

Article

# Machine Learning Techniques for Non-Terrestrial Networks

Romeo Giuliano <sup>\*,†</sup>  and Eros Innocenti <sup>†</sup> 

Department of Engineering Science, Guglielmo Marconi University, Via Plinio 44, 00198 Rome, Italy

\* Correspondence: r.giuliano@unimarconi.it

† These authors contributed equally to this work.

**Abstract:** Traditionally, non-terrestrial networks (NTNs) are used for a limited set of applications, such as TV broadcasting and communication support during disaster relief. Nevertheless, due to their technological improvements and integration in the 5G 3GPP standards, NTNs have been gaining importance in the last years and will provide further applications and services. 3GPP standardization is integrating low-Earth orbit (LEO) satellites, high-altitude platform stations (HAPSs) and unmanned aerial systems (UASs) as non-terrestrial elements (NTEs) in the NTNs within the terrestrial 5G standard. Considering the NTE characteristics (e.g., traffic congestion, processing capacity, oscillation, altitude, pitch), it is difficult to dynamically set the optimal connection based also on the required service to properly steer the antenna beam or to schedule the UE. To this aim, machine learning (ML) can be helpful. In this paper, we present novel services supported by the NTNs and their architectures for the integration in the terrestrial 5G 3GPP standards. Then, ML techniques are proposed for managing NTN connectivity as well as to improve service performance.

**Keywords:** non-terrestrial networks; machine learning; unmanned aerial vehicle; high-altitude platform station; low-Earth orbit satellite; 3GPP release 17; 3GPP release 18

## 1. Introduction

Non-terrestrial networks (NTNs) have been gaining importance in the last years due to their technological improvements and the integration in the 3GPP standards. According to [1], direct satellite-to-device connectivity will have more than 25 million subscribers by the end of 2023 due to the investments of several important companies, such as Apple, Globalstar, SpaceX/Starlink and T-Mobile, just to cite a few, to which will be added Inmarsat, Iridium and Samsung.

Traditionally NTNs are used for a limited set of applications, such as TV broadcasting and communication support during disaster relief. Moreover, the natural service of global communications easily implemented by NTNs has been used to a limited extent worldwide, usually replaced by terrestrial communications systems. Previously, NTN communications were considered just a connection between a terrestrial terminal (both fixed or mobile) with a satellite, usually in medium-Earth orbit (MEO) as for Iridium or in geostationary-Earth orbit (GEO). The link distance and the difficulties in providing indoor coverage have limited the applicability of the satellite and the NTN service diffusion.

Recent technological advances, such as chip-set integration, terrestrial and in-orbit equipment improvement and mobility management strategies as well as the cost reduction in the equipment and in launch, have favored NTN interest for the research and business community. In addition to MEO and GEO satellites, low-Earth orbit (LEO) satellites, high-altitude platform stations (HAPSs) and unmanned aerial vehicles (UAVs) are under investigation to be integrated in the NTNs. Then, novel applications and services related to *service continuity*, *throughput enhancement* and *ubiquitous connectivity* can now be provided, worldwide. Nevertheless, some intrinsic limitations of NTN elements remain due to their latency and indoor coverage.



**Citation:** Giuliano, R.; Innocenti, E. Machine Learning Techniques for Non-Terrestrial Networks. *Electronics* **2023**, *12*, 652. <https://doi.org/10.3390/electronics12030652>

Academic Editors: Nurul I. Sarkar and Juan-Carlos Cano

Received: 16 December 2022

Revised: 20 January 2023

Accepted: 24 January 2023

Published: 28 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

According to the authors, the following aspects are fundamental for the NTN diffusion and to exploit their potentials:

- The integration of terrestrial networks (TNs) and NTNs; deploying two (or more) separated networks will increase the service availability.
- The adoption of a unique terminal for both NTN and TN communications; people connect to TNs most, whilst they rarely connect to NTNs. In case it is necessary to switch between the terrestrial equipment and the NTN enabled equipment, this will discourage the NTN diffusion.
- Overcoming of the NTN limits in terms of throughput availability and delays. Even with narrow-beams, very high-altitude satellites (i.e., MEO and GEO) cover a large area limiting the available throughput for a single user. Moreover, large distances (from 8000 km to 36,000 km) causes inadequate user experience, not only for video and real-time applications, such as gaming, but also for voice service, experiencing latencies from about 200 ms up to 800 ms (including processing) [2].

These aspects are being overcome from one side by the adoption of the new radio (NR) radio interface, which envisages the integration of the NTN in the 3GPP standards, and from the other side, the use of dedicated non-terrestrial elements (NTEs), such as UAV, HAPS and LEO satellites on novel orbits, which are more suitable for the current near real-time applications.

Considering the characteristics of the NTEs, it is difficult to dynamically set the optimal connection, also taking into consideration the required service in terms for example of required latency and throughput. For example, the NTN system should anticipate the switch from one NTE to another according to the mobility of the user equipment (UE) or to properly steer the antenna beam. To this aim, machine learning (ML) can be helpful [3].

As will be detailed in the following sections, ML is a key enabler for NTN networks. In fact, traditional terrestrial 5G approaches could be inadequate for an architecture, which includes ground, air and space elements. Satellite communications have to tolerate long delays, while aerial networks suffer from limited capacity and link instability, and both of them have to deal with high mobility patterns. On the other hand, traditional terrestrial networks have higher throughput and reliability. ML comes into play with the aim to take the best from these technologies in order to efficiently complement each other, thus improving the overall quality at each altitude from the users' perspective. More specifically, ML can be applied from the physical layer, link and medium access control layers, network and eventually in the application layer. In general, ML proved to be a valid alternative to complex traditional algorithms, offering a superior outcome in terms of accuracy and response time. Moreover, in highly dynamic scenarios, self-adaptable models can evolve and therefore tackle unseen issues.

The authors' contributions introduced in this paper are reported in the following.

- Presentation of the innovative services that novel NTN may provide. They are provided in terms of the most suited NTE and required data rate as well as being linked with the 5G categories.
- Description of the NTN architecture envisaged in the 3GPP Release 17 and in the forthcoming Release 18. Involved NTEs and functionalities within 5G system are detailed in terms of architectures and protocol stacks over the 5G network. It is also proposed how multi-connectivity (both in hybrid terrestrial/non-terrestrial and non-terrestrial/non-terrestrial modes) and the deployment of multi-access edge computing (MEC) should be implemented in NTN.
- Adoption of the most suited ML techniques in order to guarantee the novel services in NTN by supporting their requirements. Proposed ML techniques are also provided in order to enhance the UE connectivity or to optimize the system performance in NTN. Furthermore, some insights are also reported in order to properly investigate the adoption of the ML techniques in selected situations typical of NTN.

This paper is organized as follows. In Section 2, a brief review of the literature is proposed in terms of communication provided by satellites or aerial elements as well as the adoption of ML to support the connectivity. In Section 3, it is reported the general aspects of a NTN network and the typical parameters of any NTE type. In Section 4, we describe the services and the network architectures of NTNs. Some details on 3GPP standardization are also given. In Section 5, the general aspect of ML is introduced, while in Section 6, ML techniques are reported both for NTN services and NTN performance improvements in each protocol stack layer. In Section 7, conclusions and open research issues are reported.

## 2. Related Works

Networks involving satellites and other aerial vehicles have been largely studied in the literature. See these reviews [4–6] and their references. Most of them focus on the adoption of satellites or aerial vehicles to support cellular communications or internet of things (IoT) connections in some use cases [7]. Usually they provide connectivity just assuming a transceiver is posed on the satellite, thus without considering any base station constraints (e.g., energy, weight and database). In [8,9], some use cases are proposed for IoT data collection or coverage extension but without considering any standard for communication, thus reducing the impact of the work. Most of them propose strategies for traffic optimization or balancing between terrestrial and satellite networks as in [10] or link performance analysis, such as in mmWave [11] or in optical links [12]. The authors in [13] provide performance analysis for IoT in rural areas using low power wide area (LPWA) technologies. Some testbeds are now available [14]. Interesting results related to aerial link characterization and antenna orientation for IoT sensor data collection are in [15], useful for any coverage dimensioning. Any access procedure was not considered for any proper service delivery. Other works, such as [16,17], investigate the optimization of the scheduling strategy or NTE relay with respect to energy consumption. Security aspects emerging in a space network subject to eavesdropping are analyzed in [18] in which appropriate secrecy performance metrics are also reported. Further NTN challenges, such as coexistence, terminal requirements and spectrum policies are analyzed in [19].

The authors in [20] report the preliminary efforts of 3GPP to integrate satellite within the 5G explaining only the flow segmentation between the NTE based on the role of the NTN infrastructure. In [21], the authors envisage possible evolution steps towards 6G. Differently from available works in literature, this paper reports the NTN architectures that can be integrated in the 3GPP and proposes new services with the related requirements that can be supported by the NTN integration in the NR. Scenarios and systems that do not consider any communication standard and any standardized wireless system architecture in service provisioning are out of the scope. Protocol stack split between NTEs is also investigated.

ML has been widely implemented in 5G and 6G networks. Supervised, unsupervised and reinforcement learning techniques are applicable to cellular networks in order to optimize aspects, which are hard to manage with traditional algorithmic approaches. In [22], an extensive review of advantages and disadvantages of popular ML approaches is presented. Refs. [23,24] also proposed an overview and a survey, respectively, on this study subject. The authors in [25] investigated the adoption of federated learning (FL) in LEO-based satellite communication networks in order to improve the performance and reduce overheads and latencies. However, in this work, only space elements are taken into consideration, while the airborne ones are excluded. Adding intelligence to IoT sensor networks, especially in remote areas, is also an issue under study. In [26], the authors present a review of ML technologies for enhanced NTNs, at every altitude, specifically for IoT scenarios. This paper draws attention to the prediction and decision-making procedures, which are essential to achieve self-adaptation, thus enhancing the considered performance metrics.

### 3. General Aspects of NTN

In this section, general aspects related to the NTN deployment are reported in terms of characteristics of the ground, aerial and space elements and their parameters. In Figure 1, it is reported the generic NTN architecture. The user terminal (i.e., the UE) is able to transparently connect to a terrestrial node as well as to one of the NTE, through the service link. The NTE is then connected to the 5G core network (5GC) through the feeder link to the ground station, namely NTN gateway. In the same figure, it is reported the possibility to have a multi-hop connectivity through the inter-satellite link (ISL) with one of the NTEs (i.e., any satellite but also to an UAV or a HAPS).

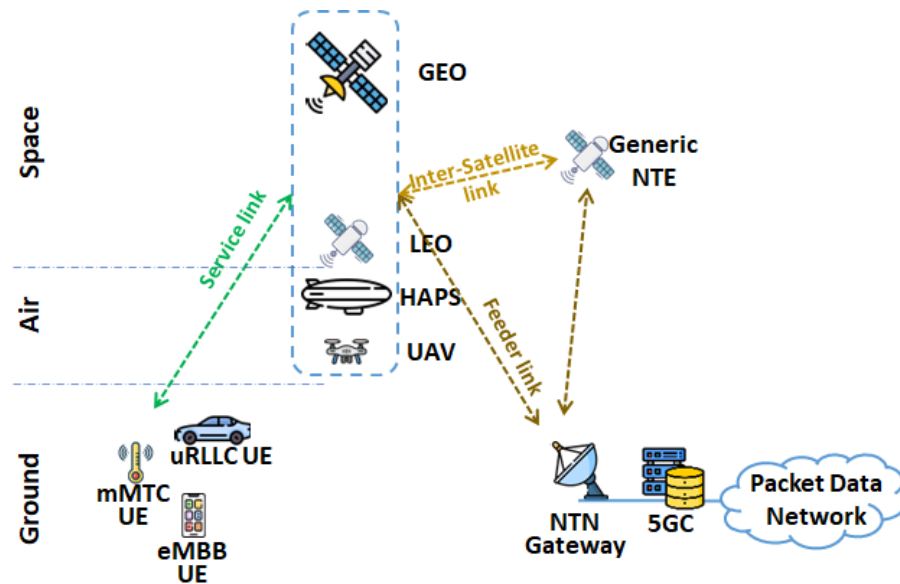


Figure 1. Reference architecture of the non-terrestrial network.

Depending on the NTE, the experienced radio parameters can be different. In Table 1, the most important parameters of NTN are reported. In particular, for each NTE, its altitude from the ground, the typical size of its beam footprint, the maximum distance  $D_{max}$  between the NTE and UE located at cell edge at its minimum elevation angle (typical 10 deg) and the maximum round trip time (RTT) delay with respect to the gNB are reported. In the RTT evaluation, only the propagation delay is considered. Depending of the location of the gNB, RTT can be double the service link propagation delay (regenerative case) or double the service and feeder link propagation delay (transparent case). These two cases are better defined in Section 4.

Finally, it is worth noting the difference in the propagation path loss between TN and NTN. In the case of TN, the main source of attenuation in radio signal propagation is due to the blockage of hills and buildings (of course the distances are reduced with respect to those in NTN). In the case of NTN, we can consider line-of-sight (LoS) propagation conditions for low-altitude NTEs (e.g., UAV and HAPS), while for high-altitude NTEs (e.g., MEO and GEO), the radio signal propagation is dominated by atmospheric path loss due to gases and rain fades. See [27,28] for details and parameters for the link budget.

**Table 1.** Parameters for NTN types and scenarios [2,27].

NTE	Altitude	Footprint Size	$D_{max}$	RTT
GEO	35,786 km	200–3500 km	40,581 km	541.46 ms (service and feeder links) 270.73 ms (service link only)
MEO	7000–25,000 km	100–1000 km	16,000 km ( $h_0 = 8000$ km)	213 ms (8000 km) (service and feeder links) 107 ms (8000 km) (service link only)
LEO	300–1500 km	100–1000 km	1932 km ( $h_0 = 600$ km) 3131 km ( $h_0 = 1200$ km)	25.77 ms (600 km), 41.77 ms (1200 km) (service and feeder links) 12.89 ms (600 km), 20.89 ms (1200 km) (service link only)
HAPS	8–50 km	5–200 km	115 km ( $h_0 = 20$ km)	15.53 ms (20 km) (service and feeder links) 0.77 ms (20 km) (service link only)
UAV	1–10 km	1–50 km	30 km ( $h_0 = 5$ km)	<67 $\mu$ s (service and feeder link), <33 $\mu$ s (service link)

#### 4. Non-Terrestrial Network Architectures

In this section, the services and the corresponding network architectures of NTNs are reported. Moreover, it is also analyzed the standardization process of NTN within the 3GPP organization.

##### 4.1. Services for NTN

Due to satellite and HAPS characteristics (e.g., latency, pathloss and Doppler shift), classical three use cases envisaged by 5G (i.e., enhanced multi-broad band, eMMB, massive machine type communications, mMTC, and ultra-reliable and low latency communications, uRLLCs) and defined in [29] are slightly modified in [30], providing revised and adjusted requirements for satellite use cases. They are reported in Table 2, where it is assumed a bandwidth of up to 30 MHz over one satellite beam.

**Table 2.** Requirements in IMT-2020 for satellite radio interface [29].

Requirement	Value
Peak data rate	70 Mbit/s (in downlink), 2 Mbit/s (in uplink)
User experienced data rate	1 Mbit/s (in downlink), 100 kbit/s (in uplink)
Area traffic capacity	8 kbit/s/km <sup>2</sup> (in downlink), 1.5 kbit/s/km <sup>2</sup> (in uplink)
Latency	10 ms (user plane), 40 ms (control plane)
Latency for SRI in high-altitude satellites	650 ms (user plane), 1150 ms (control plane)
Connection density	500 devices per km <sup>2</sup>
Reliability	0.999
Mobility	250 km/h (car), 500 km/h (train), 1200 km/h (airplane)
Mobility interruption time	50 ms

In the following, we present the most-suited services that an NTN system is able to provide based on its characteristics and grouped within the three novel areas: *service continuity*, *throughput enhancement* and *ubiquitous connectivity*. After their description, we also provide possible service requirements, which is the most suited NTE to provide it. Based on this grouping, telecommunication operators may properly implement changes or integration with terrestrial network elements in order to provide services in a specific vertical market.

**Moving cell connectivity.** Several platforms used for public/private transport are moving objects worldwide, including ships, airplanes, trains, and river boats. Connectivity service is not provided to their passengers or provided intermittently, such as in high-speed trains. The adoption of NTNs can overcome the coverage issue, thus providing the basic service of *ubiquitous connectivity*. Since the considered mobile platforms are located in open spaces and travel at considerable speed, LEO and HAPS are of greater interest for the connectivity service. It is considered to connect each passenger with a data rate of 50 Mbit/s.

**Unserved/Under-served and isolated areas.** There exist several areas in the world that have limited or no access to the internet. NTN can solve the *ubiquitous connectivity* but its applicability depends on the specific context in which the connectivity is provisioned. Three main cases or applications are considered. The first one is the connectivity of remote access points hard to be reached out by fiber optics. In this case, NTN acts as an eMMB backhauling to the locally deployed gNB, where UEs are easily connected with NR radio interface. UEs can experience eMMB shared services (e.g., 50 Mbit/s or more depending of the simultaneous active connections). The second case is when the NTN allows the connectivity of several sensors deployed in a wide area, such as remote and isolated areas (e.g., North Scandinavia, Siberia and long coasts). Sensors' throughput is in the order of tens of kbit/s. The third case is the monitoring of critical infrastructures (such as ports, bridges, airports and dams). This application assumes to have some sensors and actuators deployed in a remote but limited area that are able to locally communicate with a local access point (acting as a concentrator). The local links may exploit different wireless or wired technologies (e.g., ZigBee, Wi-Fi or other IEEE 802.15.x standards) according to the required data rates and sensors' constraints. The local access point connects with the NTN using NR. Differently from the first two cases, where HAPS and LEO may be the most suited NTEs, in the third case, also UAVs can be adopted, depending on the given criticisms of the monitored area.

**Throughput increase.** In some regions (e.g., suburban), the terrestrial network is available but the area served by a single gNB is wide, and thus the provided capacity is shared among all the active UEs. This limits the quality of experience (QoE) of the users above all in congested hours. The NTN can provide a second link to the users in the area, thus enabling a *throughput enhancement* in their connection. A stringent collaboration should be performed by TN and NTN in the dual connectivity for the data rate split of each user. The stream merging from the TN link and from the NTN link is managed by the UE in downlink and by the gNB acting as a master cell in uplink. Acknowledgments and retransmissions are managed separately on the two links. For this service, HAPS can be normally used, but in some cases, such as traffic jams, vehicle accidents, or open-air gathering events, the throughput enhancement can be provided by deploying a dedicated UAV over the interested area.

**Secondary link for backup.** In some cases, connectivity is used to provide services to lone workers, trains, and small ships that should be guaranteed in wide areas. For example, the high-speed trains as well as regional trains have the necessity to constantly communicate with their remote-control center for sending their updated position and receiving the authorization to move on the rail. Connectivity provided by TNs cannot be available in rural and mountain areas since the base station deployment of the telecommunication operator is limited to urban/suburban zones. In order to guarantee the *service continuity* and to optimize the rail infrastructure, NTN can cover the unserved areas through LEO sats or HAPSs. The NTN link is just as a backup complementing the terrestrial coverage and offering connectivity from few tens of kbit/s up to 1 Mbit/s per moving object.

**Disaster relief.** The adoption of the satellites has been largely investigated by the literature in the case that the TN has suffered significant damage from earthquakes, explosions, tsunamis or other natural disasters. NTN is able to fast provide connectivity (i.e., *throughput enhancement*) to rescue teams and civil protection up to 500 Mbit/s (or even more) in the affected area. Moreover, it can help missing people and survivors for their localization and limited communications, providing them *ubiquitous connectivity*. In the beginning, LEO satellites can be adopted in the affected area, but the radio capacity can be later improved by fast deploying UAVs and HAPSs. It is important to ensure that people can connect with their mobile phone rather than resetting communications only for law enforcement and military.

**Broadcasting/ Multicasting.** This service exploits the natural advantages of the satellites. High-definition (HD) and ultra-HD video channels can be broadcast or multicast to people subscribing the service. The average video data rate varies according to the image resolution, frame rate, coding and quality level, having, for example, a bit rate of about 7 Mbit/s for 4K (i.e.,  $3840 \times 2160$ ) using HEVC x.265 codec, 50 Hz, 8/10 quality level [31]. The *throughput enhancement* can be obtained above all in the case of limited terrestrial connectivity (e.g., fiber) or for live events. NTN can easily interconnect the final user with the content provider network, bringing the contents to the edge of the network.

Table 3 summarizes the NTN services, the most-suited NTE and related requirements in terms of data rates. We also consider possible 5G categories where they can fall: eMBB, mMTC, and uRLLC. Note that uRLLC in Table 3 should not be considered as the classical way in 5G services (i.e., requiring connectivity lower than 5 ms). For NTN, we consider only the reliability concept based on the provision of coverage and then connectivity in the (even wide) service area with limited requirements on latency.

**Table 3.** Service requirements for NTN networks.

NTN Service	5G Category	Suited NTEs	Required Data Rate
Mobility cell connectivity Unserved/Under-served and isolated areas:	eMMB	HAPS, LEO	50 Mbit/s/person
<ul style="list-style-type: none"> <li>Fixed cell connectivity</li> <li>Wide area IoT</li> <li>Sensor/ actuators in local remote areas</li> </ul>	eMMB mMTC mMTC	HAPS, LEO UAV, HAPS, LEO	50 Mbit/s/person 0.1–100 kbit/s/sensor 1–10 Mbit/s per local hot spot
Throughput increase	eMMB	UAV, (HAPS)	100–200 Mbit/s/hot spot
Secondary link for backup	uRLLC *	HAPS, LEO	0.01–1 Mbit/s
Disaster relief	eMMB, uRLLC *	UAV, HAPS	10–500 Mbit/s
Broadcasting/ Multicasting:			
<ul style="list-style-type: none"> <li>TV</li> <li>Alerting UEs for possible disaster warning</li> </ul>	eMMB mMTC	LEO, (HAPS) LEO, HAPS	e.g., video 4K 10 kbit/s

#### 4.2. NTN Architectures

In order to provide the services presented above, several NTN architectures are proposed and detailed in the following.

Depending on the position of the gNB functionalities, it is possible to consider three basic cases: the transparent architecture, the regenerative architecture and the split architecture. In the last case, the gNB functionalities are split in the distributed unit (DU) and in the central unit (CU). The DU is located on-board and the CU is left on the ground. Finally, for each of them, it is also reported their protocol stack architectures both for user plane (UP) and control plane (CP), allowing to know in which NTE each protocol function is closed.

##### 4.2.1. Transparent NTN Architecture

In the transparent architecture, gNB is located at the ground, thus after the NTN ground station (see Figure 2). The NTE is transparent to the signal sent by the UE, which is retransmitted to the NTN gateway without any processing on-board (except possible frequency translation). The gNB is connected to the 5GC and then the signal is sent to the external packet data network (PDN).

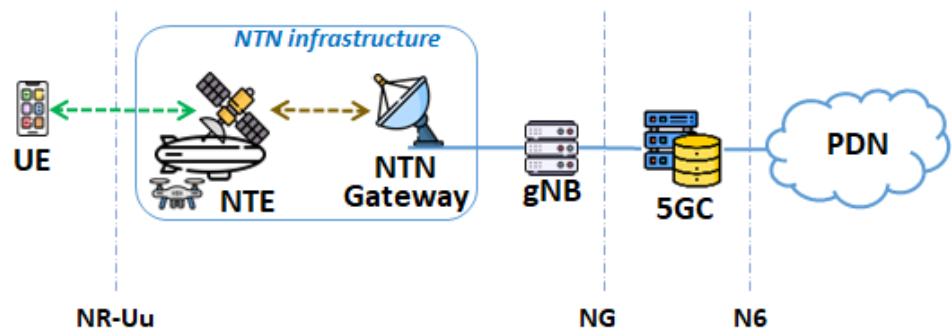


Figure 2. Transparent NTN architecture.

In Figures 3 and 4, the protocol stack of this NTN architecture for UP and CP, respectively, are reported. Since the NTE (and the NTN gateway) are transparent, protocol messages are forwarded to the terrestrial gNB (from the physical layer (PHY) to the service data adaptation protocol (SDAP)) and to the user plane function (UPF) (i.e., upper layers, such as IP, TCP and applications). In the CP, protocols from PHY to radio resource control (RRC) close to gNB. The non-access stratum (NAS), which is made up by the mobility management (MM) and the session management (SM) functionalities, closes to the access and mobility function (AMF) and session management function (SMF) in the 5GC, respectively. In the same figure, the CP protocol stack is also completed by other two functions/nodes of 5GC as the policy control function (PCF) and application function (AF). The AF enables the external service by interfacing with the NTN or in general with the 5G system through the N5 interface. Then, over-the-top services, such as those related to vehicular, health or industry, can be provided to users. Finally, note that between gNB and UPF in the UP, the generic tunneling protocol (GTP) over the user datagram protocol (UDP) is adopted to create a tunnel, while between the gNB and the AMF in the CP, the NG application protocol (NG-AP) over the stream control transport protocol (SCTP) is adopted for tunneling in the N2 interface.

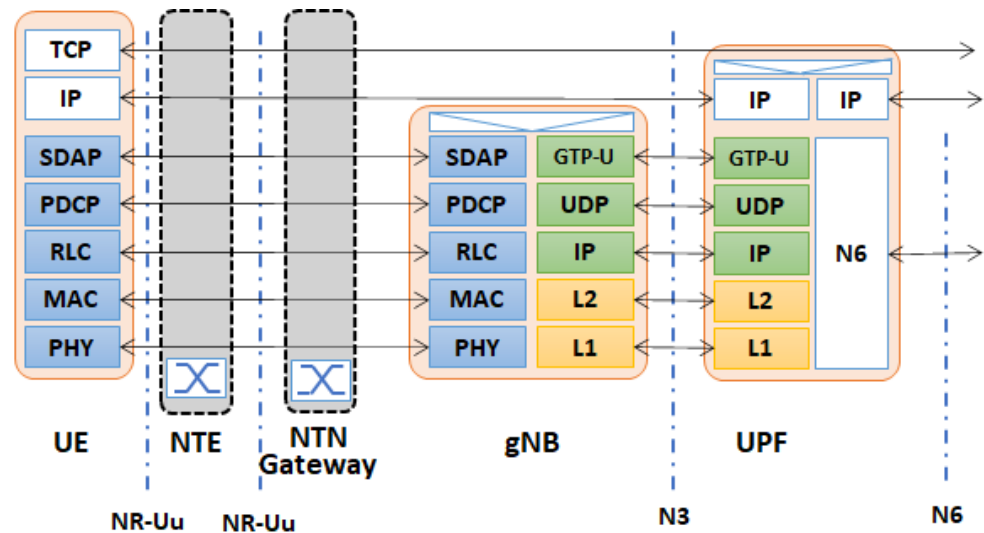


Figure 3. Protocol stack of the user plane in the transparent NTN architecture [27].



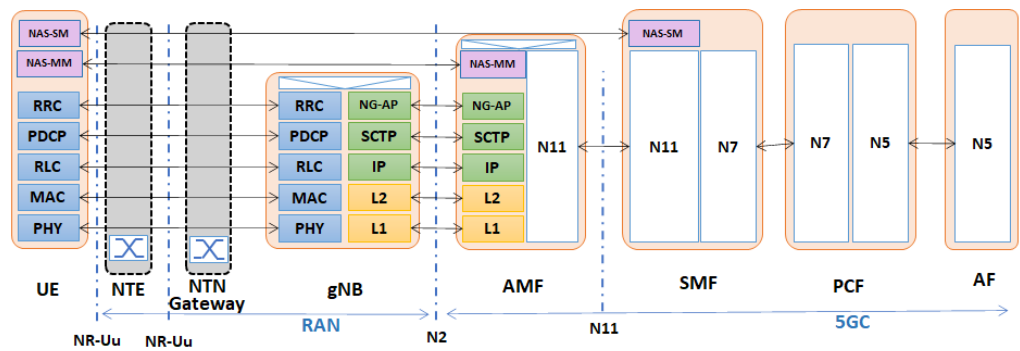


Figure 4. Protocol stack of the control plane in the transparent NTN architecture [27].

#### 4.2.2. Regenerative NTN Architecture

In Figure 5, the regenerative NTN architecture is reported. In this case, the gNB is mounted on-board to the NTE, thus improving the NTN performance. Differently from the transparent NTN case in Figure 2, where the satellite radio interface (SRI) in the feeder link is based on 5G-Uu, for regenerative NTN, the SRI is a transport link used to transmit both user and control data from the NTE to the NTN gateway on the ground.

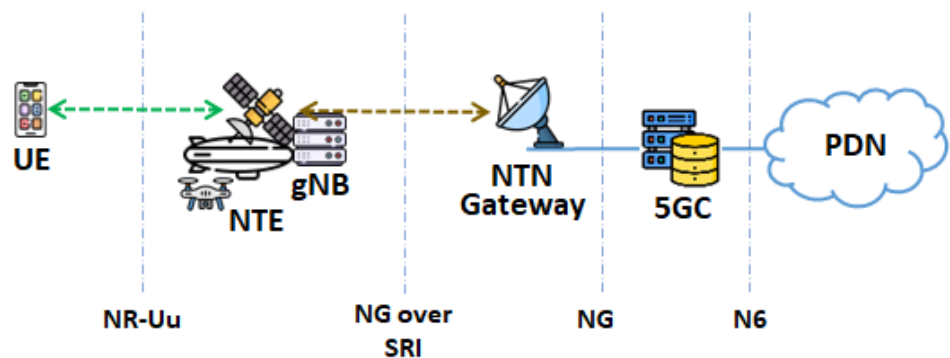


Figure 5. Regenerative NTN architecture.

The structure of the protocol stack for regenerative NTN is reported in Figures 6 and 7 for UP and CP, respectively. The NTE is not transparent, and the typical processing of the gNB is performed on-board, where the access stratum protocols close. The NTN gateway is transparent, and it does not affect the GTP tunnel from the gNB to UPF.

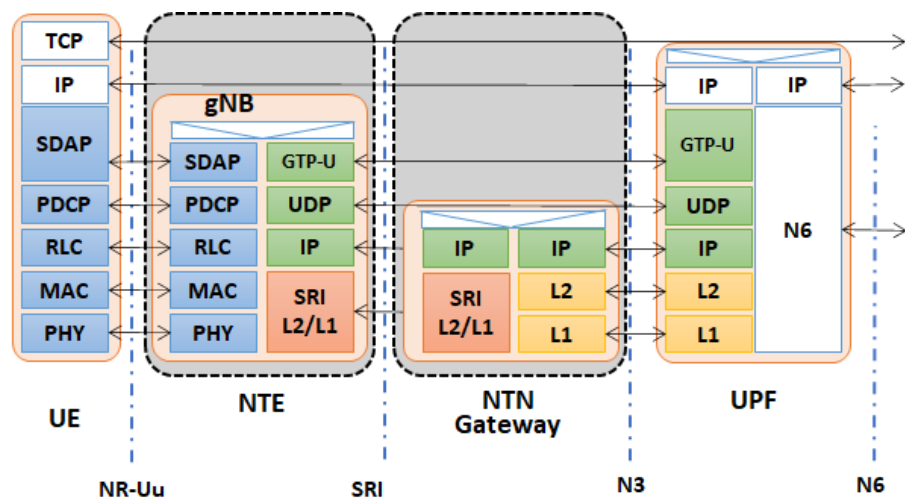


Figure 6. Protocol stack of the user plane in the regenerative NTN architecture [27].

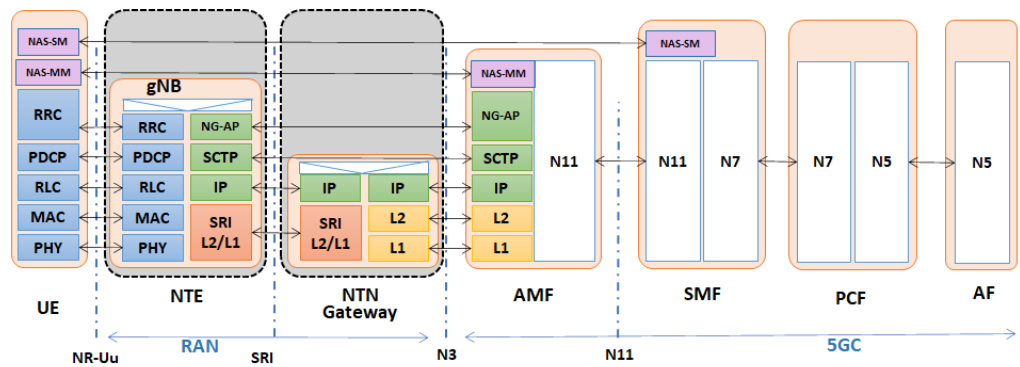


Figure 7. Protocol stack of the control plane in the regenerative NTN architecture [27].

#### 4.2.3. On-Board Distributed Unit NTN Architecture

In the last case, the NTN can consider the gNB functional split into distributed unit (DU) and central unit (CU) as reported in Figure 8. The DU is mounted on-board, while the CU is on the ground after the NTN gateway [32]. The feeder link is based on the F1 interface protocols as highlighted in the protocol stacks for UP in Figure 9 and for CP in Figure 10. Of course, this NTN configuration is a trade-off between the transparent case, where no processing is performed on-board and the regenerative case, where all protocols of NR-Uu close.

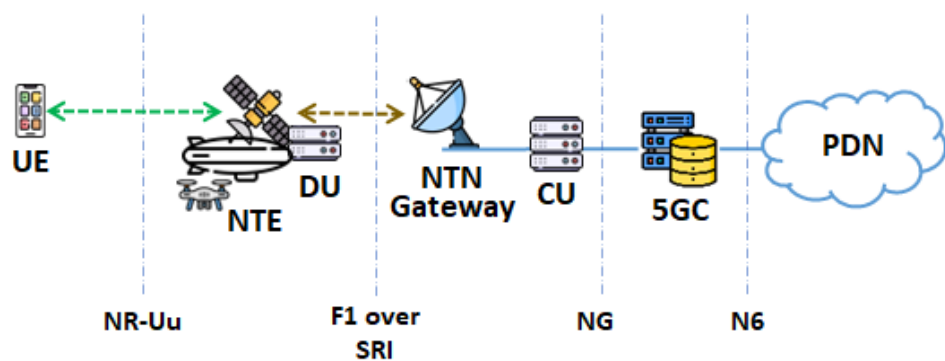


Figure 8. On-board distributed unit NTN architecture.

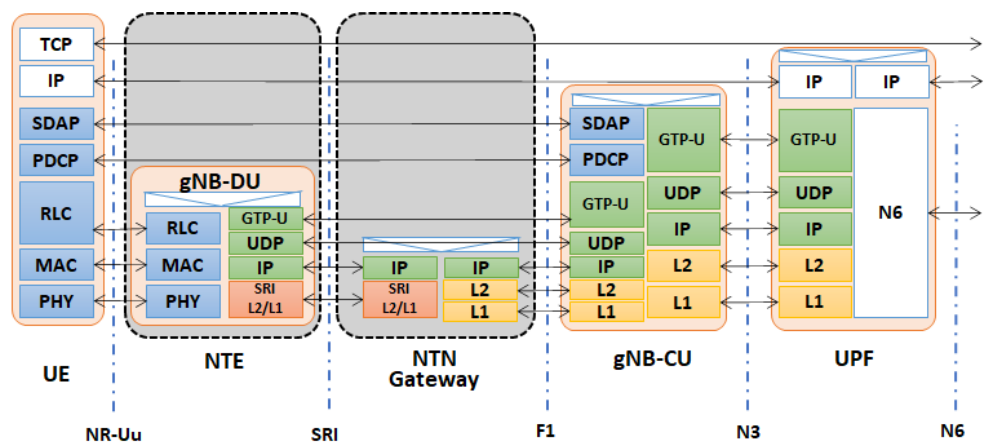
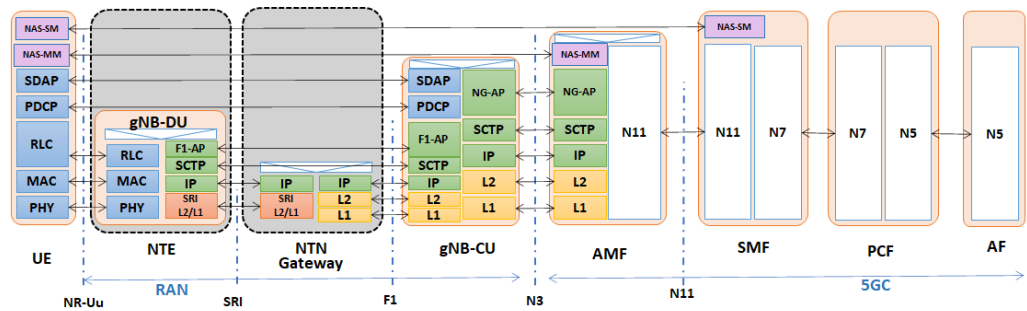


Figure 9. Protocol stack of the user plane in the NTN architecture with on-board distributed unit [27].



**Figure 10.** Protocol stack of the control plane in the NTN architecture with on-board distributed unit [27].

### 4.3. Impact of NR in NTN

NTN standardization in 3GPP started in 2017, and it is fundamental to gain global diffusion and for the adoption of a unique UE, both for connecting TN and NTN simultaneously or depending on the accessed area. According to the technical specification (TS) 38.821 [27], NTN can operate in the two bandwidth groups: FR1 (i.e., at frequency  $f < 6$  GHz) and FR2 (i.e., at frequency  $f > 6$  GHz). In FR1, the operating frequency for the service link is 2 GHz (S-band) with a maximum channel bandwidth of 30 MHz, while in FR2, the operating frequencies are 20 GHz in downlink and 30 GHz in uplink (Ka-band) with a maximum channel bandwidth up to 1 GHz.

Due to large delays and Doppler shifts, some changes are proposed in the NR. One important aspect is related to the UE mobility management. Due to the NTE moving with respect to the ground (except for the GEO and low-altitude UAV), there exists the shift in the antenna beam on the ground, leading to have moving tracking areas (TAs). This causes frequent location area and tracking area update on UEs, even if there are fixed or slightly moving UEs. One possible solution is to implement fixed TAs obtained by properly steering the antenna beam of the NTE, causing just a feeder link setting when a LEO, for example, goes down to the horizon and another covers the TA. In the case of fixed TAs, the mobility is managed according to similar procedures as for terrestrial NG-RAN. Anyway, moving TAs can also be implemented.

NR defined in Release 15 and Release 16 is a good basis to support the NTN connectivity. Nevertheless, some changes are needed: in layer 1, the optimization of some timers, adapting the HARQ procedure in terms of number of processes and feedback, and definition of switching procedure for the feeder link; and in layer 2, adaptation in the random access (RA) procedure, such as for preamble ambiguity and extension of the RA response (RAR) window, timing advance for the access and synchronization of the downlink and uplink frames, the discontinuous reception (DRX), scheduling request (SR), extending the status report in radio link control (RLC), sequence numbers in packet data convergence protocol (PDCP), definition of specific system information block (SIB) for NTN (i.e., SIB-19), and additional information should be provided to UEs in the case of moving cells (i.e., moving TAs), such as satellite Ephemeris information to favor its cell reselection. For example, it is considered a common timing advance defined as an offset between the gNB and a reference point in order to favor the synchronization and scheduling activities, as well as inserting in SIB-1 the information of whether the radio access network is terrestrial or non-terrestrial [33].

### 4.4. Multi-Connectivity

Multi-connectivity (MC) is defined in [34], and it is based on the bearer split at PDCP layer in UP. In some use cases envisaged in NTN, MC can be of interest, above all with TN. Examples are (i) in under-served areas where the terrestrial coverage is limited, as is the throughput for UE; (ii) in the case that the UE needs to have an increased throughput and TN cannot provide it; (iii) and to support vehicle operations (e.g., regional trains) along remote lines where TN coverage is highly variable.

According to [34], the two sites (i.e., the master node (MN), which provides the CP, and the secondary node (SN)) should be connected through the Xn interface. Two groups of MC can be set: the first case involves two NTEs, while the second involves one NTE and one terrestrial gNB. Depending on the location of the gNB (i.e., the PDCP layer, where the bearer split occurs), proper delays of one flow with respect to the other should be considered according to the channel latencies. Details are in [20,27]. As a basic idea, in case the gNB is located at the ground (for example, when the MC is performed by two transparent NTEs or by two NTEs with on-board DU but CUs on the ground), the Xn interface is not critical. Similarly, with two on-board gNB (i.e., regenerative NTN), an ISL can provide the required connectivity between the two NTEs. Additionally, in the case of hybrid MC (i.e., involving terrestrial gNB and one NTE), the MC is not critical in case the NTE is transparent or with the CU on the ground (see Figure 8). More challenging is the case when MC involves a NTE with an on-board gNB (i.e., regenerative NTN) and a terrestrial gNB since a connectivity should be set between the MN and SN, and delays should be properly aligned due to the highly different path. Finally, MC can be provided also by two different types of NTE (e.g., one UAV and one LEO). Similar considerations apply related to the location of the gNB.

In Figure 11, it is reported the typical architecture of the DC. We report two main cases. In the first case, we consider DC between one gNB in the TN and one in the NTN (see the HAPS), while in the second case, we consider two different NTEs, i.e., the HAPS and the satellite [20,27]. Easier implementations are when gNBs are located both on-board or both at the ground, while more challenging is when gNB is on-board on an NTE and the second is in the TN. Anyway, in all cases, the Xn interface must be available to split the flows between the two involved gNBs.

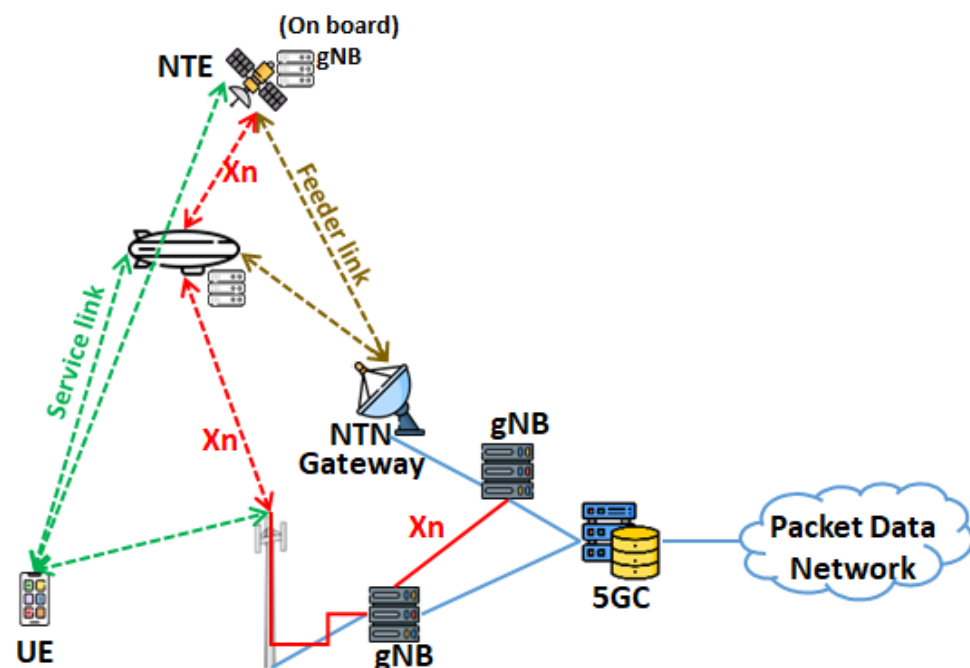


Figure 11. Typical dual connectivity architecture, involving one TN and one NTN or two NTNs.

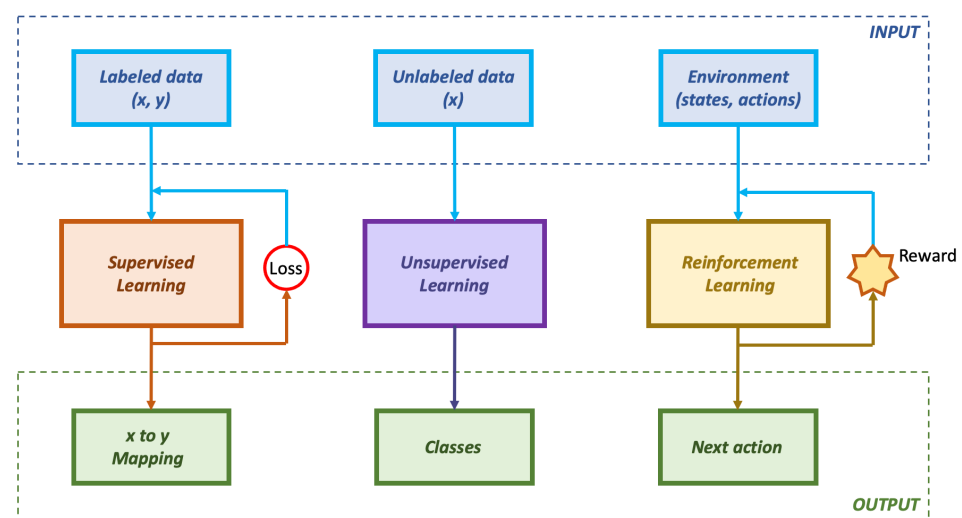
## 5. Machine Learning General Aspects

As opposed to the conventional engineering approach, which relies on finding an algorithmic solution to the problem, ML leverages on building the domain knowledge from examples. The training process leads to the production of a *model*, which represents the rules governing the specific object of study. The assumption is that, after some epochs of training, the model will accurately represent the reality and, therefore, it can guarantee to behave at least as the engineered model. Some of the advantages of machine learning models are as

follows: (i) they can adapt better to new scenarios, (ii) they perform well, even in exceptional cases, and (iii) they can spot trends and patterns, which are otherwise difficult to capture. However, many drawbacks have to be considered as well. ML requires a lot of training data, and the model accuracy heavily depends on the quality of the training set. Moreover, it is necessary that the training set is also balanced between each class it represents. Having a large amount of high quality data typically leads to better performance. Nevertheless, processing high data volumes can be time- and energy-consuming. These kinds of problems are slowing down the adoption of ML, especially on edge- and battery-powered devices.

### 5.1. Machine Learning Methods Taxonomy

Machine learning techniques can be classified in three main branches: (i) supervised learning, (ii) unsupervised learning and (iii) reinforcement learning. In Figure 12, the three types are depicted, highlighting their inputs and outputs.



**Figure 12.** Types of machine learning: supervised, unsupervised and reinforcement.

#### 5.1.1. Supervised Learning

In supervised learning, the training set is a collection of input and expected output pairs. The aim is to learn the relation between input and output spaces. Supervised learning can be further split in two sub-categories: (i) regression and (ii) classification. Regression makes use of sigmoid or linear functions in order to approximate a real value. Conversely, classification has the aim to categorize data samples into one of several classes. In recent years, classification is a sub-field, where the use of neural networks (NNs) has excelled. With the continuous increase in computational resources, deep neural networks (DNNs) proved that this task can be solved with higher accuracy, using large datasets and more complex neural network architectures. Some of the most important techniques in classification are support vector machines (SVMs), decision trees (DTs) and decision tree ensembles, such as random forests (RFs). Concerning NN-based methods, multi-layer feed-forward networks (FNNs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are often used to solve this task, and more specifically, in network communications. Some examples of supervised learning applications in NTN are remote sensing, decoding, demodulation, and network traffic control.

#### 5.1.2. Unsupervised Learning

In unsupervised learning, the training set consists of a collection of unlabeled inputs, without knowing the related expected output. The aim in this case is to find a small set of properties, which summarizes the entire data set. Some of the most important techniques in this field are K-means clustering, and the Gaussian mixture models (GMMs) that are used to determine, for every example, to which normal distribution it belongs

to. Unsupervised learning is also used for the dimensionality reduction purpose. In this area, principal component analysis (PCA) is often used to compress input datasets through feature extractions, thus revealing redundancies. Another recent development of unsupervised learning is the advent of *generative models*. These networks, once trained, can sample synthetic examples from the learned distribution or give a score according how similar two examples are. Unsupervised learning can be used in NTN in a variety of scenarios, such as dimensionality reduction in order to reduce the computational costs, and also security or multicast transmissions.

### 5.1.3. Reinforcement Learning

The previous discussed techniques (i.e., supervised and unsupervised learning) are also sometimes called *offline* learning methods. As a matter of fact, in supervised and unsupervised learning, the training phase takes place upfront, when the machine learning algorithm is disconnected from the environment. Reinforcement learning (RL) instead is applied to an agent, which interacts with the environment. At each time step, the agent receives observations from the environment and must choose an action which should bring it closer to the main goal. The environment, in response to the action, gives a reward which tells the agent how good the taken action was. This kind of processes can be mathematically described by Markov decision processes (MDPs), assuming that the environment is fully observable. Conversely, partial observable Markov decision processes (POMDPs) are used when the agent cannot directly observe the underlying environment state. Different algorithms are used in this field. One of the most important is Q-learning, where the maximum reward for a set of action is used as the learning factor. A recent advance of this algorithm is deep Q-networks (DQNs), where, as the name suggests, the Q-learning technique is empowered by the adoption of a DNN. DQNs fall into the deep reinforcement learning (DRL) techniques. State action reward state action (SARSA) is another algorithm based on Q-Learning, where the next action decision is not dependent on the past rewards and states. The name comes from the quintuple used for the state update, which is composed by the initial state-action, the observed reward after choosing the new state-action pair. Given that DQNs make use of neural network models to predict the next correct action, the entire RL task can be further classified in model-free methods (e.g., Q-Learning, SARSA) and model-based methods (e.g., deep Q-Networks). In NTN, RL can be used to optimize resource allocation, employ non-orthogonal multiple access (NOMA) methods, dynamic power allocation and computation offloading/caching to name a few.

### 5.2. Data Collection

As stated in the previous sections, supervised learning techniques require a lot of data to train a model. In many cases, the amount of data is directly proportional to the resulting accuracy. Although having a large training dataset is recommended, it is also necessary that the data are balanced, and the whole dataset is representative of the environment under study. In communication networks, there are many data sources, and most of them can be used to train a ML model. Starting from the physical layer, baseband signals or channel state information can be obtained. Moving to the medium access control (MAC)/link layer, several dimensions may be observed, such as the frame error rate (FER), throughput over time, and also random access load and latencies in various scenarios. Additionally, in the network layer, much information can be gathered. As an example, UE battery levels, mobility patterns, outage rates, global positioning system (GPS) location, traffic load and statistics are a few. Eventually, in the application layer, more structured data can be collected. Content demands, service subscription information, traffic statistics, user behavior patterns, user preferences and metrics from the quality of service (QoS) process.

## 6. Machine Learning Techniques in NTN

The adoption of ML enables improvements in NTN performance. Each ML technique should be carefully evaluated based on input parameters and the functionality to be

supported in a given NTE, thus exploiting the data availability. Proposed ML techniques are provided in the following, grouped per services or functions available in the layers of the system protocol stack.

### 6.1. ML for NTN Service

In the following, for each NTN service presented in Section 4, the adoption of some ML techniques is provided.

**Moving cell connectivity.** Deploying a reliable service for private/public transport, especially in high-speed scenarios, such as airplanes or trains, can be challenging. As stated before, LEO or HAPS are NTE that should be considered to solve this problem. RL Q-Learning algorithms can be used to dynamically adjust the position of HAPS with the aim of maximizing the network capacity and minimize the transmission latency. As an example, the trajectory planner has to position the HAPS according to the total number of users under the footprint, with their service requirements, while predicting their trajectory. Moreover, the nearest LEOs positions and trajectories have to be taken into consideration as well if space links are required to fulfill the service. RL-based approaches are also used to control the disconnectivity time and handover rate by taking into account the UAV battery lifetime and the remaining time to accomplish the requested task. In [35], the use of RL improved the NTN performance by reducing the handover rate compared to traditional strategies. More specifically, the authors used an objective function  $f(T_{tr}, P_u, T_d, \mathcal{R}_h)$  composed of several factors, such as the task completion time  $T_{tr}$ , the power consumption  $P_u$ , the disconnectivity time  $T_d$  and the handover rate  $\mathcal{R}_h$ . According to their simulation results, the adoption of RL reduced the handover rate by 50%, compared to a blind strategy. In [35], only UAVs are taken into account, but NTNs could benefit from the RL-powered handover rate reduction as well. In this case, many optimizations are possible, for example, between the same type of NTEs (e.g., LEO) or between different NTEs (e.g., LEO and HAPS).

**Unserved/Under-served and isolated areas.** This scenario is partially similar to the previous one, and therefore, reinforcement Q-Learning techniques with the aim of finding the best position of the NTE can be exploited, too. In this case, UAVs can be adopted as well, and thus, new challenges should be considered. Special attention should be given to the energy resource management. Scheduling to ground UEs can be assisted by RL in order to minimize the Age of Information (AoI) and minimize the energy consumption. Low power devices, such as mMTC sensors, can also benefit by ground-air multi-access edge computing (MEC) by offloading computation depending on the task resource requirements. In order to obtain better QoS, multi-user MEC offloading should be properly designed. As an example, in [36], a fully distributed game-based ML algorithm is used to obtain the Nash equilibrium without any information exchange. This work showed that applying ML leads to better offloading strategies and therefore better overall performance. In this work, the authors considered the average response time and the average energy consumption as performance metrics. Simulations proved that by using ML, the average cost per device can be reduced by 50–60% with respect to a random or even offloading strategy.

**Throughput increase.** Dedicated NTEs can be used to temporarily increase the network capacity adding extra links and therefore improving the quality of experience (QoE) of the served users. During outdoor events, dedicated UAVs can be deployed above crowded areas. As discussed before, ML and especially RL can help in 3D positioning, but in this particular scenario, UAVs can also help by content caching/storage, offering significant improvements in latency performance compared to remote cloud servers. Caching in the user proximity can ease the burden on the network at the backhaul, core and network levels. However, several research problems arise, such as what, when and how to cache contents. ML can help in estimating future user requests, based on several learnable factors, and producing optimized caching decisions.

Multi-connectivity, which has the aim of maximizing the throughput, is another technique that can benefit from the optimization provided by ML in the hybrid NTN/TN

environment. For example, a RL algorithm could learn to select the most convenient couple of transmitters between terrestrial (i.e., gNB) or non-terrestrial (i.e., UAV, HAPS, and LEO) or among the NTEs available in a given area. Parameters that should be taken into account by the RL are the NTE and UE trajectories, congestion estimation of both NTEs and terrestrial gNBs, signal quality received by the target UE or coverage estimation. Referring to the architecture described in Section 4.4, the RL technique should take into account the position of the transmitting gNBs (on the ground, on air or on space) and the related latencies on the Xn interface.

**Secondary link backup.** Service outages are frequent in today's internet communications, most of which are due to a variety of reasons (e.g., out-of-coverage, network failures, provider maintenance, and equipment failures). As described in Section 4, NTN can be exploited to quickly restore the link by deploying NTEs, such as UAVs, HAPS and LEOs. However, deploying an UAV or HAPS may take lot of time, and it often require to be planned in advance. ML can shorten the downtime by classifying the failure, thus helping the network operations to solve the problem in less time. In [37], the authors developed a classifier for 15 different kinds of failures by using bidirectional forwarding detection (BFD). Results showed that using these features, a 0.99 F1-Score is achievable for link failures. Radio link outages can also be predicted by using ML. More specifically, in [38], the author proposed a combination of long short-term memory (LSTM) and neural network (NN) to find a correlation between radio measurements, such as signal power and quality, and radio link failures. Training the system on a dataset obtained from a 5G testbed produced a model which was able to predict failures with a 0.94 F1-Score performance. A second aspect arises in cases when a secondary backup link is deployed with the sole aim of guaranteeing service continuity. Then, when the primary and also the backup connections are present, it can be convenient to exploit the secondary link for load-balancing purposes. Traditional load-balancing techniques can be used to decide which link to use. Many of them are based on network metrics, such as packet loss, round-trip latency, or remaining available link bandwidth. However, when several UAVs are deployed as multiple MEC nodes, load balancing can be challenging to be implemented, and in many situations more robust solutions, such as deep RL-based techniques, are recommended. For example, in [39], a deep RL algorithm is conceived for solving the task scheduling problem among  $N$  UAVs acting as MEC nodes, providing service for  $K$  IoT nodes. Regarding the performance metric, the authors measured the average task slowdown  $t_d \geq 1$ , where it is evaluated considering the positions of the generic UAV and of the IoT node, formulated as the real task completion time over the ideal task completion time. Simulation results highlighted a 65–75% average slowdown reduction compared to the traditional first-come first-served (FCFS) scheduling scheme.

**Disaster relief.** In emergency situations, such as natural disasters, NTEs are used to quickly bring connectivity in order to support rescue teams or enable survivors to call for help. In these scenarios, the UAVs position should be optimized in order to maximize the coverage and minimize the power consumption, thus extending the flight time. In [40], the authors proposed multi-layer perceptron (MLP) and LSTM approaches to optimize the UAVs position according to the user service requirements. Results showed that the combined use of MLP–LSTM reported a 98% user throughput maximization accuracy. UAVs can also scan the area at different altitudes, or track rescue teams by using computer vision techniques. In search and rescue, a drone can use object detection to spot people and alert the rescue team. As for the isolated areas scenario, an UAV can offload the expensive computation (e.g., image processing) to an air-space MEC or a terrestrial one. Deep learning-based object tracking can also be used to follow rescue teams and in addition to bring them connectivity; they also scan the surrounding area.

**Broadcasting/Multicasting.** ML has a significant role in optimizing cache performance. Existing network architectures struggle in handling mass content delivery with low response time. Having the contents at the edge of the network is vital to enable interactive broadcast services. To this aim, a ML framework can decide to fetch contents in advance



and proactively push it to the edge reducing the delay and the network burden, especially in peak hours. The learning process in this case is based on content popularity, user access and mobility. A problem that arises by introducing caching services is the replacement strategy. Traditional cache replacement techniques, such as first-in first-out (FIFO), least frequently used (LFU) or least recently used (LRU) do not fit well in a non-deterministic scenario. To mitigate this problem, a (deep) Q-Learning RL network could be applied to take cache replacement/eviction decisions based on the Q-Value or reward obtained from the environment [41]. Concerning the performance metric, the authors measured the system throughput as a function of the cluster size under different replacement schemes. In their test-bed implementation, the adoption of a dynamic RL-powered policy improved the throughput by 4–5 times compared to a static policy.

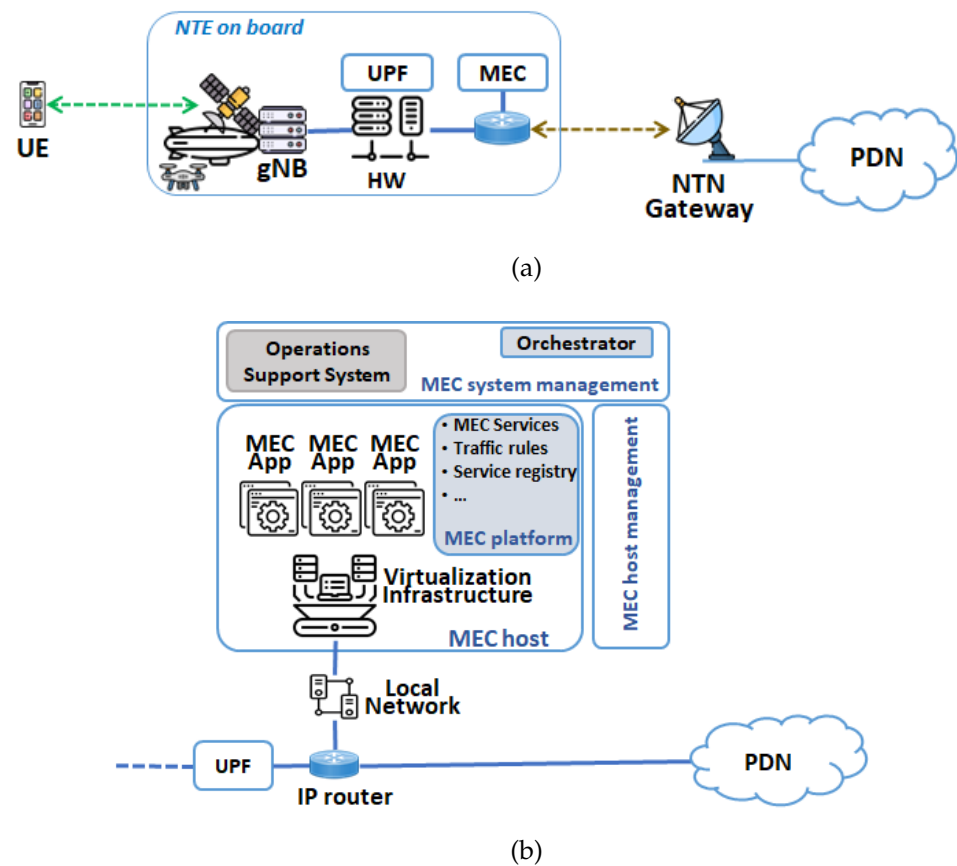
Table 4 reports some of the ML techniques that can be applied to each NTN service presented in Section 4.1.

**Table 4.** ML applications for NTN services.

NTN Service	ML Applications	Target NTE
Mobility cell connectivity	3D Positioning, Trajectory planning, handover rate minimization	HAPS, LEO
Unserved/Under-served and isolated areas	3D Positioning, Trajectory planning, Energy management, Multi-user MEC offloading	UAV, HAPS, LEO
Throughput increase	Content caching, selection of the Tx elements in Multi-connectivity	UAV, HAPS
Secondary link for backup	Network failures classification, Radio link outage prediction, load balancing	HAPS, LEO
Disaster relief	UAV position optimization, Object detection and tracking, MEC offloading	UAV, HAPS
Broadcasting/Multicasting	Content caching fetching/eviction strategy, MEC offloading	LEO, HAPS

As a final remark, it is worth noting that the assessment of the deployment position of a MEC is important for two main aspects. First, a MEC cannot be placed everywhere in the NTN architecture. In fact, it is necessary to consider the protocol stack of the system since MEC has to be deployed in (terrestrial or non-terrestrial) network elements able to read the user IP address at least in order to properly route the user packets to it and to retrieve contents (see [42]). This is possible only for UPF in the 5G architecture if we look at the UP. Theoretically, it is possible to deploy a gNB and (light) UPF on-board of an NTE, even if not considered in the protocol stack architectures reported in Section 4.2. However, this brings us to the second point. Installing a MEC in a UAV can be challenging due to energy constraints, processing capacity limitation and storage issues. It may be possible on a HAPS or LEO. Proper design should be considered for deploying a MEC on board NTE. Of course, its advantages will be many in terms of latency, experienced throughput and traffic offload [43]. In Figure 13a, a possible deployment of the MEC is reported. In this case, we considered only the on-board NTE deployment, which is more challenging in terms of weight and energy consumption but at the same time better performing for all NTN services since it provides the highest data rates and the lowest delays. In Figure 13b, it is reported the general scheme of the MEC architecture [44].

The MEC hosts its applications running over virtual machines on the virtualization infrastructure. The MEC platform supports the application configuration. It is provided by the MEC host manager, which, for example, configures traffic rules, policies, and solves possible conflicts. The MEC system management is in charge of instantiating the applications requested by UEs (i.e., the customers) through the operation support system and supported by the MEC orchestrator. Finally, in order to provide the connectivity of the MEC of another mobile network operator, it should be provided by connecting the two MEC hosts through the two MEC platforms.



**Figure 13.** Structure of the multi-access edge computing: (a) on-board deployment; (b) architecture.

## 6.2. ML for Protocol Stack Layers

In the next subsections, we will analyze some benefits of adopting ML in different layers.

### 6.2.1. Physical Layer

In the physical layer, ML can improve performance in many ways. As an example, in [45], a CNN is used to classify outputs of the receiver among possible constellations of modulation schemes, such as 16QAM and 64QAM. The authors used AlexNet, which is a DNN made of 650 thousand neurons and 60 million parameters. Results show that AlexNet can be more accurate than the traditional support vector machine (SVM) based techniques. More specifically, the authors measured the accuracy of the classifier as a function of the SNR (signal-to-noise ratio) of four modulation types: QPSK, 8PSK, 16QAM and 64QAM. Simulation results show that in the low SNR region, QPSK and 8PSK identification obtained an accuracy greater than 90%, and about 60% for 16QAM and 64QAM. Nevertheless, on average, this technique showed to be sufficiently (>80%) accurate at an  $\text{SNR} \geq 4$  dB. Propagation loss prediction is another field where ML can help. Statistical models such as Okumura–Hata require less computational power, but at the same time are less accurate than the deterministic ones, such as those based on ray tracing methods. However, these models require more computational power and more detailed input data. In [46], the authors presented a NN-based procedure to predict the path loss in an urban environment. The main goal is to approximate the ray tracing performance while reducing the computation cost. Results showed an accuracy of  $\pm 2.5$  dB for uniformly built-up environments and 4.9 dB for non-uniform environments.

### 6.2.2. Link and Medium Access Control Layers

In traditional wireless networks, ML can be used for efficiently and proactively allocating the spectrum. In [47], a DNN is adopted to solve the resource allocation problem in

real time. The authors considered the weighted sum-rate (WSR) maximization problem subject to the transmit power constraints. Results show that the approximation obtained by a DNN, trained on input/output pairs of a resource allocation algorithm, can speed up the computation time by orders of magnitude compared to a weighted minimum mean square error (WMMSE) method. It is fair to point out that these results are referred to a Python implementation of both WMMSE and DNN. However, the performance improvement obtained by the Python DNN compared to a C implementation of WMMSE is about 7% on average. It is worth noting that the same resource allocation problem can be formulated by considering the spectrum allocation of the LEO beams or even between different NTEs. As an example, a DNN model can be used to optimize the area throughput (defined as the sum of the throughput of all users belonging to a specific area), by properly allocating the spectrum among different NTEs, or even dynamically adjust the size of the LEO beam footprint, for example enlarging the footprint in low populated areas while reducing it in the most populated areas. This can provide a greater spectrum available for the most requiring users. Then, several inputs for the enhancement by ML can be used, such as the user density in a given area, the grade of user mobility, NTE and UE trajectories or the requested traffic type (e.g., MMB or MTC).

Another application related to this layer is the UE traffic prediction. ML can take the traffic load and spectrum usage as inputs and predict the traffic generated by UEs in a dynamic way, tolerating high variations on these input factors. In [48], an auto-encoder based model is used for spatial modeling and LSTM units for temporal modeling. The authors have chosen the average error (i.e., prediction vs actual) as their performance metric calculated as mean squared error (MSE), mean absolute error (MAE) and log loss. Simulation results showed a 30.8%, 20.5%, 33.1% reduction, respectively, compared to the support vector regression (SVR) and 40.4%, 28.4%, 18.5% reduction, respectively, compared to the auto-regression integrated moving average (ARIMA). While considering ground, air and space elements in a standalone way, these techniques are still applicable to NTN. However, NTNs introduce novel specific issues, which can be mitigated by the adoption of ML techniques. ML can help in the NTE selection for a given service, taking into account the network congestion and the requested QoS/QoE, for example, choosing the one with the shortest delay or the least congested.

### 6.2.3. Network Layer

Finding the optimal path between two or more nodes of a network can be challenging if highly varying network conditions, such as overloaded routers, malfunctions or network outages, have to be considered. Traditional routing algorithms have to deal with a heavy computation load to assure good end-to-end transmission performance. Moreover, obtaining good local performance does not imply that the global transmission performance improves, too. In NTNs, the routing problems it is even more difficult due to the heterogeneous nature of the network. In [49], the authors proposed a supervised learning approach, which improves the routing performance in terms of signaling overhead, throughput and per hop delay compared to the traditional Open Shortest Path First (OSPF) protocol. Simulation results proved that the ML approach outperforms OSPF in the three metrics. More specifically, this technique lowered the signaling overhead by 70%, increasing the throughput of 2% and reducing the hop-delay of 90%.

### 6.2.4. Application Layer

ML can be used in the application layer in several ways. As an example, internet traffic classification can be difficult due to the encrypted traffic streams. Traditionally, the feature extraction phase is performed by field experts, and the classification is made on the base of the obtained features. ML can streamline the process by performing both feature extraction and classification in one system. In [50], the authors presented a system which first characterizes the traffic and then identifies the specific application, obtaining a precision and recall for plain and VPN-encapsulated traffic greater than 90%. In NTNs,

traffic classification may be used to properly select the NTE, when more of them are available in the area. As an example, when a VoIP call is identified, it may be better to route the traffic over UAV or HAPS, and leave GEO satellites to fewer sensitive applications, such as unidirectional video streaming.

Another interesting use of ML at application layer consists in optimizing the data rate at application level by estimating the end-to-end round trip time (RTT). Popular media streaming services (e.g., YouTube or Netflix) suffer from high RTT because of the TCP congestion control strategies (i.e., higher RTT values lower end-to-end data rate [43]). Predicting future values of users' RTT can help in directly improving the throughput, filling the buffer of the user that is expected to experience a larger RTT or properly modify in advance the coding strategy in order to avoid video quality worsening. Due to the air and space elements, instantaneous RTT readings may not be representative because NTEs are subject to RTT fluctuations. In this case, having an estimation in advance is even more important to properly implement the most suitable countermeasure.

Table 5 reports some of the ML techniques that can be applied to each layer in the NTN protocol stack based on the implementations in Section 4.2.

**Table 5.** ML applications for layers in the NTN protocol stack.

NTN Layer	ML Applications	Target NTE
Physical Layer	Constellation modulation schemes detection, pathloss prediction	UAV, HAPS, LEO
Link and Medium Access Control Layers	Resource allocation, beams dynamic footprint size adjustment, dynamic spectrum allocation between NTEs, traffic prediction, NTE selection based on network congestion	UAV, HAPS, LEO
Network Layer	ML powered routing protocol	UAV, HAPS
Application Layer	Traffic classification, RTT estimation for throughput improvement	HAPS, LEO

## 7. Conclusions and Open Research Issues

In this paper, we proposed a review of NTNs from the services and architectures point of view. In the first part, we detailed how NTNs work and the services that these networks can provide. We summarized the required data rates for each service in eMMB, mMTC and uRLLC categories of 5G, as well as the suited NTEs for each of them. Afterward, we detailed the most important architectures and its protocol stacks considering the transparent NTN architecture, the regenerative architecture and eventually the on-board distributed unit architecture.

Due to the heterogeneity of NTNs, these networks introduce several challenges compared to terrestrial networks, which can be solved or mitigated at least by the joint use of ML techniques. In the last two sections, we carried out an analysis of the most promising applications of ML, referring to the services mentioned in Section 4, and afterwards organizing them by protocol stack layers.

Considering the interest of NTN and its integration with terrestrial NR networks, the performance should be accurately investigated in order to tailor the most suitable integration point based on a given service (e.g., collecting IoT sensor data in wide areas or providing dual connectivity to a moving platform) in terms of peak and average data rate, latency and channel reliability. Moreover, the location of the analytics platform within the integrated network can be further investigated considering the availability of big data, the interoperability of different types of sensors (or in general UEs) connecting to two different mobile network operators and the connectivity to a given NTE.

**Author Contributions:** Conceptualization, R.G.; methodology, R.G.; software, E.I.; validation, E.I.; formal analysis, R.G.; investigation, R.G.; resources, R.G. and E.I.; data curation, R.G. and E.I.; writing—original draft preparation, R.G. and E.I.; writing—review and editing, R.G. and E.I.; visualization, R.G.; supervision, R.G.; project administration, R.G.; funding acquisition, R.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Analysis Mason's Research Prediction for 2023. Available online: [https://www.analysismason.com/contentassets/bd58910f9777465aae2543a9220bf2f7/analysys\\_mason\\_research\\_predictions\\_2023\\_dec2022.pdf](https://www.analysismason.com/contentassets/bd58910f9777465aae2543a9220bf2f7/analysys_mason_research_predictions_2023_dec2022.pdf) (accessed on 5 January 2023).
2. 5G Americas. 5G & Non-Terrestrial Networks. February 2022. Available online: <https://www.5gamericas.org/5g-and-non-terrestrial-networks/> (accessed on 24 January 2023).
3. Morocho-Cayamcela, M.E.; Lee, H.; Lim, W. Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access* **2019**, *7*, 137184–137206. [\[CrossRef\]](#)
4. Kodheli, O.; Lagunas, E.; Maturo, N.; Sharma, S.K.; Shankar, B.; Montoya, J.F.M.; Duncan, J.C.M.; Spano, D.; Chatzinotas, S.; Kisseleff, S.; et al. Satellite communications in the new space era: A survey and future challenges. *IEEE Commun. Surv. Tutorials* **2021**, *23*, 70–109. [\[CrossRef\]](#)
5. Geraci, G.; Garcia-Rodriguez, A.; Azari, M.M.; Lozano, A.; Mezzavilla, M.; Chatzinotas, S.; Chen, Y.; Rangan, S.; Di Renzo, M. What will the future of UAV cellular communications be? A flight from 5G to 6G. *IEEE Commun. Surv. Tutorials* **2022**, *24*, 1304–1335. [\[CrossRef\]](#)
6. Araniti, G.; Iera, A.; Pizzi, S.; Rinaldi, F. Toward 6G Non-Terrestrial Networks. *IEEE Netw.* **2022**, *36*, 113–120. [\[CrossRef\]](#)
7. Vaezi, M.; Azari, A.; Khosravirad, S.R.; Shirvanimoghaddam, M.; Azari, M.M.; Chasaki, D.; Popovski, P. Cellular, Wide-Area, and Non-Terrestrial IoT: A Survey on 5G Advances and the Road Toward 6G. *IEEE Commun. Surv. Tutorials* **2022**, *24*, 1117–1174. [\[CrossRef\]](#)
8. Zeng, Y.; Zhang, R.; Lim, T.J. Wireless communications with unmanned aerial vehicles: Opportunities and challenges. *IEEE Commun. Mag.* **2016**, *54*, 36–42. [\[CrossRef\]](#)
9. Burleigh, S.C.; De Cola, T.; Morosi, S.; Jayousi, S.; Cianca, E.; Fuchs, C. From connectivity to advanced Internet services: A comprehensive review of small satellites communications and networks. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 6243505. [\[CrossRef\]](#)
10. Niephaus, C.; Kretschmer, M.; Ghinea, G. QoS provisioning in converged satellite and terrestrial networks: A survey of the state-of-the-art. *IEEE Commun. Surv. Tutorials* **2016**, *18*, 2415–2441. [\[CrossRef\]](#)
11. Giordani, M.; Zorzi, M. Satellite Communication at Millimeter Waves: A Key Enabler of the 6G Era. In Proceedings of the 2020 International Conference on Computing, Networking and Communications (ICNC), Big Island, HI, USA, 17–20 February 2020; pp. 383–388. [\[CrossRef\]](#)
12. Kaushal, H.; Kaddoum, G. Optical communication in space: Challenges and mitigation techniques. *IEEE Commun. Surv. Tutorials* **2017**, *19*, 57–96. [\[CrossRef\]](#)
13. Wang, D.; Traspadini, A.; Giordani, M.; Alouini, M.-S.; Zorzi, M. On the Performance of Non-Terrestrial Networks to Support the Internet of Things. In Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 30 October–2 November 2022.
14. Tsuji, H.; Miura, A.; Carrasco-Casado, A.; Toyoshima, M. R&D for Satellite Communications and Non-terrestrial Networks toward Beyond-5G in Japan. In Proceedings of the 2022 27th Asia Pacific Conference on Communications (APCC), Jeju Island, Republic of Korea, 19–21 October 2022; pp. 178–181. [\[CrossRef\]](#)
15. Ahmed, N.; Kanhere, S.S.; Jha, S. On the importance of link characterization for aerial wireless sensor networks. *IEEE Commun. Mag.* **2016**, *54*, 52–57. [\[CrossRef\]](#)
16. Koulali, S.; Sabir, E.; Taleb, T.; Azizi, M. A green strategic activity scheduling for UAV networks: A sub-modular game perspective. *IEEE Commun. Mag.* **2016**, *54*, 58–64. [\[CrossRef\]](#)
17. Choi, D.H.; Kim, S.H.; Sung, D.K. Energy-Efficient Maneuvering and Communication of a Single UAV-Based Relay. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 2320–2327. [\[CrossRef\]](#)
18. Li, B.; Fei, Z.; Zhou, C.; Zhang, Y. Physical-layer security in space information networks: A survey. *IEEE Internet Things J.* **2020**, *7*, 33–52. [\[CrossRef\]](#)
19. Al-Hraishawi, H.; Chougrani, H.; Kisseleff, S.; Lagunas, E.; Chatzinotas, S. A Survey on Non-Geostationary Satellite Systems: The Communication Perspective. *IEEE Commun. Surv. Tutorials* **2022**. [\[CrossRef\]](#)
20. Rinaldi, F.; Maattanen, H.-L.; Torsner, J.; Pizzi, S.; Andreev, S.; Iera, A.; Koucheryavy, Y.; Araniti, G. Non-Terrestrial Networks in 5G & Beyond: A Survey. *IEEE Access* **2020**, *8*, 165178–165200. [\[CrossRef\]](#)
21. Azari, M.M.; Solanki, S.; Chatzinotas, S.; Kodheli, O.; Sallouha, H.; Colpaert, A.; Montoya, J.F.M.; Pollin, S.; Haqiqatnejad, A.; Mostaani, A.; et al. Evolution of Non-Terrestrial Networks from 5G to 6G: A Survey. *IEEE Commun. Surv. Tutorials* **2022**. [\[CrossRef\]](#)
22. Rekkas, V.P.; Sotiroudis, S.; Sarigiannidis, P.; Wan, S.; Karagiannidis, G.K.; Goudos, S.K. Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends. *Electronics* **2021**, *10*, 2786. [\[CrossRef\]](#)

23. Muscinelli, E.; Shinde, S.S.; Tarchi, D. Overview of Distributed Machine Learning Techniques for 6G Networks. *Algorithms* **2022**, *15*, 210. [CrossRef]
24. Bartsiokas, I.A.; Gkonis, P.K.; Kaklamani, D.I.; Venieris, I.S. ML-Based Radio Resource Management in 5G and Beyond Networks: A Survey. *IEEE Access* **2022**, *10*, 83507–83528. [CrossRef]
25. Chen, H.; Xiao, M.; Pang, Z. Satellite Based Computing Networks with Federated Learning. *arXiv* **2021**, arXiv:2111.10586. <https://doi.org/10.48550/arXiv.2111.10586>.
26. Michailidis, E.T.; Potirakis, S.M.; Kanas, A.G. AI-Inspired Non-Terrestrial Networks for IIoT: Review on Enabling Technologies and Applications. *IoT* **2020**, *1*, 21–48. [CrossRef]
27. 3GPP. Solutions for NR to Support Non-Terrestrial Networks (NTN) (Release 16). TR 38.821 V16.1.0 Release 16 May 2021. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3525> (accessed on 24 January 2023).
28. 3GPP. Study on New Radio (NR) to Support Non-terrestrial Networks (Release 15). TR 38.811 v15.4.0, September 2020. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3234> (accessed on 24 January 2023).
29. ITU. MT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond. Report ITU-R, M.2083-0. September 2015. Available online: <http://www.itu.int/publ/R-REC/en> (accessed on 24 January 2023)
30. ITU. Vision, Requirements and Evaluation Guidelines for Satellite Radio Interface(s) of IMT-2020. Report ITU-R, M.2514-0. September 2022. Available online: <https://www.itu.int/pub/R-REP-M.2514-2022> (accessed on 24 January 2023)
31. Tan, T.K.; Weerakkody, R.; Mrak, M.; Ramzan, N.; Baroncini, V.; Ohm, J.-R.; Sullivan, G.J. Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 76–90. [CrossRef]
32. 3GPP. Study on New Radio Access Technology: Radio Access Architecture and Interfaces. 3GPP TR 38.801 V14.0.0, Release 14 March 2017. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3056> (accessed on 24 January 2023).
33. 3GPP. NR and NG-RAN Overall Description Stage 2. TS 38.300 V17.2.0 Release 17 September 2022. Available online: [https://www.etsi.org/deliver/etsi\\_ts/138300\\_138399/138300/16.02.00\\_60/ts\\_138300v160200p.pdf](https://www.etsi.org/deliver/etsi_ts/138300_138399/138300/16.02.00_60/ts_138300v160200p.pdf) (accessed on 24 January 2023).
34. 3GPP. Multi-Connectivity, Stage 2. 3GPP TS 37.340 V17.2.0, Release 17 September 2022. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198> (accessed on 24 January 2023).
35. Azari, M.M.; Arani, A.H.; Rosas, F. Mobile Cellular-Connected UAVs: Reinforcement Learning for Sky Limits. In Proceedings of the 2020 IEEE Globecom Workshops (GC Wkshps), Taipei, Taiwan, 7–11 December 2020. [CrossRef]
36. Wang, Y.; Yang, J.; Guo, X.; Qu, Z. A Game-Theoretic Approach to Computation Offloading in Satellite Edge Computing. *IEEE Access* **2020**, *8*, 12510–12520. [CrossRef]
37. Evang, J.M.; Ahmed, A.H.; Elmokashfi, A.; Bryhni, H. Crosslayer network outage classification using machine learning. In *Proceedings of the Workshop on Applied Networking Research (ANRW '22)*; Association for Computing Machinery: New York, NY, USA, 2022. [CrossRef]
38. Boutiba, K.; Baga, M.; Ksentini, A. Radio Link Failure Prediction in 5G Networks. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021; pp. 1–6. [CrossRef]
39. Yang, L.; Yao, H.; Wang, J.; Jiang, C.; Benslimane, A.; Liu, Y. Multi-UAV-Enabled Load-Balance Mobile-Edge Computing for IoT Networks. *IEEE Internet Things J.* **2020**, *7*, 6898–6908. [CrossRef]
40. Munaye, Y.Y.; Lin, H.-P.; Adege, A.B.; Tarekegn, G.B. UAV Positioning for Throughput Maximization Using Deep Learning Approaches. *Sensors* **2019**, *19*, 2775. [CrossRef] [PubMed]
41. Lee, M.-C.; Feng, H.; Molisch, A.F. Dynamic Caching Content Replacement in Base Station Assisted Wireless D2D Caching Networks. *IEEE Access* **2020**, *8*, 33909–33925. [CrossRef]
42. ETSI. Multi-Access Edge Computing (MEC) MEC 5G Integration. ETSI GR MEC 031 V2.1.1, October 2020. Available online: [https://www.etsi.org/deliver/etsi\\_gr/MEC/001\\_099/031/02.01.01\\_60/gr\\_MEC031v020101p.pdf](https://www.etsi.org/deliver/etsi_gr/MEC/001_099/031/02.01.01_60/gr_MEC031v020101p.pdf) (accessed on 24 January 2023).
43. Ciccarella, G.; Giuliano, R.; Mazzenga, F.; Vatalaro, F.; Vizzarri, A. Edge cloud computing in telecommunications: Case studies on performance improvement and TCO saving. In Proceedings of the 2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC), Rome, Italy, 10–13 June 2019; pp. 113–120
44. ETSI MEC ISG. Mobile Edge Computing (MEC); Framework and Reference Architecture. ETSI, DGS MEC 003, April 2016. Available online: [http://www.etsi.org/deliver/etsi\\_gs/MEC/001\\_099/003/01.01.01\\_60/gs\\_MEC003v010101p.pdf](http://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/01.01.01_60/gs_MEC003v010101p.pdf) (accessed on 5 January 2023).
45. Peng, S.; Jiang, H.; Wang, H.; Alwageed, H.; Yao, Y.-D. Modulation classification using convolutional Neural Network based deep learning model. In Proceedings of the 2017 26th Wireless and Optical Communication Conference (WOCC), Newark, NJ, USA, 7–8 April 2017; pp. 1–5. [CrossRef]
46. Sotiroidis, S.P.; Siakavara, K.; Sahalos, J.N. A Neural Network Approach to the Prediction of the Propagation Path-loss for Mobile Communications Systems in Urban Environments. *Piers Online* **2007**, *3*, 1175–1179. [CrossRef]
47. Sun, H.; Chen, X.; Shi, Q.; Hong, M.; Fu, X.; Sidiropoulos, N.D. Learning to optimize: Training deep neural networks for wireless resource management. In Proceedings of the 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, 3–6 July 2017; pp. 1–6. [CrossRef]

48. Wang, J.; Tang, J.; Xu, Z.; Wang, Y.; Xue, G.; Zhang, X.; Yang, D. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In Proceedings of the IEEE INFOCOM 2017—IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9. [\[CrossRef\]](#)
49. Kato, N.; Fadlullah, Z.M.; Mao, B.; Tang, F.; Akashi, O.; Inoue, T.; Mizutani, K. The Deep Learning Vision for Heterogeneous Network Traffic Control: Proposal, Challenges, and Future Perspective. *IEEE Wirel. Commun.* **2017**, *24*, 146–153. [\[CrossRef\]](#)
50. Lotfollahi, M.; Zade, R.S.H.; Siavoshani, M.J.; Saberian, M. Deep packet: A novel approach for encrypted traffic classification using deep learning. *arXiv* **2017**, arXiv:1709.02656.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.