SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 18, No. 3, October 2021, 397-426

UDC: 004.7:004.8 **DOI: https://doi.org/10.2298/SJEE2103397D**

Application of Deep Learning Algorithms and Architectures in the New Generation of Mobile Networks

Dejan Dašić[1,2,3], Miljan Vučetić[1], Nemanja Ilić[1,4], Miloš Stanković[1,2], Marko Beko[5]

Abstract: Operators of modern mobile networks are faced with significant challenges in providing the requested level of service to an ever increasing number of user entities. Advanced machine learning techniques based on deep architectures and appropriate learning methods are recognized as promising ways of tackling the said challenges in many aspects of mobile networks, such as mobile data and mobility analysis, network control, network security and signal processing. Having firstly presented the background of deep learning and related technologies, the paper goes on to present the architectures used for deployment of deep learning in mobile networks. The paper continues with an overview of applications and services related to the new generation of mobile networks that employ deep learning methods. Finally, the paper presents practical use case of modulation classification as implementation of deep learning in an application essential for modern spectrum management. We complete this work by pinpointing future directions for research.

Keywords: Deep learning, Mobile networks, Mobile data analysis, Network security, Drone-based communications, Signal processing, Modulation Classification.

1 Introduction

Generation of massive volumes of heterogeneous data is a feature of modern mobile networks. In order to classify such data or make predictions or decisions based on them, implementation of deep learning (DL) models and algorithms, as opposed to traditional machine learning (ML), is advised. This advice is based on the complexity of DL models being sufficient to account for the underlying complexity of phenomena within the area of mobile networks applications. Moreover, performance of deep learning algorithms greatly benefits from the

[1]Vlatacom Institute, Belgrade, Serbia; E-mail: dejan.dasic@vlatacom.com; miljan.vucetic@vlatacom.com
[2]Singidunum University, Belgrade, Serbia; E-mail: milos.stankovic@singidunum.ac.rs
[3]COPELABS, Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal
[4]College of Applied Technical Sciences, Kruševac, Serbia; E-mail: nilic@asss.edu.rs
[5]Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal; Email: beko.marko@gmail.com

vastness of data available for training, which act as natural regularizers and prevent over-fitting [1]. These huge amounts of data, although extending the duration of the training process, do not represent fundamental obstacle for the successful design of deep learning models. This is due to the inherent scalability of the used training algorithms, which typically represent variants of the Stochastic Gradient Descent (SGD) method, feeding themselves on small chunks of data at each training step [2].

By nature, mobile networks generate data which are typically noisy and heterogeneous and which exhibit complex spatial and temporal patterns [3]. These characteristics go in favor of emphasizing another advantage of deep learning compared to traditional ML algorithms, which is the inherent ability of extraction of the feature hierarchy from the data of complex structure with inner correlations [1]. Human effort needed for manual feature engineering from such data would make such an endeavor expensive [4].

In addition, greater interest for building and implementing DL models over traditional ML in the next generation of mobile networks arises from the fact that deep learning provides numerous scalable and effective methods. Modern deep learning approaches enable use of unlabeled data for the purposes of extraction of useful patterns while mitigating the cost of annotation. The importance of these algorithms for the next generation of mobile systems, which also generate semi-labeled or unlabeled data [3] is also the subject of consideration in this work.

Deep learning requirements have forced designers of mobile networks to consider numerous architectures in order to support new generation of services. Various infrastructure models for implementation of deep learning applications have been recognized, including cloud, fog, edge and mobile computing resources [5]. These computing techniques significantly impact the availability, scalability, latency and network communication capabilities of the deep learning based applications in mobile systems of next generation. Thus, 5G and 6G networks can be augmented using cognitive capacity of DL in different network environments.

This paper identifies the research potential between DL techniques and the new generation of mobile networks, by presenting the current state of achievements made by implementing DL models in various aspects of mobile networks. The main aim of this survey is to showcase how complex problems in this field requiring intense computations can be tackled by DL schemes. The work presented in this paper is an extended version of our previously published conference paper [6]. Structure of this paper is as follows. In the following section (Section 2), the principles behind the deep learning models are discussed. Section 3 focuses on DL architectures and algorithms referenced in the subsequent sections. Section 4 presents the areas within mobile networking that have been in the focus of DL researchers. A succinct overview of recent advances in these application areas is also provided. Section 5 presents one use case of modulation classifications as an application of DL of great importance for spectrum management. Section 6 concludes the paper, summarizing the emerging challenges and stating future research directions.

2 Deep Learning Basics

Deep Learning represents an approach to Artificial Intelligence (AI); more specifically, it is a subfield of Machine Learning, a technique that enables intelligent agents to improve their performance on a given task with data and experience (i.e., to learn) [1]. ML has proved itself as the only viable AI system able to successfully operate in complex real-world environments, with explicit programming being replaced by the learning process. Deep Learning, unlike classical ML which typically relies on the construction of hand-designed features for the learning task, automates feature design process in a consistent way. It serves as a paradigm for the creation of algorithms that exhibit great performance and flexibility, owing to the successful representation of the considered problem at hand. This representation involves a nested hierarchy of concepts, with growing complexity and increasing abstraction level.

An illustration of relationship between AI, ML and DL is given in Fig. 1.

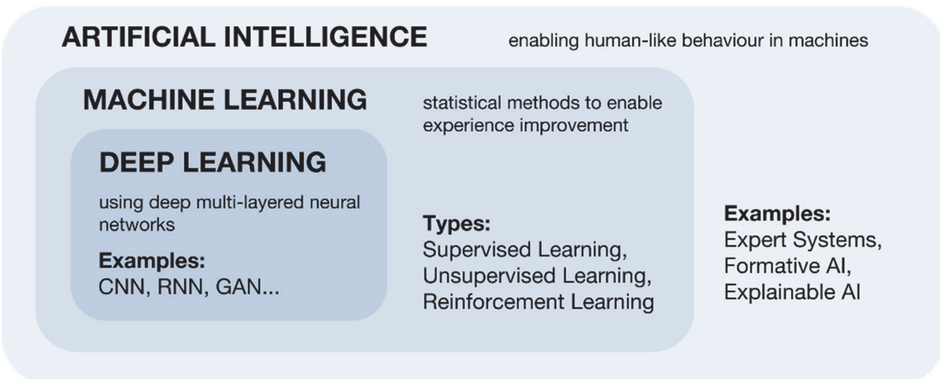


Fig. 1 – *Illustration of relationship between Artificial Intelligence, Machine Learning and Deep Learning.*

The algorithms applied in machine learning can be categorized by the type of experience they are allowed to have during training process into three broad classes: supervised learning, unsupervised learning and reinforcement learning.

Supervised learning assumes that, during training, the model is fed with labeled data pairs (feature data associated with data labels or targets) in order to learn a desired mapping from feature data to data labels. Unsupervised learning works without the labeled data by learning useful properties of the structure of data it is presented with. Within the deep learning context, unsupervised learning usually means that the whole probability distribution that generated input data is being learned. Combination of labeled and unlabeled data in a single dataset is also possible and represents an input for the so-called semi-supervised learning. Reinforcement learning (RL) does not experience a given input dataset, but instead assumes algorithms that interact with the environment, changing the environment's state and receiving rewards for the actions taken. The goal of RL is to design an action plan for the algorithm which would maximize an expected cumulative reward obtained in interaction with the environment [7]. It is safe to say that usage of deep learning and neural network models as function approximation tools, forming what is known as deep reinforcement learning (DRL), has brought a renaissance of its kind to the field of RL.

The prototypical example of a deep learning model is the feedforward deep neural network, or the so-called MultiLayer Perceptron (MLP). With its history connected to biological neurons and synapses, MLP architecture is built by stacking layers of neurons, and connecting them by weighted links. The resulting set of weights represents the set of parameters to be found in accordance with the given task. Therefore, the training process is aimed at configuring these network parameters (e.g. in a way that would ensure that the set of wanted inputs maps to the set of wanted outputs [8]). Within this context, the term "deep" when followed by neural network actually refers to neural networks with a (sufficiently large) number of hidden layers. It should be noticed that other multilayered structures, beside deep neural networks, such as deep random forests [9], deep Gaussian processes [10], etc., depending on the researchers, may (or may not) be considered part of the deep learning architectures spectrum.

MLPs assume all-to-all connections between neurons in adjacent layers (the so-called fully connected layers). This approach, although very popular in the past [11], suffers from problems arising from its high computation requirements. This issue becomes important when training deep architectures, especially when dealing with high-dimensional inputs such as images. As an alternative, the so-called Convolutional Neural Networks (CNNs) are widely used today [12]. They adopt several restrictions on the architecture that allow for a more appropriate treatment of spatially correlated structures such as images and for a significant reduction in the number of parameters (weights). Namely, each neuron is connected to a small part of the adjacent input layer (neuron's receptive field), and groups of neurons share the same sets of weights (corresponding to the so-called convolution filter weights). CNNs have achieved extraordinary performance in imaging applications, with their state-of-the-art results even surpassing human-level performance in image classification tasks.

Although contemporary Deep Learning algorithms often draw strength by leveraging large labeled datasets, there exists a corpus of algorithms that focus on unsupervised learning techniques and on generalizing from small datasets. One of the most famous examples is a structure called Deep Belief Network (DBN) [13]. A DBN is made of a stack of Restricted Boltzmann Machines (RBM), a type of energy-based undirected graphical models which include a visible layer and a hidden layer in which each unit can only assume binary values. The efficient training of DBNs by the strategy called “greedy layer-wise pre-training” has served as one of the key impulses in the past for the resurgence of interest for the DL paradigm within AI research community.

One of the most popular unsupervised learning structures are Auto-Encoders (AEs), representing an emulation of the copying process between their inputs and their outputs over multilayered structure between them. This structure is performing dimensionality reduction in the process, aimed at learning representations of lower level than the inputs [14]. Variational Auto-Encoders (VAEs) represent AEs with a probabilistic twist, which enables them to generate novel data samples that are similar to the inputs [15].

Recurrent Neural Networks (RNNs) generalize forward signal flow in MLPs to structures with feedback loops, so that their current states depend on the inputs as well as on the (temporally) previous states of the network. In that way, neuron states can be considered as memory units, comprising an architecture with a potential to successfully model sequential data [16]. Unfortunately, these networks failed in practice to fulfill the performance expectations when they needed to learn long-term dependencies over the sequences. This main drawback was a consequence of vanishing/exploding gradients phenomenon present in the training process. In order to overcome this issue, the authors of the so-called Long Short-Term Memory (LSTM) networks introduced special units essentially trying to keep constant gradient flow, in addition to gating mechanisms controlling the memory processes [17]. In [18] the authors proposed networks with Gated Recurrent Units (GRUs), representing, roughly speaking, light and trimmed LSTM alternative, showing that complex architecture of LSTMs may not be necessary for obtaining state-of-the-art results. More recently, the focus when working with sequential data has been more and more shifting from RNNs to the so-called Transformer models [19]. Transformers heavily rely on the Attention mechanism [20], which weights the data depending on its importance within a specified context. RNNs can also benefit from the introduction of the Attention mechanism, but they have one significant disadvantage: feedback connections make them inherently unparallelizable, and thus much less convenient in terms of training on very large datasets.

Many deep learning models have great problems when applied on the adversarial data examples (inputs obtained by applying carefully chosen small perturbations to original data), outputting a wrong answer with high confidence [21]. This phenomenon has motivated a research which has resulted in a framework called Generative Adversarial Network (GAN). GANs assume two networks (generative and discriminative) in a game theoretic setting, where discriminative network aims to differentiate synthetic from natural data, while generative network aims to generate synthetic data appearing as natural as possible in order to fool the discriminative network [22]. The two networks are trained simultaneously, which often represents a significant difficulty since gradient-based training methods are not very suitable for the seesaw-like GANs' training process.

It is worth noting that most of the mentioned architectures have been trained or used by researchers both as part of supervised, unsupervised or reinforcement learning, depending on the desired outcome or the functionality and performance of the implemented scheme. A brief overview of features of machine learning approaches is given in **Table 1**.

Table 1

Overview of features of machine learning types.

	Machine Learning type		
feature	**Supervised**	**Unsupervised**	**Reinforcement**
premise	learns by using labelled data	trained by using unlabelled data	interaction with the environment
problem type	classification, regression	clustering, association	exploitation, exploration
purpose	calculate outcome	discover underlying pattern	learn a series of actions
data type	labelled	unlabelled	not predefined
supervision	yes	no	no
algorithms	Nearest Neighbour, Naive Bayes, Decision Trees, Linear Regression, Support Vector Machines, Neural Networks…	K-means clustering, C - means clustering, Association Rules…	Q-Learning, SARSA, Deep Adversarial Networks...
application	forecasts, risk evaluation…	anomaly detection, recommender systems...	vehicular control, healthcare, gaming…

3 Deep Learning Architectures for Mobile Networks

Conventional mobile communication systems do not provide considerable computing resources for data-driven and deep learning applications. The next generation of mobile systems will rely on the following requirements: (a) ultra-low latency (smaller than 1ms); (b) very high communication speed (>1Gbps); (c) low energy consumption; (d) high reliability and service availability [23]. In order to meet criteria for the practical implementation of Big Data and ML applications in the fifth generation of cellular networks, new network architectures are required.

The standard service-driven architecture in 5G networks is based on the cloud implementation. All services, access, transport, core functions and intelligent data analysis are running in the remote data center referred to as the Cloud. This facility coordinates multiple services that require high computing power with latency-aware resource control. Hence, data processing for deep learning applications (e.g. augmented reality on smartphones) in remote cloud data centers cannot meet real-time requirements. In addition, a large number of devices (smartphones, Internet of Things (IoT) devices, tablets, wearables) might affect the quality of service/experience (QoS/QoE) for users when simultaneously using massive data-driven applications (e.g. deep learning based video models) [24].

In order to better support intelligent applications relying on deep neural networks, distributed cloud standards such as fog computing and mobile edge computing (MEC) have been recognized for addressing real-time analysis, scalability, throughput and energy requirements. These standards bring the computational processing power, storage and network resources closer to the end user and data sources in mobile systems. In contrast to the centralized approach, fog computing is an edge-driven concept providing faster responses in data processing tasks and higher quality in the distributed cloud environment hosted on local servers closer to the end-user than the central Cloud servers. Fog nodes extend and complement the cloud concept through distributing heterogeneous resources and managing them in a decentralized way, often acting as filters or pre-processors for the cloud-intended content. The fog computing paradigm has influenced different intelligent services based on deep learning techniques in the next generation mobile systems such as anomaly detection [25]. MEC plays a key role in introducing ML and Big Data applications in the next generation mobile networks. Unlike fog architecture which extends the edge of the network, MEC resources are located within the edge of network (even closer to the devices than the fog and cloud, e.g. base stations). This architecture has shown to be a cornerstone together with ML algorithms and software-defined networks (SDN) for practical implementation of intelligent services within multi-tenant cellular networks. Deep learning models running on the MEC can be used to address different problems such as prediction of number of users in base stations [23], traffic prediction [26], channel state information estimation [27], fault detection for providing low latency and reliable communication [28], cache optimization for mitigating challenges in transporting the big volume of data [29, 30], anomaly detection in the network traffic [31] handover prediction for avoiding errors and improving user experience [32], resource allocation [33]. Fig. 2 presents the network layers and the entities involved in Cloud, Fog, Edge and Mobile computing concepts.

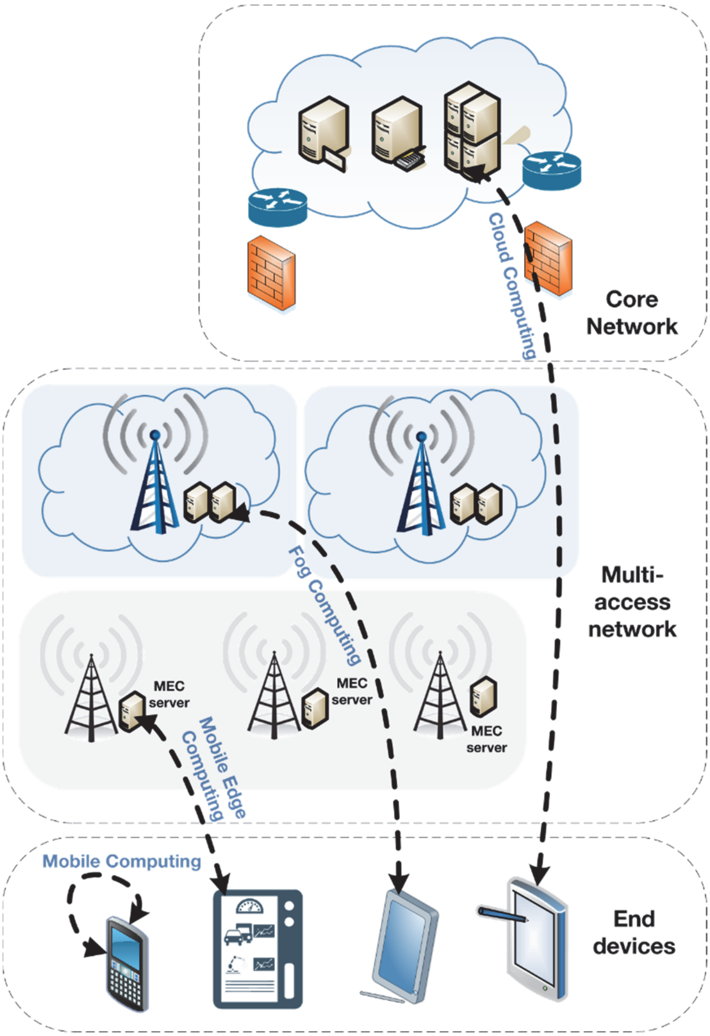


Fig. 2 – *Overview of Mobile, Edge, Fog and Cloud computing.*

Recent innovations have enabled deployment of deep learning models on the end devices known as mobile computing. While previously described computing paradigms rely on the cloud resources, machine learning on-device is a technique which can process data locally. Mobile computing provides quick response time, improved data privacy and lower communication bandwidth, but requires special hardware architectures such as mobile graphic processing units and field programmable gate arrays integrated on end devices (i.e. smartphones, laptops, wearables). Nowadays, many platforms have been used for running deep learning application on end devices such as TensorFlow Lite, PyTorch Mobile, Mobile AI

Compute Engine, Paddle Lite, Core ML, etc. Deep learning on end devices has the potential to create different services related to augmented reality, security, personal health assistance, biometric data analysis and user activity recognition [34]. Bringing AI capabilities to end devices has targeted them as an important platform for future machine learning projects.

4 Deep Learning Applications in Mobile Networks

DL is applied in many aspects of mobile networking. Various methods of implementation of DL techniques are viewed by researchers as ways of dealing with the issues seen at the physical, MAC and network layers of future cellular networks [35]. DRL alone has seen implementations in applications of network access and adaptive rate control, proactive caching and data offloading, network security and connectivity preservation, traffic routing, resource scheduling and data collection [36]. The text that follows describes the aspects of mobile networking which have seen advances and significant benefits by implementation of deep learning. Such aspects of mobile networking, as identified in this paper, are shown in Fig. 3. Figs. 4 – 12 will present the lists of references named in this paper related to the areas of interest shown in Fig. 3.

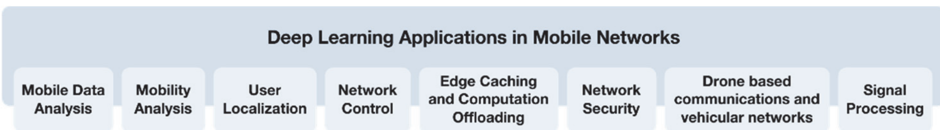


Fig. 3 – *Subject areas in mobile networks that have seen implementation of deep learning applications.*

4.1 Mobile data analysis

A great number of entities originating from and associated to mobile networks present storage points of various, usually huge amount of data. This data can be found in sensors and applications on end devices, user profiles, end user devices themselves, radio information, call records, network performance indicators, metadata of infrastructure elements, etc. All of this makes mobile networks a true source of big data, which is partly stored in the cloud and partly stored at the network edge. Based on this location of data, mobile network data can be categorized into a group of network-level data and application-level data [37].

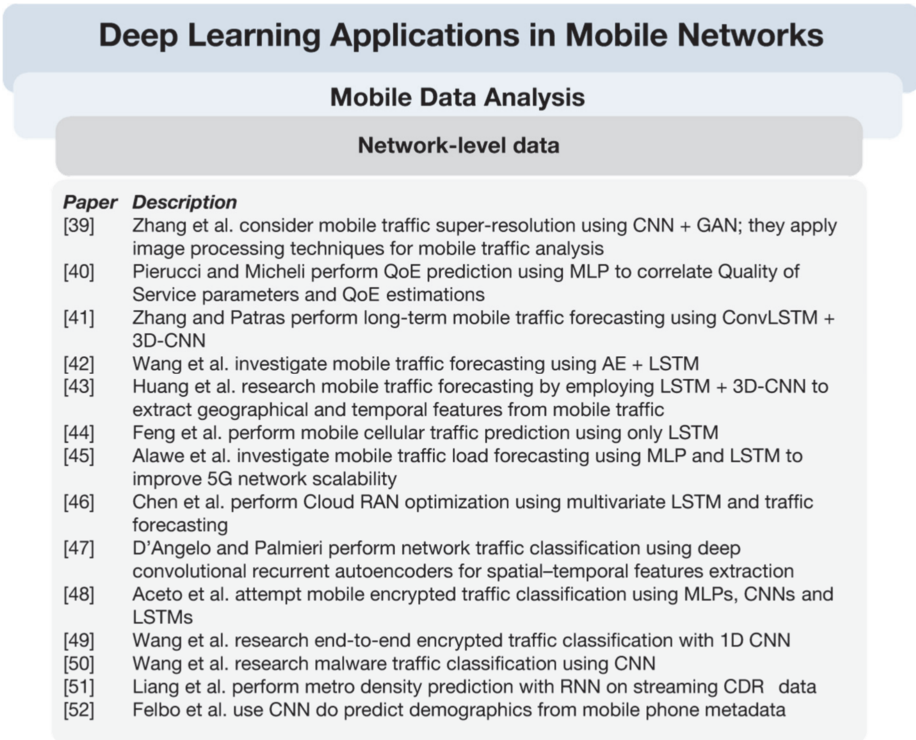
Network-level data

Users' changing behavior is part of the reason for spatial and temporal variations in network-level data [38]. It is exactly this feature that makes such data good basis for the user mobility analysis and public transportation planning [39], as well as for the network analysis and management.

Network-level data are also used to make network state predictions. Such predictions mean the use of historical measurements or other related data in order to estimate certain network performance indicator or traffic requirement. In line with this, authors in [40] employ MLPs and objective metrics such as average user throughput, number of active users in a cell, average data volume per user, and channel quality indicators as their inputs to demonstrate high fidelity of Quality of Experience (QoE) parameter prediction. Deep learning is also applied with the goal of forecasting mobile traffic on a city-scale by using spatio-temporal correlations of geographic mobile traffic measurements [41]. Spatial features extracted by AEs are further processed for temporal modeling by LSTMs in order to provide a definite forecast of the traffic load of a certain cell by authors in [42]. The performance of thusly proposed setup is superior to that of more classical approaches such as Autoregressive Integrated Moving Average (ARIMA) models or Support-Vector Machines (SVM) when tested on real-world datasets. Deep learning applied to CNNs and LSTMs also shows accuracy higher than ARIMA when employed by some authors ([43 – 46]) in order to forecast mobile traffic. Inspired by techniques of image super resolution, authors in [39] have investigated the possibility of combining CNNs with GANs so as to achieve very fine granularity of prediction of network-wide mobile traffic consumption. They used coarse-grained measurements in a setting aimed at measurement overhead reduction and have achieved improvement in measurement granularity of up to 100 times in experiments with real-world datasets.

Another area of research interest of deep learning applications related to network-level data is mobile traffic classification. Authors in [47] employ and compare several deep network combinations, including AEs, CNNs, RNNs and LSTMs, in the task of network traffic classification based on spatial and temporal features extraction. Encrypted traffic classification has been attempted using MLPs, CNNs and LSTMs in [48]. Authors in [49] propose a 1D CNN for the same goal and establish it to be a promising architecture because of its reduced complexity. Treating mobile data as images has enabled the use of CNN structures for the purpose of malware traffic identification in [50]. Within that context, representation learning inherent to CNNs has proved itself helpful in classifying unusual patterns which represent malware traffic. Further advances in application of deep learning for the purposes of spatio-temporal data mining is only to be expected due to the rise of importance of location based services.

Call Data Register (CDR) is only one of the reservoirs of network-level data in a mobile network. Mining of this data can transform it into information. Being well suited to sequential data analysis, RNNs are employed to estimate passenger density in a metro system by considering user trajectories from streamed CDR data as location sequences in [51]. CDR data can also be used for user's gender and age estimation, as is illustrated in [52], where authors employ CNNs for these purposes.



Deep Learning Applications in Mobile Networks

Mobile Data Analysis

Network-level data

Paper	Description
[39]	Zhang et al. consider mobile traffic super-resolution using CNN + GAN; they apply image processing techniques for mobile traffic analysis
[40]	Pierucci and Micheli perform QoE prediction using MLP to correlate Quality of Service parameters and QoE estimations
[41]	Zhang and Patras perform long-term mobile traffic forecasting using ConvLSTM + 3D-CNN
[42]	Wang et al. investigate mobile traffic forecasting using AE + LSTM
[43]	Huang et al. research mobile traffic forecasting by employing LSTM + 3D-CNN to extract geographical and temporal features from mobile traffic
[44]	Feng et al. perform mobile cellular traffic prediction using only LSTM
[45]	Alawe et al. investigate mobile traffic load forecasting using MLP and LSTM to improve 5G network scalability
[46]	Chen et al. perform Cloud RAN optimization using multivariate LSTM and traffic forecasting
[47]	D'Angelo and Palmieri perform network traffic classification using deep convolutional recurrent autoencoders for spatial–temporal features extraction
[48]	Aceto et al. attempt mobile encrypted traffic classification using MLPs, CNNs and LSTMs
[49]	Wang et al. research end-to-end encrypted traffic classification with 1D CNN
[50]	Wang et al. research malware traffic classification using CNN
[51]	Liang et al. perform metro density prediction with RNN on streaming CDR data
[52]	Felbo et al. use CNN do predict demographics from mobile phone metadata

Fig. 4 – *Overview of references in the field of network-level data analysis.*

Application-level data

We call the data collected from various end user devices in a mobile network – application-level data. It only makes sense to use end user devices themselves for collection and preprocessing of such data as they are the carriers of all the sensory equipment used to produce said data [53]. Sensors appearing on modern day end devices include cameras, GPS sensors, accelerometers, but can also include health or body function monitors. These sensors produce heterogeneous data which can be collected through crowd-sourcing schemes in order to be further analyzed. User preferences, communicational patterns and mobility information can be obtained from such data as the data reflects the user profile and thus the user behavior [54]. The potential for use in user targeted marketing systems as well as recommender systems is obvious. There is a specific challenge related to extraction of useful patterns from end-user sensing devices without raising some concerns related to user privacy. This issue represents an ongoing research and debate topic.

In general, application-level mobile data analysis can be implemented using two approaches. One approach is cloud-based computing and the other is edge-based computing; differences between the two arise from different localizations of the actual object performing the computing. While in cloud-based computing the inference is done in a cloud and the results are sent to edge devices, in edge-based computing, edge devices themselves deploy the models and perform related calculations locally.

Application-level data is used in mobile health applications in which physical condition of the users or their surroundings caught by the appropriate sensors are analyzed in order to warn disabled persons (i.e. with hearing disabilities) of dangerous situations [55]. Other applications employ deep learning networks to perform user lifestyle classification based on information captured by the sensory information of end-user devices, or to perform edge-based health monitoring.

Deep Learning Applications in Mobile Networks

Mobile Data Analysis

Application-level data

Paper	*Application*	*Description*
[54]	Network optimization	Zheng et al. investigate big data-driven mobile network optimization framework in 5G networks to enhance QoE performance
[55]	Mobile healthcare	Liu and Du propose a deep learning-driven platform to aid deaf people's awareness of emergencies
[56]	Pattern recognition	Tobias et al. use CNN schemes on mobile devices for objection recognition tasks
[57]	Activity recognition	Almaslukh et al. use AE for cloud-based human activity recognition
[58]	Activity recognition	Bhattacharya and Lane use edge-based deep learning algorithms to perform smart watch-based activity recognition
[59]	Mobile healthcare	Zhao et al. estimating 3D human skeleton from radio frequently signal using CNN in an cloud-based deployment
[60]	Speech recognition	Yoshioka et al. deploy CNNs in a cloud-based environment for the purpose of speech recognition
[61]	Audio sensing	Georgiev et al. employ MLP in an edge-based environment for the purpose of multi-tasking in the process of mobile audio sensing

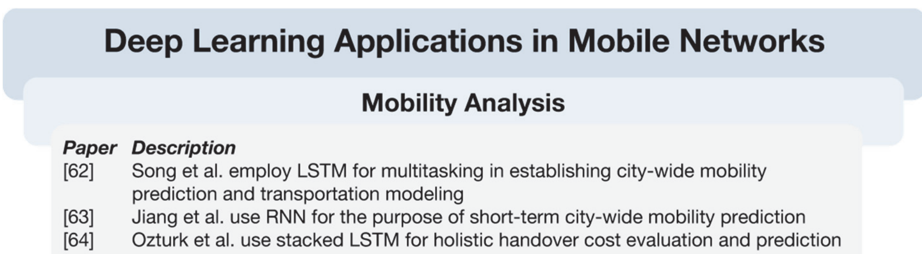
Fig. 5 – *Overview of references covering the area of mobile data analysis of application-level data.*

The area of pattern recognition has attracted a significant level of interest and application-level data is used there for either activity recognition or object classification. For the purpose of object recognition performed on mobile devices some researchers [56] investigate the possibility of employing CNNs. Comparing the task performance metrics on mobile platform to those achieved on GPUs and CPUs, authors demonstrate the possibility of successful mobile device implementation of such a platform. The field of activity recognition in the context of deep learning encompasses a wide range of topics: human activity is recognized by offline analysis of datasets gathered by accelerometers and gyroscopes on smart phones in [57], data collected by sensors on smart watches is performed using RBMs in [58] and a virtual "X-ray" machine based on radio frequency signals analyzed by a 4D CNN framework is designed for the purpose of for human pose recognition in [59].

Another attractive field of interest of researchers related to application of deep learning techniques that make use of application-level mobile data is Automatic Speech Recognition (ASR). Researchers in [60] deploy a specific architecture within a context of CNNs and use multi-microphone devices set in noisy environments to facilitate ASR. A novel deep learning framework is suggested in [61] in order to perform multi-task audio sensing. Algorithms deployed on smartphones show great accuracy while performing the tasks of ambient scene analysis, speaker identification, emotion recognition and stress detection, and achieving high runtime, memory and energy efficiency in the process.

4.2 Mobility analysis

Processes of urban planning, public service provisioning, and network resource management could profit greatly if networks would properly utilize the data of movement patterns of individuals and groups that can be extracted from mobile network data. Deep learning techniques are well suited for such mobility analysis, due to their inherent ability to extract spatial dependencies in sequential data. This quality is used by authors in [62] where they model the joint movement patterns of a large group of people and vehicles on a city-wide scale by employing LSTM networks. Deep RNNs for short-term mobility forecasting have been applied in [63], based on a real-world deployment dataset, showing accuracy superior to traditional shallow approaches. Application of stacked LSTM model as a recurrent deep learning architecture, within the Control/Data Separation Architecture (CDSA) is considered in evaluation of the holistic cost of data-driven handover prediction in [64].


Deep Learning Applications in Mobile Networks

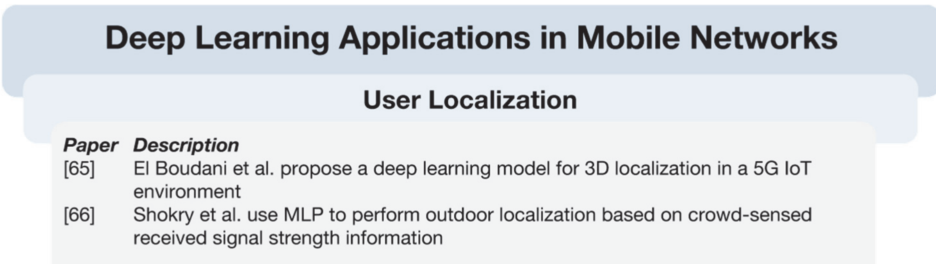
Mobility Analysis

Paper	Description
[62]	Song et al. employ LSTM for multitasking in establishing city-wide mobility prediction and transportation modeling
[63]	Jiang et al. use RNN for the purpose of short-term city-wide mobility prediction
[64]	Ozturk et al. use stacked LSTM for holistic handover cost evaluation and prediction

Fig. 6 – *Overview of references covering the area of mobility analysis.*

4.3 User localization

Real time individual user positions are required for location-based applications and services in mobile network environments. This positioning becomes even more required with the deployment of new mobile networks of fifth generation and beyond. These networks encompass location-based IoT services which can range from targeted advertising to individual users in malls, through finding doctors and patients inside hospitals to localizing firefighters in a building or structure of interest. In [65] authors propose a machine learning model for a new deep learning-based co-operative architecture implemented for improving 3D localization in a 5G IoT environment. While the area of indoor localization does take the majority of research interest, some researchers deal with outdoor localization in cellular networks. Authors in [66] try to learn an MLP of the correlation between the cellular signals and the location of the users and they obtain geo-tagged received signal strength information by using crowd-sensing schemes.


Deep Learning Applications in Mobile Networks

User Localization

Paper	*Description*
[65]	El Boudani et al. propose a deep learning model for 3D localization in a 5G IoT environment
[66]	Shokry et al. use MLP to perform outdoor localization based on crowd-sensed received signal strength information

Fig. 7 – *Overview of references covering the area of user localization.*

4.4 Network control

Improvements to the standard reinforcement and imitation learning processes achieved by incorporation of deep learning paradigms, thus exploiting its powerful function approximation abilities, opened up new horizons to applications in network control, an area known for its complexity. Besides reinforcement learning and imitation learning, network control also applies the so-called analysis-based control. Unlike the reinforcement and imitation learning, analysis-based approach to control does not output actions directly, but rather provides an agent with useful information extracted from the available data for further actions.

Important part of network control processes is network optimization, focusing on management of network resources toward improving overall performance of the entire network. In [54] authors propose big data driven framework for network optimization which analyzes mobile as well as network data and predicts spatial and temporal changes in the network resource requirements. This enables network operators optimal allocation of radio resources which leads to efficient fulfilment of user demand. Traffic load predictions at base stations in ultra-dense networks performed by employing LSTMs are used by authors in [67] to dynamically change the resource allocation policies in order to avoid congestion. Deep RNNs are used for traffic prediction in dynamic spectrum assignment within mobile networks in [68]. Distributed power allocation schemes enabling implementation in large networks and real-world scenarios applicable to 5G networks based on Deep Reinforcement Learning (DRL) have been developed in [69]. Also based on DRL, the notion of powering the small base stations on or off for the purpose of power consumption reduction in ultra-dense networks (UDNs) has been treated in [70].

Network operation under adversarial circumstances also represents an interesting research topic. Authors in [71] propose anti-jamming DRL algorithm to tackle the issue of communication in unknown dynamic adversarial environments. When the algorithm is applied to the proposed recursive CNN with only limited prior knowledge of the environment and raw spectrum information as input, the performance of such an arrangement outperforms the reference deep Q-network setting.

Heterogeneous nature of modern and future mobile networks brings about a set of diverse requirements for quality of service that a single network needs to fulfil. One of the ways that networks tackle this issue is virtualization in the form of network slicing. Network slicing is the separation of the network into several virtual parts, each of which is designed and optimized to fulfil a certain service requirement. Authors in [72] propose a DRL-based scheme and the appropriate architecture for network slicing optimization in order to control allocation of user requests for network slices.

Deep Learning Applications in Mobile Networks

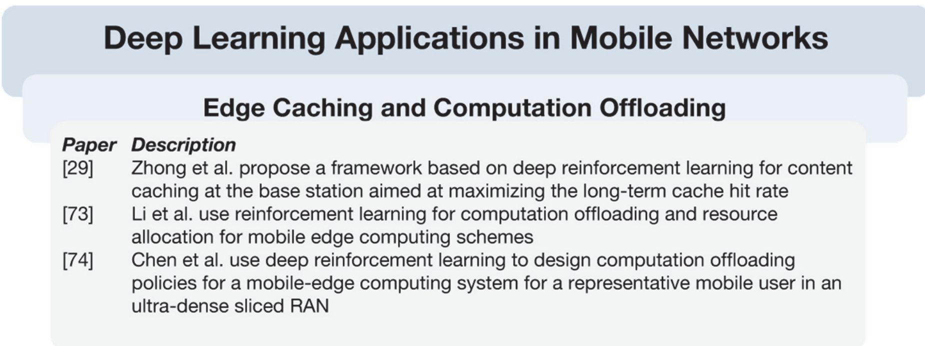
Network Control

Paper	*Description*
[67]	Zhou et al. investigate analysis-based radio resource assignment using LSTM
[68]	Rutagemwa et al. use RNN to perform dynamic spectrum assignment
[69]	Nasir and Guo investigate distributed power allocation schemes in 5G networks based on deep reinforcement learning
[70]	Li et al. research dynamic resource allocation for self-powered ultra-dense networks based on deep reinforcement learning
[71]	Liu et al. employ reinforcement learning in order to establish anti-jamming communications in dynamic and unknown environment
[72]	Xiong et al. propose a DRL-based scheme for network slicing optimization

Fig. 8 – *Overview of references covering the area of network control.*

4.5 Edge caching and computation offloading

In order to support the demand of mobile applications for processing power and data access, modern mobile networks employ both computation offloading as well as data caching at their edge resulting in an improvement quality of service (QoS) and overall energy efficiency. Edge caching of frequently requested content is employed to shorten latency and relieve the network backbone link of excessive traffic loads. Authors in [29] use DRL and try to maximize the long-term cache hit rate for cache replacement decisions within a single base station, which, as an edge node, serves content to multiple mobile users. Offloading of computational tasks from the user equipment by employing resources at the edge of the network is considered in both single server [73] and multi-server environments [74] by using DRL to minimize long-term expected cost of the arrangement.


Deep Learning Applications in Mobile Networks

Edge Caching and Computation Offloading

Paper	Description
[29]	Zhong et al. propose a framework based on deep reinforcement learning for content caching at the base station aimed at maximizing the long-term cache hit rate
[73]	Li et al. use reinforcement learning for computation offloading and resource allocation for mobile edge computing schemes
[74]	Chen et al. use deep reinforcement learning to design computation offloading policies for a mobile-edge computing system for a representative mobile user in an ultra-dense sliced RAN

Fig. 9 – *Overview of references covering the area of edge caching and computation offloading.*

4.6 Network security

With almost ubiquitous presence of modern mobile networks, occurrences of malicious attacks and information leakage are nothing short of expected. The functionality which deals with these challenges and with general protection of network resources and entities from unauthorized access is network security. Such an important function is accomplished by implementation of a combination of software and hardware resources in intrusion detection systems, firewalls and anti-virus software. Such cyber security systems and resources of today employ deep learning in both the unsupervised and the supervised form. Undesired signatures and patterns from previous experience are easily recognized by supervised learning platform, thus enabling detection of future attempts of intrusion. Unsupervised learning platforms, on the other hand, enable identification of patterns which clearly differ from regular and expected network behavior. However, as the security systems improve, so do the activities of the attackers, as they can also employ deep learning schemes for their benefit. This antagonism then resembles the game between generative and discriminative networks in GANs.

Network events that fall out of the category of ordinary or expected behavior might indicate an attack, so the research at the infrastructure level of network security goes in the direction of identifying such events. Network traffic can accurately be categorized using stacked AEs [75], which, when combined with MLPs, can be used for feature selection and extraction [76]. AEs are in the focus also in [77], where they help the detection of abnormal usage of the spectrum.

Issues of botnets in modern mobile networks and smartphone malware are addressed at the software level of network security. These threats can be detected and analyzed by using deep learning schemes. RBMs trained on both labeled and unlabeled mobile applications data are used in [78] to classify Android malware. Some approaches detect malware by means of analysis of certain essential application features. To this end, DBNs can be used to extract the malware features and SVMs to classify the applications quickly and with high accuracy as the authors demonstrate in [79]. Malware can also be detected by CNNs. In [80] CNNs are fed disassembled byte-code of an application as an input text for analysis from which they can learn to detect sequences indicative of malware. LSTMs and MLPs are shown in [81, 82] to be able to detect botnets through extraction of features from behavior of mobile botnets.

Preservation of user privacy is also a concern of network security. In this aspect, deep learning can be employed to ensure privacy is not breached in the use of data for the processes of training and evaluation of the used models. Partitioning of deep learning scheme to run partly in the cloud and partly on mobile devices, whereas the feature extraction is performed locally on the mobile device and the classification is performed in the cloud, is demonstrated to ensure strong enough guaranty of privacy [83].

As mentioned, deep learning networks can be employed not only to safeguard, but also to attack networks. Researchers in this field have used DL, especially on GAN architectures, to enact cyber-attacks aimed at compromising private user information and predicting passwords. Authors in [84] try to perform side channel key recovery attacks using various deep learning architectures. Deep learning based methods, when compared to other template machine learning attacks, prove to be better suited to break the unprotected or protected implementations of the Advanced Encryption Standard (AES). One interesting research [85] shows that the list of active applications run on a smartphone can be very accurately guessed by the CNNs using only data on orientation and data from a magnetometer. Research further shows that in case data from motion sensors is also used, the accuracy levels rise to 98%, thus posing a potential threat to user privacy. However, injection of Gaussian noise to the used data is shown to reduce inference accuracy significantly, which does mitigate the user privacy threat.

Deep Learning Applications in Mobile Networks

Network Security

Paper	Description
[75]	Thing uses stacked AE for IEEE 802.11 network anomaly detection and attack classification
[76]	Aminanto and Kim use MLP and stacked AE to detect Wi-Fi impersonation attacks
[77]	Feng et al. employ AE for spectrum anomaly detection
[78]	Yuan et al. use RBM for Android malware detection in app stores
[79]	Su et al. employ DBN and SVM to perform Android malware detection in app stores
[80]	McLaughlin et al. employ supervised learning of CNNs for Android malware detection in Android MalwareGenome project and Google Play Store
[81]	Torres et al. use LSTM for botnet detection
[82]	Alauthaman et al. use MLP for botnet detection
[83]	Ossia et al. employ CNN for privacy preservation in mobile analytics by offloading feature extraction from cloud to mobile device
[84]	Maghrebi et al. employ various deep learning architectures for the purpose of cryptographic attacks
[85]	Ning et al. research the application of CNN for mobile apps sniffing

Fig. 10 – *Overview of references covering the area of network security.*

4.7 Drone based communications and vehicular networks

Implementation of 5G networks of today has brought about conceptual work on the mobile networks beyond 5G (B5G), including the sixth generation (6G) of mobile networks. These networks are envisaged to offer further more capacity expansion, both in the terms of ultra-high throughput, as well as in the terms of ultra-low latency, so as to deliver new services like haptic applications, augmented and virtual reality, vehicular networks, etc [86]. One in a set of new technologies that will shape the future of new mobile networks, besides THz communications, programmable meta-surfaces, backscatter communication and tactile internet are drone-based communications.

Several operational scenarios that include the use of unmanned aerial vehicles (UAVs) or drones have already been developed for implementation in future networks. There are several ways that drones are seen to play a part in wireless communication of tomorrow. They can be Cellular-enabled Drones (CEDs), operated as user equipment (UE) that can be deployed in high risk applications of mining, oil and gas, transportation, surveying and monitoring, etc. In other applications, drones can be Wireless Infrastructure Drones (WIDs), which extend network capabilities by either extending capacity or coverage or both. Based on their functionality WIDs can be Drone base stations, Aerial Relays or Aerial Backhaul for Cellular Networks [87].

Control of UAVs is one of the major applications of ultra-reliable low-latency communications (URLLC) [88]. A novel framework based on DRL aimed at provision of model-free URLLC in the downlink of an orthogonal frequency division multiple access (OFDMA) system is proposed in [89].

Besides drone based communication, ML, and DL especially, is seen as a promising technology to deal with several aspects of Vehicular Networks that are developed as part and alongside 6G mobile networks – vehicular communication, vehicular networking and vehicular security [90].

Deep Learning Applications in Mobile Networks

Drone based communications and vehicular networks

Paper	*Description*
[88]	Ali et al. speak of machine-learning aide control of UAV from the aspect of CoCoCo (communication and control co-design) aspect
[89]	Kasgari and Saad propose a deep reinforcement learning framework for model-free ultra reliable low latency communication
[90]	Tang et al. provide a survey on various ML techniques applied to communication, networking, and security aspects of vehicular networks and envision the ways of enabling AI toward a future 6G vehicular network

Fig. 11 – *Overview of references covering the area of drone based communication and vehicular networks.*

4.8 Signal processing

Signal processing domain, mostly the area of Multiple-Input-Multiple-Output (MIMO) processing and modulation, is proven to be fertile ground for application of DL techniques. They can help optimize performance of MIMO systems based on environmental conditions. Research in [91] shows that MLP based networks, when considered as possible estimators of transmitted vectors in a MIMO channel, achieve high accuracy while requiring relatively small computation. CNNs as well as MLPs are also used by researchers for channel estimation and signal detection purposes.

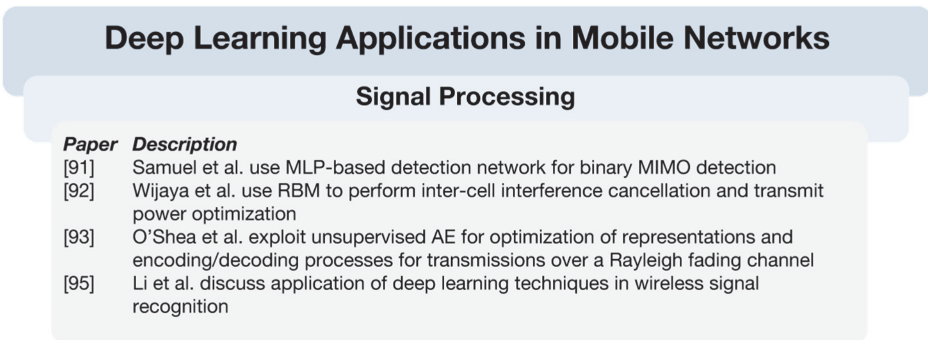
The notion that control of transmit power of base stations by using non-iterative neural networks can prevent inter-cell interference is discussed in [92]. It assumes that a neural network estimates the transmit power optimal for the transmission of every packet, and simulations confirm the performance to be better than that of a belief propagation algorithm that has been commonly used.

Application of DL in network physical layer design is investigated by researchers in [93]. They consider incorporating a deep AE into a single-user end-to-end MIMO system, in which the transmitter is formed of MLP followed by a normalization layer. Channel is presented by a Gaussian noise layer and another

MLP forms the receiver. Such an AE system is shown to considerably outperform the alternative Space Time Block Code approach in the experiment.

Ever-increasing demand for resources by the new 5G and IoT networks poses a significant challenge to the spectrum utilization [94]. For the purpose of spectrum management, spectrum sensing schemes, as well as signal recognition mechanisms gain considerable interest [95]. Wireless signal recognition (WSR) usually means modulation recognition (MR), or automatic modulation classification (AMC), and wireless technology recognition (WTR).

When the task of modulation classification is concerned, CNNs have been shown to provide satisfying accuracy. However, tests show that DL platforms employing LSTMs are the best candidates for performing this task. In the chapter that follows, we will present a use case of modulation classification based on such DL platforms.


Deep Learning Applications in Mobile Networks

Signal Processing

Paper	Description
[91]	Samuel et al. use MLP-based detection network for binary MIMO detection
[92]	Wijaya et al. use RBM to perform inter-cell interference cancellation and transmit power optimization
[93]	O'Shea et al. exploit unsupervised AE for optimization of representations and encoding/decoding processes for transmissions over a Rayleigh fading channel
[95]	Li et al. discuss application of deep learning techniques in wireless signal recognition

Fig. 12 – *Overview of references covering the area of signal processing.*

5 Deep Learning-based Modulation Classification Use Case

In this section, we shall present one solution to a very popular problem of modulation classification of radio signals using deep neural network models. This task is well suited for showcasing different DL architectures and their advantages over shallow ML approaches. DL methods allow for the so-called "blind" temporal learning, wherein time samples of the signals are directly fed into the network. We shall train our models on publicly available time series dataset RadioML2016.10a [96] and build upon some of the existing results ([97, 98]) to offer a more comprehensive comparative evaluation of different DL models as well as an insight into the possibility of improving these results by ensemble averaging.

The problem we are solving is modulation recognition among 11 possible modulation types in the dataset of complex (with real and imaginary parts) time-domain signals 128 samples long. Based on the dataset, we differentiate between

20 signal-to-noise ratios (SNRs). The 2×128 vectors are fed as inputs to different DL models which are trained as convenient end-to-end solutions. We use Google Colab's TensorFlow framework and Keras library for designing and training our models. We compare performance of the following neural networks:

- Baseline convolutional model (labelled as CNN), with two convolutional layers consisting of 64 1×3 and 16 2×3 filters, respectively, followed by a fully connected layer with 128 units before a softmax layer. The model uses dropout regularization layers with 0.5 probability in between the aforementioned layers.
- Deeper convolutional model (labelled as CNN_deep), with four convolutional layers consisting of 256 1×3, 256 2×3, 80 1×3 and 80 1×3 filters, respectively, followed by a fully connected layer with 128 units before a softmax layer. This model also uses dropout regularization layers, similarly as the previous model.
- Deeper convolutional model with an LSTM layer (labelled as CNN_deep_LSTM), with four convolutional layers consisting of 128 1×3, 64 2×3, 32 1×3 and 16 2×3 filters, respectively, followed by an LSTM layer with 64 units and a fully connected layer with 128 units before a softmax layer. The dropout regularization layers are also used.
- Deeper convolutional model with an LSTM layer and ensemble averaging (labelled as CNN_deep_LSTM_ens_cons), which represents an ensemble of three CNN_deep_LSTM models, each trained on a different augmented dataset of same size as the original one, obtained by time domain approaches of window cropping combined with flipping [99]. The final results are obtained by averaging the individual softmax results; the averaging procedure can be efficiently distributed in the case of a large number of models used in the ensemble by an ad-hoc consensus algorithm, following similar line of thought as in [100].

The obtained results are given in Fig. 13. We show the top-1 classification accuracy on test data (30% of the original dataset) for different models and different SNRs of the modulated signals. It can be seen that the introduction of the LSTM layer improves the classification performance and that ensemble averaging can further increase the accuracy. We have also experimented with a network with residual layers but were unable to successfully tune its parameters that would give comparable performance results. In order to obtain a better insight into the performance, we show the confusion matrix for the classification results of the ensemble model in the case of the highest SNR in Fig. 14. It can be seen that the obtained results are very good and that further improvement is possible, since e.g. QAM16 and QAM64 signals are relatively often mistaken for each other, as are WBFM for AM-DSB.

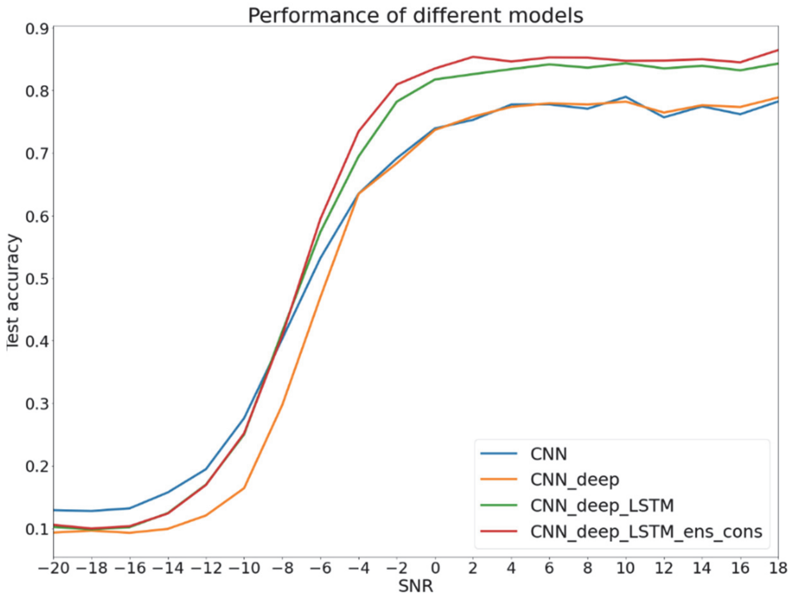


Fig. 13 – *Performance comparison of different deep learning models with respect to different SNRs.*

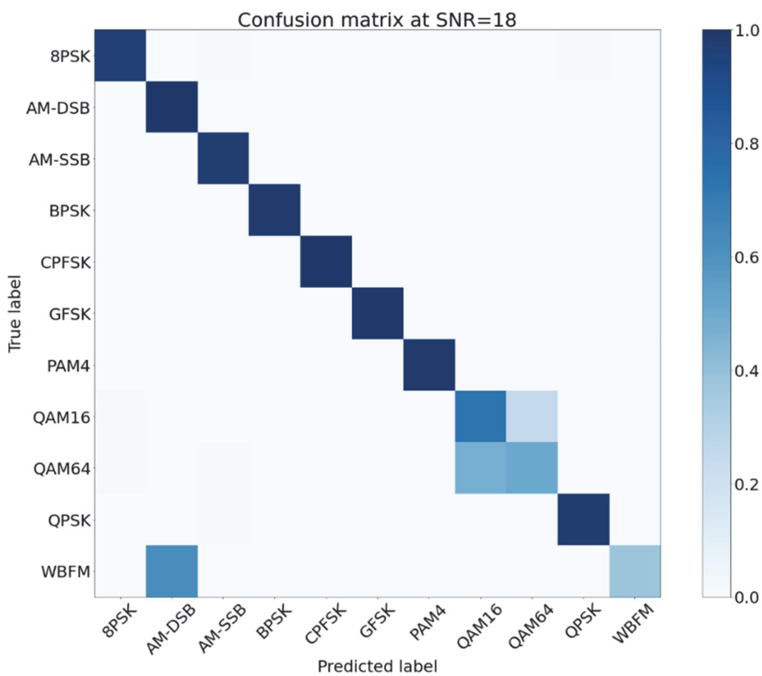


Fig. 14 – *Performance of CNN deep_LSTM_ens_cons model with respect to different modulation types.*

It is to be emphasized that our intention has been only toward showing the potential of the DL approaches in applications related to cognitive radios. The obtained results clearly show the benefits of using DL models, since the manually designed features based on advanced signal processing techniques combined with standard ML models [97] have been shown to have inferior performance. Therefore, DL approaches currently present themselves as one of the primary tools for radio signals classification.

6 Conclusion

This paper provides a comprehensive overview of main ideas of DL concepts in new generation of mobile networks. In addition, we present architectures for the implementation of applications and services in mobile networks. Our idea is to provide information and help researchers and practitioners in designing new generation of services in mobile environments based on deep learning techniques. As advantageous as deep learning techniques may be in several named areas of mobile networking, some drawbacks do restrict its applicability. Deep learning models are known for low interpretability and require massive volumes of data for training.

However, besides the well-known application of deep learning in creation of massive datasets of high quality, we see other rising areas in which deep learning is seeing implementation. Because of the importance of location based services, advances in the field of application of deep learning for the purposes of spatio-temporal data mining can be expected. Due to the emerging idea of drone based networking and vehicular networks as part of next generation of ubiquitous mobile networks, we foresee a rise in contributions of deep learning research to both network and vehicle control applications. Due to the proclaimed notion of omnipresence of future mobile networks, we expect the extension of efforts of deep learning applications in areas of fog and edge computing, as well as network security and signal processing.

As a practical use-case scenario related to the latter, and in order to investigate the efficiency of deep learning compared to more traditional setups, we have shown through simulation, that, when applied for the purpose of modulation classification, implementation of a deep CNN model with an LSTM layer and ensemble averaging significantly improves the results obtained by using shallower neural network schemes.

7 Acknowledgments

This work was supported by Vlatacom Institute and in part by the Science Fund of the Republic of Serbia under Grant 6524745, AI-DECIDE.

8 References

[1] I. Goodfellow, Y. Bengio, A. Courville: Deep Learning, The MIT Press, Cambridge, 2016.

[2] C. Zhang, P. Patras, H. Haddadi: Deep Learning in Mobile and Wireless Networking: A Survey, IEEE Communications Surveys & Tutorials, Vol. 21, No. 3, Thirdquarter 2019, pp. 2224 – 2287.

[3] M.A. Alsheikh, D. Niyato, S. Lin, H.- P. Tan, Z. Han: Mobile Big Data Analytics Using Deep Learning and Apache Spark, IEEE Network, Vol. 30, No. 3, June 2016, pp. 22 – 29.

[4] Pedro Domingos: A Few Useful Things to Know About Machine Learning, Communications of the ACM, Vol. 55, No. 10, October 2012, pp. 78 – 87.

[5] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, J. P. Jue: All One Needs to Know About Fog Computing and Related Edge Computing Paradigms: A Complete Survey, Journal of Systems Architecture, Vol. 98, September 2019, pp. 289 – 330.

[6] D. Dašić, M. Vučetić, G. Hew A Kee, M. Stanković: Deep Learning Applications in Mobile Networks, Proceedings of the International Scientific Conference on Information Technology and Data Related Research (Sinteza 2019), Novi Sad, Serbia, April 2019, pp. 553 – 560.

[7] R.S. Sutton, A.G. Barto: Reinforcement Learning: An Introduction, 2nd Edition, The MIT Press, Cambridge, 2018.

[8] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.- C. Chen, L. Hanzo: Machine Learning Paradigms for Next-Generation Wireless Networks, IEEE Wireless Communications, Vol. 24, No. 2, April 2017, pp. 98 – 105.

[9] Z.- H. Zhou, J. Feng: Deep Forest: Towards an Alternative to Deep Neural Networks, Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI-17), Melbourne, Australia, August 2017, pp. 3553 – 3559.

[10] A.C. Damianou, N.D. Lawrence: Deep Gaussian Processes, Proceedings of the 16th International Conference on Artificial Intelligence and Statistics, Scottsdale, USA, April 2013, pp. 207 – 215.

[11] R. Collobert, S. Bengio: Links between Perceptrons, MLPs and SVMs, Proceedings of the 21st International Conference on Machine Learning (ICML '04), Banff, Alberta, CA, July 2004, pp. 177 – 184.

[12] A. Krizhevsky, I. Sutskever, G. E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, Communications of the ACM, Vol. 60, No. 6, June 2017, pp. 84 – 90.

[13] G. E. Hinton, S. Osindero, Y.- W. Teh: A Fast Learning Algorithm for Deep Belief Nets, Neural Computation, Vol. 18, No. 7, July 2006, pp. 1527 – 1554.

[14] Y. Bengio: Learning Deep Architectures for AI, Foundations and Trends® in Machine Learning, Vol. 2, No. 1, January 2009, pp. 1 – 127.

[15] C. Doersch: Tutorial on Variational Autoencoders, ArXiv, January 2021, pp. 1 – 23, Available at: https://arxiv.org/abs/1606.05908.

[16] I. Sutskever, O. Vinyals, Q. V. Le: Sequence to Sequence Learning with Neural Networks, Proceedings of the 27th International Conference on Neural Information Processing Systems - Advances in Neural Information Processing Systems (NIPS 2014), Montreal, Ca, Vol. 2, December 2014, pp. 3104 – 3112.

[17] F.A. Gers, J. Schmidhuber, F. Cummins: Learning to Forget: Continual Prediction with LSTM, Proceedings of the 9th International Conference on Artificial Neural Networks (ICANN ’99), Edinburgh, UK, September 1999, pp. 850 – 855.

[18] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio: Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, October 2014, pp. 1724 – 1734.

[19] T. Wolf et al.: Transformers : State-of-the-Art Natural Language Processing, Proceedings of the Conference on Empirical Methods in Natural Language Processing: Systems Demonstrations (EMNLP 2020), Online, October 2020, pp. 38 – 45.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L, Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin: Attention is All you Need, Proceedings of the 31st International Conference on Neural Information Processing Systems - Advances in Neural Information Processing Systems (NIPS 2017), Long Beach, USA, Vol. 30, December 2017, pp. 6000 – 6010.

[21] I.J. Goodfellow, J. Shlens, C. Szegedy: Explaining and Harnessing Adversarial Examples, Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, USA, May 2015, pp. 1 – 11.

[22] M. Arjovsky, S. Chintala, L. Bottou: Wasserstein Generative Adversarial Networks, Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, August 2017, pp. 214 – 223.

[23] M. Polese, R. Jana, V. Kounev, K. Zhang, S. Deb, M. Zorzi: Machine Learning at the Edge: A Data-Driven Architecture with Applications to 5G Cellular Networks, IEEE Transactions on Mobile Computing, Vol. 20, No. 12, December 2021, pp. 3367 – 3382.

[24] M. McClellan, C. Cervelló-Pastor, S. Sallent: Deep Learning at the Mobile Edge: Opportunities for 5G Networks, Applied Sciences, Vol. 10, No. 14, July 2020, pp. 1 – 27.

[25] L. Fernández Maimó, A. Huertas Celdrán, M. Gil Pérez, F. J. García Clemente, G. Martínez Pérez: Dynamic Management of a Deep Learning-Based Anomaly Detection System for 5G Networks, Journal of Ambient Intelligence and Humanized Computing, Vol. 10, No. 8, August 2019, pp. 3083 – 3097.

[26] X. Zhao, K. Yang, Q. Chen, D. Peng, H. Jiang, X. Xu, X. Shuang: Deep Learning Based Mobile Data Offloading in Mobile Edge Computing Systems, Future Generation Computer Systems, Vol. 99, October 2019, pp. 346 – 355.

[27] M. Mehrabi, M. Mohammadkarimi, M. Ardakani, Y. Jing: Decision Directed Channel Estimation based on Deep Neural Network k-Step Predictor for MIMO Communications in 5G, IEEE Journal on Selected Areas in Communications, Vol. 37, No. 11, November 2019, pp. 2443 – 2456.

[28] P. Hu, J. Zhang: 5G-Enabled Fault Detection and Diagnostics: How Do We Achieve Efficiency?, IEEE Internet of Things Journal, Vol. 7, No. 4, April 2020, pp. 3267 – 3281.

[29] C. Zhong, M.C. Gursoy, S. Velipasalar: A Deep Reinforcement Learning-Based Framework for Content Caching, Proceedings of the 52nd Annual Conference on Information Sciences and Systems (CISS 2018), Princeton, USA, March 2018, pp. 1 – 6.

[30] B. Serbetci, J. Goseling: On Optimal Geographical Caching in Heterogeneous Cellular Networks, Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, USA, March 2017, pp. 1 – 6.

[31] B. Hussain, Q. Du, S. Zhang, A. Imran, M.A. Imran: Mobile Edge Computing-Based Data-Driven Deep Learning Framework for Anomaly Detection, IEEE Access, Vol. 7, September 2019, pp. 137656 – 137667.

[32] H. Abdah, J.P. Barraca, R.L. Aguiar: Handover Prediction Integrated with Service Migration in 5G Systems, Proceedings of the IEEE International Conference on Communications, Dublin, Ireland, June 2020, pp. 1 – 7.

[33] J. Wang, L. Zhao, J. Liu, N. Kato: Smart Resource Allocation for Mobile Edge Computing: A Deep Reinforcement Learning Approach, IEEE Transactions on Emerging Topics in Computing, Vol. 9, No. 3, July 2021, pp.1529 – 1541.

[34] A. Ignatov et al.: AI Benchmark: All About Deep Learning on Smartphones in 2019, Proceedings of the International Conference on Computer Vision Workshop (ICCVW 2019), Seoul, Korea, October 2019, pp. 3617 – 3635.

[35] R. Shafin, L. Liu, V. Chandrasekhar, H. Chen, J. Reed, J.C. Zhang: Artificial Intelligence-Enabled Cellular Networks: A Critical Path to Beyond-5G and 6G, IEEE Wireless Communications, Vol. 27, No. 2, April 2020, pp. 212 – 217.

[36] N.C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.- C. Liang, D.I. Kim: Applications of Deep Reinforcement Learning in Communications and Networking : A Survey, IEEE Communications Surveys & Tutorials, Vol. 21, No. 4, Fourthquarter 2019, pp. 3133 – 3174.

[37] D.Z. Yazti, S. Krishnaswamy: Mobile Big Data Analytics: Research, Practice, and Opportunities, Proceedings of the IEEE 15th International Conference on Mobile Data Management, Brisbane, Australia, Vol. 1, July 2014, pp. 1 – 2.

[38] D. Naboulsi, M. Fiore, S. Ribot, R. Stanica: Large-Scale Mobile Traffic Analysis: A Survey, IEEE Communications Surveys & Tutorials, Vol. 18, No. 1, Firstquarter 2016, pp. 124 – 161.

[39] C. Zhang, X. Ouyang, P. Patras: ZipNet-GAN: Inferring Fine-Grained Mobile Traffic Patterns via a Generative Adversarial Neural Network, Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '17), Incheon, Korea, December 2017, pp. 363 – 375.

[40] L. Pierucci, D. Micheli: A Neural Network for Quality of Experience Estimation in Mobile Communications, IEEE MultiMedia, Vol. 23, No. 4, October 2016, pp. 42 – 49.

[41] C. Zhang, P. Patras: Long-Term Mobile Traffic Forecasting Using Deep Spatio-Temporal Neural Networks, Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc '18), Los Angeles, USA, June 2018, pp. 231 – 240.

[42] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, D. Yang: Spatiotemporal Modeling and Prediction in Cellular Networks: A Big Data Enabled Deep Learning Approach, Proceedings of the IEEE Conference on Computer Communications, Atlanta, USA, May 2017, pp. 1 – 9.

[43] C.- W. Huang, C.- T. Chiang, Q. Li: A Study of Deep Learning Networks on Mobile Traffic Forecasting, Proceedings of the IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, Canada, October 2018, pp. 1 – 6.

[44] J. Feng, X. Chen, R. Gao, M. Zeng, Y. Li: DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction, IEEE Network, Vol. 32, No. 6, November 2018, pp. 108 – 115.

[45] I. Alawe, A. Ksentini, Y. Hadjadj-Aoul, P. Bertin: Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach, IEEE Network, Vol. 32, No. 6, November 2018, pp. 42 – 49.

[46] L. Chen, D. Yang, D. Zhang, C. Wang, J. Li, T.- M.- T. Nguyen: Deep Mobile Traffic Forecast and Complementary Base Station Clustering for C-RAN Optimization, J. Netw. Comput. Appl., Vol. 121, November 2018, pp. 59 – 69.

[47] G. D’Angelo, F. Palmieri: Network Traffic Classification Using Deep Convolutional Recurrent Autoencoder Neural Networks for Spatial–Temporal Features Extraction, Journal of Network and Computer Applications, Vol. 173, January 2021, pp. 1 – 16.

[48] G. Aceto, D. Ciuonzo, A. Montieri, A. Pescapé: Mobile Encrypted Traffic Classification Using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges, IEEE Transactions on Network and Service Management, Vol. 16, No. 2, June 2019, pp. 445 – 458.

[49] W. Wang, M. Zhu, J. Wang, X. Zeng, Z. Yang: End-To-End Encrypted Traffic Classification with One-Dimensional Convolution Neural Networks, Proceedings of the IEEE International Conference on Intelligence and Security Informatics: Security and Big Data (ISI 2017), Beijing, China, July 2017, pp. 43 – 48.

[50] W. Wang, M. Zhu, X. Zeng, X. Ye, Y. Sheng: Malware Traffic Classification Using Convolutional Neural Network for Representation Learning, Proceedings of the International Conference on Information Networking (ICOIN), Da Nang, Vietnam, January 2017, pp. 712 – 717.

[51] V.C. Liang, R.T.B. Ma, W. Siong Ng, L. Wang, M. Winslett, H. Wu, S. Ying, Z. Zhang: Mercury: Metro Density Prediction with Recurrent Neural Network on Streaming CDR Data, Proceedings of the IEEE 32nd International Conference on Data Engineering (ICDE), Helsinki, Finland, May 2016, pp. 1374 – 1377.

[52] B. Felbo, P. Sundsøy, A. Pentland, S. Lehmann, Y.- A. de Montjoye: Using Deep Learning to Predict Demographics from Mobile Phone Metadata, Proceedings of the 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, May 2016, No. 1, pp. 1 – 16.

[53] X. Cheng, L. Fang, X. Hong, L. Yang: Exploiting Mobile Big Data: Sources, Features, and Applications, IEEE Network, Vol. 31, No. 1, January 2017, pp. 72 – 79.

[54] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, W. Xiang: Big Data-Driven Optimization for Mobile Networks Toward 5G, IEEE Network, Vol. 30, No. 1, January 2016, pp. 44 – 51.

[55] S. Liu, J. Du: MobiEar-Building an Environment-Independent Acoustic Sensing Platform for the Deaf Using Deep Learning, Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services Companion (MobiSys'16), Singapore, Singapore, June 2016, pp. 50.

[56] L. Tobias, A. Ducournau, F. Rousseau, G. Mercier, R. Fablet: Convolutional Neural Networks for Object Recognition on Mobile Devices: A Case Study, Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, December 2016, pp. 3530 – 3535.

[57] B. Almaslukh, J. Al-Muhtadi, A. Artoli: An Effective Deep Autoencoder Approach for Online Smartphone-Based Human Activity Recognition, International Journal of Computer Science and Network Security, Vol. 17, No. 4, April 2017, pp. 160 – 165.

[58] S. Bhattacharya, N. D. Lane: From Smart to Deep: Robust Activity Recognition on Smartwatches Using Deep Learning, Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), Sydney, Australia, March 2016, pp. 1 – 6.

[59] M. Zhao, Y. Tian, H. Zhao, M. Abu Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, A. Torralba: RF-Based 3D Skeletons, Proceedings of the Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '18), Budapest, Hungary, August 2018, pp. 267 – 281.

[60] T. Yoshioka et al.: The NTT CHiME-3 System: Advances in Speech Enhancement and Recognition for Mobile Multi-Microphone Devices, Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), Scottsdale, USA, December 2015, pp. 436 – 443.

[61] P. Georgiev, S. Bhattacharya, N. D. Lane, C. Mascolo: Low-Resource Multi-Task Audio Sensing for Mobile and Embedded Devices via Shared Deep Neural Network Representations, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 3, September 2017, pp. 1 – 19.

[62] X. Song, H. Kanasugi, R. Shibasaki: DeepTransport: Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level, Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), New York, USA, July 2016, pp. 2618 – 2624.

[63] R. Jiang, X. Song, Z. Fan, T. Xia, Q. Chen, S. Miyazawa, R. Shibasaki: DeepUrbanMomentum: An Online Deep-Learning System for Short-Term Urban Mobility Prediction, Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI 2018), New Orleans, USA, February 2018, pp. 784 – 791.

[64] M. Ozturk, M. Gogate, O. Onireti, A. Adeel, A. Hussain, M. A. Imran: A Novel Deep Learning Driven, Low-Cost Mobility Prediction Approach for 5G Cellular Networks: The Case of the Control/Data Separation Architecture (CDSA), Neurocomputing, Vol. 358, September 2019, pp. 479 – 489.

[65] B. El Boudani, L. Kanaris, A. Kokkinis, M. Kyriacou, C. Chrysoulas, S. Stavrou, T. Dagiuklas: Implementing Deep Learning Techniques in 5G IoT Networks for 3D Indoor Positioning: DELTA (DeEp Learning-Based Co-operaTive Architecture), Sensors, Vol. 20, No. 19, October 2020, pp. 1 – 20.

[66] A. Shokry, M. Torki, M. Youssef: DeepLoc: A Ubiquitous Accurate and Low-Overhead Outdoor Cellular Localization System, Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '18), Seattle, USA, November 2018, pp. 339 – 348.

[67] Y. Zhou, Z. M. Fadlullah, B. Mao, N. Kato: A Deep-Learning-Based Radio Resource Assignment Technique for 5G Ultra Dense Networks, IEEE Network, Vol. 32, No. 6, November 2018, pp. 28 – 34.

[68] H. Rutagemwa, A. Ghasemi, S. Liu: Dynamic Spectrum Assignment for Land Mobile Radio with Deep Recurrent Neural Networks, Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, USA, May 2018, pp. 1 – 6.

[69] Y. S. Nasir, D. Guo: Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks, IEEE Journal on Selected Areas in Communications, Vol. 37, No. 10, October 2019, pp. 2239 – 2250.

[70] H. Li, H. Gao, T. Lv, Y. Lu: Deep Q-Learning Based Dynamic Resource Allocation for Self-Powered Ultra-Dense Networks, Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, USA, May 2018, pp. 1 – 6.

[71] X. Liu, Y. Xu, L. Jia, Q. Wu, A. Anpalagan: Anti-Jamming Communications Using Spectrum Waterfall: A Deep Reinforcement Learning Approach, IEEE Communications Letters, Vol. 22, No. 5, May 2018, pp. 998 – 1001.

[72] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang, L.- C. Wang: Deep Reinforcement Learning for Mobile 5G and Beyond: Fundamentals, Applications, and Challenges, IEEE Vehicular Technology Magazine, Vol. 14, No. 2, June 2019, pp. 44 – 52.

[73] J. Li, H. Gao, T. Lv, Y. Lu: Deep Reinforcement Learning Based Computation Offloading and Resource Allocation for MEC, Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, April 2018, pp. 1 – 6.

[74] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, M. Bennis: Optimized Computation Offloading Performance in Virtual Edge Computing Systems via Deep Reinforcement Learning, IEEE Internet of Things Journal, Vol. 6, No. 3, June 2019, pp. 4005 – 4018.

[75] V.L.L. Thing: IEEE 802.11 Network Anomaly Detection and Attack Classification: A Deep Learning Approach, Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, USA, March 2017, pp. 1 – 6.

[76] M.E. Aminanto, K. Kim: Detecting Impersonation Attack in WiFi Networks Using Deep Learning Approach, Proceedings of the 17th International Workshop on Information Security Applications (WISA 2016), Jeju Island, Korea, August 2017, pp. 136 – 147.

[77] Q. Feng, Y. Zhang, C. Li, Z. Dou, J. Wang: Anomaly Detection of Spectrum in Wireless Communication via Deep Auto-Encoders, The Journal of Supercomputing, Vol. 73, No. 7, July 2017, pp. 3161 – 3178.

[78] Z. Yuan, Y. Lu, Z. Wang, Y. Xue: Droid-Sec: Deep Learning in Android Malware Detection, ACM SIGCOMM Computer Communication Review, Vol. 44, No. 4, October 2014, pp. 371 – 372.

[79] X. Su, D. Zhang, W. Li, K. Zhao: A Deep Learning Approach to Android Malware Feature Learning and Detection, Proceedings of the IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, August 2016, pp. 244 – 251.

[80] N. McLaughlin et al.: Deep Android Malware Detection, Proceedings of the 7th ACM on Conference on Data and Application Security and Privacy (CODASPY '17), Scottsdale, USA, March 2017, pp. 301 – 308.

[81] P. Torres, C. Catania, S. Garcia, C.G. Garino: An Analysis of Recurrent Neural Networks for Botnet Detection Behavior, Proceedings of the IEEE Biennial Congress of Argentina (ARGENCON), Buenos Aires, Argentina, June 2016, pp. 1 – 6.

[82] M. Alauthaman, N. Aslam, L. Zhang, R. Alasem, M.A. Hossain: A P2P Botnet Detection Scheme based on Decision Tree and Adaptive Multilayer Neural Networks, Neural Computing and Applications, Vol. 29, No. 11, June 2018, pp. 991 – 1004.

[83] S. Ali Osia, A.S. Shamsabadi, S. Sajadmanesh, A. Taheri, K. Katevas, H.R. Rabiee, N.D. Lane, H. Haddadi: A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics, IEEE Internet of Things Journal, Vol. 7, No. 5, May 2020, pp. 4505 – 4518.

[84] H. Maghrebi, T. Portigliatti, E. Prouff: Breaking Cryptographic Implementations Using Deep Learning Techniques, Proceedings of the 6th International Conference on Security, Privacy, and Applied Cryptography Engineering, Hyderabad, India, December 2016, pp. 3 – 26.

[85] R. Ning, C. Wang, C.S. Xin, J. Li, H. Wu: DeepMag: Sniffing Mobile Apps in Magnetic Field through Deep Convolutional Neural Networks, Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, March 2018, pp. 1 – 10.

[86] L. Bariah, L. Mohjazi, S. Muhaidat, P.C. Sofotasios, G.K. Kurt, H. Yanikomeroglu, O.A. Dobre: A Prospective Look: Key Enabling Technologies, Applications and Open Research Topics in 6G Networks, IEEE Access, Vol. 8, August 2020, pp. 174792 – 174820.

[87] M. Mozaffari, W. Saad, M. Bennis, Y.- H. Nam, M. Debbah: A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems, IEEE Communications Surveys & Tutorials, Vol. 21, No. 3, Thirdquarter 2019, pp. 2334 – 2360.

[88] S. Ali et al.: 6G White Paper on Machine Learning in Wireless Communication Networks, ArXiv, April 2020, pp. 1 – 29, Available at: https://arxiv.org/abs/2004.13875.

[89] A.T.Z. Kasgari, W. Saad: Model-Free Ultra Reliable Low Latency Communication (URLLC): A Deep Reinforcement Learning Framework, Proceedingsa of the IEEE International Conference on Communications (ICC), Shanghai, China, May 2019, pp. 1 – 6.

[90] F. Tang, Y. Kawamoto, N. Kato, J. Liu: Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches, Proceedings of the IEEE, Vol. 108, No. 2, February 2020, pp. 292 – 307.

[91] N. Samuel, T. Diskin, A. Wiesel: Deep MIMO Detection, Proceedings of the IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, July 2017, pp. 1 – 5.

[92] M.A. Wijaya, K. Fukawa, H. Suzuki: Intercell-Interference Cancellation and Neural Network Transmit Power Optimization for MIMO Channels, Proceedings of the IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, USA, September 2015, pp. 1 – 5.

[93] T.J. O'Shea, T. Erpek, T.C. Clancy: Deep Learning Based MIMO Communications, ArXiv, July 2017, pp. 1 – 9, Available at: https://arxiv.org/abs/1707.07980.

[94] D. Dašić, M. Vučetić, M. Perić, M. Beko, M. Stanković: Cooperative Multi-Agent Reinforcement Learning for Spectrum Management in IoT Cognitive Networks, Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics, Biarritz, France, June 2020, pp. 238 – 247.

[95] X. Li, F. Dong, S. Zhang, W. Guo: A Survey on Deep Learning Techniques in Wireless Signal Recognition, Wireless Communications and Mobile Computing, Vol. 2019, February 2019, ID 5629572, pp. 1 – 12.

[96] T.J. O'Shea, N. West: Radio Machine Learning Dataset Generation with GNU Radio, Proceedingas of the 6th GNU Radio Conference, Boulder, USA, September 2016, pp. 1 – 6.

[97] T.J. O'Shea, J. Corgan, T.C. Clancy: Convolutional Radio Modulation Recognition Networks, Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN 2016), Aberdeen, UK, September 2016, pp. 213 – 226.

[98] N.E. West, T.O'Shea: Deep Architectures for Modulation Recognition, Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, USA, March 2017, pp. 1 – 6.

[99] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, H. Xu: Time Series Data Augmentation for Deep Learning: A Survey, ArXiv, September 2021, pp. 1 – 9, Available at: https://arxiv.org/abs/2002.12478.

[100]D. Dašić, N. Ilić, M. Vučetić, M. Perić, M. Beko, M.S. Stanković: Distributed Spectrum Management in Cognitive Radio Networks by Consensus-Based Reinforcement Learning, Sensors, Vol. 21, No. 9, May 2021, pp. 1 – 20