Deep Learning for Primary Sector Prediction in FR2 New Radio Systems

ADEL ALDALBAHI[1], (Member, IEEE), **MOHAMMED A. JASIM**[2], (Member, IEEE),
FARZAD SHAHABI[3], (Member, IEEE), **ASIM MAZIN**[4], (Member, IEEE),
NAZLI SIASI[5], (Member, IEEE), AND **DIOGO OLIVEIRA**[6], (Member, IEEE)

[1]Department of Electrical Engineering, College of Engineering, King Faisal University, Al-Ahsa 31982, Saudi Arabia
[2]School of Engineering, University of Mount Union, Alliance, OH 44601, USA
[3]Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA
[4]Moffitt Cancer Center and Research Institute, University of South Florida, Tampa, FL 33620, USA
[5]Department of Physics, Computer Science and Engineering, Christopher Newport University, Newport News, VA 23606, USA
[6]School of Information, Florida State University, Tallahassee, FL 32306, USA

Corresponding author: Mohammed A. Jasim (jaesimad@mountunion.edu)

ABSTRACT Millimeter wave communications technology is an essential component of the *new radio* (NR) standard for standalone 5G networks, i.e., *frequency range* 2 (FR2) bands. This technology provides contiguous bandwidth at the detriment of high path loss and blockage sensitivity. Hence beamforming architectures are leveraged to compensate for the channel impairments. However, beamforming introduces significant challenges in terms of initial beam access and beam adaptation requirements. Namely, the *base station* (BS) and *mobile station* (MS) are compelled to search the entire spatial directions to specify the pointing directions (optimum beamforming and combining vectors) that yield in the strongest impinged signal levels. This search process results in a high computational complexity, extended delay periods, high power consumption and energy inefficiency. Hence this paper proposes a novel sector (beam) prediction scheme that leverages the synergistic combination of *convolutional neural networks* (CNN) and *long short-term memory* (LSTM). The goal is to propose ultra-low access times for FR2-bsaed 5G networks, thus enhancing mmWave bands to operate independently as a standalone network without reliance on *Frequency Range* 1 (FR1) bands, e.g., dual-connectivity. The proposed scheme here predicts the primary sector with the highest popularity class at the BS, which is affiliated with the mostly used beamforming vector. This retrieves information about the sector locations with the highest MS traffic, thereby the BS can eliminate the spatial search over locations of scarce MS densities. Consequently, this process reduces the beam scanning search at the BS, while performing conventional search at the MS. The proposed scheme yields in reduced complexity and access times as compared to prominent existing methods.

INDEX TERMS Beamforming, convolutional neural networks, deep learning, initial access, long short-term memory, millimeter wave communications, uniform planar array.

I. INTRODUCTION

Millimeter wave (mmWave) bands comprise an essential component of the Phase II specifications for *standalone* (SA) 5G systems, i.e., *n*257-*n*261 bands. Namely, the *frequency range* 2 (FR2) of the *new radio* (NR) interface that includes the 24.25-52.6 GHz range with 0.5-3 GHz channelization chunks. The FR2 deployment is aimed as a dual-connectivity with FR1 and as standalone network, as specified by *Third*

Generation Partnership Project (3GPP) in Release 16 [1]. The consideration for mmWave bands supports high data rates for *enhanced mobile broadband* (eMBB), which is attributed to the abundant contiguous channel bandwidths. However, these bands suffer from various challenges. Foremost, path and penetration losses, diffraction, and blockage sensitivity, thus resulting in aggregated signal degradation. Moreover, the use of wide channelization here yields in a high noise power, which cannot be compensated by increasing the transmitted power, attributed to RF safety measures and power transmission regulations.

The associate editor coordinating the review of this manuscript and approving it for publication was Olutayo O. Oyerinde.

Along these lines, beamforming architectures are utilized to compensate for the path loss and noise power by introducing additional link gains. Now this compels the BS and MS entities to deploy directional transmission and reception, as omni-directional mode becomes impractical in FR2 networks due to its reduced directivity. In turn, this requires the BS and MS to search (scan) over all spatial directions, i.e., to specify the optimum beamforming and combining vectors that yield in the highest received signal level. Consequently, this mandates a higher number of beam measurements (computational complexity) and prolonged access times during the initial access procedure, which is required to perform prior to the control- and data-planes.

Nevertheless, initial beam access schemes need to attain reduced times to achieve short control-plane latencies. This is essential in order to meet the latency limits specified by the *International Mobile Telecommunications* (IMT) framework, which is 1 ms for eMBB in 5G systems [2]. Upon this, beam access and adaptation arise as a significant challenge in standalone mmWave systems. Hence it is essential to develop beam access schemes that meet the aforementioned requirements, while consuming reduced power and energy levels.

Various efforts in the literature discusses the problem of initial access in mmWave systems. This includes conventional and training-based schemes, see [3]–[30]. These schemes (surveyed in Section II) lack adopting vital specifications set by the vital FR2 NR standard of 5G systems. For instance, few models consider an omni-directional transmission at the MS, while limiting the beam search process to the BS only. Other proposals are reliant upon multiple BSs, geographical location of the MS, or legacy (sub-6 GHz) bands, hence limiting the standalone operation of mmWave networks, i.e., independently from legacy bands. Further, these schemes yield in relatively minor enhancement to conventional techniques (such as the beam sweeping and exhaustive search). Finally, the schemes lack modeling access times and power consumption metrics in the control-plane. Therefore, it is important to consider the beam access delays in standalone beamforming-based mmWave systems.

Now prevalent 5G systems adopt legacy sub-6 GHz bands as the main carrier, while utilizing mmWave bands as a supplementary carrier, e.g., carrier aggregation (dual bands). However, mmWave networks are projected to operate independently at one stage, without dependence on FR1 (legacy) bands. Nevertheless, this is conditioned upon the development of efficient mmWave radio solutions that exhaust low latencies and provide high reliability. Hence this paper aims to enhance mmWave networks as a standalone component of the 5G FR2 in the NR standard, i.e., a key practical application for upcoming cellular networks.

Along this, this paper proposes a deep learning framework that is solely specific for FR2 networks without reliance on FR1 (sub-6 GHz) bands. Namely, the work aims to develop beam (sector) prediction at reduced access times and power consumption, i.e., each sector is covered by one beam and is represented by an index. The model operates in learning and

training modes, where the goal here is to estimate the best sector index that is associated with a unique beam direction at the BS.

The goal of this work is to predict the sector at the BS with the highest MS densities, termed as the primary sector. This is performed by leveraging deep learning networks. Specifically, the work combines 1D CNN and LSTM networks to leverage sectors densities during past time periods to predict future sectors densities. Hence it predicts the primary sector that contains the highest number of traffic requests. Consequently, during the initial access process, the BS starts the beam scanning process at the primary (highest-popular) sector, then moves to the scarce (least-popular) sectors. The primary sector exhibits the highest MS densities. Therefore, when a new MS joins the network, its high-likely to exist in this primary sector. The prediction process here yields in a reduced beam scanning search at the BS, which consequently reduces the computational complexity and access times, thereby accelerating the initial access process.

The deployment scenarios assume that the BS beamformer radiates non-overlapping patterns, where each pattern covers a sector, and each sector serves multiple MSs of variable traffic volumes distributions. The sectors are affiliated with unique pointing directions, represented by unique beamforming vectors (beam indices), as per the sector model in Figure 1. Moreover, the MS beamformer generates patterns at different directions, that are represented by the combining vectors. In Figure 1, each sector is represented by the unique beam pointing direction of a specific beamwidth, thus creating partially overlapped beams that cover the entire spatial domain. Overall, the result is adjacent pentagon shaped sectors that cover a specific geographical area.
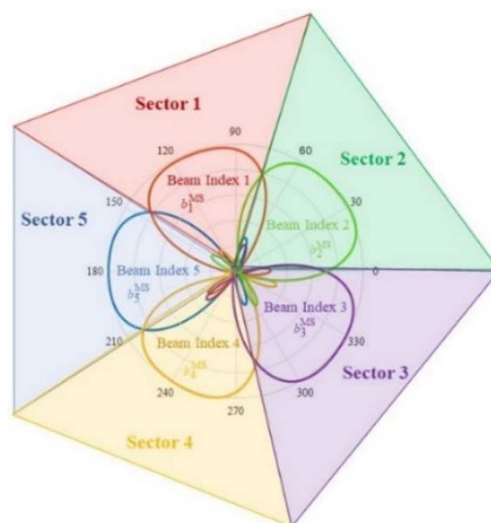

FIGURE 1. Sector Model at the BS.

This paper is structured as follows. existing work on initial beam access schemes is surveyed in Section II. Then, Section III introduces the system model, comprised of the beamforming, signal and channel models. This is followed

TABLE 1. Existing Research on Deep Learning Methods for Beam Management in mmWave Communications.

Beam Access Scheme	Studied Metrics	Advantages	Limitations
Deep neural network for beam prediction at BS [16]	• Algorithm accuracy	• Improves beam search at the BS versus exhaustive scheme	• Omni-directional mode at the MS • No beam discovery at the MS • Lacks time and power models
Deep neural network [17]	• Beam selection accuracy	• Ray-tracing simulations and over-the-air experiments	• Dependent on sub-6 GHz bands • It lacks time and power models
Approximation theory, deep neural network [18]	• Beam-prediction accuracy • Spectral efficiency	• Empirical models for the beam prediction performance	• Dependent on sub-6 GHz bands • It lacks delay measurements
Deep learning model, multi-fingerprint database plurality mechanism [19]	• Instantaneous RSS for various traffic densities • Fingerprint probabilities	• High spectral efficiency	• Lack of delay and power analysis. • Need for multiple BSs • Limited to vehicular networks
Neural network, beam management, interference coordination [20]	• Sum-rate • Time ratio	• High sum-rates • Reduced computational complexity	• Limited to indoor environment • Lack of outdoor scalability • Lack of latency and power simulation.
Random forest classifiers and multilayer perceptrons [21]	• Access times •Prediction accuracy	• Reducing access times • Prediction accuracy	• Requirements for position estimation for UE reliance on lower-frequency links for position feedback signaling.
Super-resolution method for an analog beam selection scheme [22]	• Signal-to-noise ratio (SNR) • Delay overhead	• Low overheads • Accurate beam quality	• Comparison is limited to only brute-force sequential beam searching method
Deep learning for vehicle-to-infrastructure [23]	• Average reward	• Consideration for mobility	• Not tested for outdoor cellular networks • Lack of downlink performance
Federated learning for V2I mmWave [24]	• Prediction accuracy • Throughput	• Reduced network overhead	• Requirements for LIDAR to retrieve side information in the access scheme
Energy-angle domain access and frame structure for V2X [25]	• Probability of false alarm • Probability of detection • Access time and throughput	• Reduced access time • High throughput levels	• Limited to V2X context and lack of beamforming models.
Deep neural network [26]	• Achievable and success rates for various SNR	• Success rates • Consideration for power leakage	• Lack of beamforming, time and power models
Deep neural networks [27]	• Network convergence • Spectral efficiency	• Multi-user beam alignment • Online prediction	• Lack of time and power models/simulation
Machine learning and situational awareness [28]	• Root mean squared and regression errors	• Multi-user beam Alignment	• Location-dependent estimation • Lack of beam delay model
Deep Neural Network [29]	• Effective achievable rate	• High achievable rates at different dataset size	• Use of omni-directional antennas at the MS • Requirements for multiple BSs • Lack of beam estimation at the MS
Convolutional neural network [30]	• Training and validation loss of different depths	• High accuracy for different SNR	• Dependence on sub-6 GHz bands • Lack of beamforming models • Similar spatial features between microwave and mmWave bands

by the proposed deep learning-based access scheme in Section IV. Simulation results and performance evaluation are presented in Sections V and VI, respectively. Finally, conclusions are discussed in Section VII.

II. RELATED WORK

Multiple studies here investigates the problem of initial access and beam management in mmWave communications.

These studies can be classified as *conventional* and *training-based* schemes. The first set of schemes depend on beamforming, network architectures and geometries, whereas the latter depends on machine learning algorithms. Consider the details.

Conventional beam access schemes mandate the BS and MS to perform search over all spatial directions. For example, authors in [3] propose a hierarchical codebook beam structure

that conducts iterative search using wide beams in the initial stage, then refinement is performed in the following stages deploying narrow beams.

The search process is conducted using a compressive sensing algorithm to capture signal clusters in the sparse mmWave channel structure. The work adopts a hybrid beamforming architecture composed of analog phase shifters that feature constant modulus and quantized phases. However, this technique suffers from reduced directivity and blockage sensitivity, attributed to the low gains present in the initial codebook stage. Further, the techniques in [4]–[7] leverage meta-heuristic optimization techniques to accelerate times and minimize the power consumption levels. These techniques are limited to analog beamformers and likewise suffer from a reduced directivity. Also, sidelobe information is exploited in [8] to retrieve the direction of the main lobe. Nevertheless, it is restricted to *line-of-sight* (LoS) configurations of single-ray channels. Grating lobes are purposely generated in [9] for simultaneous transmission to enhance the directivity, where beam coding is implemented to overcome interference and distinguish signals of interest. However, the grating lobes approach exhibit a sophisticated beamforming architecture that is comprised of a high number of antennas, i.e., needed to formulate the multiple array sections. Furthermore, the beam access scheme in [10] leverages a *global positioning system* (GPS) at the MS to estimate the BS location. The GPS here assists the MS to acquire geolocation *context information* (CI), after which it gauges the expected *angle-of-arrival* (AoA) and points the combining beam vector in the direction acquired from the CI. The work adopts an analog beamforming based on *phase shifters network* (PSN) architecture to reduce the angular errors in the acquired CI. However, this method is restricted to outdoor environments in order to facilitate a good GPS connectivity for an improved signal quality. It further compels the BS to perform an exhaustive beam search to find the location of the MS, since the CI-based method cannot be applied at the BS due to the association with multiple MSs of different locations.

Authors in [11] propose a beam training method for multiuser access based on a single RF chain instead for multiple RF chains as the case in massive MIMO, i.e., in efforts to reduce the energy consumption and computational complexity. It proposes *downlink* (DL)-*uplink* (UL) and DL-DL beam training techniques to establish a subset group which is trained in one time-slot. Specifically, one BS maps all the MSs in subset groups according to their request. Here the BS has a single RF chain connected to multiple linear antennas to realize the beamforming structure for multi-user transmission. Further, each MS is supplied with one RF chain that is connected to multiple linear array antennas.

Finally, the work in [12] presents a codebook design based on subarray-cooperation scheme. It features a beam alignment method to reduce the search overhead. Digital beamforming is used for coupling between different subarrays to achieve flat beam patterns. Further, the beamformers here are

configured as equal-size *discrete Fourier transform* (DFT) vectors for hardware simplification. The scheme dynamically chooses initial layers based on a wide range of concurrently detected *signal-to-noise* (SNR) levels. It then aligns the gauged beam pairs to form a single dominant channel. Overall, these schemes still suffer from large computational complexity, excess times and energy inefficiencies.

As earlier discussed, the second set of studies leverages deep learning algorithms for the mmWave beam access and adaptation problem, as summarized in Table 1.

An earlier study in [13] deploys deep learning for beam prediction in *standalone* (SA) beamforming-based mmWave communications. It estimates the subsequent beam for use at the MS and BS in the next time step based on the highest signal level. Along this, the trained network estimates the best beams while eliminating beam scanning at the BS and MS, thus achieving low access times. As opposed to [13] that adopts a single beam prediction mechanism, this paper adopts a sector prediction approach to improve the scalability of initial access, where multiple users can exist in same sector (geographical distribution), i.e., predicting the most popular (congested) sector based on context-popularity modeling. Further, the work in [13] leverages the deep learning at both MS and BS, which adds complexity and power constraints at the MS unit. Hence, the work here limits the adoption of deep learning at the BS only, while leveraging analog beamforming and codebook-based access at the MS. Finally, this paper extends the performance analysis to include popularity classes and prediction success, whereas only the time and energy are studied in [13].

The work in [14] leverages *gated recurrent units* (GRU) deep learning at the BS to predict the link experiencing blockage in the next time frames based upon previous blockage observations. This promotes a proactive handover scheme without the loss of communication sessions. Similarly, authors in [15] use GRU to introduce a dynamic beam scanning sequence. However, it is limited to dataset predictions, lacks beamforming models and operational modes at the BS during beam search. In addition, it lacks studying the impact on access times.

Furthermore, authors in [16] deploy *deep neural network* (DNN) to estimate the best pointing direction at the BS, while adopting an omni-directional antenna at the MS. The goal is to reduce the search complexity as compared to the brute-force approach that scans all directions sequentially. A key limitation here is the use of omni-directional transmission that can reduce the transmission range and degrade the link quality during channel blockage. Another issue is the use of a limited number of beams (24 beams only) at the BS, which yields in wide beams as opposed to the requirements of narrow beams. This is compared to 64 beams in this paper, along with extended performance analysis that includes access times, power and energy consumption. In addition, the results are compared versus the fastest beam access schemes reported in literature. Also, this paper proposes beamforming models at the MS and BS, after which

the deep learning network is used, thus eliminating omni-directional transmission mode.

A DNN-based beam selection strategy is proposed in [17] to reduce the time overhead by employing microwave (sub-6 GHz) channels to retrieve information regarding the best pointing directions. Specifically, the DNN algorithm predicts the *power delay profile* (PDP) of the sub-microwave channel, which is used thereafter as an input for training. The reliance on microwave links to establish mmWave links can limit the latter to operate as an independent network, i.e., impeding the efforts of FR2 in the NR standard. Also, it assumes that the microwave link is established in advance, without incorporating the time required to accomplish this link acquisition. This also yields in an incomplete beam access time model. Therefore, the access delay needs to be gauged from the time an MS joins the network to the initiation of the data-plane. Likewise, the proposal in [18] applies a deep learning framework based on microwave bands to accomplish both initial beam access and blockage recovery in mmWave networks. Mapping functions here are predicted directly from the sub-6 GHz channel, i.e., in efforts to minimize the learning overhead. However, this prediction can be complicated as it requires a large neural network to achieve high levels of accuracy. Further, the reliance on sub-6 GHz bands here burdens the transceivers with an aggregated power demand.

Deep learning is also leveraged for beam alignment in [19] for mmWave vehicular communication. A fingerprinting approach is developed at the BS based on a set of beam pairs that determine a specific fingerprint for each location. Also, a plurality mechanism is developed for the beams that return a high signal strength, thus enabling multiplexing and diversity.

Furthermore, DNN is applied for beam management in indoor *wireless local area networks* (WLAN) such as the IEEE 802.11ay (indoor) mmWave standard in [20]. Namely, beam training is used to establish the directional links between the *mobile access point* (MAB) and *stationary access point* (SAP). These links are then used to generate training data for DNN to mitigate interference, i.e., controlling power levels. This is contingent upon achieving comparable sum rates to the conventional access methods that do not use DNN. Nonetheless, the deep learning network is not applied here for initial beam access, where the association is based on beam training. Further, the implementation is limited to indoor scenarios and not applied for outdoor settings at longer separation distances. Authors in [21] utilize machine learning techniques such as *random forest classifiers* (RFC) and *multilayer perceptron* (MLP) for beam alignment by predicting the optimal *access point* (AP) and a set of candidate beams at the *user equipment* (UE). Further, the positions of the UE are gauged by localization techniques, given the GPS coordinates of the UE. After which the coordinates are fed to the AP and UE as well UE easily for positioning feedback on lower-frequency links. Note that ray-tracing is used to generate, train and evaluate the models. The goal here is to reduce the number of candidate beams needed to acquire the optimal AP, thus reducing access times at high accuracy.

Another use for DNN is applying it in a super-resolution method for an analog beam selection scheme [22]. Beam estimation is concluded based on partial beam measurements that select the beam with the highest SNR. The DNN network here is trained on past beam sweepings to achieve low overheads. Further, the estimation is gauged using a codebook of different beamwidths that cover the entire spatial domain. Results show that this scheme features accurate beam quality estimations at high prediction probability versus brute-force sequential beam search method.

Deep learning is also used in beam access for mmWave-based vehicular networks. Foremost, authors in [23] develop a dataset for beam selection in *vehicle-to-infrastructure* (V2I) networks that use mmWave links. A method for data generation is presented for mobility scenarios using a ray-tracing simulator. However, the context here is specific for V2I networks, i.e., irrelevant to cellular networks (MS and BS association). Further, this work lacks the analysis of the downlink performance, as opposed to this paper that addresses standalone mmWave networks, while analyzing link performance with beamforming paradigms at the BS and MS. Federated learning is leveraged in [24] for V2I that mount LIDAR sensors to assist in the beam selection process. A shared neural network is trained by the connected vehicles leveraging the LIDAR data in the learning stage. Hence, the side information extracted from the LIDAR help the neural network to reduce the network overhead during link configuration. Authors in [25] propose an energy-angle domain access and transmission frame structure for mmWave *vehicle-to-everything* (V2X) communications. The access scheme labels signals by different directions with multi-power level to account for transmission interruptions attributed to blockage. The performance here is evaluated for the probability of false alarm, probability of detection, along with access time and throughput tradeoff, i.e., an optimal access time is achieved that returns high throughput levels.

Authors in [26] leverage deep learning for beam training in mmWave massive MIMO system in efforts to enhance the success and achievable rates at reduced overheads. It predicts the best beam combinations that return the highest signal levels based upon probability vectors. Here the nonlinear properties of power leakage in the channel are leveraged in the estimation process. Key limitations here include the lack of latency and power consumption models, and lack of beamforming models at the BS and MS.

Furthermore, DNN is also applied in [27] to propose a partial-beam alignment method for mmWave massive MIMO system. The work aims to enhance the spectral efficiency and minimize the training overhead versus the hierarchical and compressed sensing-based techniques. Namely, offline training is performed for the channel model, followed by the online prediction stage to estimate distribution vectors using partial beams. Authors in [28] combine deep learning and situational awareness to estimate the power and optimal

beam index. Initially, the AoAs are predicted based upon the location. Then, this information is leveraged as an input to the DNN for beam selection. The dependency on user location information (as a prior knowledge) for training purposes weakens the proposed algorithm and can increase the system complexity as well.

The work in [29] introduces a joint beamforming mechanism between distributed BSs that deploy machine learning to concurrently serve a MS. The latter sends a location signature on a single UL training sequence using an omni (or quasi-omni) beam to the participating BSs. The signatures are then deployed at the deep learning network to predict the beamforming vectors at the BSs, i.e., to reduce the training overhead. However, the use of wide beams here yields in low link gains, which can make the mechanism inefficient during beam blockage scenarios. Finally, a dual (microwave and mmWave) band approach is leveraged in [30] for deep learning-based beam prediction. The scheme aims to minimize the training complexity by predicting the best mmWave link based on out-of-band information from the microwave link. However, the scheme emphasizes on network accuracy without accounting for beamforming and channel models. Another limitation here is assuming analogous spatial characteristics between the channels of the dual bands. This can lead to a reduced accuracy given the dynamics and fluctuations on mmWave links.

Machine learning is also applied for beam access for *unmanned aerial vehicles* (UAVs). Foremost, a beam management method in [31] utilizes *Gaussian process machine learning* (GPML) to predict UAVs positions in efforts to reduce inter-UAV information exchange and reduce network delays. A clustering algorithm of UAVs is proposed for coarse-*angular domain information* (ADI) acquisition. Following the UAV position prediction and clustering, the ML algorithm is then applied again for beam pattern selection for different clusters at different time instants, i.e., to enhance spectrum efficiency.

In other studies, initial access is carried for inactive UE prior to their active mode (joining the network) in [32], i.e., to reduce the delay of uplink beam sweeping. Along this, an uplink multi-beam sweeping technique is proposed using digital beamforming. Further, the multi-beam approach is leveraged to facilitate a rapid beam recovery using backup beam pair. Overall, the work aims to reduce the sweeping cycle and outage probability. A sensitivity study in [33] investigates beam selection at the UE prior to the deterioration of the best beam at the BS (gNodeB). It defines the period over which the BS beam provides a high link quality, termed as the average *time-of-stay* (ToS). This is studied at various mobility ranges and channel dynamics, i.e., set at FR2 operational frequency of 28 GHz. Finally, a beam sweeping procedure in [34] determines the subsequent beam sequence to reduce the average discovery time during initial access. It exploits the position and the correlation among *angles-of-departure* (AoD) of the incoming UE and its closest connected UE. The nearest-neighbor beam search method starts from the beam of the closest UE, and thereafter diverges farther to find the best direction.

III. SYSTEM MODEL
A. ANALOG BEAMFORMING AT THE MS
Various geometries exist for antenna arrays at the MS beamforming structure, such as linear, circular and planar arrangements. The latter provides symmetric radiation patterns with low *side lobe levels* (SLL), enhanced directivity and 2D spatial coverage. This is in contrast to the 1D linear and circular geometries of low directivity.

Moreover, the MS leverages a multi-resolution cascaded codebook composed of $c = 1, 2, \ldots, C$, stages. Here the initial stage performs search using wide beams, then further filtering is carried out in following stages using narrow beams. Each codebook stage c is composed of $i = 1, 2, \ldots, I$, beam directions $\in [0, 2\pi]$, generated i.e., $I = 2\pi/\Phi_{MS}^c$, where Φ_{MS}^c is the *half-power beamwidth* (HPBW) at stage c at the MS.

Hence consider a MS equipped with a *uniform planar array* (UPA), composed of $N_{MS}^c \times M_{MS}^c$ total number of microstrip patch antennas in a square grid along the x and y planes, respectively. Here, $n_{MS} \in N_{MS}^c$ and $m_{MS} \in M_{MS}^c$ are uniformly spaced at d_x and d_y in the x and y directions, respectively. The spacing is set to achieve maximum radiation and directivity, as well as avoiding grating lobes and mutual coupling, i.e., $d_x = \lambda/2$ and $d_y = \lambda/2$, where λ is the wavelength, $\lambda = \alpha/f_\alpha$, where α is the speed of light and f_α is the operating frequency. Now each antenna is associated with a single analog phase shifter, thus forming antenna and phase shifter grids. Then these grids are connected to a single RF chain and baseband unit, i.e., enabling an analog beamformer at the MS that radiates a single beam pattern, as depicted in Figure 2.
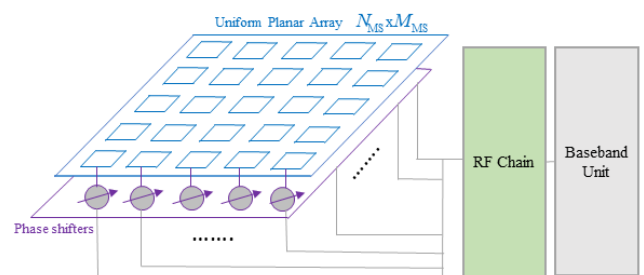

FIGURE 2. Analog Beamformer at the MS.

At each stage c, the overall radiation pattern at the MS is specified by the combining response vector u_{MS}^i which is specified by the array factor, $A_{MS}(\Theta_{MS}^i)$, expressed as,

$$A_{MS}(\Theta_{MS}^i)$$
$$= \Omega \sum_{m_{MS}=0}^{M_{MS}-1} \exp(j\eta(m_{MS}d_x \sin\theta_{MS}^i \cos\phi_{MS}^i + \beta_x),$$
$$\sum_{n_{MS}=0}^{N_{MS}-1} j\eta(n_{MS}d_y \sin\theta_{MS}^i \cos\phi_{MS}^i + \beta_y), \quad (1)$$

where Θ_i^{MS} is the principal main lobe peak direction, Ω is the uniform amplitude excitation for the elements, and η denotes the wavenumber, i.e., $\eta = 2\pi/\lambda$.

The principal direction is expressed by the azimuth θ_i^{MS} and the elevation ϕ_i^{MS}, pointing directions, i.e., $\Theta_i^{MS} = \{\theta_i^{MS}, \phi_i^{MS}\}$. These pointing directions product of two uniform linear arrays, expressed as,

$$\phi_{MS}^i = \arctan \frac{\beta_y d_x}{\beta_x d_y}, \ \theta_{MS}^i$$
$$= \arcsin\left(\sqrt{\left(\frac{\beta_x}{kd_x}\right)^2 + \left(\frac{\beta_y}{kd_y}\right)^2}\right). \quad (2)$$

where the parameters β_x and β_y represent in order the progressive phase shift along the x and y axis in the grid at the MS. The radiation from each element here is combined at the RF stage to form a single beam, since the array is connected to a single RF chain, R_{MS}.

Finally, the HPBW at codebook stage c, Φ_{MS}^c, contains the elevation and azimuth HPBW planes that has the maximum radiation, represented by [35], (3), as shown at the bottom of the page, where the parameter Φ_x represents the HPBW of linear arrays along the x axis, gauged as,

$$\Phi_x = \begin{cases} \cos^{-1}\left(\frac{\lambda}{2\pi d_x}\eta d \cos\theta_{MS}^i \pm \frac{2.782}{N_{MS}^c + M_{MS}^c}\right), \\ \qquad 0 < \theta_{MS}^i \leq \pi, \\ 2\cos^{-1}(1 - 1.391\lambda/N_{MS}^c d_x), \\ \qquad for \ \theta_{MS}^i = 0, \pi. \end{cases} \quad (4)$$

Likewise for the HPBW for linear array along the y axis, Φ_y.

B. DIGITAL BEAMFORMING AT THE BS

Digital beamforming solutions are implemented at the BSs attributed to the available input power supply at the *base band units* (BBU) and *radio remote heads* (RRH). In turn this supports an increased dynamic range and multi-user connectivity. Figure 3 shows the digital beamformer comprised from a grid of uniformly spaced antennas in a planar configuration, which connects to a similar dimensional grid of phase shifters. The output is then connected to an array of RF chains, the processed signals are then fed to the baseband unit. In notations, consider a single BS with a UPA composed of M_{BS}^c and N_{BS}^c antennas in each codebook stage c that are equally spaced along the x and y planes, respectively. Further, each n_{BS} and m_{BS} antenna here($n_{BS} \in N_{BS}^c$ and $m_{BS} \in M_{BS}^c$) is connected to a single RF chain, r_{BS}(i.e., in contrast to the MS structure).

Note that the total number of antenna elements at the BS is equal to the number of RF chains, $N_{BS}^c \times M_{BS}^c = R_{BS}$.
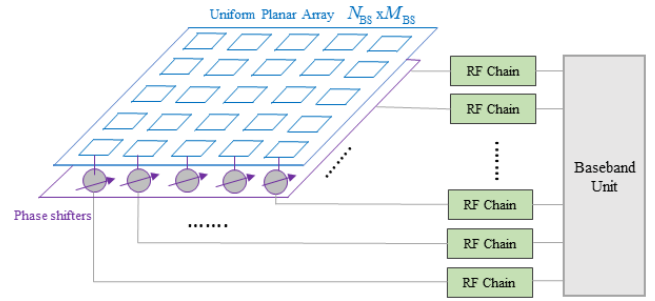

FIGURE 3. Digital Beamformer at the BS.

The overall radiation pattern from each RF chain is formulated by the beamforming vector v_{BS}^j at Θ_{BS}^j pointing direction, i.e., $j = 1, 2, \ldots, J$, where J is the total number of directions at the BS. The beamforming vector v_j^{BS} is specified by the array factor at the BS, $A_{BS}(\Theta_{BS}^j)$, composed of θ_{BS}^j azimuth and ϕ_{BS}^j elevation pointing directions (similar to the MS design), $\Theta_{BS}^j = \{\theta_{BS}^j, \phi_{BS}^j\}$. Note that each sector at the BS is represented by a beamforming vector, v_{BS}^j where the total number of sectors S, is equal to the total number of beamforming vectors, i.e., $S = J$.

C. CHANNEL MODEL

The propagating wavelength at mmWave frequencies is much smaller than objects present in the channel. This promotes the use of geometry-based stochastic channel models. The channel in this model is composed of several distinct rays of certain *angle-of-departure* (AoD) and *angle-of-arrival* (AoA). These angles are attributed to the scattering from objects in the propagation environment, e.g., diffraction, reflection, etc, with small angular spread impinging from scatterers. Here groups of rays propagate closely along a similar path forming a cluster. Note that the rays follow exponential distribution, whereas the clusters follow Poisson distribution. Overall, geometric models construct the channel to be represented by delay and directional domains. This is formed by using the bounce clusters that represent scattering objects, which exhibit low-*bistatic radar cross section* (BRCS) at mmWave frequencies. Along these lines, this channel is formulated as,

$$Q = \sqrt{\frac{(N_{MS}^c M_{MS}^c)(N_{BS}^c N_{BS}^c)}{PL}} \sum_{k=1}^{K}\sum_{l=1}^{L} q_l U_{MS}^c (V_{BS}^c)^{H}, \quad (5)$$

where the term PL stands for the path loss, q_l denotes the gain of the l-th path, as part of L independent and identically distributed (*i.i.d.*) total number of paths. These paths are observed in K clusters. Further, the path gains here follow

$$\Phi_{MS}^c = \frac{\Phi_x \Phi_y}{\cos^2\theta_{MS}^i \sqrt{\left(\sin^2\phi_{MS}^i + \frac{\Phi_y^2}{\Phi_x^2}\cos^2\phi_{MS}^i\right)\left(\sin^2\phi_{MS}^i + \frac{\Phi_x^2}{\Phi_y^2}\cos^2\phi_{MS}^i\right)}}, \quad (3)$$

Rayleigh distribution, $q_l \sim (0, \overline{P}_r)$, where $\overline{P}_r$ is the average received power. Also, U_{MS}^c and V_{BS}^c in Eq. (5) denote the combining and beamforming matrices composed of u_{MS} u_{MS}^i and combining and v_{BS} beamforming vectors, i.e., expressed by the array factors at $(\theta_{BS}^j, \phi_{BS}^j)$ and $(\theta_{MS}^i, \phi_{MS}^i) \in [0,2\pi]$ directions. The beamforming matrix here is composed of the baseband V_{BB} and RF (analog) V_{RF} stages, $V_{BS}^c = V_{BB}V_{RF}$. Furthermore, the term PL is calculated using the floating intercept model based on the least-square fit regression approach [36] for outdoor environments as,

$$PL[dB] = \alpha + \rho 10\log_{10}(r) + \xi, \tag{6}$$

where α, ρ and ξ denote the floating intercept, slope, and shadow fading, respectively, i.e., $\xi \sim \log N(0, \sigma_s)$ with σ_s standard deviation.

D. SIGNAL MODEL

Assume a single BS and MS operate in a *time division duplexing* (TDD) mode of known *channel state information* (CSI). Along this, the DL signal at the analog combiner at the MS is expressed as,

$$y_{RF} = V_{BS}^c Qz + w, \tag{7}$$

where the variables Q and z the complex channel and reference control signal, respectively. Moreover, the variable w in Eq. (7) represents the additive white Gaussian noise (AWGN), i.e., $w \sim N(0, \sigma_w^2)$, with σ_w^2 variance. Meanwhile, the received signal at the analog combiner at the MS, U_{MS}^c, is expressed as,

$$y_{BB} = P_{tr} \left(U_{MS}^c \right)^H V_{BS}^c Qz + \left(U_{MS}^c \right)^H w, \tag{8}$$

where P_{tr} denotes the transmitted signal power.

IV. CNN-LSTM DEEP LEARNING ALGORITHM FOR PRIMARY SECTOR PREDICTION

In traditional network operation without deep learning, the BS periodically broadcasts synchronization signal blocks sweeping over all directions using BS beams, where the beams correspond to sectors. Likewise, the MS sweeps the beam directions and measures the reference signal received power for each beam pair combination. Following this process (BS-MS beam sweeping over all the combinations), the best beam pair is attributed to the best sector at the BS, where it is called the primary sector at the BS.

Meanwhile with deep learning, instead of performing conventional scanning at the BS, it relies on deep learning to estimate the primary sector or the best beam, whereas the MS still conducts conventional scanning. Hence the goal of the paper is to eliminate scanning at the BS only. However, extending the deep learning model to the MS is also feasible in similar settings.

Along this, the LSTM model predicts the received signal power variations of each sector in advance. Hence the BS starts pointing its beam towards the primary sector, i.e., the sector that resulted in the highest received signal at the BS

during the period of the ground truth. The model then determines the specific sector at which the MS exists and thus the BS points its beam towards the MS. However, the MS still needs to scan the spatial directions to points toward the BS, which is the main contributor to the added complexity and prolonged access times in the results. Due to the high density of MS users in highly dense sectors, more MS users are likely to transmit signals that are detected at the BS, as compared to sectors with less density.

Hence, the decision is made based on the traffic density and the likelihood of the highest received signal level detected at the BS from congested sectors. Hence, when a new MS joins the network, it is anticipated it will join the most congested (primary) sector that has the strongest demodulation reference signal during cell search. An example here is a congested zone in a specific metropolitan area. Eventually, the BS will direct one of its arrays towards the sector that has the highest MS density that radiate the most signals. Meanwhile, other sectors may not return any signal level due to the lack of MS users there.

The prediction scheme in this work combines *convolutional neural networks* (CNN) and *long short-term memory* (LSTM), proposed in [37] and [38], respectively. This allows a time series estimation of variable sequence lengths. First, CNN features local spatial coherence in the input that allows fewer weights, i.e., the convolution process here allows the extraction of relevant local information at a low computational cost. Despite this saliency, CNN fails to capture long-term dependencies. Therefore, it is enhanced with LSTM to leverage the use of memory gates. Hence, the network is composed of CNN and LSTM combined layers. Further, the selection of LSTM technique is attributed to its robustness to time gap length, as compared to conventional *recurrent neural networks* (RNN) that suffers from gradient vanishing problem, such as hidden Markov generative models. In addition, it can learn long-term dependencies of the network states based upon the long-term memory, i.e., cell memory state.

Overall, this synergistic combination enables complex time series prediction, where the CNN layer extracts different features between several variables that affect the sector prediction. It further considers correlation between multivariate variables. Meanwhile, the LSTM layer models temporal information of irregular behavior in the time series components. This combination supports LSTM to acquire additional contextual information about the beam index as compared to using conventional LSTM only. Along these lines, the prediction scheme combines CNN and LSTM in a four-layers chain to acquire an accurate prediction solution based on variable sequence lengths.

Now consider a BS with $s = 1, 2, \ldots, S$ number of sectors covered by $v_{BS}^c = 1, 2, \ldots, V_{BS}^c$ total beamforming vectors, serving $e = 1, 2, \ldots, E$ number of MSs exist within each sector. The purpose is to identify the primary sector s_{pri} with the highest traffic volumes, $\Delta(s)$, $s_{pri} = max\{\Delta(s)|\forall s \in S\}$. Hence, when a MS joins the network, it is anticipated that it will be joining the primary sector, at which the highest

received signal level is received. Thus, the BS predicts future primary sectors and their affiliated primary beamforming vectors. As a result, the proposed scheme reduces the requirements for sequential or random beam scanning.

The CNN-LSTM scheme continues to recursively process sequences of dynamic lengths during various time periods. For training over subsequent periods such as the $(t + 2)$th period, the MS distributions can change dramatically over time in real-time implementation. Then the model will require retraining using new dataset. When the training occurs to estimate outcomes for the $(t + 1)$th period, the network also records a new ground truth to be used latter for predicting the outcomes over the $(t + 2)$th period. Overall, the prediction results of the previous periods are not used as new dataset for the next periods to achieve a higher accuracy by avoiding aggregate errors, i.e., despite the low loss model achieved in this paper. Therefore, the model needs to retrain a new ground truth to account for new changes in the traffic patterns at any given sector, when the system is implemented in real-time scenario.

A. PROBLEM FORMULATION

The primary sector problem is formulated as an estimation of the best sector at $(t + 1)$th time step of Δ_{t+1} traffic volume. This value is gauged by knowing the sector status G_t at the (t)th time step of Δ_t traffic volume. The latter represents the ground truth, i.e., the vector of the traffic volume in each sector. Along this, the goal is to increase the success probability of sector prediction at the BS, formulated as $\mathbb{P}[\hat{G}_t = G_t]$.

B. SECTOR PREDICTION OPERATION

The BS aims to predict the sector that is most likely to accommodate the highest traffic volume, at which a new MS will probably be located when joining the network at the next time step, $(t+1)$th. This is accomplished by knowing the sequence of sectors with highest traffic volumes over $t = 1, 2, \ldots, T$ time steps. Overall, the network operates in the two learning and training modes.

Learning Mode (Mode I): The network runs in normal settings, where beam scanning is conducted using conventional schemes such as codebook-based refinement. During this mode, the BS feeds the primary sector index information at every time step to the CNN-LSTM network to be used in Mode II. After the model is well-trained, the BS leverage it to estimate the next best sector, as explained next in the deep learning model in Mode II.

Training Mode (Mode II): Following the completion of Mode I, now the sequences of beam indices (sectors) with the highest selection over time t are available. Along this, the BS next predicts the subsequent primary sector that is most likely to be used at the subsequent time period $(t+1)$th. Specifically, the prediction scheme deploys parametric information from previous time steps, after which it labels the subsequent steps, thus predicting the sector index that yields in the strongest signal level. The CNN-LSTM method here recursively processes

beam sequences at every time step of the input. Afterwards, it keeps a hidden state that is a function of the previous state and the current input. The detailed operation of the model in Mode II is now presented.

C. DEEP LEARNING MODEL

The key elements of the proposed reconfigurable deep learning (CNN-LSTM) model are comprised of the input and training phases, as presented next.

The network learns the input data from Mode I, upon which spatial characteristics of the beam indices (sectors) are retrieved. Namely, these multivariate variables are acquired from the convolution and pooling layers of the CNN layer and passed to the LSTM layer. The LSTM layer models the irregular time information using the transmitted spatial features. Specifically, it consolidates memory units that update the previous hidden state, thus preserving long-term memory. The memory unit (cell) in LSTM stores the information across extended time steps controlled by the adaptive multiplicative gates. The input and output flows of a cell activation is modulated by the input and output gates of a memory cell. Further, a forget gate is used to reset the self-recurrent value when it becomes irrelevant. Note that the forget gate uses binary settings to delete (0) or retain (1) values for the subsequent step by multiplying it with a memory cell [39].

Input Phase: This phase achieves labeled data and is composed of the input vector, i.e., detected primary sector from Mode I over time, and the output vector that indicates the sector popularity class. Namely, the input to the deep learning model at each time index (step) is the primary sector. This is denoted by an input index, φ_t. Then the model starts with a layer than maps every index to an input vector γ_t^{in},

$$\gamma_t^{in} = embed(\varphi_t), \tag{9}$$

where the term *embed* denotes the lookup table developed during the training mode.

Training Phase: The CNN-LSTM network forms the processing unit/core of the deep learning model. It is comprised of an input layer that accepts sector information as inputs, an output layer that extracts features to LSTM, along with several hidden layers. The latter consists of a convolution layer, a rectified linear unit (ReLU) layer, an activation function and a pooling layer.

The convolution layer applies the convolution operation to the incoming multivariate time series sequences, after which it passes the results to the next layer. The convolution operation emulates the response of individual neurons to visual stimulation. Each convolution neuron processes sector data only for the receptive field. Overall, the convolutional operation here reduces the number of parameters and make the CNN-LSTM network deeper [40].

Furthermore, the convolution layer leverages the pooling layer to merge the output of a neuron cluster in one layer into a single neuron in the next layer. This in turn minimizes the space size of the representation. This in turn reduces the number of parameters, along with computation cost and

complexity in the network. Namely, a filter of size κ in the pooling layer is run across the input to achieve a convolved output. Subsequently, each feature map of the convolved is subsampled with a mean or *max* pooling layer to extract the important features from the data input in the latter. The convolutional layer is then followed by a *max* pooling layer. The output is then flattened to feed into LSTM layers. The model here consists of four hidden LSTM layers followed by a dense layer to provide the output.

Each LSTM cell consists of a cell state that is comprised of input, input modulation, forget and output gates. These gates specify the type of information in the cell state through recursive multiplication operations. Here the recursive non-linear looping operations are implemented to permit information from previous intervals to be stored with in the LSTM cell. Figure 4 depicts the LSTM structure composed of the memory cell and gates that regulate the information flow in and out of the memory, i.e., represented by the cell state vector. The latter undergoes changes via forgetting old memory using the forget gate and adding new memory using the input gate. The figure also shows gates that are sigmoid layer followed by pointwise multiplication operator. Consider the detailed operation of the LSTM unit.
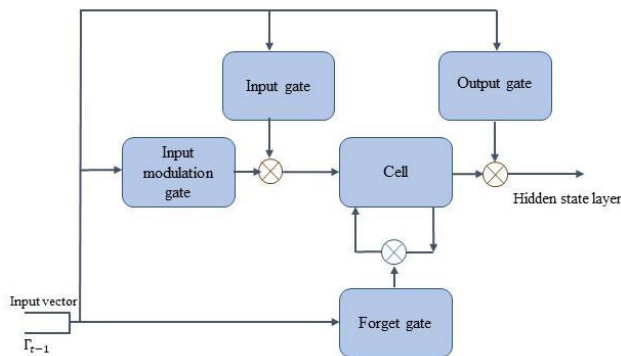

FIGURE 4. LSTM cell structure.

Initially, the cell state at time t, ς_t, determines information carried to the next sequence. It is modified by the forget gate, g_t^f (remember vector) in the sigmoid layer placed underneath it. This gate in turn is adjusted by the input modulation gate, g_t^{mod} that delivers the new candidate cell state. The forget gate receives hidden state vector, Γ_{t-1} (output vector of the LSTM unit) at $(t-1)$th time period, and input vector at time t, γ_t^{in}, as its inputs. It then generates an output number between 0 and 1 for each number in the previous cell state at time $t-1$, ς_{t-1}. Note that the output of the forget gate here leads the cell state, where information is discarded by multiplying 0 to its location in the matrix. Further, if the output of the forget gate is 1, then the information is reserved in the cell state, where a sigmoid function σ_g is applied to the weighted input/observation and preceding hidden state. Along this, Eqs. (10)-(14) represent the cell state ς_t, forget gate g_t^f, input gate g_t^{in}, input modulation gate g_t^{mod}, and output gate g_t^{out}

(all at time t), respectively,

$$\varsigma_t = g_t^f \varsigma_{t-1} + g_t^{in} \widehat{\varsigma_t}, \tag{10}$$

$$g_t^f = \sigma_g \left(W_f \left[\Gamma_{t-1}, \gamma_t^{in} \right] + b_f \right), \tag{11}$$

$$g_t^{in} = \sigma_g \left(W_{in} \left[\Gamma_{t-1}, \gamma_t^{in} \right] + b_{in} \right), \tag{12}$$

$$g_t^{mod} = tanh \left(W_\varsigma \left[\Gamma_{t-1}, \gamma_t^{in} \right] + b_\varsigma \right), \tag{13}$$

$$g_t^{out} = \sigma_g \left(W_{out} \left[\Gamma_{t-1}, \gamma_t^{in} \right] + b_{out} \right). \tag{14}$$

The parameters W_f, W_{in}, W_{out}, and W_c are the weight matrices, whereas the variables b_f, b_{in}, b_{out}, and b_ς denote the bias vectors for the forget, input and output gates and cell state, respectively (developed during the training mode). Finally, the hidden state layer output, Γ_t (also known as the working memory) is modeled as $\Gamma_t = g_t^{out}.tanh(\varsigma_t)$.

The main parameters for the model here are the logistic *Sigmoid* and the hyperbolic *tanh* nonlinear activation function for each gate, i.e., used to predict probability of the output. The input gate is a *sigmoid* function with range $\in [0,1]$ that only adds memory without the ability to forget memory, as the cell state equation is a summation of the prior states. Consequently, the input modulation gate is activated with a *tanh* activation function with a $[-1, 1]$ range that allows the cell state to forget memory.

V. DEEP LEARNING SIMULATION RESULTS

The deep learning network model is now simulated using the dataset in [41] as part of the *Big Data Challenge*. It is a dataset collection campaign for the distribution of MS users in cellular networks. The dataset comes from various cellular providers that adopt different standards. Note that their spatial distribution irregularity is aggregated in a grid with cells to allows comparisons between different areas and eases the geographical management of the data as specified in [41]. The varying traffic densities for the MS users in the dataset is composed of 5 activities, i.e., SMS-in,SMS-out, call-in,call-out, and internet activities. The dataset also includes location and time stamp per request activity. Namely, these activities are recorded over time steps inside a square ID of 200 meters, i.e., geographical grids forming a single sector. Along this, the popularly class of each sector is determined by the activities amount of MS users in each sector over a particular time period. Hence, each beam sector is associated with a grid and the selection of adjacent grids was made randomly to account for random MS distributions.

A. NETWORK TRAINING

The training settings for the deep learning network include four hidden layers and 50 LSTM units in each layer. Also, the drop-out regularization rate used in each cell as a regulizer is set at 0.2. The model is trained with 350 epochs over a duration that ranges from 200 minutes to 2 weeks. Along this, a data structure with 60 time steps and a single output is created (each of 10 minutes), since the cells store long short-term memory state. Therefore, there are 60 previous set

elements in each training stage for each taken sample. As a result, the first 60 samples are vital to achieve an accurate prediction about the rest of the traffic distribution.

Along this, the training objective here is to compute the *embed* values, weight matrices, and the bias vectors that minimize the MSE for all training time instances. Figures 5-8 show the prediction performance for the test set of the CNN-LSTM network for 4 of the sectors (results for Sector 5 are not shown here).
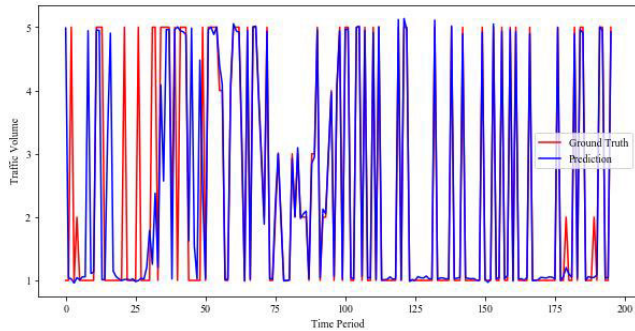

FIGURE 5. Traffic volume in Sector 1.
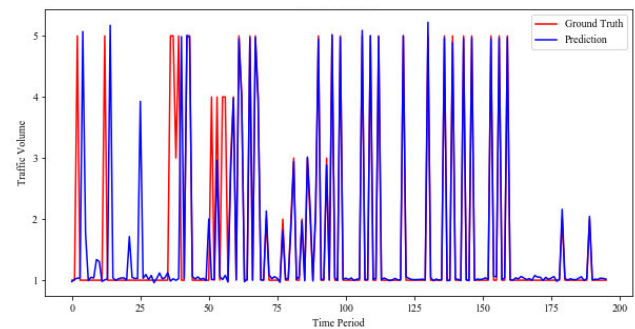
FIGURE 6. Traffic volume in Sector 2.

High approximation is achieved here between the ground truth G_t and prediction $\widehat{G}_t$ for the various sectors over the training period T that extends up to 200 minutes. First, Figure 5 shows the traffic volume in Sector 1, where the BS here detects the highest number of received signal levels from the MS, i.e., high traffic intensity and signal levels exist in Sector 1. The model suffers from some errors for a short time period (below 50 minutes), albeit increased accuracy between 50-200 minutes, during which high accuracy is achieved between the ground truth and the prediction results. Further, Figure 6 shows the prediction output for Sector 2 for the same time period (200 minutes). The model here performs better at the early stage with noticeably reduced error. Thereafter, the model achieves a high level of precision and determines an accurate prediction. Note that when the BS uses the beam associated with Sector 2, a smaller number of received signal levels are achieved here, i.e., less MS are sending signals to the BS in this location as compared to Sector 1.

Traffic density increases again in Sector 3 as shown in Figure 7, albeit less than Sector 1. Likewise, the model here

exhibits reduced error and enhances the match between the ground truth and prediction, where high accuracy is achieved at approximatively 55 minutes. Finally, Figure 8 shows that Sector 4 features the least number of received signal levels affiliated with less dense Sector area, i.e., high precision starting at 30 minutes.
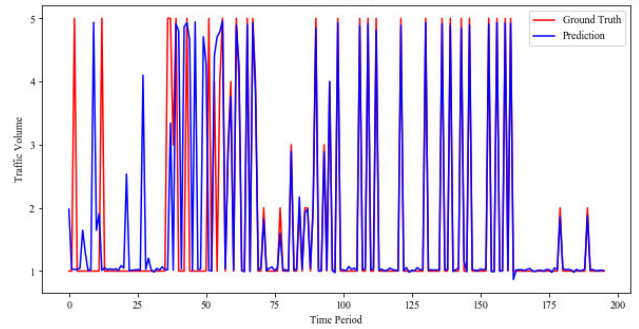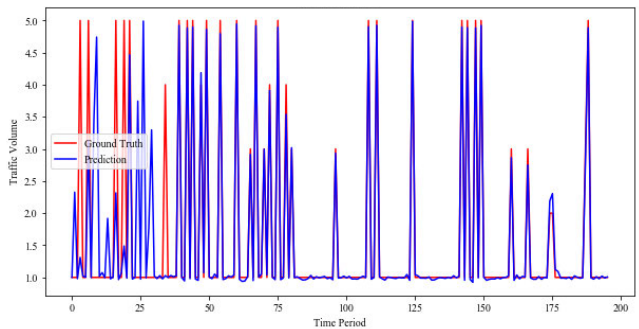

FIGURE 7. Traffic volume in Sector 3.

FIGURE 8. Traffic volume in Sector 4.

It can be noted that Sector 1 includes the highest traffic densities, with the highest MS requests from which it receives the highest number of signal levels. Therefore, Sector 1 is considered as the primary sector. This is followed by Sectors 2, 3 and 4. Hence, when a MS transfers from sleep-mode to idle- or active-mode, then there is a high probability that this MS will exist in Sector 1, at which the highest received signal is recorded. Consequently, the BS starts searching over the beam index that covers Sector 1, then Sector 2 as the subsequent one to be searched if needed, etc. hence reducing computational complexity and search times. This is alternative to randomly searching over all beam directions (without prior knowledge). See Table 2 for the overall model settings showing the hyper-parameters chosen for layers of the CNN- LSTM network.

B. LOSS FUNCTION

The training objective here aims to achieve a reduced loss function. The *mean square error* (MSE) is adopted between the prediction vector of the proposed model $\widehat{G}_t$ and the actual ground truth G_t in the upcoming time step. This is gauged over predictions generated from a sample of data points on

TABLE 2. Deep learning model parameters.

Class	Parameter	Value
CNN Layers	No. of filters	5
	Pool size	2
	Kernel size	3
	Activation Function	ReLU
LSTM layers	Number of cells in LSTM layers	[50,50,50,50]
	Activation Function	ReLU
Other Hyper-parameters	Batch size	16
	Epochs	350
	Optimizer	rmsprop
	Dropout rate	0.2

all variables during T time period. Note that both G_t and $\hat{G}_t$ feature similar distributions over the same time period. This loss function model at every time t is formulated as,

$$Loss\left(G_t, \hat{G}_t\right) = \frac{1}{T}\sum_{t=1}^{T}\left(G_t, \hat{G}_t\right)^2. \qquad (15)$$

Figures 9-12 visualize and validate the prediction loss for each sector based on hyperparameter epoch iterations, where the number of epochs specifies the iterations that the learning model operates during the defined training period. Overall, results show that the proposed scheme yields in low errors. First, Figure 9 illustrates that the prediction error for Sector 1 is very low, i.e., ranging between 0.03 and 0.06.
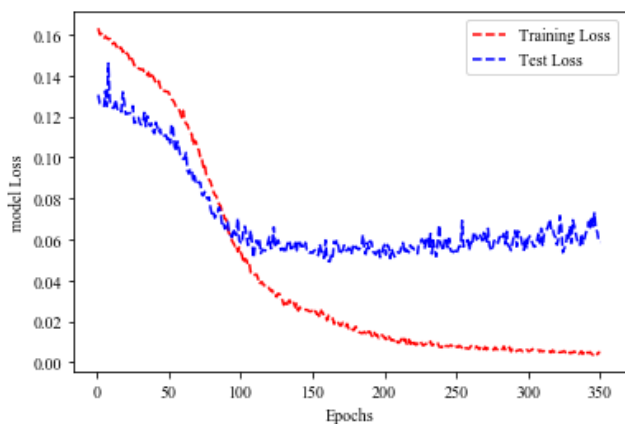

FIGURE 9. Loss Model for Sector 1.

Low MSE is achieved for Sector 2 as depicted in Figure 10, i.e., 0.035 at the start then converges to approximately 0.02 at 350 epochs. It is clear that the risk function here achieves reduced model losses with increased epochs. Namely, a low MSE indicates more centralized data (less skewed), i.e., high dispersion towards central moment. For instance, the model loss approaches 0.06 and 0.02 at 350 epchoes for Sectors 1 and 2, respectively. Furthermore, Figure 11 shows the MSE output at a slighter higher error that converges to 0.124 at 350 epchoes. Lastly, Figure 12 depicts the loss model for Sector 4 that ranges between 0.04 and 0.05. Overall, the
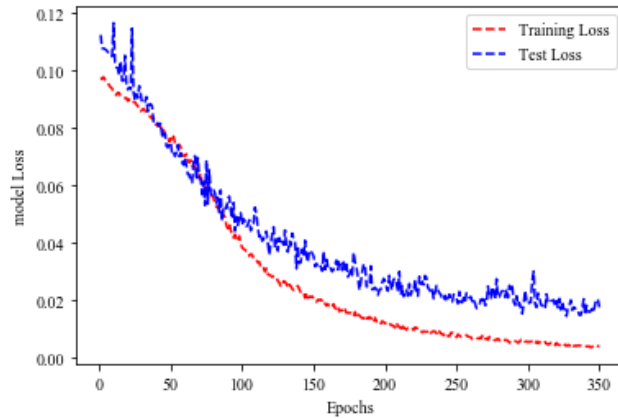

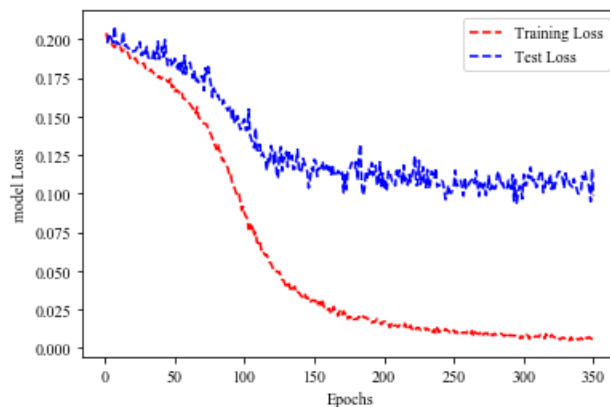
FIGURE 10. Loss Model for Sector 2.

FIGURE 11. Loss Model for Sector 3.

low MSE achieved here for the four sectors concludes that the proposed scheme/estimator can predict observations of the sector parameters with an acceptable accuracy.
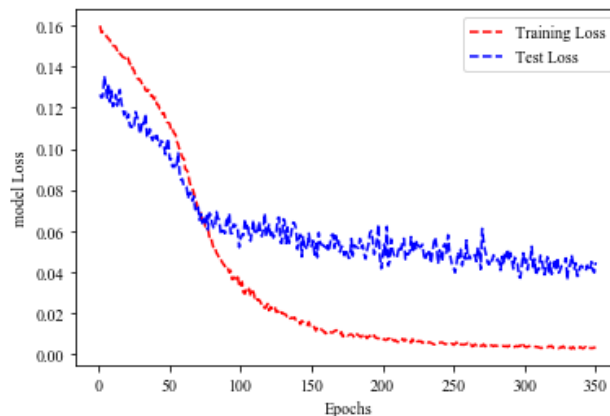

FIGURE 12. Loss Model for Sector 4.

1) SECTOR POPOLUARITY
This parameter represents the relative popularity or obscurity class of a sector as a member of the traffic (population).

This popularity class follows a *Zipf* distribution, modeled as $X \sim Zipf(\mu, S)$, where X is the popularity random variable. The variable μ denotes the skewness popularity that models probability variations among different sectors. Here large μ represents a small amount of popularity (high diversity among MSs distribution), highly right-skewed histogram distribution. Meanwhile, small μ (flat-skewed histogram) represents a high popularity. Recall that the variable S represents the number of sectors in the traffic.

The popularity of a sector $s \in S$ is defined by the number of traffic requests generated by $m \in M$ number of MSs that are covered by this sector, $m \to s$. Hence this popularity is formulated as $P_M(s) = |\Xi_M(s)|$, where $\Xi_M(s) = \{m \in M, m \to s\}$. Overall, the set of all sectors that traffic is generated from is modeled as $S = \{s: \exists m \in M, m \to s\}$. The probability distribution here over different sectors and their associated popularities is defined by $\mathbb{P}_s(\mu)$. For each sector $s \in S$, the probability that its popularity $P_M(s)$ is equal to μ is given by,

$$\mathbb{P}_s(\mu) = \frac{1}{S} \sum_{s \in S} \mu, \quad \mu \in [0, F], \tag{16}$$

where F is the total traffic (population) generated from terminals in the sectors.

Figure 13 shows the popularity class for the different sectors, i.e., correpondsing to popular sectors with highest MSs traffic densities. It is shown that Sector 1 (affiliated with beam Index 1) yeilds the highest popularity class, $\mathbb{P}_1(\mu) = 0.60$, i.e., approximately 60% of the incoming traffic generates from MSs located in this sector. This is followed by Sector 2 (beam Index 2) with 20% of the overall generated traffic. This is followed by Sector 3 which accommodates 9% of the traffic. Meanwhile, the scarce Sectors 4 and 5 feature the lowest-popular with 6% and 5%, respectively.
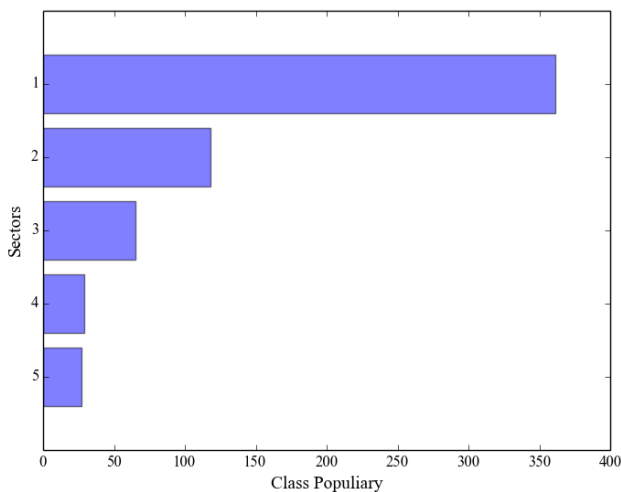

FIGURE 13. Sector (Beam Index) Popularity Class.

Moreover, the *complementary cumulative distribution function* (CCDF) or the survivor function is plotted in Figure 14. The CCDF here analyzes the reliability of the proposed prediction scheme for the popularity classes of the various sectors. For instance, when a MS joins the network, the association probability with Sector 1 is approximately 0.60, as compared to 0.20 and 0.09 probabilities with Sectors 2 and 3, respectively. The CCDF achieved here determines the BS association probabilities and connection success rates based upon popularity classes (rates).
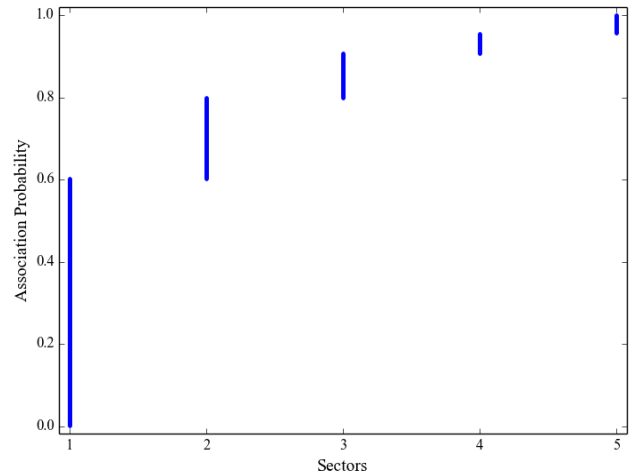

FIGURE 14. The CCDF of the sector popularity.

C. SUCCESS RATES
To provide a comprehensive performance evaluation for the proposed sector prediction scheme, the success rates are also recorded in Figure 15. It shows the successful prediction probability over extended time periods of increased dataset training sizes. The success probability of the proposed scheme proportionally improves at increased time periods. This is attributed to the increase in the dataset size, which relates information on prior sectors and their popularity rates. For instance, the CNN-LSTM scheme features 72% success rates when dataset size is measured over 200 minutes. Thereafter, the performance gradually improves, i.e., approaching 93% success rate at 2000 minutes. This in turn increases the robustness of the sector prediction concept.

VI. PERFORMANCE EVALUATION
The proposed deep learning model is now leveraged to predict the primary sector for incoming MS. This is now analyzed in terms of computational complexity and access times during access phase, See Table 3 for the overall parameter settings.

A. COMPUTATIONAL COMPLEXITY
In general, the computational complexity to determine the primary sector for incoming MSs is gauged by the total number of signal level evaluations at each beamforming and combining vectors (beam indices). Consider the BS and MSs use a hierarchical codebook of $c = 1, 2 \ldots, C$ stages, then the total number of measurements at each codebook stage c, Ψ_c, using the proposed learning is gauged by the measurements at the BS Ψ_c^{BS} and MS Ψ_c^{MS}, i.e., $\Psi_c = \Psi_c^{BS} + \Psi_c^{MS} = 1 + I$.
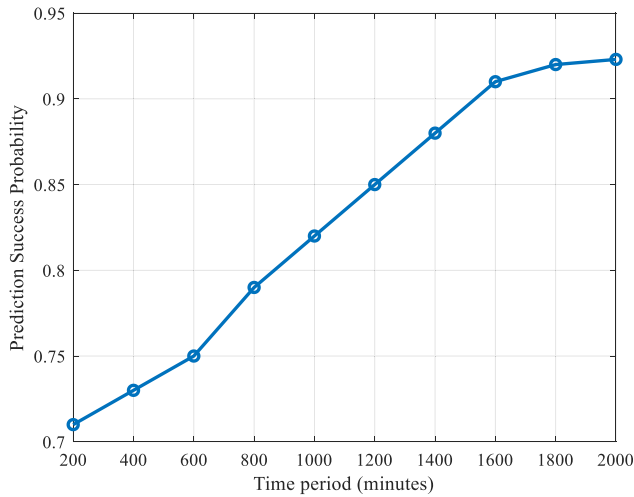
FIGURE 15. Prediction success rates at extended time periods.

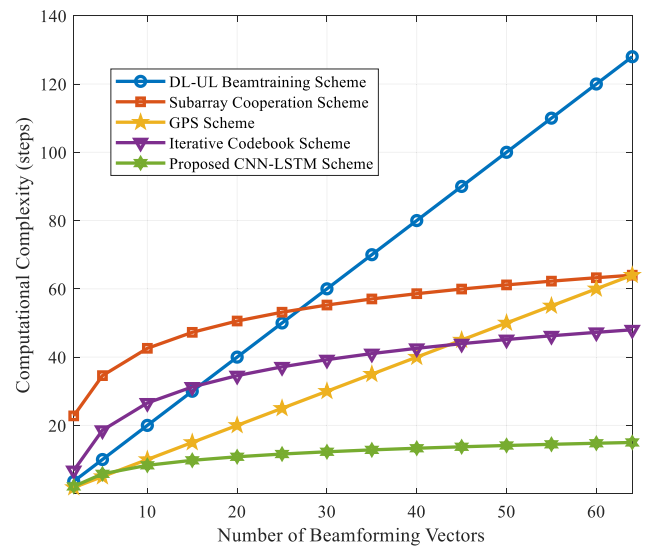| Parameter settings	Variable	Value
Entities	S, I	5, 5
Channel	f_α(GHz), BW(MHz), σ_w^2, ζ, K, L	28, 800, 1, 3, 3, 5
Beamforming	$C, d_x, d_y, R_{MS}, R_{BS}, N_{MS}^1, N_{MS}^2, N_{MS}^3, N_{BS}^1, N_{BS}^2, N_{BS}^3$	3, 0.005, 0.005, 1, 76, 4, 18, 54, 4, 18, 54
Power Consumption	$Q_m^{ps}, Q_m^{ant}, Q_{LNA}, Q_{ADC}, Q_{BB}, Q_{MIX}, Q_{LO}, Q_{LPF}, Q_{AMP}$	78, 5.4, 20, 20, 20, 19, 5, 14, 5

FIGURE 16. Computational Complexity for the proposed scheme.

Note that $\Psi_c^{BS} = 1$, as the model learns the best sector, i.e., measuring signal level at the primary sector only (scanning is not performed at the BS). However, the MS still needs to scan over all directions I, in order to specify the best combining vector that returns highest signal level, y_{high}. Moreover, the complexity model is scaled by $O\left(\Psi_c log_4(I^2)\right)$ at higher number of beamforming and combining vectors using narrower beams.

Figure 16 shows the computational complexity for the proposed scheme versus major existing access methods. The figure depicts the increased computational complexity as a function of the number of beamforming (combining) vectors. Overall, the proposed learning model achieves significant reduction in computational complexity at various beamforming vectors. When using wide beams at the BS and MS, low number of measurements is required to cover the entire spatial domain due to the large HPBW, which exhibits less significance in the access procedure due to the small number of signal measurements carried over the spatial domain. However, this becomes significant when transmitting at pencil beams, e.g., using 64 beams. Here the proposed scheme requires 15 steps to detect the primary sector and its affiliated best beamforming vector at the BS, and the combining vector at the MS. This is compared to 48 for the iterative codebook [3] scheme, 64 steps for both the GPS [10] and subarray cooperation [12] schemes, and 128 steps for the DL-UL beam training [11] scheme.

The iterative search in [3] features the least complexity among other schemes due to the adoption of a codebook structure. The search is performed by wide beams to reduce the number of beams covering the entire spatial domain. Then the beam that returns the highest signal is refined to select a single narrow beam direction. Meanwhile, the GPS approach in [10] adopts an exhaustive search fashion that starts with a narrow beam, albeit reduced search complexity at the BS using the GPS that eliminates the search process. Meanwhile, the directions are divided into multiple subarrays

in [12] that simultaneously cooperate to find the best direction. Finally, the DL-UL approach in [11] requires multiple stages of beam-training at the BS to establish the beams subset group. However, this approach is limited to the BS and thus the MS is still compelled to perform spatial search, which keeps the computational complexity relatively high, as compared to the proposed scheme.

B. ACCESS TIMES
The time delay for the initial access process, T_{acc}, is the duration needed to determine the best directions that yields in the strongest signal level during control-plane, e.g., during MS transition from sleep- to active-mode. This is modeled as,

$$T_{acc} = CT_{cont}t_{RS}, \qquad (17)$$

where T_{cont} is the number of time-slots occupied during control signals exchanged between the BS and MS, and t_{RS} is the reference signal duration transmitted in every beam vector.

Figure 17 plots the aggregated access times for sector prediction and beam association against versus beamforming (combining) directions. The figure shows that the proposed scheme features significant reduction in access times as compared to (prominent) existing access methods.

Overall, the proposed scheme yields at least 70% reduction in access times. Namely, 3 ms are required to acquire a pencil beam when using 64 beamforming vectors. This is compared to 10 ms for the iterative codebook scheme [3] attributed to the high number of codebook stages needed to achieve a pencil beam. Further, 13 ms are required for the GPS scheme [10], as the GPS is only used at the MS. Therefore, the BS conducts exhaustive beam search to allocate the MS, thus extending the search period. Likewise, 13 ms is required for the subarray cooperation [12] scheme, since it spends time in the refinement of the initial layers based upon various simultaneous SNR levels. Finally, the DL-UL beam training approach [11] spends 26 ms to estimate the primary sector, i.e., attributed to the use of a single RF chain at the BS.
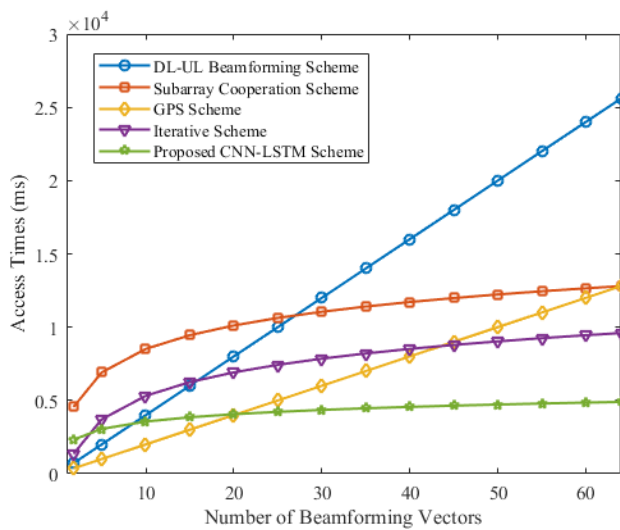

FIGURE 17. Access times for the proposed scheme.

Overall, the reduced access times achieved here can enable *ultra-reliable low latency communications* (URLLC) over the air interface, which can be delivered by the core and radio access networks of 5G. This strengths the latter to provide mission critical applications and other delay-sensitive services such as online gaming and autonomous driving.

Overall, the low access delays acquired here support a short control-plane latency, which meets the ultra-low delays, as defined by the 3GPP (1 ms). This in turn improves the *quality-of-service* (QoS) and enforces the implementation of SA mmWave networks. Furthermore, the technique can be adopted in other networks that implement mmWave links such as IEEE 802.11ay WLAN standard for V2X communication. Moreover, the deep learning scheme allows the deployment of highly directional beams without depending on wide beam codebooks. Consequently, this can eliminate the vulnerability of beam blockage due to low directivity. The MS here can use narrow beams when switching from sleep to idle (or active) mode in the control-plane. This in turn helps in shortening beam access times. Further, high data rates can be supported here by leveraging the narrow

beams, i.e., high channel capacities can be achieved from the aggregated antenna gains.

C. POWER AND ENERGY CONSUMPTION

In order to provide a complete assessment for the proposed sector prediction scheme, the power consumption requirement is gauged at the BS for a single user. It is determined by the number of used antennas, phase shifters, filters, amplifier and other RF circuitry components. Further, power consumption for deep learning networks is critical to determine whether training is feasible, i.e., in particular at the MS that possesses limited battery lifetime. The deep learning is implemented at the BS only in this paper, albeit the same approach and model can be applied at the MS. The power consumption of the deep learning network is specified by the processing hardware unit.

Traditionally, deep learning networks are implemented using hardware accelerators such as *graphical processing units* (GPU) that allow parallel pipelining as opposed to *central processing unit* (CPU). However, GPU can still yield in high power demands, thus an alternative accelerator is adopted to achieve less power requirements. Namely, a *deep learning accelerator unit* (DLAU) is adopted to realize the proposed method. This unit proposed in [42] is a scalable accelerator architecture for large-scale deep learning networks based on *field-programmable gate array* (FPGA). It speeds up the kernel computational parts of deep learning algorithms. It employs three pipelined processing units to minimize memory transfer operations and reuse the computing units to implement the large size neural networks. Experimental results in [42] show that the DLAU accelerator achieves up to $36.1\times$ speedup comparing to the Intel Core2 processors, with a power consumption at 234 milliwatts. It is composed of three fully pipelined processing units, including *tiled matrix multiplication unit* (TMMU), *part sum accumulation unit* (PSAU), and *activation function acceleration unit* (AFAU). Note that the architecture model for the proposed scheme is assumed to run on ALDU, whereas the architectures for the compared schemes [3], [10]–[12] run on a DSP/CPU of a lower power consumption [43]. Along these lines, the power consumption at the BS Q_{BS} for a single user is gauged as,

$$Q_{\text{BS}} = Q_{\text{BS}}^{ps} \left[N_{ps} \right] + \left[Q_{ant} + Q_{LNA} \right] \\ \times \left[N_{ant} \right] + R_{\text{BS}} \left[Q_{RF} + Q_{ADC} + Q_{BB} \right], \quad (18)$$

where the variables $Q_m^{ps}, Q_m^{ant}, Q_{LNA}, Q_{RF}, Q_{ADC}, Q_{BB}$ denote the power consumption (in milliwatts) for the phase shifters, antenna elements (rectangular microstrip patch antenna), the *low noise amplifier* (LNA), RF chain, ADC, and the *baseband combiner* (BB). Here Q_{RF} and Q_{ADC} are calculated respectively by [44],

$$Q_{RF} = Q_{MIX} + Q_{LO} + Q_{LPF} + Q_{AMP}, \quad (19)$$
$$Q_{ADC} = E_{ADC}^{step} Sr_{ADC} 2^B, \quad (20)$$

where Q_{MIX}, Q_{LO}, Q_{LPF}, and Q_{AMP} denote the power consumption for the *mixer* (MIX), *local oscillator* (LO), *low pass filter* (LPF), and the *baseband amplifier* (AMP). Further, the variable E_{ADC}^{step} symbolizes the energy consumption per conversion in the *analog-to-digital converter* (ADC). The variable Sr_{ADC} accounts for the sampling rate and B denotes the total number of bits. The power consumption for the different components is also summarized in Table 3, as derived from studies in [44]. Overall, the energy consumption at the BS is gauged by calculating the power consumption during the initial access time, i.e., $E_c = Q_{BS}\tau_a$.
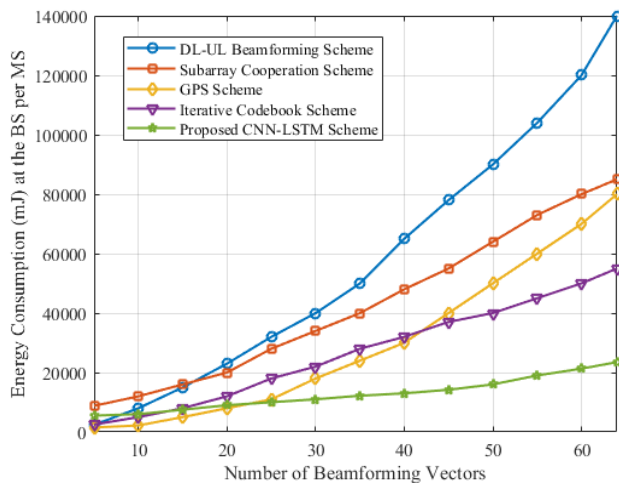

FIGURE 18. Energy Consumption at the BS.

Finally, Figure 18 illustrates the energy consumption rates for the proposed deep learning and various schemes as a function of the number of used beam directions. The figure shows that the proposed scheme exhibits low energy levels versus other schemes. The energy efficiency achieved here is mainly due to the shortened usage times for the power demanding RF chains, as the deep learning method predicts the primary sector in advance and thus eliminates extended search time. For example, the BS consumes 3 and 20 Joules to determine the primary sector using the deep learning approach, when radiating with 10 and 64 beams, respectively. This compared to 58 Joules for the iterative schemes [3], where the rise here is due to the extended usage time of the hybrid beamformer that possesses higher number of RF chain, thus increasing the power consumption. The GPS scheme [10] requires 80 Joules, due to high power associated with the GPS connectivity and excess search time at the BS. Moreover, the high number of subarrays mandate higher power that approaches 85 Joules for the cooperation scheme in [11]. Lastly, the prolonged time required to establish the subset groups at the BS yields in high energy consumption at 140 Joules for the DL-UL scheme [12]. Overall, the CNN-LSTM scheme achieves 65% higher energy efficiency and 50% reduced access times versus the closest scheme, i.e., iterative codebook search.

VII. CONCLUSION

In this work, a novel prediction scheme is proposed for the initial beam access problem in beamforming-based standalone millimeter wave cellular networks. This scheme leverages convolutional neural and long short-term memory networks to predict the primary sector and its affiliated beam index that includes the highest mobile station densities. The scheme reduces the number of beam measurements required to allocate the mobile stations. This yields in reduced computational complexity and access times. Moreover, the prediction scheme features low model loss and high success rates with the increased number of data samples over extended time periods. Future efforts will investigate the deep learning method to predict user mobility and blockage effects.

REFERENCES

[1] Third Generation Partnership Project (3GPP), "Technical specification group services and system aspects," Release 16, Meeting #86, Specification #21.916, Sitges, Spain, Dec. 2019.

[2] International Mobile Telecommunications, *Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface*, document ITUR Study Group 5, R15-SG05-C-0040, Feb. 2017, pp. 6–7.

[3] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[4] M. Jasim and N. Ghani, "Generalized pattern search for beam discovery in millimeter wave systems," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–5.

[5] M. Jasim and N. Ghani, "Hooke Jeeves search method for initial beam access in 5G mmWave cellular networks," in *Proc. IEEE 28th Annu. Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–7.

[6] H. Xu, J. Yu, and S. Zhu, "Ping-pong optimization of user selection and beam allocation for millimeter wave communications," *IEEE Access*, vol. 7, pp. 133178–133189, 2019.

[7] M. Jasim and N. Ghani, "Adaptive initial beam search for sparse millimeter wave channels," in *Proc. 26th Wireless Opt. Commun. Conf. (WOCC)*, Newark, NJ, USA, Apr. 2017, pp. 1–6.

[8] M. Jasim and N. Ghani, "Sidelobe exploitation for beam discovery in line of sight millimeter wave systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 234–237, Apr. 2018.

[9] M. Jasim, J. E. Pezoa, and N. Ghani, "Simultaneous multi-beam analog beamforming and coded grating lobes for initial access in mmWave systems," in *Proc. Conf. Electr., Electron. Eng., Inf. Commun. Technol. (CHILECON)*, Pucon, Chile, Oct. 2017, pp. 1–6.

[10] W. B. Abbas and M. Zorzi, "Context information based initial cell search for millimeter wave 5G cellular networks," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2016, pp. 111–116.

[11] X. Zhao, A. M. A. Abdo, Y. Zhang, S. Geng, and J. Zhang, "Single RF-chain beam training for MU-MIMO energy efficiency and information-centric IoT millimeter wave communications," *IEEE Access*, vol. 7, pp. 6597–6610, 2019.

[12] R. Zhang, H. Zhang, W. Xu, and X. You, "Subarray-cooperation-based multi-resolution codebook and beam alignment design for mmWave backhaul links," *IEEE Access*, vol. 7, pp. 18319–18331, 2019.

[13] A. Aldalbahi, F. Shahabi, and M. Jasim, "BRNN-LSTM for initial access in millimeter wave communications," *Electronics*, vol. 10, no. 13, p. 1505, Jun. 2021.

[14] A. Alkhateeb, I. Beltagy, and S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Anaheim, CA, USA, Nov. 2018, pp. 1055–1059.

[15] A. Mazin, M. Elkourdi, and R. D. Gitlin, "Accelerating beam sweeping in mmWave standalone 5G new radios using recurrent neural networks," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Chicago, IL, USA, Aug. 2018, pp. 1–4.

[16] T. S. Cousik, V. K. Shah, J. H. Reed, T. Erpek, and Y. E. Sagduyu, "Fast initial access with deep learning for beam prediction in 5G mmWave networks," 2020, *arXiv:2006.12653*.

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

[18] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, Sep. 2020.

[19] K. Satyanarayana, M. El-Hajjar, A. A. M. Mourad, and L. Hanzo, "Deep learning aided fingerprint-based beam alignment for mmWave vehicular communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10858–10871, Nov. 2019.

[20] P. Zhou, X. Fang, X. Wang, Y. Long, R. He, and X. Han, "Deep learning-based beam management and interference coordination in dense mmWave networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 592–603, Jan. 2019.

[21] Y. Heng and J. G. Andrews, "Machine learning-assisted beam alignment for mmWave systems," *IEEE Trans. Cognit. Commun. Netw.*, early access, May 7, 2021, doi: 10.1109/TCCN.2021.3078147.

[22] H. Echigo, Y. Cao, M. Bouazizi, and T. Ohtsuki, "A deep learning-based low overhead beam selection in mmWave communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 682–691, Jan. 2021.

[23] A. Klautau, P. Batista, N. Gonzalez-Prelcic, Y. Wang, and R. W. Heath, Jr., "5G MIMO data for machine learning: Application to beam-selection using deep learning," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2018, pp. 1–9.

[24] M. Boloursaz Mashhadi, M. Jankowski, T.-Y. Tung, S. Kobus, and D. Gunduz, "Federated mmWave beam selection utilizing LIDAR data," 2021, *arXiv:2102.02802*.

[25] S. Huang, Y. Gao, W. Xu, Y. Gao, and Z. Feng, "Energy-angle domain initial access and beam tracking in millimeter wave V2X communications," *IEEE Access*, vol. 7, pp. 9340–9350, 2019.

[26] C. Qi, Y. Wang, and G. Y. Li, "Deep learning for beam training in millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, early access, Sep. 22, 2020, doi: 10.1109/TWC.2020.3024279.

[27] W. Ma, C. Qi, and G. Y. Li, "Machine learning for beam alignment in millimeter wave massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 875–878, Jun. 2020.

[28] Y. Wang, M. Narasimha, and R. W. Heath, Jr., "MmWave beam prediction with situational awareness: A machine learning approach," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Kalamata, Greece, Jun. 2018, pp. 1–5.

[29] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.

[30] K. Ma, P. Zhao, and Z. Wang, "Deep learning assisted beam prediction using out-of-band information," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, Antwerp, Belgium, May 2020, pp. 1–5.

[31] W. Xu, Y. Ke, C.-H. Lee, H. Gao, Z. Feng, and P. Zhang, "Data-driven beam management with angular domain information for mmWave UAV networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7040–7056, Nov. 2021.

[32] C. Gao, H. Zhang, Y. Liang, and P. Hao, "Efficient uplink multi-beam initial access scheme for inactive users in mmWave networks," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Istanbul, Turkey, Sep. 2019, pp. 1–6.

[33] F. Fernandes, C. Rom, J. Harrebek, and G. Berardinelli, "Beam management in mmWave 5G NR: An intra-cell mobility study," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Helsinki, Finland, Apr. 2021, pp. 1–7.

[34] S. Tomasin, C. Mazzucco, D. De Donno, and F. Cappellaro, "Beam-sweeping design based on nearest users position and beam in 5G mmWave networks," *IEEE Access*, vol. 8, pp. 124402–124413, 2020.

[35] C. Balanis, *Arrays: Linear, Planar, and Circular, Antenna Theory: Analysis and Design*, 3rd ed. Hoboken, NJ, USA: Wiley, 2005, pp. 290–304.

[36] S. Sun, G. R. MacCartney, and T. S. Rappaport, "Millimeter-wave distance-dependent large-scale propagation measurements and path loss models for outdoor and indoor 5G systems," in *Proc. 10th Eur. Conf. Antennas Propag. (EuCAP)*, Davos, Switzerland, Apr. 2016, pp. 1–5.

[37] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[39] G. Swapna, S. Kp, and R. Vinayakumar, "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals," *Proc. Comput. Sci.*, vol. 132, pp. 1253–1262, Mar. 2018.

[40] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.

[41] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri, "A multi-source dataset of urban life in the city of Milan and the province of trentino," *Sci. Data*, vol. 2, no. 1, pp. 1–15, Dec. 2015.

[42] C. Wang, L. Gong, Q. Yu, X. Li, Y. Xie, and X. Zhou, "DLAU: A scalable deep learning accelerator unit on FPGA," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 3, pp. 513–517, Mar. 2017.

[43] S. Kaxiras, G. Narlikar, A. D. Berenbaum, and Z. Hu, "Comparing power consumption of an SMT and a CMP DSP for mobile phone workloads," in *Proc. Int. Conf. Compil., Archit., Synth. Embedded Syst. (CASES)*, Atlanta, GA, USA, 2001, pp. 211–220.

[44] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2015, pp. 90–97.

ADEL ALDALBAHI (Member, IEEE) received the B.S. degree in electrical engineering from Virginia Commonwealth University, Richmond, VA, USA, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from the New Jersey Institute of Technology, Newark, NJ, USA, in 2013. He is currently an Assistant Professor in electrical engineering, the Vice Dean of Academic Affairs, and the Director of the Industrial Relation and Technology Transfer Unit, King Faisal University, Al-Ahsa, Saudi Arabia. His current research interests include visible light communication, millimeter wave networks, and deep learning.

MOHAMMED A. JASIM (Member, IEEE) received the B.S. degree in electronics and communications engineering from Applied Science University, Jordan, in 2010, the M.S. degree (Hons.) in wireless communications systems from Brunel University London, U.K., in 2011, and the Ph.D. degree in electrical engineering from the University of South Florida (USF), USA, in 2018. He is currently an Assistant Professor in electrical and computer engineering with the University of Mount Union, OH, USA, and previously at Valparaiso University, Valparaiso, IN, USA. His research interests include millimeter wave communications, phased arrays, fog computing, and network virtualization.

FARZAD SHAHABI (Member, IEEE) received the B.S. degree in electrical engineering from the University of Kerman, Iran, and the M.S. degree in electrical engineering from the University of South Florida, in 2019, with a focus on data analytics and security. He specializes in image classification, time series regression, and statistical inference solutions. His research interests includes enhancing and developing human-in-the-loop machine learning algorithms and deep learning for human activity recognition area.

ASIM MAZIN (Member, IEEE) received the B.S. degree in electronics and telecommunications engineering from the College of Industrial Technology Misurata, Libya, in 2007, the M.S. degree in electrical engineering from Southern Illinois University, in 2013, and the Ph.D. degree from the University of South Florida, in 2019. He is currently a Postdoctoral Research Fellow with the Moffitt Cancer Center & Research Institute at USF. His research interests include physical layer security, multiple access in M2M IoT networks, and applying machine learning in 5G wireless networks.

NAZLI SIASI (Member, IEEE) received the B.S. degree in information technology and computer engineering from the Amirkabir University of Technology, 2011, the M.S. degree in information technology and computer engineering from Tabriz State University, in 2014, and the M.S. and Ph.D. degrees in electrical engineering from the University of South Florida (USF), in 2019 and 2020, respectively, with a focus on systems and security. She is currently an Assistant Professor in cybersecurity with Christopher Newport University, VA, USA. Her research interests include virtualization, blockchain, fog networks, and disaster recovery.

DIOGO OLIVEIRA (Member, IEEE) received the master's degree in computer engineering from the Federal University of Goias, Brazil, and the Ph.D. degree in electrical engineering from the University of South Florida (USF), in 2018. He is currently an Assistant Professor with the School of Information, Florida State University (FSU), Tallahassee, FL, USA. His research interests include cybersecurity, virtualization, blockchain, optimization, and millimeter wave communications.

• • •