switching to new spectrum sub-band without interrupting transmission that can be very long.

5.1.4 Machine learning approaches applied to 5G Cell fault management challenges

According to [186], ML techniques in cell fault management can be divided into two principal categories: analytics and active techniques. Analytical techniques are used for detection, diagnosis and root cause analysis, whereas active techniques have been exclusively applied to the compensation stage. The analytical techniques are composed of the following attributes: output type (continuous, discrete), supervision mode (supervised, unsupervised), training

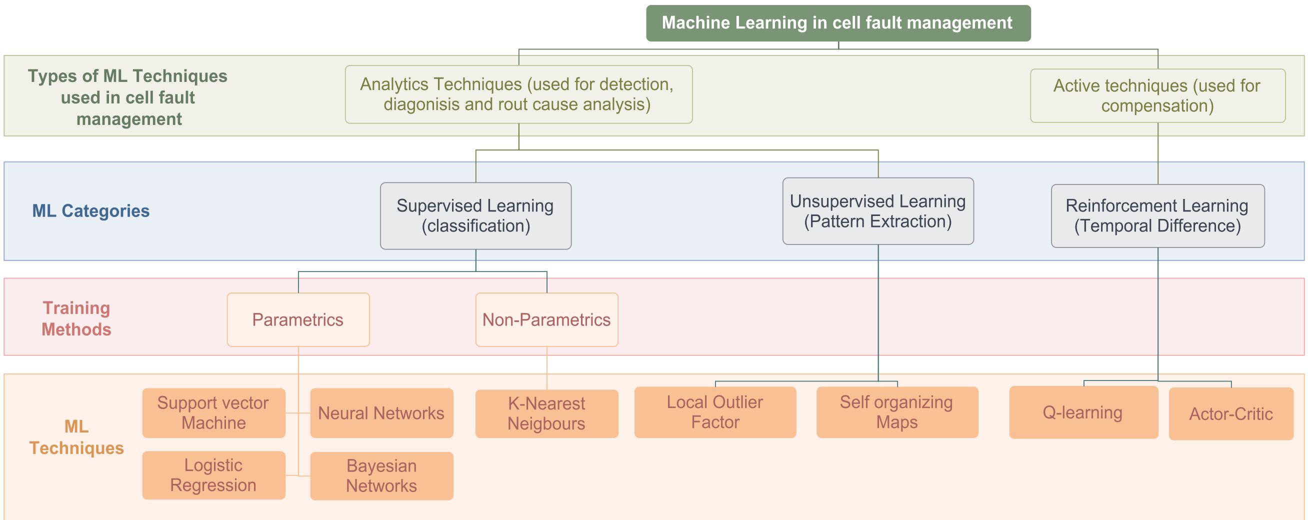

Fig. 7 ML applied in fault management [186]

method (parametric, non-parametric), scope (global, local). The parametric method adapts a model using the training data all along with a training phase. The training data set in supervised mode is composed of input data and output data values that can be the class labels in supervised classification. In order to reduce the cost metric, the fitted model adjusts their parameters. Then, it will be used to classify live data. The non-parametric method directly used the training data without a fitted model. The parametric method can use various ML techniques such as SVM, NN, Logistic Regression, Bayesian Networks. NN can be used in detection and diagnosis, it manages several input data attributes without the need for high preprocessing. It requires high computing power and enormous datasets to train in complexity case. Bayesian Networks is used for diagnosis the understandable fault causes and historic data accessible situations. Bayesian Networks have limitations like the one cause must be diagnosed at a time and each cause must use an independent symptom than other cause. Logistic Regression, K-NN and SVM are used in detection, SVM and K-NN propose a binary classification using low dimensional data input, but in another hand, the computing can be expensive with K-NN and SVM due to enormous training sets. In addition, K-NN offers a low performance in case of using high dimensional data input. Logistic Regression and K-NN can be used also for diagnosis. Although Logistic Regression does not need high computing resources, it can not handle heterogeneous classes. In unsupervised mode, SOM is used to diagnosis. SOM can represent a data cluster and design high dimensional datasets onto 1D and 2D discrete dataset, but it requires additional cluster validation algorithms. The active type is based on RL techniques, like QL that used for compensation. It takes a long time to converge. The application of ML in fault cell management is summarized in Fig. 7. In general, ML solutions can operate a high number of input features, it can autonomously enhance and retrain itself. It uses standard libraries that help to solve fault cell management problems. In the other hand, it needs to collect sufficient fault data which is a difficult task for ML in fault management. Several interesting techniques are provided to tackle the fault management problem in 5G [187–189]. Cell Fault management aims to reduce the number of errors such as cell outage in order to optimize the QoS. Detection of fault is performed to find degraded cells or cell outage. In 5G HetNets, cell outage is considered as a critical problem. Due to the heterogeneity and complexity of 5G cellular network, the risk of hardware and/or software failures and misconfiguration of parameters caused by human error are high which cause a total or partial cell outage.

The cell outage can be considered as an anomalous behaviour. To address this issue, the authors in [190] proposed a solution based on HMM named detecting outage by hidden Markov model (DOHM) to effectively detect cell outage in 5G HetNet. This solution classifies 5G bases station states into four different states: healthy, degraded, crippled and catatonic. HMM has the advantage to automatically predict the BS's states based on the current states classification to probabilistically estimate if there exists a possible failure. Results show that the proposed solution is capable to predict a BS state with $\simeq 80\%$ accuracy and a cell outage detection with $\simeq 95\%$ accuracy of the time. We extract some advantages and disadvantages of HMM utilisation in this work. Using HMM, a high correctly detection of cell outage is resulted with $\simeq 95\%$. In addition, the precision of detection is improved while the number of users is augmented. Moreover, HMM can offer an excellent accuracy of healthy macro BS estimation with $\simeq 99\%$. In the other hand, the use of HMM does not help to well estimate the degraded cell and to provide a high prediction accuracy of cell state, only $\simeq$ 80%.

In [191], the authors presented a new self-organized cell outage detection architecture to detect the cell outage in 5G H-CRAN. This architecture is evaluated under the handover data analysis use case. Each user served by an outage RRH will be offloaded to its neighbour cells. The rapid decrease of cell's power transmission during the Handover indicates that cell outage has occurred. The solution uses modified Local Outlier Factor (LOF) to detect the anomalous behaviours of handover in cell outage. The results evaluation indicate the effectiveness of the solution. The authors compared the performance of two ML solutions M-LOF and LOF. With LOF and M-LOF, 3% of the chance that an outage cell is not recognized from the outage cells which is a low percentage. In addition, M-LOF has a better performance than LOF, M-LOF presents 6% of chance that a normal cell is recognized as the outage cell from all normal cells, while LOF presents 12%. We observe that M-LOF is more correct.

The cell outage rates are expected to be higher with the increase of 5G cellular network complexity and potential conflicts between SON functions, the authors of [192] suggest a solution that analyzes and evaluates the behaviour of the cellular network and the effects of the arrival of faults on the reliability behaviour of the cellular network. Using Continuous Time Markov Chain (CTMC) ML solution, a reliable model is created that can predict the expected time of the first occurrence of the fault and the long term behaviour of BS, considering the diverse possible fault scenarios. The diverse possible fault scenarios are estimated from the analysis and evaluation of previous behaviour scenarios. CTMC can offer a reduction on diagnosis time and compensation time thanks to starting on the diagnosis before the approach of the expected suboptimal behaviour or outage time. Thus, the network will be more reliable and provides high QoE.

5.2 Machine learning approaches applied to 5G interference challenges

5.2.1 Machine learning approaches applied to D2D interference challenges

the interference is among the major challenging problems in D2D communication. The reuse of spectrum between D2D users that are controlled by the BS and the sharing of the same radio resources between D2D users and cellular system cause three types of interferences. It is necessary to control and minimize the interference between D2D users communications, the interference from BS communications and D2D users communications and vice versa [193, 194]. For example, in [195], authors propose an interference management technique such as power control based on ML at D2D users. Besides, power control minimizes the interference between D2D caused by multiple users which use the same frequency band, optimize the energy-efficiency and improves capacity in cellular systems. Moreover, the user can adjust his own transmit power through power control which ensures a good QoS and an adequate signal-to-interference ratio (SIR) at the base station [196]. Thus, the authors of [195] provided two power control algorithms based on distributed QL and CART DT algorithms. In fact, QL is used to reach the global optimal policy. DT classifier is used to minimize the time complexity of QL and it is based on the generation of training samples from QL. The results presented by four main algorithms: QL, DT, close loop control and open-loop control are compared. The results present a good performance in D2D throughput and system throughput, D2D energy efficiency and system energy efficiency when using QL and DT algorithms compared with close-loop control and open-loop control. In the other hand, QL consumes a high time because of the high number of iterations.

Other D2D interference solution [197] proposes hierarchical extreme learning machine (H-ELM) neural network with the same power control technique as in [195] to manage the severe interference in D2D communications. The QL algorithm in the previous work is a high time-consumed algorithm because of the high number of iterations. The H-ELM has the advantage to be time-efficient ML algorithm. It brings two main benefits: low computational delay and interference minimization between D2D users. Compared to the ML used by [195, 196], H-ELM has better performance, since it performs with less time for data classification in real-time than QL. Like the previous work, QL and H-ELM are compared with DT and Close loop control and open-loop control. H-ELM and QL present a lot of label, the same system throughputs and D2D throughput. In addition, they present the best performance. H-ELM presents the less time-consuming for the real-time power control task, then DT and finally QL. In the other hand, compared with close loop control and open-loop control, it presents a high time-consuming for the real-time power control task.

5.2.2 Machine learning approaches applied to interference in UDSC challenges

although the deployment of small cell ensures high capacity, it comes with energy consumption and inter-cell interference problems. The inter-cell interference can be reduced by using an intelligent power control algorithms based on ML techniques. To consider these issues, authors in [198] have proposed UL clustering approach called Affinity Propagation Power Control (APPC). APPC performs data analysis and extracts the knowledge and behaviour of the system under complex environment. APPC mechanism includes Affinity propagation clustering algorithm that allows grouping small cells into clusters by identifying a centre at each cluster and determines the number of clusters and the corresponding cluster centres. Indeed, the cluster centre generates the strongest interference compared to other members of this cluster. Thus, it is necessary to adjust the power transmission of cluster centre in order to reduce the co channel interference with neighbouring cells. The APPC clustering and power control present a high total throughput and energy efficiency in the cluster centre than K-means clustering approaches, always BSs are ON and BSs are ON/OFF transmission approaches. In addition, the proposed solution offers low time execution than other solutions. The APPC algorithm can reach $\simeq 95\%$ of energy efficiency but it is not the best. However, minimizing interference causes performance degradation resulting in poor throughput to the cell edge users of the cluster centre. This work adds another Victim-Aware Channel Rearrangement algorithm (VACR) mechanism to improve the signal quality of edge users by rearranging channels used in the interfering cells in order to minimize the interference of these victimized users. Thus, APPC and VACR can enhance energy-efficiency and throughput in the plug-and-play UDSC scenario. This article proposes a Data-Driven Resource Management (DDRM) framework included APPC and VACR. It is necessary to collect various operation data from UDSC such as the number of users, transmission power per small cell, reference signal received power (RSRP), channel usage per cell by using a central controller HeNB management system (HMS).

In addition to inter-cell interference, the interference caused by femtocells to neighbouring cells is an additional problem in dense 5G HetNet which causes a degradation of performance. To intelligently manage the interference in UDSC, RL-based methods are well-studied and exploited to cope with resource allocation problem. For instance, Amiri et al. [169] have used cooperative QL based power control to increase the QoS of users in femtocells without considering the channel variations. The proposed system model has the following architecture: a single macro cell with Macro Base Station (MBS) including M femtocells with M Femto Base Station (FBS). Each FBS serves a single Femto User Equipment (FUE) and MBS also serves a single Macro User Equipment (MUE). MUE receives a signal from MBS including interference from FBS and it is the same for FUE, it receives a signal from its FBS including interference from MBS and other femtocell base stations. The signal is characterized by the channel gain and the power transmitted by the base station towards its user equipment. Their goal is to implement cooperative QL in order to manage the interference via distributed resource allocation optimization. This solution provides an efficient power allocation between FBSs to maximize the sum capacity of the FUEs ensuring QoS for each one of them. Therefore, it uses a new reward function to satisfy the required QoS and contribute to the desired solution of the optimization problem. QL provides better fairness throughout for the whole network. QL approach satisfies the QoS (capacity of MUE) for MUE until arriving a high number of FBS, the QoS decreases. Thus the densification of FBS influences the capacity of MUE. The sum capacity of FUE is increasing when the number of FBS increase and higher than other approaches. In addition, it provides the same number of iterations for each number of FBS, consequently a better understanding of time duration. When reducing the complexity of resource allocation, the proposed QL approach can serve all users FUE.

5.2.3 Machine Learning approaches applied to Fog-HetNet for interference mitigation

the lack of coordination between small cells results in severe interference. Therefore, fog network investigation is important in this context to provide control and coordination between small cells. Although the integration of fog computing principles, by upgrading some small cells to fog nodes in HetNet, can enhance coordination and decrease the interference in HetNets, the location selection of fog nodes among many small cells and the determination of connection between small cells and fog nodes are challenging design problem.

The article [199] proposes an approach that helps to select the locations of edge nodes that will be upgraded from small cells. The main concept of this approach is to dynamically determine the fog nodes among many small cells depending on the channel conditions and the fog nodes are upgraded from small cells to solve the performance problems, then, using Water-Filling Clustering algorithm to cluster many small cells to around fog node head. This algorithm is adapted to ML in order to maximize the throughput and minimize the interference. The advantage of this work is that spectral efficiency increases and the latency decreases with the increase of signal-to-noise ratio, in addition, the use of water-filling helps to improve the spectral efficiency and latency diminishing compared with Voronoi tessellation model. The water-filling clustering presents a low latency with high bandwidth. However, the biggest limitation of this work is: the augmentation of fog node number decreases the spectral efficiency because of the poor channel presented by some fog node, and increases the latency.

5.2.4 Machine learning approaches applied to H-CRAN interference challenges

other inter-cell interference solution applied in H-CRAN, like [200, 201], propose an allocation of resource blocks (RBs) with considering the user's priority and power scheme in H-CRAN based on online learning in order to minimize the inter-cell interference between macro BSs and RRHs and improve energy-efficiency spectral efficiency and data rate. Comparing with standard online learning, the proposed online learning algorithm increases the convergence speed. In the other hand, the average energy efficiency decreases when SINR threshold of MUE is increasing.

5.3 Machine learning approaches applied to 5G handover challenges

Although the network densification enhances the capacity and increases the data rate, it causes mobility problems: handover problem such as too-late HO, too-early HO, HO to wrong cell, ping-pong HO, and unnecessary HO, and to mitigate these problems, the work [202] suggests a data-driven handover optimization (DHO) approach based on NN ML algorithm. The NN estimates the relationship between the weighted average of the ratios of these mobility problems and the HO parameters such as the HandOver Margin (HOM) and time-to-trigger (TTT). This solution aims to minimize the weighted average mobility problem ratios and to measure the mitigation of mobility problems in order to optimize handover between densified cells. This NN solution has several advantages such as the efficient mitigation of HO problems and HO parameters optimization. The simulation shows that this approach may effectively mitigate mobility problems. However, the considered NN, to estimate the KPI function, requires a large diversity of training for real-world operation, which presents severe constraint for the mobile networks with densification.

Radio link failures (RLF) is another problem which can be occurred as a result of wrong HO decision, congested cells or even poor coverage. To reduce RLFs within HO, Khunteta et al. [203] have proposed a recurrent neural network (RNN) or long short-term memory network (LSTM) ML algorithms in order to predict signal conditions, then, the behaviour of these signals conditions will be the inputs to K-nearest neighbour (K-NN). Based on the signal conditions and the status of the handovers that happened in the past, K-NN allows predicting RLF in advance by classifying whether the HO will be successful or not. This advance classification will help users to take decisions in order to reduce the possible link failure in HO. In 5G cellular network, the users can autonomously control HO decisions in order to reduce failure link. They control links based on enough data about the link conditions and assistance from network in order to decide to perform HO. The main advantage of RNN application is that it uses its internal memory to perform input sequence and presents in their neurons an internal directed cycles. This work presents a comparison between RNN+classification results and LSTM+classification results and it concludes that LSTM presents a high accuracy classification compared to RNN, almost 99% true positive classification and 1% misclassified for HO fail event. Recent research has considered an early initiation of HO in order to improve energy efficiency by switching off the serving BS resources earlier [204]. By considering the energy efficiency of HO process, the reduction of the signalling overheads during HO is a promising future direction.

5.4 Machine learning approaches applied to 5G large data volumes challenges

Caching contents is designed for the problem of storage capacity in a 5G network. The primary goal of caching is to improve the QoE of the end-users by ensuring fast adaptation to radio link conditions. In [172], the authors provide an interesting work of edge caching, to maximizing the backhaul offloading, using TL-based approach. This approach uses the user's content requests to estimate the popularity matrix of the content. TL offers the capability to learn and transfer hidden latent features extracted from the source domain (D2D-based Social network) to the target domain (Cache-enabled small base station.) in order to optimally cache most popular contents. In other words, TL transfers the knowledge from prior relative tasks to a latter task accelerating the training process. A key advantage of TL is that it offers a satisfaction ratio and backhaul offloading gains nearly similar to the ground truth baseline in high storage sizes case. Moreover, TL has a satisfaction gain more than Collaborative Filtering method up to 22% and a backhaul offloading gain more than Collaborative Filtering up to 5% in high storage case. Although the use of bottlenecks, the satisfaction gain and backhaul offloading gain with TL decrease, they still more than Collaborative Filtering. The increase in capacity of backhaul avoids the use of bottlenecks, in this case, TL achieves an important backhaul offloading gain compared to Collaborative Filtering. However, this approach may not be applicable in practical systems where the content set or user preference are dynamic since the content popularity is assumed to be static.

In [205], another ML TL-based approach achieves better prediction on caching decision while reducing network cost. In fact, MEC is a solution that minimizes latency by storing data in edge nodes network. The authors designed a new proactive caching mechanism named Learning-based Cooperative Caching (LECC) strategy that minimizes the cost of transmission and optimizes QoE. The authors use TL technique to estimate the content popularity, then they propose a nonlinear programming formulation to optimize cache content placement which is a greedy heuristic algorithm. First, a content items classification based on access feature correlation is established using K-means clustering, then TL uses the historical popularity information gathered from the target and source domain to estimate the cache content popularity.

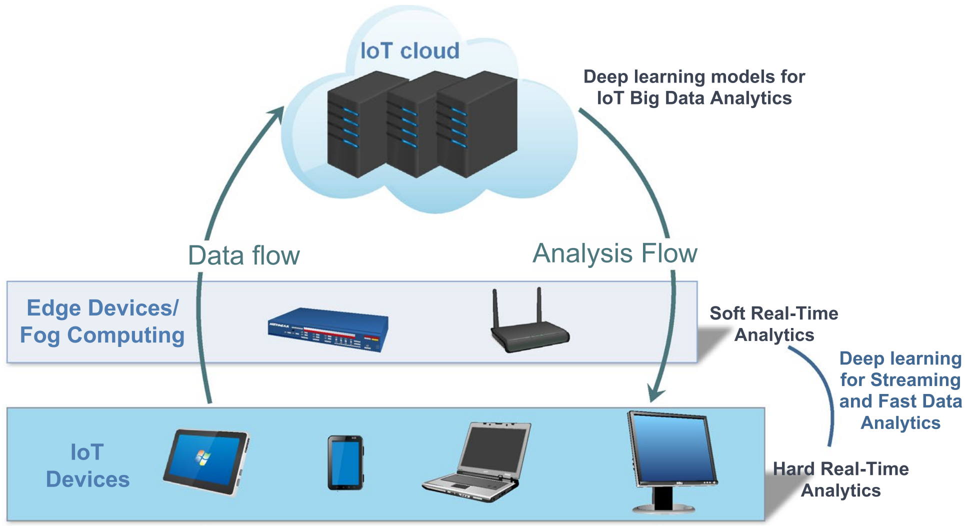

Fig. 8 Generation of IoT data [211]

Due to the limitation of SBS caching size, article [206] mentioned the need to maximize caching-efficiency in order to improve QoE and optimize the backhaul offloading. To maximize the cache efficiency with high accuracy and low complexity, the authors propose two ML techniques to tackle this problem. K-means clustering algorithm can fully uncover hidden spatio-temporal patterns of content requests at SBSs, and achieve personalized inter-cluster cache and predictive intra-cluster cache. The second one is the use of K-NN classification algorithm to categorize the new contents and periodically cache them in the corresponding cluster with high accuracy and low complexity. The K-means solution presents a low backhaul load than Model predictive control (MPC) strategy with 20% of cache size used. With K-means, user satisfaction increases when the cache size is increasing with higher performance than MPC strategy. Other advantages of K-NN application can be outlined, is that the backhaul is lower than K-means clustering solution, nearly to 99% of load in 40% of cache size thanks to the use of the corresponding cluster to cache the new contents. K-NN can satisfy users better than K-means clustering.

Another solution for caching optimization include the work in [207], it proposes a new caching solution in the 5G network based on Reinforcement-learning especially QL. Therefore, QL is implemented to find the optimal caching policy in order to intelligently prefetch popular files across time and space. The intelligent prefetch is based on learning of what and when SBS has to cache, regarding SBS memory limitations, the massive number of available contents, the unknown popularity profiles, in addition to the space-time popularity dynamics of user file requests. QL solution provides faster convergence and low complexity and memory requirements.

The article [208] proposes a new proactive caching architecture based on big data platform, which processes a huge amount of data on a big data platform and estimates the content popularity using Collaborative Filtering ML. The content caching decision is performed by storing the contents greedily in SBS up to no space available. Collaborative Filtering presents user satisfaction less than ground truth and a backhaul load high than ground truth until arrives at 79% of storage size. Besides, user satisfaction increases with ground truth until it reaches 100% in 79% of storage size. This work presents several disadvantages of Collaborative Filtering, a performance gap between the ground truth and Collaborative Filtering is high in backhaul load and the evolution of users' request satisfaction with respect to the backhaul capacity ratio due to the estimation error.

6 Machine learning applied to 5G applications challenges

Several problems related to IoT, healthcare, smart grid and smart home and Internet of Tactile applications have appeared and must be solved especially with the emerged of 5G. In this section, we mention some application challenges the effective of ML in overcoming these challenges.

6.1 Machine learning approaches applied to internet of things challenges

Employing AI and ML in 5G is imperative to address the challenges of IoT applications and to deliver an appropriate decision-making process. The emerging applications of IoT such as medical, smart city and transportation applications require an intelligent learning mechanism for prediction, data mining and data analytics. Among the several ML techniques, DL is actively used in many IoT applications. This is obviously due to the fact that DL has the ability to effectively work on heterogeneous data and to address the emerging analytic needs of IoT systems. Figure 8 illustrates the management of IoT data generation and deep learning models to knowledge work automation. In this respect, several successful applications of DL in IoT domains can be found in [209]. In addition, the emerging challenge for realizing the efficient IoT is the pressure it puts on the existing communication infrastructures, requiring transfer of enormous data volumes. Thus, mobile tasks require DL methods in order to achieve efficient transmission of enormous data volumes with high QoS and user satisfaction. With the increase in the number of devices, it will be necessary to equally use multiple channels that assigned to each link in order to balance the heavy loads of integrated traffic in switches.

These channels must be frequently changed to adjust the high dynamic changement of traffic load. Without SDN central controller, the high dynamic of traffic load can affect the channels assignment, the suspension time of the network will be limited by the distributed setting. Recently, IoT becomes based on SDN to deal with heterogeneous resources and structure. In this context, the work [210] proposes DL solution performed in SDN central controller. SDN can process all channel assignment without the necessity to frequent changement of channels. By learning the previous channel assignment processes, DL can predict the future traffic load and help SDN to intelligently assign the multiple channels to each link. The results show that DL can minimize the iteration times of channel assignment process and data loss than without using DL. In addition, it can offer throughput higher than without DL. It offers high accuracy with low number of nodes, but the accuracy is degrading with the augmentation in number of nodes.

6.2 Machine learning approaches applied to healthcare challenges

The major part of the huge volume of big data is generated by IoT devices in healthcare environment. Several works have proposed 5G systems based on ML techniques to remotely provide a high-quality and continuous monitoring of the patient's health situations and supply treatment services without visiting the doctors.

In this context, especially in the diabetes detection context, the existing systems of diabetes detection are obsolete, uncomfortable and they lack intelligence, continuation in the supervision of multi-dimensional physiological indicators of diabetes patients, data sharing mechanism and continuous approaches of the prevention and supervision strategies for treatment of diabetes. The article [212] suggests 5G-Smart Diabetes system that detects and analyze the status of diabetic patients. Besides, it provides a sustainable, cost-effective and intelligent diagnosis of diabetes with post-hospitalization prevention treatment of diabetes. Effective analyses are taken to supervise the real-life and exercise of users. A check on the overall conditions of the user is done in a long-term and continual way. The authors propose ML solutions to achieve analyzing and predicting the disease. The authors used a DT, SVM, and ANN to contribute different models (different number of layers) for the public diagnosis of diabetes (public data come from the hospital diabetes big datasets with the removal of users' privacy and sensitive information) in order to evaluate the system performance. DT offers a high accuracy when the number of DT layers increases. ANN offers high accuracy when the number of hidden layers arrives at 110 layers. DT offers high accuracy with number of layers lower than ANN hidden layers. In general, SVM offers high accuracy compared with DT and ANN.

Another challenging issue in the Healthcare system is remote detection of a patient's' emotional state. To overcome this challenge, [213] proposes a health-system based on 5G cognitive system (5GCS) which remotely detects the patients' emotional state in order to analyze the psychological diseases, understands the emotional and psychological situation for human beings and rapidly provide treatment for physiological diseases. Authors require the intelligent 5G cognitive system to provide ultra-low latency and ultra-high reliability to the proposed system and to intelligently analyze the healthcare big data of patient's health status. The intelligent analyze is based on cognitive engine incorporation which provides the cognitive ability to the system using the ML and DL techniques along with cloud computing. This ability is relevant to particular healthcare requirements. SVM is used to predict the user's emotion. In the other way, the use of ML and especially DL techniques has an important role in the decision-making process even with very large volume of data.

The work [214] proposes to apply ML in 5G network to early detect heart diseases. It proposes a scalable three levels architecture based on IoT which is interconnected to cloud computing. This architecture aims to store and process the enormous amount of data generated by the health monitoring system based on IoT. It is divided into three-tier: the sensor data collection using the 5G network, storing of the huge volume of portable IoT sensor data in the cloud and development of the logistic regression-based prediction model in case of heart diseases. The third architecture is based on a prediction model using logistic regression.

6.3 Machine learning approaches applied to smart grid and smart homes challenges

To provide good management of electricity and energy consumption, the smart grid can use 5G technologies. Users can use the smart grid to remotely control smart homes appliances in smart phones. This control can adjust the energy consumption, deliver economic benefits and provide a real-time access to energy usage data. In fact, smart grid and smart home technologies trigger security problems making the grid vulnerable to physical damages and commercial privation.

In this context, the article [215] proposes an architecture which can secure smart grid communications to detect price integrity or load alteration attacks in order to safe smart grid. In addition, the AMI introduction affects the grid and make it accessible and vulnerable to DoS attacks or data injection attack against their different parts. The proposed network architecture integrates IDS to detect the attacks. IDS is distributed in different placements of the network, such as Home Area Network and Neighbourhood Area Network in Advanced Metering Infrastructure, to build a collaborative IDS in order to detect the above attacks. The attacker concurrently performs many high consumption loads. Based on this point in order to detect the attack, this IDS uses signature based on ML or statistical techniques to learn the statistics of voltage and frequency of the Alternating current power at different distribution placements. Thus, IDS intelligently measures the demand in real-time using the statistics and verify if there are sudden abnormal behaviour in the demand.

6.4 Machine learning approaches applied to internet tactile challenges

Another 5G application challenge which is focused by [216], it proposes a telesurgery robot using the 5G tactile Internet and AI. It is based on human-machine interaction data which provides the tactile perception of the doctors during the surgery process. This solution is proposed to enhance intelligently the treatment efficiency of various diseases and it can be released using the hardware virtualization, SDN, ML, DL, IoT technologies.

7 Synthesis

In this section, we will synthesize the previous works above in order to provide a comprehensive comparison among them. Each work choices ML solution according to the desired goal that solving a specific problem in 5G. The work contributions are summarized in Fig. 9 and Table 4. The works, in anomaly detection subsection, propose various ML solutions to achieve efficient anomaly detection with high accuracy. To optimize the intrusion detection, [102] and [176] use the same technology IDS although they apply different ML solutions. In contrast to [102], authors in [176] give serious attention to the time factor with the accuracy. Indeed, K-means in [176] causes a high processing time compared to DL solution that begins with quick classification to achieve a low time response gain. In [102], DL can offer a low execution time but they replicate some structure to improve the parallelism contrary in [176] that not need to replication. Due to the amount of traffic data, [102] sacrifices some accuracy contrary to K-means that gives more importance to the accuracy by intelligently learning rules of a large data amount. Another anomaly detection system is proposed in [180], based on a mixture of botnet traffic features. To reduce the high number of features, authors propose PCA in order to create an efficient learning module. Then, Class Sorter ML solution is chosen to classify flows and generate a reaction policy. All the classification results indicate that the selected class sorter ML, random forest is the most precise, the best in terms of area under the curve, classification accuracy and recall compared to K-NN, Naive Bayes, logistic regression. Moreover, this solution presents a better classification accuracy than [176]. Based on RL solution, the authors of [185] propose learning the current state and detecting the jammer signal. The goal is to decide future actions by selecting the sub-band policy in order to secure the edge caching procedure. The proposed selection solution can achieve up to 73% of the above-mentioned maximum possible performance which is better than random selection policy solution. Alias et al. [190] and Yu et al. [191] focus on the detection of cell outage. Based on the current states of BS, [190] proposes HMM to proactively estimate the future cell outage which can decrease the latency. While [191] proposes to reactively identify the cell outage based on the rapid decrease of cell's power transmission during a Handover. The solution in [190] presents an accuracy of cell outage detection higher than the detection of other cell states. Thus, this solution can be more effective with cell outage detection than the detection of cells in good condition. The works [190] and [192] have almost the same objective that the estimation of future behaviour of the cell. In the D2D interference section, [197] focuses on performance amelioration of D2D management and minimize the time consumption in a better way than the solution in [195]. To minimize the time consumption, H-ELM algorithm trains using the labelled data generated by QL algorithm and DT trains by collecting the data generated in QL and determining the training samples, feature space and output space. The results show that the use of QL and H-ELM provides better performance and lower time consumption than QL and DT. In [198], the proposed solution enhances the performance of UDSC although the complicated environment including several interferences and dynamical traffic variations. Like [169, 198] focuses on solving power control problem. The solution of [198] focuses on the effectiveness of APP Clustering and power control to minimize the severe interference. [198] concentrates on VACR algorithm in quality of edge users' signals improvement by rearranging channels used in interfering cells. In contrast to [169] which uses the QL to

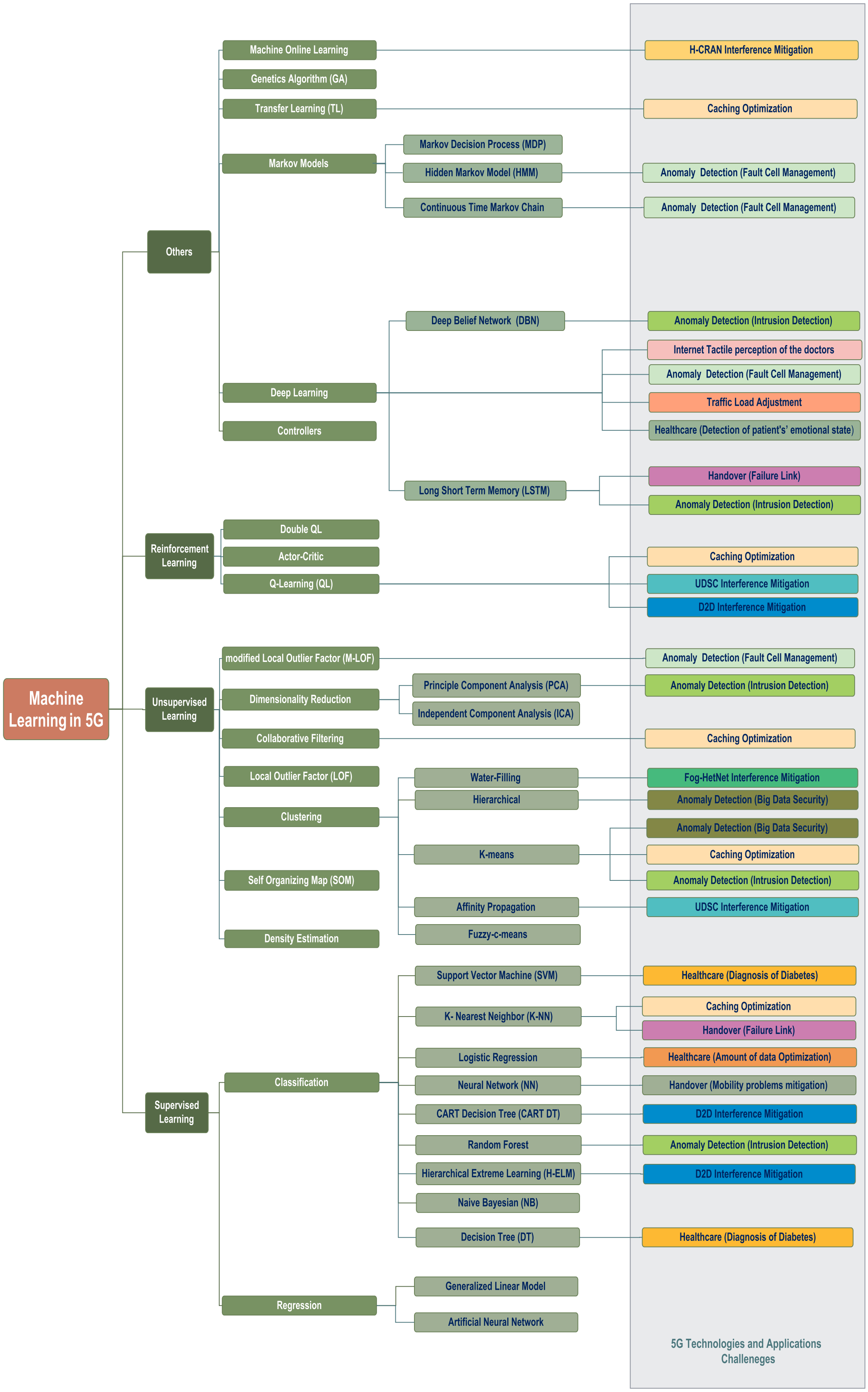

Fig. 9 Machine learning paradigm in 5G

Table 4 Summary table

Domain	Challenge	Ref+Y	ML solution	Objective	Main conclusion
Anomaly detection	Intrusion detection	[102] 2017	DL model	Intrusion detection optimization in 5G mobile network	The proposed solution can rapidly and reliably identifies cyber threats in 5G mobile networks
		[178] 2018	Different DL models	Intrusion detection optimization in 5G mobile network	The proposed solution aims to quickly adapting to new data while using minimal computation resources.
		[176] 2017	Random forest and k-means ++ with adaboost	Learn the rules of a large data amount and detect intrusions	The proposal can achieve a higher accuracy reached 97.3% using NSLKDD dataset than other solutions and without sacrification of high time complexity
		[180] 2017	Dimensionality reduction module, principal component analysis and class sorter random forest algorithm	Intrusion detection prevision	The proposal proposes to avoid the Curse of Dimensionality problems
	Big data security	[184]2017	K-means clustering and hierarchical clustering	Detection and and analysis of anomaly behavior using big data	The proposal achieves an accurate prediction of anomalies with minimum mean square error between predicted and real anomalous data
	Mobile edge caching security	[185]	QL	Secure offloading to the edge nodes against jamming attacks	The edge node can learn the current state, in order to decide their actions including the security complexity and defense levels. The proposed selection solution can achieve up to 73% of the above-mentioned maximum possible performance
	Fault cell management	[190]2016	HMM	Performant estimation of cell outage in 5G heterogeneous wireless networks	The proposal can predict a BS state with 80% accuracy and a cell outage detection with 95% accuracy of the time.
		[191]2018	DL model and M-LOF	Cell outage identification based on anomalous behaviours of handover detection in 5G H-CRAN	M-LOF presents 6% of chance that a normal cell is recognized as the outage cell from all normal cells, while LOF presents 12%.
		[192]2015	CTMC	Analysis the effects of faults arrival on the reliability behavior of ultra-dense Heterogeneous Complex cellular network	The network will be more reliable and provides a high QoE

Table 4 (continued)

Domain	Challenge	Ref+Y	ML solution	Objective	Main conclusion
Interference mitigation	D2D Interference mitigation	[195] 2017	Distributed QL and CART DT	Interference control by optimizing energy-efficiency and system capacity	The proposal can achieve good performance in D2D thoughput and system throughput, D2D energy efficiency and system energy efficiency
		[197] 2018	H-ELM and QL	Interference mitigation with high throughput, energy-efficiency and limited time consumption	The proposal limits the time consumption and minimize the interferance between D2D user
	UDSC interference mitigation	[198] 2018	Affinity propagation clustering	Inter-cell interference mitigation and energy-efficiency improvement	The proposal can reach 95% of energy efficiency
		[169] 2018	QL model	Power allocation between FBSs to minimize the interference	The proposal provides an efficient power allocation between FBSs to maximize the sum capacity of the FUEs ensuring QoS for each one of them.
	Fog-HetNet interference mitigation	[199] 2018	Water-Filling clustering	Selection of fog nodes to minimize interference and increase the overall data rate	Spectral efficiency increases and the latency decreases
	H-CRAN interference mitigation	[200] 2018, [201] 2017	Online learning	Inter-cell interference mitigation between macro BSs and RRHs and energy-efficiency improvement	The proposal increases the convergence speed
Handover	Mobility problems mitigation	[202]2016	NN ML	Handover optimization between densified cells	The proposal optimizes the HO parameters to minimize the weighted average mobility problem ratios.
	Failure link	[203]2017	RNN, LSTM and K-NN	Failure link minimization	The proposal can autonomously make HO decisions by controlling links in order to reduce failure link

Table 4 (continued)

Domain	Challenge	Ref+Y	ML solution	Objective	Main conclusion
Large data volumes	Backhaul offloading and content caching optimization	[172] 2015	TL	Optimally caching of most popular contents in order to maximize the backhaul offlading	The best satisfaction gain and backhaul offloading gain
		[206] 2017	K-means clustering and K-NN	System backhaul load minimization and QoE improvement	The proposal facilates the caching of the most accurate contents in SBSs considering the highly random and continual change content demands
		[207] 2018	*QL*	Prefetch popular files across time and space to reduce SBS memory limitations	The proposal provides a faster convergence and low complexity and memory requirement
	Content caching estimation	[208] 2016	Collaborative Filtering	Backhaul offloading and QoE optimization	Collaborative Filtering presents user satisfaction less than ground truth and a backhaul load high than ground truth until arrives to 79% of storage size
	Content popularity estimation and proactive caching optimization	[205] 2018	*TL*	Cache content placement optimization	Transmission cost minimization and QoE improvement
Internet of things	Traffic load adjustments	[210] 2018	DL algorithm	Predict the future traffic load and assign channels assignment to each link	The proposal offers less iteration time of channels assignment process and less suspension time caused by channels assignment
Healthcare	Amount of data optimization in IoT sensor devices	[214] 2018	Logistic regression model	Sensor data processing and the identification of the most significant clinical parameters to obtain heart disease	Respiratory Rate, Heart rate, Blood Pressure and Systolic Range and Body Temperature parameters help to indicate the heart disease.
	Sustainable, efficient and intelligence diabetes diagnosis	[212] 2018	SVM, ANN ,DT	Efficient diagnosis and treatment of diabetes	SVM offers the high accuracy compared with DT and ANN.
	Remote detection of patient's' emotional state	[213] 2017	ML and DL algorithms.	Remotly detection of the patients' emotional state in order to analyze the psychological diseases	The proposal provides ultra-low latency and ultra-high reliability in the health-system
Internet Tactile	Tactile perception of the doctors during the surgery process	[216] 2018	ML and DL algorithms	Provides the tactile perception of the doctors during the surgery process	The proposal achieves a powerful treatment of various diseases

mitigate the interference and enhance the QoS of users in femtocell without considering the channel variation. Interference minimization is the biggest objective of [198, 199]. The main difference between these two works consists in the use of water-filling clustering in[199] to create the fog nodes clustering. In fact, fog nodes clustering provides coordination among small cells to minimize inter-cell interference, while [198] is based on power control to minimize the interference. In the other hand, [200, 201] concentrates on the interference minimization between macro cells and RRHs.

To facilitate the HO procedure, [202, 203] present two different solutions. [202] uses NN to estimate the weighted average of different mobility problem ratios in order to mitigate HO problems. The work [203] uses the K-NN to predict the RLF problem in order to decide the HO in the future without focusing on solving the RLF problem. The good classification of HO fail or not facilitates solving RLF problem.

In large data volume subsection, [172, 205] uses the same ML solution but for different objectives. In [172], TL is used to optimally cache most popular contents and in [205], TL is used to estimate content popularity in order to proactively optimize the cost of transmission/latency. Similar to [205] objective, [207] propose QL to proactively learn the caching policy to prefetch popular files considering the memory limitations of BSs. Hou et al. [205] aims to minimize the cost of transmission, and [207] aims to minimize the complexity in memory requirements in addition to latency minimization. Comparing with [205, 208] also aims to estimate content caching using Collaborative Filtering solution. In general, Collaborative Filtering learning techniques are suboptimal because of ML challenges such as data sparseness and cold-star problems. In fact, the content popularity matrix is large and sparse with few users ratings which make collaborative filtering inefficient. TL can overcome these problems by exploiting the available data from other rich information sources such as from source domain.

8 Open issues and future trends

This penultimate section focuses on the future 5G-applied AI/ML challenges. We present some techniques that may ameliorate previous related works. The limitations of ML techniques that applied in large scale communication network, especially in 5G network. Moreover, we provide a deep insight into Beyond 5G and 6G key technologies corresponding with AI/ML challenges.

8.1 Opportunities and challenges for future works

The proposed works above had suggested several varieties of future opportunities and challenges. In this section, we underline the important research issues presented by previous works emerging 5G networks.

- 5G intrusion detection The authors in [102] have been conducted to measure the performance of using CPU/GPU by calculating the number of feature vectors processed per second. According to experiments, this performance is presented with batch size and vector of 128 features. Researchers had noticed that the gap between CPU and GPU performance increases relative to the number of layers. Furthermore, the results showed that the maximum of six-layer GPU performance presents 2.47 million feature vectors per second. Thus, this result can be enhanced, especially with the huge increase of data in 5G networks by adapting of aggregation pace and using more virtualized resources. Besides, two levels of ASD and NAD can be trained using real data to evaluate the accuracy of the anomaly detection architecture as a whole. The paper [180] presents another future work in intrusion detection, it proposed a novel system to detect the botnet intrusions, it suggests to still working on obtaining as several datasets to test and evaluate the proposed system. The enough collected data can also help to select the optimal classification algorithm which will enhance the proposed system.
- 5G mobile network big data: The authors in [184] propose an extension of their work that concerns the analysis of the user users' contextual information in envisioned smart cities. The developed algorithms of resource allocation may improve its utilization. Energy meters tools can be used to collect data and to provide an insight into its utilization in smart cities.
- 5G mobile edge caching: In future, mobile edge caching will need more security. Indeed, it will be more difficult to precisely and rapidly estimate the attack state. Thus, the inaccurate and delayed state information impacts must be examined and the MEC security solutions must be enhanced by implementing advanced RL techniques. Furthermore, the paper [95] proposes the use of TL techniques to help RL using data mining to explore existing defence experiences and reduce the risk of trying bad defence policies in RL. On the other hand, 5G must use Backup protocols designed for it to avoid the security disaster that comes bad decision learning from rogue edge connection. Another approach that uses caching contents solution in 5G edge network is presented in[172]. It can be performed using real traces to know hidden problems.
- Interference in 5G UDSC: In this context, [198] suggests a DDRM solution to reduce inter-cell interference. This solution can also be used to examine the fast handover and cross-tier interference in the heterogeneous networks.
- 5G fault detection: The authors [192] can analyze the efficiency of their solution based on past failure logs col-

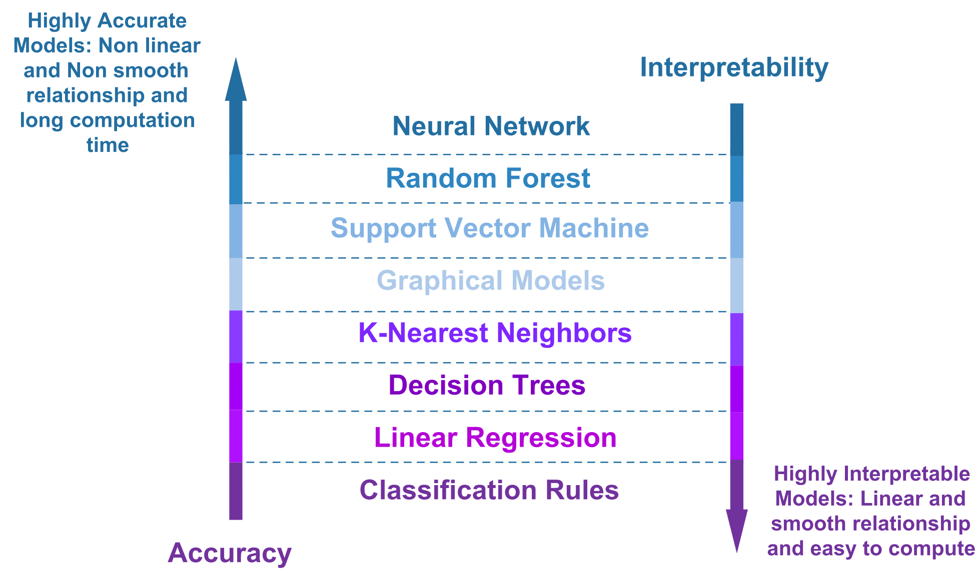

Fig. 10 Trade-off between interpretability and accuracy of some relevant ML techniques [165]

lected from in real network. The solution can be extended with non-exponential distribution for failures and recovery times.

- 5G healthcare: The security is inevitable challenge [213]. Unified API implementation for the system to simplify the installation of the application and the system accuracy improving with low time complexity are two important research directions. More significant attention to the optimization in the time of data analysis is required.

8.2 Computational issues of ML techniques for network systems

Network operators call for improve some performance and financial metrics such as: QoE, reliability, throughput, resource utilization, per-bit costs, and latency (for specific applications or overall) [217]. However, adopting an ML technique by network operator is complex decision that is constrained by several conditions. It depends on the type of the problems to solve, implementation complexity, the financial benefits and the operator affordability. Accordingly, the following limitations should be understood:

- Poverty of datasets and labeling: To offer an accurate ML models with exact results and useful insights and decisions, it is necessary to use a large number of available datasets with high quality. Currently, this goal is difficult to achieve, for 5G network, because of the scarcity of datasets, in addition to inaccessibility of existing operators' data due to privacy issues. For example, in DL, the high accuracy depends on the high volume of trained data. The supervised and semi-supervised learning is more complex when needing to labelled data to perform algorithms. Moreover, the expensive annotated data can not encourage operators to use it.
- Difficult AI/ML interpretation: AI/ML interpretation is the capacity to explain how to make decisions based on input data. Low accuracy has resulted when the interpretation of AI/ML models is complex. In fact, the high capacity in the interpretation of AI/ML models can provide full automation, reliability and transparency in Zero-touch network and service management (ZSM) [218] and enforces the reliability in AI enabled systems. In contrast, DL presents a compromise between high accuracy and complex interpretation of AI/ML models. Figure 10 taken from [165] illustrates the trade-off between interpretability and accuracy of some relevant ML models. Highly interpretable algorithms such as classification rules, or linear regression, are often inaccurate. While more accurate DNNs are as black boxes.
- High real-time management operations and high demand for computation, memory and energy resources: Among 5G requirements is the high accuracy decisions making with real-time management operations. Although DL technique can offer high accuracy decisions, it expands high training time to achieve this accuracy, especially the dynamic environment of 5G network. In this respect, ZSM [218] is defined to never need to handcraft a custom ML algorithm.

The major conditions for the application of ML are described as follows. The first condition for adopting ML is that its category (i.e., regression problems, classification problems, clustering problems and Markov decision making problems) should be suitable with problem to handle. For instance, content caching needs to get content request probabilities of users. The category of this problem can be seen as a regression problem, the user profile can be the input and the content request probabilities can be the output. Besides, K-means clustering algorithms can naturally handle the BS clustering, while many resource management and mobility management are modelled as a Markov decision making problem that can be efficiently solved by RL [219]. The second condition for the application of ML is the training data availability [219]. The type of data (i.e., labelled, or unlabelled) is key enabler to decide which type of learning to use, particularly in the deployment of applications for 5G use cases. The data used for training the ML system is collected in different way depending on tackled problems. For example, authors in [220] use one mobile device equipped with an IWL 5300 NIC that can read CSI data from the slightly modified device driver. In [221], a feature vector in the training data set under a specific resource allocation scenario is composed of time-variant parameters, such as the number of users and CSI, and these parameters can be collected and stored by the cloud.

In 5G mobile communications, the scarcity of available real datasets is considered as one of the greatest challenges for researchers and ML practitioners. For this respect, 5G research groups, academics, and key industry partners are defining and developing 5G infrastructure to generate their own datasets for research [222–224]. Moreover, [225] offers an open archive that stores wireless trace data from many contributing locations to develop ML algorithms and analyze the data.

The third condition that needs to be addressed is that the time cost should meet the requirement of the applications. Two time metrics can be considered training time and response time [4] The training time metric represents the amount of time that ensures a machine learning algorithm is fully trained. It helps to make accurate predictions for future inputs (for supervised and unsupervised ML) and to learn a good strategy or policy (for reinforcement ML). The response time metric refers to the time needed to output a prediction given an input, while it refers to the time needed to output an action for a trained reinforcement learning model. For instance, in resource management applications, where decisions should be made on a timescale of milliseconds, there is a stringent requirement about response time. For example, Neural Network techniques can be adopted to make resource management decisions in time for power allocation approaches [226, 227].

The fourth condition for the application of machine learning is that implementation complexity should be acceptable. A powerful algorithm can find a globally optimal solution at a low computational complexity and storage. For instance, ML techniques, like fuzzy QL, that involve simple mathematical operations in their learning process and their inputs are common network KPIs (like call dropping ratio), are considered as ML with low implementation complexity [219].

8.3 Beyond 5G or 6G

From 2020 to 2030, it is expected that 55% of the data traffic will be annually generated until it reaches 5,016 ExaByte (EB) by the year 2030. At that moment, 5G will reach its limits and 6G will take its place to satisfy user requirements [112, 228].

Therefore, 6G or Beyond 5G (B5G) or 5G+ will meet several stricter requirements [27, 112, 228–232] such as:

- Deliver an individual data rate 100 × higher than 5G and downlink data rate more than 1 Tbps.
- Provide volumetric spectral and energy-efficiency 100 × higher than 5G and support battery-free IoT devices.
- Provide energy, bandwidth-efficient and computing capabilities of devices much better than those of 5G.
- Provide U-plane latency less than 0.1 ms and C-plane latency less than 1 ms.
- Mobility that arrives up to 1000 km/h.
- Explore frequency bands beyond sub-6 GHz.
- Operating frequency arrives up to 1 THz.
- Provide 10 ns of Radio-Only Delay Requirements and processing delay.
- Provide 10^{-9} frame error rate of reliability.
- Provide 1 cm on 3D of localization precision compared by 5G which provides 10 cm on 2D.

Generally, new services are evolved based on 6G requirements [230, 231, 233]. Besides, they can not be provided by 5G development like as:

- Mobile broadband reliable low latency communication (MBRLLC).
- Massive URLLC.
- Human-centric services (HCS).
- Multi-purpose 3CLS and energy services (MPS).
- Holographic communications that use multi-view cameras with a data rate of the order of Tbps.
- High-precision manufacturing with very low delay jitter, in the order of 1 μsec.
- Sustainable development and smart environments which provide battery-free communications with efficiency in the order of 1 pJ/bit.

Building upon the 5G vision, 6G will use the same 5G technologies and integrates new other technologies, for example, Tiny cells deployment with a radius of a few tens of meters, large intelligent surfaces (LISs) based on holographic radio frequency (RF) and holographic MIMO, Edge AI that provides MPS to 6G users and forms 3D radio environment maps, Terrestrial, Airborne, and Satellite networks, the Visible Light Communications (VLC) to enhance data rate, FD to enhance the spectral-efficiency, wireless optical communications to minimize backhaul congestion issue, ML-based schemes for a cyber-security solution to enhance 6G security.

As 5G, these technologies help to serve some 6G applications, for instance, haptic communication for virtual and augmented reality, massive IoT integrated smart city, automation and manufacturing, M2M, industry 4.0, industrial avatars, vehicle-to-everything (V2X), connected eXtended reality (XR) services: augmented, virtual and mixed reality ecosystem (AR/MR/VR), holographic telepresence, connected robotics and autonomous systems (CRAS), wireless brain–computer interactions (WBCI), blockchain and distributed ledger technologies (DLT) [112, 228, 232, 233].

- AI and ML concepts applied in 6G: To manage intelligently the high quantity of data, 6G will rely on AI, especially ML, in the performance optimization including data-rate, reliability, heterogeneous networks and in IoT and M2M communications, besides, it will address the access congestion problem in IoT. To reach optimality, in the physical layer, 6G aims to enhance the optimization in the E2E communication system. It uses ML to automate the traditional physical layer and make it more intelligent to learn and optimize E2E communication. Moreover, 6G can use RL and Deep Reinforcement Learning (DRL) to ameliorate the action of the decision maker which based on the physical system feedback. DRL will be generally used to cope with 6G challenges such as adaptive modulation, wireless caching and data offloading. Besides, 6G is expected to rely on intelligence more than the softwarization. Furthermore, intelligence based AI and ML services will be enabled by providing a seamless platform that includes several advanced technologies. [112, 234, 235].

9 Conclusion

In this paper, we have provided a comprehensive survey involving ML techniques applied to 5G cellular networks. To do this end, we have described the major components of the 5G systems (i.e., architectures, key enabler technologies, associated challenges, etc.). After that, we have reviewed the basic of machine learning, including major categories and representative algorithms. Furthermore, we have applieda variety of successful applications and use scenarios relyin on machine learning approaches in 5G network systems. We have highlighted how ML techniques bring fruits for building intelligent future networks. In this respect, we have provided useful references for researchers who are interested in applications of ML in future 5G networks. Moreover, we have concluded by poinpointing the conditions of ML applications in terms of computational concerns. Also, we have presented number of issues that need further investigations for the forthcoming networks. While many research directions have been studied to design autonomous and intelligent networks, several key problems remain open: How to reach a fully autonomous network? How to best explore the trade-offs between cost/complexity and efficiency?

References

1. Monserrat JF, Mange G, Braun V, Tullberg H, Zimmermann G, Bulakci Ö (2015) Metis research advances towards the 5G mobile and wireless system definition. EURASIP J Wirel Commun Netw 2015(1):53
2. Al-Falahy N, Alani OY (2017) Technologies for 5G networks: challenges and opportunities. IT Prof 19(1):12–20
3. Onoe S (2016) 1.3 evolution of 5G mobile technology toward 1 2020 and beyond. In: 2016 IEEE international solid-state circuits conference (ISSCC). IEEE, pp 23–28
4. Valente KP, Imran MA, Onireti O, Souza RD (2017) A survey of machine learning techniques applied to self organizing cellular networks. IEEE Commun Surv Tutor 19:2392–2431
5. Intelligence G (2014) Understanding 5G: perspectives on future technological advancements in mobile. White paper, pp 1–26
6. Agiwal M, Roy A, Saxena N (2016) Next generation 5G wireless networks: a comprehensive survey. IEEE Commun Surv Tutor 18(3):1617–1655
7. Talwar S, Choudhury D, Dimou K, Aryafar E, Bangerter B, Stewart K (2014) Enabling technologies and architectures for 5G wireless. In: 2014 IEEE MTT-S international microwave symposium (IMS2014). IEEE, pp 1–4
8. Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong AC, Zhang JC (2014) What will 5G be? IEEE J Sel Areas Commun 32(6):1065–1082
9. Osseiran A, Boccardi F, Braun V, Kusume K, Marsch P, Maternia M, Queseth O, Schellmann M, Schotten H, Taoka H et al (2014) Scenarios for 5G mobile and wireless communications: the vision of the metis project. IEEE Commun Mag 52(5):26–35
10. Li QC, Niu H, Papathanassiou AT, Wu G (2014) 5G network capacity: key elements and technologies. IEEE Veh Technol Mag 9(1):71–78
11. Hossain E, Hasan M (2015) 5G cellular: key enabling technologies and research challenges. IEEE Instrum Meas Mag 18(3):11–21
12. Papadopoulos H, Wang C, Bursalioglu O, Hou X, Kishiyama Y (2016) Massive mimo technologies and challenges towards 5G. IEICE Trans Commun 99(3):602–621
13. Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive mimo for next generation wireless systems. IEEE Commun Mag 52(2):186–195
14. Zhang Y, Yu R, Nekovee M, Liu Y, Xie S, Gjessing S (2012) Cognitive machine-to-machine communications: visions and potentials for the smart grid. IEEE Netw 26(3):6–13
15. Goudar SI, Hassan S, Habbal A (2017) 5G: The next wave of digital society challenges and current trends, Journal of Telecommunication. Electron Comput Eng (JTEC) 9(1–2):63–66

16. Alnoman A, Anpalagan A (2017) Towards the fulfillment of 5G network requirements: technologies and challenges. Telecommun Syst 65(1):101–116
17. Wu G, Yang C, Li S, Li GY (2015) Recent advances in energy-efficient networks and their application in 5G systems. IEEE Wirel Commun 22(2):145–151
18. Mell P, Grance T et al (2011) The nist definition of cloud computing
19. Zhang Q, Cheng L, Boutaba R (2010) Cloud computing: state-of-the-art and research challenges. J Internet Serv Appl 1(1):7–18
20. Nguyen V-G, Brunstrom A, Grinnemo K-J, Taheri J (2017) Sdn/nfv-based mobile packet core network architectures: a survey. IEEE Commun Surv Tutor 19(3):1567–1602
21. Abdelwahab S, Hamdaoui B, Guizani M, Znati T (2016) Network function virtualization in 5G. IEEE Commun Mag 54(4):84–91
22. Rangan S, Rappaport TS, Erkip E (2014) Millimeter wave cellular wireless networks: potentials and challenges. arXiv preprint arXiv:1401.2560
23. Ma Z, Zhang Z, Ding Z, Fan P, Li H (2015) Key techniques for 5G wireless communications: network architecture, physical layer, and mac layer perspectives. Sci China Inf Sci 58(4):1–20
24. Zheng G (2015) Joint beamforming optimization and power control for full-duplex mimo two-way relay channel. IEEE Trans Signal Process 63(3):555–566
25. Mao Y, You C, Zhang J, Huang K, Letaief KB (2017) A survey on mobile edge computing: the communication perspective. IEEE Commun Surv Tutor 19(4):2322–2358
26. Wang S, Zhang X, Zhang Y, Wang L, Yang J, Wang W (2017) A survey on mobile edge networks: convergence of computing, caching and communications. IEEE Access 5:6757–6779
27. Letaief KB, Chen W, Shi Y, Zhang J, Zhang Y-JA (2019) The roadmap to 6g-AI empowered wireless networks. arXiv preprint arXiv:1904.11686
28. Arfaoui G, Vilchez JMS, Wary J-P (2017) Security and resilience in 5G: current challenges and future directions. In: 2017 IEEE Trustcom/BigDataSE/ICESS. IEEE, pp 1010–1015
29. Klautau A, Batista P, Prelcic N, Wang Y, Heath R (2016) 5G mimo data for machine learning: application to beam-selection using deep learning. In: 2018 proceedings of information theory and applications workshop (ITA), pp 1–9
30. Kafle VP, Fukushima Y, Martinez-Julia P, Miyazawa T (2018) Consideration on automation of 5G network slicing with machine learning. In: 2018 ITU Kaleidoscope: machine learning for a 5G future (ITU K). IEEE, pp 1–8
31. International Telecommunication Union (ITU) (2017) Focus Group on Machine Learning for Future Networks including 5G (FG-ML5G). https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx. Accessed Nov 2017
32. 5G Public Private Partnership (5G-PPP) (2015) CogNet-building an intelligent system of insights and action for 5G network management. http://www.cognet.5g-ppp.eu/. Accessed 1 July 2015
33. Wang X, Li X, Leung VC (2015) Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges. IEEE Access 3:1379–1391
34. Kibria MG, Nguyen K, Villardi GP, Zhao O, Ishizu K, Kojima F (2018) Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks. IEEE Access 6:32328–32338
35. Long F, Li N, Wang Y (2017) Autonomic mobile networks: the use of artificial intelligence in wireless communications. In: 2017 2nd international conference on advanced robotics and mechatronics (ICARM). IEEE, pp 582–586
36. Moysen J, Giupponi L (2018) From 4G to 5G: self-organized network management meets machine learning. Comput Commun 129:248–268
37. Pérez-Romero J, Sánchez-González J, Sallent O, Agustí R (2016) On learning and exploiting time domain traffic patterns in cellular radio access networks. In: International conference on machine learning and data mining in pattern recognition. Springer, pp 501–515
38. Alsharif MH, Nordin R (2017) Evolution towards fifth generation (5G) wireless networks: current trends and challenges in the deployment of millimetre wave, massive mimo, and small cells. Telecommun Syst 64(4):617–637
39. Li S, Da Xu L, Zhao S (2018) 5G internet of things: a survey. J Ind Inf Integr 10:1–9
40. Kazi BU, Wainer GA (2019) Next generation wireless cellular networks: ultra-dense multi-tier and multi-cell cooperation perspective. Wirel Netw 25(4):2041–2064
41. Gupta A, Jha RK (2015) A survey of 5G network: architecture and emerging technologies. IEEE Access 3:1206–1232
42. Marsch P, Da Silva I, Bulakci O, Tesanovic M, El Ayoubi SE, Rosowski T, Kaloxylos A, Boldi M (2016) 5G radio access network architecture: design guidelines and key considerations. IEEE Commun Mag 54(11):24–32
43. Elijah O, Leow CY, Rahman TA, Nunoo S, Iliya SZ (2016) A comprehensive survey of pilot contamination in massive mimo-5G system. IEEE Commun Surv Tutor 18(2):905–923
44. Ahmed I, Khammari H, Shahid A, Musa A, Kim KS, De Poorter E, Moerman I (2018) A survey on hybrid beamforming techniques in 5G: architecture and system model perspectives. IEEE Commun Surv Tutor 20(3060):3097
45. Chin WH, Fan Z, Haines R (2014) Emerging technologies and research challenges for 5G wireless networks. IEEE Wirel Commun 21(2):106–112
46. Zhang C, Patras P, Haddadi H (2019) Deep learning in mobile and wireless networking: a survey. IEEE Commun Surv Tutor 21:2224–2287
47. Zhu G, Zan J, Yang Y, Qi X (2019) A supervised learning based qos assurance architecture for 5G networks. IEEE Access 7:43598–43606
48. Jiang C, Zhang H, Ren Y, Han Z, Chen K-C, Hanzo L (2017) Machine learning paradigms for next-generation wireless networks. IEEE Wirel Commun 24(2):98–105
49. Javaid N, Sher A, Nasir H, Guizani N (2018) Intelligence in iot-based 5G networks: opportunities and challenges. IEEE Commun Mag 56(10):94–100
50. Li R, Zhao Z, Zhou X, Ding G, Chen Y, Wang Z, Zhang H (2017) Intelligent 5G: when cellular networks meet artificial intelligence. IEEE Wirel Commun 24(5):175–183
51. Zhang J (2016) The interdisciplinary research of big data and wireless channel: a cluster-nuclei based channel model. China Commun 13(Supplement 2):14–26
52. Bogale TE, Wang X, Le LB (2018) Machine intelligence techniques for next-generation context-aware wireless networks. arXiv preprint arXiv:1801.04223
53. Latif S, Qadir J, Farooq S, Imran MA (2017) How 5G wireless (and concomitant technologies) will revolutionize healthcare? Future Internet 9(4):93
54. Qadir J, Yau K-LA, Imran MA, Ni Q, Vasilakos AV (2015) Ieee access special section editorial: Artificial intelligence enabled networking. IEEE Access 3:3079–3082
55. Chen M, Challita U, Saad W, Yin C, Debbah M (2017) Machine learning for wireless networks with artificial intelligence: a tutorial on neural networks. arXiv preprint arXiv:1710.02913
56. Salahuddin MA, Al-Fuqaha A, Guizani M (2016) Reinforcement learning for resource provisioning in the vehicular cloud. IEEE Wirel Commun 23(4):128–135
57. Latif S, Pervez F, Usama M, Qadir J (2017) Artificial intelligence as an enabler for cognitive self-organizing future networks. arXiv preprint arXiv:1702.02823

58. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. Ann Data Sci 4(2):149–178
59. Fadlullah Z, Tang F, Mao B, Kato N, Akashi O, Inoue T, Mizutani K (2017) State-of-the-art deep learning: evolving machine intelligence toward tomorrow's intelligent network traffic control systems. IEEE Commun Surv Tutor 19(4):2432–2455
60. Bui N, Cesana M, Hosseini SA, Liao Q, Malanchini I, Widmer J (2017) A survey of anticipatory mobile networking: context-based classification, prediction methodologies, and optimization techniques. IEEE Commun Surv Tutor 19(3):1790–1821
61. Le NT, Hossain MA, Islam A, Kim D-Y, Choi Y-J, Jang YM (2016) Survey of promising technologies for 5G networks. Mob Inf Syst 2016:2676589
62. Boutaba R, Salahuddin MA, Limam N, Ayoubi S, Shahriar N, Estrada-Solano F, Caicedo OM (2018) A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. J Internet Serv Appl 9(1):16
63. Sultan K, Ali H, Zhang Z (2018) Big data perspective and challenges in next generation networks. Future Internet 10(7):56
64. Xie J, Song Z, Li Y, Zhang Y, Yu H, Zhan J, Ma Z, Qiao Y, Zhang J, Guo J (2018) A survey on machine learning-based mobile big data analysis: challenges and applications. Wirel Commun Mob Comput 2018:8738613
65. Xie J, Song Z, Li Y, Ma Z (2018) Mobile big data analysis with machine learning. arXiv preprint arXiv:1808.00803
66. Buda TS, Assem H, Xu L, Raz D, Margolin U, Rosensweig E, Lopez DR, Corici M-I, Smirnov M, Mullins R et al (2016) Can machine learning aid in delivering new use cases and scenarios in 5G? In: NOMS 2016–2016 IEEE/IFIP network operations and management symposium. IEEE, pp 1279–1284
67. Wang Y, Xu J, Jiang L (2014) Challenges of system-level simulations and performance evaluation for 5G wireless networks. IEEE Access 2:1553–1561
68. Yao M, Sohul M, Marojevic V, Reed JH (2019) Artificial intelligence defined 5G radio access networks. IEEE Commun Mag 57(3):14–20
69. Boccardi F, Heath RW, Lozano A, Marzetta TL, Popovski P (2014) Five disruptive technology directions for 5G. IEEE Commun Mag 52(2):74–80
70. Razavizadeh SM, Ahn M, Lee I (2014) Three-dimensional beamforming: a new enabling technology for 5G wireless networks. IEEE Signal Process Mag 31(6):94–101
71. Farhang-Boroujeny B, Moradi H (2016) Ofdm inspired waveforms for 5G. IEEE Commun Surv Tutor 18(4):2474–2492
72. Mitra RN, Agrawal DP (2015) 5G mobile technology: a survey. ICT Express 1(3):132–137
73. Wunder G, Jung P, Kasparick M, Wild T, Schaich F, Chen Y, Ten Brink S, Gaspar I, Michailow N, Festag A et al (2014) 5Gnow: non-orthogonal, asynchronous waveforms for future mobile applications. IEEE Commun Mag 52(2):97–105
74. Schaich F, Wild T (2014) Waveform contenders for 5G–OFDM vs. FBMC vs. UFMC. In: 2014 6th international symposium on communications, control and signal processing (ISCCSP). IEEE, pp 457–460
75. van der Neut N, Maharaj B, de Lange FH, Gonzalez G, Gregorio F, Cousseau J (2014) PAPR reduction in FBMC systems using a smart gradient-project active constellation extension method. In: 2014 21st international conference on telecommunications (ICT). IEEE, pp 134–139
76. Danneberg M, Datta R, Festag A, Fettweis G (2014) Experimental testbed for 5G cognitive radio access in 4G LTE cellular systems. In: 2014 IEEE 8th sensor array and multichannel signal processing workshop (SAM). IEEE, pp 321–324
77. Fettweis GP, Krondorf M, Bittner S (2009) GFDM-generalized frequency division multiplexing. In: VTC. Spring, pp 1–4
78. Mukherjee M, Shu L, Kumar V, Kumar P, Matam R (2015) Reduced out-of-band radiation-based filter optimization for UFMC systems in 5G. In: 2015 international wireless communications and mobile computing conference (IWCMC). IEEE, pp 1150–1155
79. Song L, Niyato D, Han Z, Hossain E (2015) Wireless device-to-device communications and networks. Cambridge University Press, Cambridge
80. Mumtaz S, Huq KMS, Rodriguez J (2014) Direct mobile-to-mobile communication: paradigm for 5G. IEEE Wirel Commun 21(5):14–23
81. Fortuna C, Mohorcic M (2009) Trends in the development of communication networks: cognitive networks. Comput Netw 53(9):1354–1376
82. Aliu OG, Imran A, Imran MA, Evans B (2013) A survey of self organisation in future cellular networks. IEEE Commun Surv Tutor 15(1):336–361
83. Zhang N, Cheng N, Gamage AT, Zhang K, Mark JW, Shen X (2015) Cloud assisted hetnets toward 5G wireless networks. IEEE Commun Mag 53(6):59–65
84. Liu Y, She X, Chen P, Zhu J, Yang F (2015) Easy network: the way to go for 5G. China Commun 12(Supplement):113–120
85. Feng Z, Qiu C, Feng Z, Wei Z, Li W, Zhang P (2015) An effective approach to 5G: wireless network virtualization. IEEE Commun Mag 53(12):53–59
86. Chowdhury NMK, Boutaba R (2009) Network virtualization: state of the art and research challenges. IEEE Commun Mag 47(7):20–26
87. Han B, Gopalakrishnan V, Ji L, Lee S (2015) Network function virtualization: challenges and opportunities for innovations. IEEE Commun Mag 53(2):90–97
88. Hawilo H, Shami A, Mirahmadi M, Asal R (2014) Nfv: state of the art, challenges and implementation in next generation mobile networks (vepc). arXiv preprint arXiv:1409.4149
89. Open Networking Foundation (ONF) (2014) OpenFlow-enabled SDN and Network Functions Virtualization. https://www.opennetworking.org/wp-content/uploads/2013/05/sb-sdn-nvf-solution.pdf. Accessed 17 Feb 2014
90. Yousaf FZ, Bredel M, Schaller S, Schneider F (2017) Nfv and sdn-key technology enablers for 5G networks. IEEE J Sel Areas Commun 35(11):2468–2478
91. Rangisetti AK, Tamma BR (2017) Software defined wireless networks: a survey of issues and solutions. Wirel Pers Commun 97(4):6019–6053
92. Goyal S, Liu P, Panwar SS, Difazio RA, Yang R, Bala E et al (2015) Full duplex cellular systems: will doubling interference prevent doubling capacity? IEEE Commun Mag 53(5):121–127
93. Hong S, Brand J, Choi JI, Jain M, Mehlman J, Katti S, Levis P (2014) Applications of self-interference cancellation in 5G and beyond. IEEE Commun Mag 52(2):114–121
94. Hu YC, Patel M, Sabella D, Sprecher N, Young V (2015) Mobile edge computing—a key technology towards 5G. ETSI White Pap 11(11):1–16
95. Xiao L, Wan X, Dai C, Du X, Chen X, Guizani M (2018) Security in mobile edge caching with reinforcement learning. arXiv preprint arXiv:1801.05915
96. Ordonez-Lucena J, Ameigeiras P, Lopez D, Ramos-Munoz JJ, Lorca J, Folgueira J (2017) Network slicing for 5G with sdn/nfv: concepts, architectures, and challenges. IEEE Commun Mag 55(5):80–87
97. Soleymani B, Zamani A, Rastegar SH, Shah-Mansouri V (2017) RAT selection based on association probability in 5G heterogeneous networks. In: IEEE symposium on communications and vehicular technology (SCVT), pp 1–6

98. Pérez JS, Jayaweera SK, Lane S (2017) Machine learning aided cognitive RAT selection for 5G heterogeneous networks. In: IEEE international black sea conference on communications and networking (BlackSeaCom), Istanbul, Turkey. IEEE, pp 1–5
99. Nadeem Q-U-A, Kammoun A, Alouini M-S (2018) Elevation beamforming with full dimension mimo architectures in 5G systems: a tutorial. arXiv preprint arXiv:1805.00225
100. Wei L, Hu RQ, Qian Y, Wu G (2014) Enable device-to-device communications underlaying cellular networks: challenges and research aspects. IEEE Commun Mag 52(6):90–96
101. Li Y, Wu T, Hui P, Jin D, Chen S (2014) Social-aware d2d communications: Qualitative insights and quantitative analysis. IEEE Commun Mag 52(6):150–158
102. Maimó LF, Clemente FJG, Pérez MG, Pérez GM (2017) On the performance of a deep learning-based anomaly detection system for 5G networks. In: 2017 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, pp 1–8
103. Bouras C, Kollia A, Papazois A (2017) SDN & NFV in 5G: advancements and challenges. In: 2017 20th conference on innovations in clouds, internet and networks (ICIN). IEEE, pp 107–111
104. Sun S, Gong L, Rong B, Lu K (2015) An intelligent sdn framework for 5G heterogeneous networks. IEEE Commun Mag 53(11):142–147
105. Chih-Lin I, Han S, Xu Z, Sun Q, Pan Z (2016) 5G: rethink mobile communications for 2020+. Philos Trans R Soc A Math Phys Eng Sci 374(2062):20140432
106. MacCartney GR, Zhang J, Nie S, Rappaport TS (2013) Path loss models for 5G millimeter wave propagation channels in urban microcells. In: 2013 IEEE global communications conference (GLOBECOM), pp 3948–3953
107. Shafi M, Molisch AF, Smith PJ, Haustein T, Zhu P, De Silva P, Tufvesson F, Benjebbour A, Wunder G (2017) 5G: a tutorial overview of standards, trials, challenges, deployment, and practice. IEEE J Sel Areas Commun 35(6):1201–1221
108. You X, Zhang C, Tan X, Jin S, Wu H (2019) Ai for 5G: research directions and paradigms. Sci China Inf Sci 62(2):21301
109. Shariatmadari H, Ratasuk R, Iraji S, Laya A, Taleb T, Jäntti R, Ghosh A (2015) Machine-type communications: current status and future perspectives toward 5G systems. IEEE Commun Mag 53(9):10–17
110. Tullberg H, Popovski P, Li Z, Uusitalo MA, Hoglund A, Bulakci O, Fallgren M, Monserrat JF (2016) The metis 5G system concept: meeting the 5G requirements. IEEE Commun Mag 54(12):132–139
111. Queseth O, Bulakci Ö, Spapis P, Bisson P, Marsch P, Arnold P, Rost P, Wang Q, Blom R, Salsano S, et al (2017) 5G ppp architecture working group: view on 5G architecture (version 2.0, December 2017)
112. Nawaz SJ, Sharma SK, Wyne S, Patwary MN, Asaduzzaman M (2019) Quantum machine learning for 6g communication networks: state-of-the-art and vision for the future. IEEE Access 7:46317–46350
113. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of things (iot): a vision, architectural elements, and future directions. Future Gener Comput Syst 29(7):1645–1660
114. Alliance N (2015) 5G white paper, Next generation mobile networks, white paper, pp 1–125
115. Perera C, Zaslavsky A, Christen P, Georgakopoulos D (2014) Context aware computing for the internet of things: a survey. IEEE Commun Surv Tutor 16(1):414–454
116. Fortino G, Guerrieri A, Russo W, Savaglio C (2014) Integration of agent-based and cloud computing for the smart objects-oriented IoT. In: Proceedings of the 2014 IEEE 18th international conference on computer supported cooperative work in design (CSCWD). IEEE, pp 493–498
117. Wang D, Chen D, Song B, Guizani N, Yu X, Du X (2018) From iot to 5G i-iot: the next generation iot-based intelligent algorithms and 5G technologies. IEEE Commun Mag 56(10):114–120
118. Ratasuk R, Prasad A, Li Z, Ghosh A, Uusitalo MA (2015) Recent advancements in M2M communications in 4G networks and evolution towards 5G. In: 2015 18th international conference on intelligence in next generation networks. IEEE, pp 52–57
119. Kumar N, Misra S, Rodrigues JJ, Obaidat MS (2015) Coalition games for spatio-temporal big data in internet of vehicles environment: a comparative analysis. IEEE Internet Things J 2(4):310–320
120. Lee JD, Caven B, Haake S, Brown TL (2001) Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway. Hum Factors 43(4):631–640
121. Oleshchuk V, Fensli R (2011) Remote patient monitoring within a future 5G infrastructure. Wirel Pers Commun 57(3):431–439
122. West DM (2016) How 5G technology enables the health internet of things. Brook Cent Technol Innov 3:1–20
123. Gungor VC, Sahin D, Kocak T, Ergut S, Buccella C, Cecati C, Hancke GP (2011) Smart grid technologies: communication technologies and standards. IEEE Trans Ind Inform 7(4):529–539
124. Jeong S, Jeong Y, Lee K, Lee S, Yoon B (2016) Technology-based new service idea generation for smart spaces: application of 5G mobile communication technology. Sustainability 8(11):1211
125. Rappaport TS, Sun S, Mayzus R, Zhao H, Azar Y, Wang K, Wong GN, Schulz JK, Samimi M, Gutierrez F (2013) Millimeter wave mobile communications for 5G cellular: it will work!. IEEE access 1:335–349
126. Simsek M, Aijaz A, Dohler M, Sachs J, Fettweis G (2016) 5G-enabled tactile internet. IEEE J Sel Areas Commun 34(3):460–473
127. Aijaz A, Dohler M, Aghvami AH, Friderikos V, Frodigh M (2016) Realizing the tactile internet: haptic communications over next generation 5G cellular networks. IEEE Wirel Commun 24(2):82–89
128. Popovski P (2014) Ultra-reliable communication in 5G wireless systems. In: 1st international conference on 5G for ubiquitous connectivity. IEEE, pp 146–151
129. Hossain E, Rasti M, Tabassum H, Abdelnasser A (2014) Evolution towards 5G multi-tier cellular wireless networks: an interference management perspective. arXiv preprint arXiv:1401.5530
130. Wang C-X, Haider F, Gao X, You X-H, Yang Y, Yuan D, Aggoune HM, Haas H, Fletcher S, Hepsaydir E (2014) Cellular architecture and key technologies for 5G wireless communication networks. IEEE Commun Mag 52(2):122–130
131. Zhang H, Jiang C, Beaulieu NC, Chu X, Wen X, Tao M (2014) Resource allocation in spectrum-sharing of dma femtocells with heterogeneous services. IEEE Trans Commun 62(7):2366–2377
132. Hao P, Yan X, Yu-Ngok R, Yuan Y (2016) Ultra dense network: challenges enabling technologies and new trends. China Commun 13(2):30–40
133. Ge X, Tu S, Mao G, Wang C-X, Han T (2015) 5G ultra-dense cellular networks. arXiv preprint arXiv:1512.03143
134. Grover J, Garimella RM (2019) Optimization in edge computing and small-cell networks. In: Edge computing. Springer, pp 17–31
135. Hong X, Wang J, Wang C-X, Shi J (2014) Cognitive radio in 5G: a perspective on energy-spectral efficiency trade-off. IEEE Commun Mag 52(7):46–53
136. ETSIV (2011) Machine-to-machine communications (m2m): functional architecture. Int Telecommun 102:690 (**Union, Geneva, Switzerland, Tech. Rep. TS**)

137. Mehmood Y, Haider N, Imran M, Timm-Giel A, Guizani M (2017) M2m communications in 5G: state-of-the-art architecture, recent advances, and research challenges. IEEE Commun Mag 55(9):194–201
138. Mehmood Y, Görg C, Muehleisen M, Timm-Giel A (2015) Mobile m2m communication architectures, upcoming challenges, applications, and future directions. EURASIP J Wirel Commun Netw 2015(1):250
139. Roh W, Seol J-Y, Park J, Lee B, Lee J, Kim Y, Cho J, Cheun K, Aryanfar F (2014) Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. IEEE Commun Mag 52(2):106–113
140. Zhang J, Ge X, Li Q, Guizani M, Zhang Y (2017) 5G millimeter-wave antenna array: design and challenges. IEEE Wirel Commun 24(2):106–112
141. Checko A, Christiansen HL, Yan Y, Scolari L, Kardaras G, Berger MS, Dittmann L (2014) Cloud ran for mobile networks—a technology overview. IEEE Commun Surv Tutor 17(1):405–426
142. Checko A, Christiansen HL, Yan Y, Scolari L, Kardaras G, Berger MS, Dittmann L (2015) Cloud ran for mobile networks—a technology overview. IEEE Commun Surv Tutor 17(1):405–426
143. Zhang H, Jiang C, Cheng J, Leung VC (2015) Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks. IEEE Wirel Commun 22(3):92–99
144. Rost P, Bernardos CJ, De Domenico A, Di Girolamo M, Lalam M, Maeder A, Sabella D, Wübben D (2014) Cloud technologies for flexible 5G radio access networks. IEEE Commun Mag 52(5):68–76
145. Pan C, Elkashlan M, Wang J, Yuan J, Hanzo L (2018) User-centric c-ran architecture for ultra-dense 5G networks: challenges and methodologies. IEEE Commun Mag 56(6):14–20
146. ETSI-European Telecommunications Standards Institute (2019) 5G; system architecture for the 5G System (5GS)(3GPP TS 23.501 version 15.5.0 Release 15). https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.05.00_60/ts_123501v150500p.pdf. Accessed Apr 2019
147. The 5G Infraestructure Public Private Partnership (2019) 5G Americas White Paper The Status of Open Source for 5G. http://www.5gamericas.org/files 6915/5070/2509/5G_Americas_White_Paper_The_Status_of_Open_Source_for_5G_Feb_2018.pdf. Accessed Feb 2019
148. ETSI-European Telecommunications Standards Institute (2018) 5G; system architecture for the 5G system (3GPP TS 23.501 version 15.2.0 Release 15). https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf. Accessed June 2018
149. Wu D, Wang J, Cai Y, Guizani M (2015) Millimeter-wave multimedia communications: challenges, methodology, and applications. IEEE Commun Mag 53(1):232–238
150. Comsa I-S, De-Domenico A, Ktenas D (2017) Qos-driven scheduling in 5G radio access networks-a reinforcement learning approach. In: GLOBECOM 2017-2017 IEEE global communications conference. IEEE, pp 1–7
151. Comşa I-S, Zhang S, Aydin M, Kuonen P, Lu Y, Trestian R, Ghinea G (2018) Towards 5G: a reinforcement learning-based scheduling solution for data traffic management. IEEE Trans Netw Serv Manag 15:1661–1675
152. Kamel A, Al-Fuqaha A, Guizani M (2014) Exploiting client-side collected measurements to perform QoS assessment of IaaS. IEEE Trans Mobile Comput 14(9):1876–1887
153. International Telecommunication Union (2017) Vocabulary for performance, quality of service and quality of experience . https://www.itu.int/rec/T-REC-P.10-201711-I. Accessed 13 Nov 2017
154. European Cooperation in Science and Technology, QoE definition, http://www.cost.eu
155. International Telecommunication Union, Reference guide to quality of experience assessment methodologies. https://www.itu.int/rec/T-REC-G.1011-201607-I/en
156. Chen Y, Wu K, Zhang Q (2014) From qos to qoe: a tutorial on video quality assessment. IEEE Commun Surv Tutori 17(2):1126–1165
157. Qiao J, Shen XS, Mark JW, Shen Q, He Y, Lei L (2015) Enabling device-to-device communications in millimeter-wave 5G cellular networks. IEEE Commun Mag 53(1):209–215
158. Pierucci L (2015) The quality of experience perspective toward 5G technology. IEEE Wirel Commun 22(4):10–16
159. Petrangeli S, Wu T, Wauters T, Huysegems R, Bostoen T, De Turck F (2017) A machine learning-based framework for preventing video freezes in http adaptive streaming. J Netw Comput Appl 94:78–92
160. Imran A, Zoha A, Abu-Dayya A (2014) Challenges in 5G: how to empower son with big data for enabling 5G. IEEE Netw 28(6):27–33
161. Wu J, Zhang Y, Zukerman M, Yung EK-N (2015) Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. IEEE Commun Surv Tutor 17(2):803–826
162. Jaber M, Imran MA, Tafazolli R, Tukmanov A (2017) Energy-efficient SON-based user-centric backhaul scheme. In: 2017 IEEE wireless communications and networking conference workshops (WCNCW). IEEE, pp 1–6
163. Jaber M, Imran MA, Tafazolli R, Tukmanov A (2016) A distributed son-based user-centric backhaul provisioning scheme. IEEE Access 4:2314–2330
164. Jaber M, Imran MA, Tafazolli R, Tukmanov A (2016) A multiple attribute user-centric backhaul provisioning scheme using distributed SON. In: 2016 IEEE global communications conference (GLOBECOM). IEEE, pp 1–6
165. Morocho-Cayamcela ME, Lee H, Lim W (2019) Machine learning for 5G/b5G mobile and wireless communications: potential, limitations, and future directions. IEEE Access 7:137184–137206
166. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2(1–3):37–52
167. Comon P (1994) Independent component analysis, a new concept? Signal Process 36(3):287–314
168. Yuan Y, Wan J, Wang Q (2016) Congested scene classification via efficient unsupervised feature learning and density estimation. Pattern Recognit 56:159–169
169. Amiri R, Mehrpouyan H, Fridman L, Mallik RK, Nallanathan A, Matolak D (2018) A machine learning approach for power allocation in hetnets considering qos. arXiv preprint arXiv:1803.06760
170. Van Hasselt H, Guez A, Silver D (2016) Deep reinforcement learning with double q-learning. In: AAAI, vol 2. Phoenix, AZ
171. Wang S, Chaovalitwongse W, Babuska R (2012) Machine learning algorithms in bipedal robot control. IEEE Tran Syst Man Cybern Part C (Appl Revs) 42(5):728–743
172. Baştuğ E, Bennis M, Debbah M (2015) A transfer learning approach for cache-enabled wireless networks. In: 2015 13th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt). IEEE, pp 161–166
173. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436
174. Muthuramalingam S, Thangavel M, Sridhar S (2016) A review on digital sphere threats and vulnerabilities. In: Combating security breaches and criminal activity in the digital sphere. IGI Global, pp 1–21

175. Mohr W (2015) The 5G infrastructure public-private partnership. In: Presentation in ITU GSC-19 meeting
176. Li J, Zhao Z, Li R (2017) Machine learning-based ids for software-defined 5G network. IET Netw 7(2):53–60
177. Fiore U, Palmieri F, Castiglione A, De Santis A (2013) Network anomaly detection with the restricted Boltzmann machine. Neurocomputing 122:13–23
178. Maimó LF, Gómez ÁLP, Clemente FJG, Pérez MG, Pérez GM (2018) A self-adaptive deep learning-based system for anomaly detection in 5G networks. IEEE Access 6:7700–7712
179. Garcia S, Grill M, Stiborek J, Zunino A (2014) An empirical comparison of botnet detection methods. Comput Secur 45:100–123
180. Zago M, Sánchez VMR, Pérez MG, Pérez GM (2016) Tackling cyber threats with automatic decisions and reactions based on machine-learning techniques. In: Proceedings of the 2nd conference on network management, quality of service and security for 5G networks, Oulu, Finland, pp 1–4
181. Chang Z, Lei L, Zhou Z, Mao S, Ristaniemi T (2018) Learn to cache: machine learning for network edge caching in the big data era. IEEE Wirel Commun 25(3):28–35
182. Baldo N, Giupponi L, Mangues-Bafalluy J (2014) Big data empowered self organized networks. In: European wireless 2014; 20th European wireless conference. VDE, pp 1–8
183. Srinivasa S, Bhatnagar V (2012) Big data analytics: first international conference, BDA 2012, New Delhi, India, December 24-26, 2012, Proceedings, vol 7678. Springer Science & Business Media
184. Parwez MS, Rawat DB, Garuba M (2017) Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network. IEEE Trans Ind Inform 13(4):2058–2065
185. Aref MA, Jayaweera SK, Machuzak S (2017) Multi-agent reinforcement learning based cognitive anti-jamming. In: 2017 IEEE wireless communications and networking conference (WCNC). IEEE, pp 1–6
186. Mulvey D, Foh CH, Imran MA, Tafazolli R (2019) Cell fault management using machine learning techniques. IEEE Access 7:124514–124539
187. Kumar Y, Farooq H, Imran A (2017) Fault prediction and reliability analysis in a real cellular network. In: 2017 13th international wireless communications and mobile computing conference (IWCMC). IEEE, pp 1090–1095
188. Mfula H, Nurminen JK (2017) Adaptive root cause analysis for self-healing in 5G networks. In: 2017 international conference on high performance computing & simulation (HPCS). IEEE, pp 136–143
189. Mismar FB, Evans BL (2018) Deep Q-learning for self-organizing networks fault management and radio performance improvement. In: 2018 52nd asilomar conference on signals, systems, and computers. IEEE, pp 1457–1461
190. Alias M, Saxena N, Roy A (2016) Efficient cell outage detection in 5G hetnets using hidden markov model. IEEE Commun Lett 20(3):562–565
191. Yu P, Zhou F, Zhang T, Li W, Feng L, Qiu X (2018) Self-organized cell outage detection architecture and approach for 5G H-CRAN. Wirel Commun Mob Comput 2018:6201386
192. Farooq H, Parwez MS, Imran A (2015) Continuous time Markov chain based reliability analysis for future cellular networks. In: 2015 IEEE global communications conference (GLOBECOM). IEEE, pp 1–6
193. Asheralieva A, Miyanaga Y (2016) Qos-oriented mode, spectrum, and power allocation for d2d communication underlaying lte-a network. IEEE Trans Veh Techno 65(12):9787–9800
194. Zhang L, Xiao M, Wu G, Alam M, Liang Y-C, Li S (2017) A survey of advanced techniques for spectrum sharing in 5G networks. IEEE Wirel Commun 24(5):44–51
195. Fan Z, Gu X, Nie S, Chen M (2017) D2D power control based on supervised and unsupervised learning. In: 2017 3rd IEEE international conference on computer and communications (ICCC). IEEE, pp 558–563
196. Rohwer JA, Abdallah CT, El-Osery A (2002) Power control algorithms in wireless communications. In: Digital wireless communications IV, vol 4740. International Society for Optics and Photonics, pp 151–159
197. Xu J, Gu X, Fan Z (2018) D2D power control based on hierarchical extreme learning machine. In: 2018 IEEE 29th annual international symposium on personal, indoor and mobile radio communications (PIMRC). IEEE, pp 1–7
198. Wang L-C, Cheng SH (2018) Data-driven resource management for ultra-dense small cells: an affinity propagation clustering approach. IEEE Trans Netw Sci Eng 6:267–279
199. Balevi E, Gitlin RD (2018) A clustering algorithm that maximizes throughput in 5G heterogeneous F-RAN networks. In: 2018 IEEE international conference on communications (ICC). IEEE, pp 1–6
200. Alqerm I, Shihada B (2018) Sophisticated online learning scheme for green resource allocation in 5G heterogeneous cloud radio access networks. IEEE Trans Mob Comput 17:2423–2437
201. AlQerm I, Shihada B (2017) Enhanced machine learning scheme for energy efficient resource allocation in 5G heterogeneous cloud radio access networks. In: IEEE symposium on personal, indoor and mobile radio communications (PIMRC), pp 1–7
202. Lin P-C, Casanova LFG, Fatty BK (2016) Data-driven handover optimization in next generation mobile communication networks. Mob Inf Syst 2016:2368427
203. Khunteta S, Chavva AKR (2017) Deep learning based link failure mitigation. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 806–811
204. Kanwal K (2017) Increased energy efficiency in lte networks through reduced early handover
205. Hou T, Feng G, Qin S, Jiang W (2018) Proactive content caching by exploiting transfer learning for mobile edge computing. Int J Commun Syst 31(11):e3706
206. Shen G, Pei L, Zhiwen P, Nan L, Xiaohu Y (2017) Machine learning based small cell cache strategy for ultra dense networks. In: 2017 9th international conference on wireless communications and signal processing (WCSP). IEEE, pp 1–6
207. Sadeghi A, Sheikholeslami F, Giannakis GB (2018) Optimal and scalable caching for 5G using reinforcement learning of space-time popularities. IEEE J Sel Top Signal Process 12(1):180–190
208. Zeydan E, Bastug E, Bennis M, Kader MA, Karatepe IA, Er AS, Debbah M (2016) Big data caching for networking: moving from cloud to edge. IEEE Commun Mag 54(9):36–42
209. LeCun Y (1998) The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/
210. Tang F, Fadlullah ZM, Mao B, Kato N (2018) An intelligent traffic load prediction-based adaptive channel assignment algorithm in sdn-iot: a deep learning approach. IEEE Internet Things J 5(6):5141–5154
211. Mohammadi M, Al-Fuqaha A, Sorour S, Guizani M (2018) Deep learning for iot big data and streaming analytics: a survey. IEEE Commun Surv Tutor 20(4):2923–2960
212. Chen M, Yang J, Zhou J, Hao Y, Zhang J, Youn C-H (2018) 5G-smart diabetes: toward personalized diabetes diagnosis with healthcare big data clouds. IEEE Commun Mag 56(4):16–23
213. Chen M, Yang J, Hao Y, Mao S, Hwang K (2017) A 5G cognitive system for healthcare. Big Data Cogn Comput 1(1):2
214. Kumar PM, Gandhi UD (2018) A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases. Comput Electr Eng 65:222–235

215. Saghezchi FB, Mantas G, Ribeiro J, Al-Rawi M, Mumtaz S, Rodriguez J (2017) Towards a secure network architecture for smart grids in 5G era. In: 13th international wireless communications and mobile computing conference (IWCMC). IEEE, pp 121–126
216. Miao Y, Jiang Y, Peng L, Hossain MS, Muhammad G (2018) Telesurgery robot based on 5G tactile internet. Mob Netw Appl 23(6):1645–1654
217. Paolini M, Fili S (2019) Ai and machine learning: Why now?
218. Benzaid C, Taleb T (2020) Ai-driven zero touch network and service management in 5G and beyond: challenges and research directions. IEEE Netw 34(2):186–194
219. Sun Y, Peng M, Zhou Y, Huang Y, Mao S (2019) Application of machine learning in wireless networks: key techniques and open issues. IEEE Commun Surv Tutor 21(4):3072–3108
220. Wang X, Gao L, Mao S (2016) Csi phase fingerprinting for indoor localization with a deep learning approach. IEEE Internet Things J 3(6):1113–1123
221. Wang J-B, Wang J, Wu Y, Wang J-Y, Zhu H, Lin M, Wang J (2018) A machine learning framework for resource allocation assisted by cloud computing. IEEE Netw 32(2):144–151
222. Le L-V, Sinh D, Lin B-SP, Tung L-P (2018) Applying big data, machine learning, and SDN/NFV to 5G traffic clustering, forecasting, and management. In: 2018 4th IEEE conference on network softwarization and workshops (NetSoft). IEEE, pp 168–176
223. de Vrieze C, Simic L, Mahonen P (2018) The importance of being earnest: performance of modulation classification for real RF signals. In: IEEE international symposium on dynamic spectrum access networks (DySPAN). IEEE, pp 1–5
224. Koumaras H, Tsolkas D, Gardikis G, Gomez PM, Frascolla V, Triantafyllopoulou D, Emmelmann M, Koumaras V, Osma MLG, Munaretto D et al (2018) 5GENESIS: the genesis of a flexible 5G facility. In: IEEE 23rd international workshop on computer aided modeling and design of communication links and networks (CAMAD). IEEE, pp 1–6
225. Kotz D, Henderson T (2005) Crawdad: a community resource for archiving wireless data at dartmouth. IEEE Pervasive Comput 4(4):12–14
226. Lee W, Kim M, Cho D-H (2018) Deep power control: transmit power control scheme based on convolutional neural network. IEEE Commun Lett 22(6):1276–1279
227. Ahmed KI, Tabassum H, Hossain E (2019) Deep learning for radio resource allocation in multi-cell networks. IEEE Net 33(6):188–195
228. Tariq F, Khandaker M, Wong K-K, Imran M, Bennis M, Debbah M (2019) A speculative study on 6g. arXiv preprint arXiv:1902.06700
229. Routray SK, Mohanty S (2019) Why 6G? Motivation and expectations of next-generation cellular networks. arXiv preprint arXiv:1903.04837
230. Strinati EC, Barbarossa S, Gonzalez-Jimenez JL, Kténas D, Cassiau N, Dehos C (2019) 6g: The next frontier. arXiv preprint arXiv:1901.03239
231. Saad W, Bennis M, Chen M (2019) A vision of 6g wireless systems: applications, trends, technologies, and open research problems. arXiv preprint arXiv:1902.10265
232. David K, Berndt H (2018) 6g vision and requirements: is there any need for beyond 5G? IEEE Veh Technol Mag 13(3):72–80
233. Li R (2018) Towards a new internet for the year 2030 and beyond
234. Zhang H, Ren Y, Chen K-C, Hanzo L et al (2019) Thirty years of machine learning: the road to pareto-optimal next-generation wireless networks. arXiv preprint arXiv:1902.01946
235. Luong NC, Hoang DT, Gong S, Niyato D, Wang P, Liang Y-C, Kim DI (2019) Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun Surv Tutor 21:3133–3174