Deep Learning-Based Visibility Region Classification for Extra-Large Aperture Arrays

Muhammad Zia Hameed, Dulaj Gunasinghe, and Gayan Amarasuriya
School of Electrical, Computer, and Biomedical Engineering, Southern Illinois University, Carbondale, IL, USA 62901
Email: {muhammadzia.hameed, dulaj.gunasinghe, gayan.baduge}@siu.edu

GLOBECOM 2024 - 2024 IEEE Global Communications Conference | 979-8-3503-5125-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/GLOBECOM52923.2024.10901212

***Abstract*—Spatial non wide-sense stationarities cause partial visibility regions (VRs), and it is a unique propagation characteristic of emerging extra-large aperture arrays (ELAAs). Thus, classification of VRs is a necessity for accurate estimation of channels and efficient design of VR-aware precoders for ELAAs. In this paper, a deep learning framework is proposed to classify VRs in ELAAs. Our objective is to boost the accuracy of classifying VRs based on the uplink pilots received at the ELAAs. Consequently, we focus on guaranteeing user-fairness in the presence of wholly/partial VRs and improving the achievable rates by adopting VR-aware channel estimation and precoding. We propose a hybrid deep learning architecture comprising one dimensional convolutional neural networks and long-short term memory to classify VRs of each user at the ELAA. To achieve a higher accuracy, we generate a diverse dataset through Monte-Carlo simulations that captures numerous combinations of VRs at the ELAA. A transmit power allocation algorithm is also proposed to achieve a common downlink rate for all users irrespective of the different VRs, and its computational complexity is discussed. A set of numerical results is presented to evaluate the performance of our proposed framework. It is efficient and accurate in classifying VRs. Thus, it can be used to enhance the estimation accuracy of ELAA channels with VRs and thereby to design VR-aware precoders to boost spectral/energy efficiency of the next-generation wireless systems.**

I. Introduction

Massive multiple-input multiple-output (MIMO) has been well established, and it is now being commercially deployed in fifth-generation (5G) wireless systems [1]. A logical evolution for this technology is to employ extra-large aperture arrays (ELAA) within the framework of massive MIMO to further boost spatial/energy efficiency [1].

In massive MIMO with compact antenna arrays, it is typically assumed that the channels are wide-sense stationary (WSS) in the spatial dimension [2]. However, recent channel measurement campaigns have shown that arrays with extremely large apertures give rise to spatial non-WSS characteristics [3]. Consequently, the statistical properties of the received signal, such as the average power, vary across ELAAs [3]–[5]. There is also a possibility that widely separated antennas do not share the same clusters of scatterers, leading to spatially non-stationary channels [6].

Scattering clusters in a wireless channel can be categorized into wholly-visible (WV), seen by the entire array, and partially-visible (PV), visible only to a portion of the array [3]–[5]. When an electromagnetic (EM) signal interacts with a PV cluster, it may be received by only a subset of antennas associated with that PV cluster. Consequently, there can be one user group served by WV clusters, while another user group may be served by PV clusters. The presence of WV/PV clusters and the spatial non-stationarities in ELAAs give rise to visibility regions (VRs) [3]–[5]. Thus, conventional spatial correlation models must be reexamined to account for the spatially non-stationary fading across ELAAs [6]–[8]. Specifically, it is a necessity to accurately classify VRs to efficiently design precoders and power allocation techniques for ELAA channels with spatial non-stationarities.

The necessity for accurate classification of VRs in ELAAs is elaborated next. In sub-6 GHz massive MIMO, the uplink channels are estimated at the base-station (BS) by using user pilots, and the precoders for the downlink are designed based on the uplink channel estimates by leveraging channel reciprocity in time-division duplexing (TDD) [2]. This approach ensures system scalability since the pilot length depends solely on the number of users, and it is independent of the number of BS antennas. The minimum mean square error (MMSE) based estimators require the knowledge of spatial covariance matrices. Such channel statistics are typically assumed to be known a-priori as they vary on a much large time-scale compared to the channel coherence interval [2]. A common practice in the current state-of-the-art is to model these covariance matrices by assuming full visibility of the BS array [9]. However, the MMSE estimates computed via covariance matrices with full visibility may not be accurate for ELAAs. Hence, spatial covariance matrices must capture VRs of ELAA channels to ensure VR-aware precoders and transmit power allocation. For instance, in [9], the perfect knowledge on VRs at the ELAA-based BS is assumed for channel estimation and precoder designs. The approach outlined in [9] requires accurate classification of VRs to encompass all potential combinations and scenarios involving multiple users served by ELAAs. However, attaining a high precision in VR classification of ELAA channels for each user is challenging. In [3], a sliding window method is proposed to identify VRs of MIMO channels, and [10] adopts an energy detector for the same task. However, such analytical techniques become prohibitively complicated for ELAAs in large antenna regime.

Machine learning, particularly deep learning-based artificial neural networks (ANNs), has garnered significant interest due to its potential to solve problems that are otherwise prohibitively complicated through analytical techniques [11]. It enables discovering effective solutions without relying on a predefined mathematical model, making it applicable to a wide-range of application domains. Recently, deep learning has become an enabler of efficient transceiver designs when the underlying models are mathematically intractable and optimal solutions are unattainable via conventional analytical frameworks [12]. Automatic modulation classification (AMC) is a prevalent task in wireless signal recognition. For such tasks, deep learning-based models such as convolutional neu-

ral network (CNN) and long-short term memory (LSTM) can be used to capture and adjust dynamic operations based on raw data from a diverse range of signal characteristics and channel effects [13]. In [14], an AMC is adopted for distributed spectrum sensing networks and a data-driven model solution is proposed via LSTM. In [12], a CNN-based multi-task learning (MTL) framework is proposed for concurrent modulation and signal classification. In [15], a hybrid deep learning model and image-based CNN architecture is proposed for multi-class modulation recognition. Inspired by deep learning-based pixel-by-pixel image classification with CNNs, we propose an antenna-by-antenna classification of VRs in ELAA channels. The contribution of this paper can be highlighted as follows. We propose a hybrid one-dimensional (1D) CNN-LSTM architecture to solve the challenging problem of precise VR classification for multi-user ELAAs. Our proposed technique makes uplink channel estimates more accurate by precisely capturing VRs and thereby facilitates VR-aware precoders and transmit power allocation techniques. Firstly, we curate a dataset employing a multi-user ELAA system. Secondly, we train a hybrid deep learning architecture for VR classification. We evaluate the system performance through a VR-aware maximum ratio transmission (MRT) precoder designed through predicted VRs and compare it with a VR-unaware precoder design. The uplink channels are estimated by using MMSE criterion. We also derive the achievable user rates in closed-form. We propose a VR-aware transmit power allocation algorithm based on statistical channel state information (CSI) to achieve a common rate for all the users irrespective of different VRs. A flow-chart for our design procedure is given in Fig. 1.

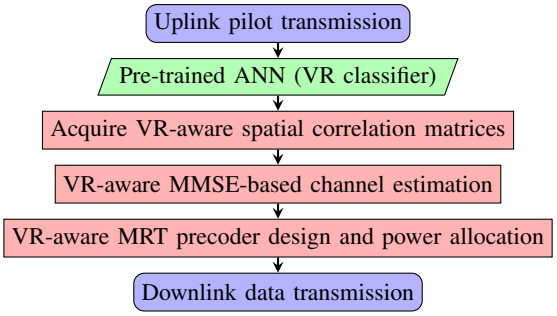

Fig. 1: A flowchart for the proposed design of ELAAs.

II. System, Channel, and Signal Models

A. System model

We consider a multi-user ELAA system in which a BS equipped with M antennas serves K single-antenna users (see Fig. 2). An ELAA, which consists of an active uniform rectangular planar array with M_1 rows and M_2 columns (where $M_1 \times M_2 = M$), is used at the BS. The uplink channel between the kth user and ELAA is denoted by $\mathbf{f}_k \in \mathbb{C}^M$ and it undergoes spatially correlated Rayleigh fading:

$$\mathbf{f}_k \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{fk}\right). \tag{1}$$

For simplicity, the large-scale fading coefficients are incorporated into the spatial correlation matrix, $\mathbf{R}_{fk}$ [2].

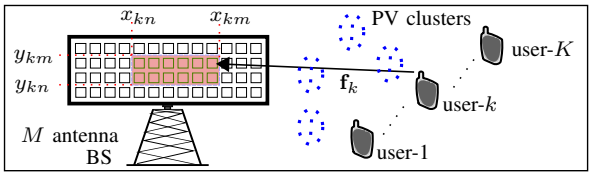

Fig. 2: An ELAA-based MIMO system with partial VRs.

1) Correlation matrices with full visibility: With full visibility, we denote $\tilde{\mathbf{R}}_{fk}$ as the correlation matrix for the channel between the BS and the kth user. The angles-of-arrival/departure (AoA/AoD) at the BS are randomly distributed in azimuth and elevation planes as $\Psi_i \sim \mathcal{N}\left(\psi_i, \alpha_i^2\right)$ and $\Phi_i \sim \mathcal{N}\left(\phi_i, \nu_i^2\right)$, respectively, with i=AoA and i=AoD denoting the AoA and AoD. The correlation matrices at the BS can be modeled as $\tilde{\mathbf{R}}_k = \zeta_k \mathbf{R}_{tk}$, where ζ_k accounts for the large-scale fading, and $\mathbf{R}_{tk} = \mathbf{R}_{tk,az} \otimes \mathbf{R}_{tk,el}$. Here, $\mathbf{R}_{tk,az}$ and $\mathbf{R}_{tk,el}$ are given by [16]

$$[\mathbf{R}_{tk,el}]_{x,y} = \mathrm{e}^{j(2\pi d_{el}(y-x)\cos(\phi_l))} \mathrm{e}^{-\frac{1}{2}(2\pi\nu_l d_{el})^2 (y-x)^2 \sin^2(\phi_l)}, \tag{2}$$

$$[\mathbf{R}_{tk,az}]_{u,v} = \delta_3^{-1/2} \mathrm{e}^{-\frac{\delta_2^2 \cos^2(\omega_l)}{2\delta_3} - \frac{(\delta_1 \alpha_l \sin(\omega_l))^2}{2\delta_3}} \mathrm{e}^{j\frac{\delta_1 \cos(\omega_l)}{\delta_3}}, \tag{3}$$

where d_{el} and d_{az} are the inter-antenna element distances. In (2) and (3), δ_1, δ_2, and δ_3 are $\delta_1 = 2\pi d_{az}(v-u)\sin(\phi_l)$, $\delta_2 = 2\pi d_{az}(v-u)\nu_l \cos(\phi_l)$, and $\delta_3 = \nu_2^2 \alpha_l^2 \sin^2(\omega_l) + 1$ [16].

2) Correlation matrices with partial visibility: Next, we define the covariance matrix at the ELAA with VRs to capture spatially non-WSS fading. Inspired by [9], we assume that the kth user visualizes the BS antennas in both horizontal (from x_{kn} to x_{km}) and vertical (from y_{kn} to y_{km}) axes (see Fig. 2)[1]. To capture the VR at the BS, two diagonal matrices $\mathbf{G}_{xk}$ and $\mathbf{G}_{yk}$ are defined as [3]–[5]

$$\mathbf{G}_{xk} = \mathrm{diag}(0, \cdots, 0, \underbrace{1, \cdots, 1}_{\text{from } x_{kn} \text{ to } x_{km}}, 0, \cdots, 0), \tag{4}$$

$$\mathbf{G}_{yk} = \mathrm{diag}(0, \cdots, 0, \underbrace{1, \cdots, 1}_{\text{from } y_{kn} \text{ to } y_{km}}, 0, \cdots, 0), \tag{5}$$

where $\mathrm{diag}(\cdot)$ generates a diagonal matrix. Then, the correlation matrix at the BS pertaining to the kth user with partial visibility (i.e., $\mathbf{R}_{fk}$ in (1)) can be written as

$$\mathbf{R}_{fk} = \left(\mathbf{G}_{xk}^{\frac{1}{2}} \mathbf{R}_{tk,az} \mathbf{G}_{xk}^{\frac{1}{2}}\right) \otimes \left(\mathbf{G}_{yk}^{\frac{1}{2}} \mathbf{R}_{tk,el} \mathbf{G}_{yk}^{\frac{1}{2}}\right), \tag{6}$$

where $\mathbf{R}_{tk,az}$ and $\mathbf{R}_{tk,el}$ are the covariance matrices with full visibility, and they are defined in (2) and (3), respectively.

B. Uplink channel estimation

The coherence interval length is set to τ_c, and τ_u symbols are allocated for orthogonal user pilots [2]. The kth user's pilot, $\boldsymbol{\psi}_k \in \mathbb{C}^{\tau_u}$ satisfies $\boldsymbol{\psi}_k \boldsymbol{\psi}_k^H = 1$ and $\boldsymbol{\psi}_k \boldsymbol{\psi}_{k'}^H = 0$ for $k \neq k'$. The received pilot signal at the BS is given as

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{\rho_u \tau_u} \mathbf{f}_k \boldsymbol{\psi}_k + \mathbf{N}, \tag{7}$$

where ρ_u is the normalized uplink signal-to-noise ratio (SNR) of pilot signals, and $\mathbf{N} \in \mathbb{C}^{M \times \tau_u}$ is an additive white Gaussian noise (AWGN) at the BS with independent and identically distributed (i.i.d.) complex Gaussian elements with zero mean and unit variance. To estimate $\mathbf{f}_k$, the received pilot signal (7) is first projected onto $\boldsymbol{\psi}_k$ and then scaled as

[1]Our deep learning framework to classify VRs is given in Section V.

$$\mathbf{y}_k = \mathbf{Y}\boldsymbol{\psi}_k^H/(\sqrt{\rho_u\tau_u}) = \mathbf{f}_k + \mathbf{n}_k, \tag{8}$$

where $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}/(\rho_u\tau_u))$. Next, we derive the MMSE estimate of $\mathbf{f}_k$ as [17]

$$\hat{\mathbf{f}}_k = \mathbf{C}_{y_k f_k}\mathbf{C}_{y_k y_k}^{-1}\mathbf{y}_k, \tag{9}$$

where $\mathbf{C}_{y_k f_k}$ and $\mathbf{C}_{y_k y_k}$ are given as [17]

$$\mathbf{C}_{y_k f_k} = \mathbb{E}\left[\mathbf{y}_k \mathbf{f}_k^H\right] = \mathbf{R}_{fk}, \tag{10}$$

$$\mathbf{C}_{y_k y_k} = \mathbb{E}\left[\mathbf{y}_k \mathbf{y}_k^H\right] = \mathbf{R}_{f_k} + (1/(\rho_u\tau_u))\mathbf{I}. \tag{11}$$

Due to MMSE criterion, the estimate and its error are orthogonal [17], the true channel $\mathbf{f}_k$ can be given in terms of its estimate (9) and estimation error as [17]

$$\mathbf{f}_k = \hat{\mathbf{f}}_k + \boldsymbol{\epsilon}_k, \tag{12}$$

where $\boldsymbol{\epsilon}_k$ is the error vector with zero-mean and a covariance matrix $\mathbf{C}_{\epsilon_k\epsilon_k} = \mathbf{R}_{f_k} - \mathbf{C}_{\hat{f}_k\hat{f}_k}$. Here, the covariance matrix of the estimator can be derived as

$$\begin{aligned}\mathbf{C}_{\hat{f}_k\hat{f}_k} &= \mathbb{E}\left[\hat{\mathbf{f}}_k\hat{\mathbf{f}}_k^H\right] = \mathbf{C}_{y_k f_k}\mathbf{C}_{y_k y_k}^{-1}\mathbb{E}\left[\mathbf{y}_k\mathbf{y}_k^H\right]\mathbf{C}_{y_k y_k}^{-1}\mathbf{C}_{y_k f_k} \\ &= \mathbf{C}_{y_k f_k}\mathbf{C}_{y_k y_k}^{-1}\mathbf{C}_{y_k f_k} = \mathbf{R}_{f_k}\mathbf{C}_{y_k y_k}^{-1}\mathbf{R}_{f_k}.\end{aligned} \tag{13}$$

C. Signal model

An MRT precoder is adopted due to its computational simplicity [2]. Due to channel reciprocity in TDD, the downlink channel from the BS to the kth user is given by $\mathbf{f}_k^H$, while its MMSE estimate is given as $\hat{\mathbf{f}}_k^H$ (9). Then, the normalized MRT precoder of the kth user can be written as

$$\mathbf{w}_k = \hat{\mathbf{f}}_k \Big/ \sqrt{\mathrm{Tr}\left(\mathbb{E}\left[\hat{\mathbf{f}}_k\hat{\mathbf{f}}_k^H\right]\right)} = \hat{\mathbf{f}}_k \Big/ \sqrt{\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_k\hat{f}_k}\right)}. \tag{14}$$

By denoting the power allocation coefficient for the kth user as η_k', we define $\eta_k = \eta_k'/\mathrm{Tr}(\mathbf{C}_{\hat{f}_k\hat{f}_k})$. The received signal at the kth user is given by

$$\begin{aligned}y_{dk} &= \sqrt{\rho_d}\sum\nolimits_{i=1}^{K}\sqrt{\eta_i'}\mathbf{f}_k^H\mathbf{w}_i x_i + n_k = \sqrt{\rho_d}\sum\nolimits_{i=1}^{K}\sqrt{\eta_i}\mathbf{f}_k^H\hat{\mathbf{f}}_i x_i + n_k \\ &= \sqrt{\rho_d\eta_k}\mathbf{f}_k^H\hat{\mathbf{f}}_k x_k + \sqrt{\rho_d}\sum\nolimits_{i\neq k}^{K}\sqrt{\eta_i}\mathbf{f}_k^H\hat{\mathbf{f}}_i x_i + n_k \\ &= \sqrt{\rho_d\eta_k}\underbrace{\mathbb{E}\left[\mathbf{f}_k^H\hat{\mathbf{f}}_k x_k\right]}_{DS_k} + \sqrt{\rho_d\eta_k}\underbrace{\left(\mathbf{f}_k^H\hat{\mathbf{f}}_k x_k - \mathbb{E}\left[\mathbf{f}_k^H\hat{\mathbf{f}}_k x_k\right]\right)}_{DU_k} \\ &\quad + \sqrt{\rho_d}\sum\nolimits_{i\neq k}^{K}\sqrt{\eta_i}\underbrace{\mathbf{f}_k^H\hat{\mathbf{f}}_i x_i}_{IUI_{ki}} + n_k,\end{aligned} \tag{15}$$

where x_k is the symbol intended for the kth user with $\mathbb{E}\left[|x_k|^2\right] = 1$, and ρ_d is the normalized SNR. In (15), $n_k \sim \mathcal{CN}(0,1)$ is the AWGN at the kth user. The first, second, and third terms in (15) are the desired signal (DS_k), decision uncertainty (DU_k), and sum of inter-user interference (IUI_{ki}), respectively. The signal-to-interference-plus-noise ratio (SINR) at the kth user can be written as

$$\gamma_k = \frac{\rho_d\eta_k|DS_k|^2}{\rho_d\eta_i|DU_k|^2 + \rho_d\sum_{i\neq k}^{K}\eta_i|IUI_{ki}|^2 + 1}. \tag{16}$$

III. Performance Analysis

A. An achievable rate

By invoking the worst-case Gaussian technique [2], an achievable rate at the kth user can be expressed via (16) as

$$\mathcal{R}_k = \left(\frac{\tau_c - \tau_u}{\tau_c}\right)\log_2\left(1 + \frac{\beta_{DS_k}}{\beta_{DU_k} + \sum_{i\neq k}^{K}\beta_{IUI_{ki}} + 1}\right), \tag{17}$$

where β_{DS_k}, β_{DU_k}, and β_{IUI_i} in (17) can be derived as [16]

$$\beta_{DS_k} = \rho_d\eta_k|DS_k|^2 = \rho_d\eta_k\left(\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_k\hat{f}_k}\right)\right)^2, \tag{18a}$$

$$\beta_{DU_k} = \rho_d\eta_k\mathbb{E}\left[|DU_k|^2\right] = \rho_d\eta_k\mathrm{Tr}\left(\mathbf{R}_{f_k}\mathbf{C}_{\hat{f}_k\hat{f}_k}\right), \tag{18b}$$

$$\beta_{IUI_{ki}} = \rho_d\eta_i\mathbb{E}\left[|IUI_{ki}|^2\right] = \rho_d\eta_i\mathrm{Tr}\left(\mathbf{R}_{f_k}\mathbf{C}_{\hat{f}_i\hat{f}_i}\right). \tag{18c}$$

From (17)-(18c), at high SNR regime ($\rho_d \to \infty$), the asymptotically achievable rate for the kth user are derived as

$$\mathcal{R}_k^{\infty} = \left(\frac{\tau_c - \tau_u}{\tau_c}\right)\log_2\left(1 + \frac{\rho_d\eta_k\left(\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_k\hat{f}_k}\right)\right)^2}{\sum_{i=1}^{K}\rho_d\eta_i\mathrm{Tr}\left(\mathbf{R}_{f_k}\mathbf{C}_{\hat{f}_i\hat{f}_i}\right)}\right). \tag{19}$$

Remark 1: To evaluate the MMSE-based channel estimate (9) and to design the precoder (14), the channel covariance matrices (6) must be known a-priori. The achievable rates in (17) and transmit power optimization in Section III-B also rely on these covariance matrices. Due to (4) and (5), these covariance matrices depend on the VRs of ELAAs. Hence, the accurate classification of the VRs is a necessity for system design and analysis. In Section VI, we further study the adverse effects of VR-unaware precoding in the downlink.

B. Transmit power allocation optimization

Different VRs lead to different achievable rates for the users [9]. Hence, we present an optimization algorithm to allocate the transmit power at the BS such that all the users achieve a common rate and thus preserving user fairness regardless of near-far and VR effects. Since $\log_2(\cdot)$ is an increasing function of its argument, we formulate our optimization problem based on statistical CSI to maximize the minimum SINR as

$$\mathcal{P}: \max_{\eta_k' \ \forall k} \ \min_{k} \ \frac{\beta_{DS_k}}{\beta_{DU_k} + \sum_{i\neq k}^{K}\beta_{IUI_{ki}} + 1}, \tag{21a}$$

$$\text{subject to} \ \ \mathrm{C}_1: \ \eta_k' \geq 0 \ \ \forall k, \tag{21b}$$

$$\mathrm{C}_2: \ \sum\nolimits_{k=1}^{K}\eta_k' \leq 1 \ , \tag{21c}$$

where $\beta_{DS_k}, \beta_{DU_k}$, and $\beta_{IUI_{ki}}$ are given in (18a)-(18c). Next, by introducing an auxiliary variable θ, which lower bounds all K user SINRs, the max-min optimization problem in (21a)-(21c) can be reformulated as

$$\mathcal{P}: \max_{\eta_k' \ \forall k} \ \theta, \tag{22a}$$

$$\text{subject to} \ \ \mathrm{C}_1: \ \eta_k' \geq 0 \ \ \forall k, \tag{22b}$$

$$\mathrm{C}_2: \ \sum\nolimits_{k=1}^{K}\eta_k' \leq 1 \ , \tag{22c}$$

$$\mathrm{C}_3: \rho_d\eta_k'\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_k\hat{f}_k}\right) \geq \sum_{i=1}^{K}\frac{\theta\rho_d\eta_i'}{\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_i\hat{f}_i}\right)}\mathrm{Tr}\left(\mathbf{R}_{f_k}\mathbf{C}_{\hat{f}_i\hat{f}_i}\right) + \theta, \ \ \forall k, \tag{22d}$$

where C_3 is obtained via (18a)-(18c) and also by utilizing $\eta_k = \eta_k'/\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_k\hat{f}_k}\right)$. The optimization problem in (22a)-(22d) is a convex optimization problem, which can be solved via geometric programming, and hence, it can also be solved via CVX [18]. Our proposed transmit power allocation optimization procedure is summarized in Algorithm-1.

Remark 2: In CVX, SDP3 solver uses interior-point algorithms to solve convex optimization problems. Hence, the computational complexity of our optimization $\mathcal{P}$ in (22a)-(22d) has a complexity in the order of $\mathcal{O}(K^{3.5})$. It is worth noting that the complexity of this algorithm does not depend

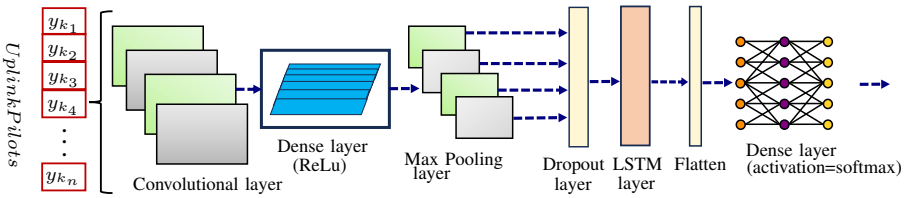

Fig. 3: The proposed deep learning architecture (N is the number of input samples).

Algorithm 1 : Transmit power allocation $(\mathcal{P})$ in (21a)-(21c)

Input: $\mathbf{R}_{f_k}$ and $\mathbf{C}_{\hat{f}_k \hat{f}_k}$ $\forall k$
Output: $\eta_k'^{*}$ $\forall k$
1: Compute $\mathrm{Tr}\left(\mathbf{R}_{fk}\mathbf{C}_{\hat{f}_i \hat{f}_i}\right)$ and $\mathrm{Tr}\left(\mathbf{C}_{\hat{f}_k \hat{f}_k}\right)$ $\forall k, i$
2: Compute $\eta_k'^{*}$ $\forall k$ by solving $\mathcal{P}$ in (22a)-(22d)

on the number of BS antennas (M). Thus, the proposed transmit power allocation algorithm is practically appealing for ELAAs in which the number of antennas can grow large.

IV. Dataset Preparation

In the dataset generation stage, we use the system, channel, and signal model presented in Section II. The received uplink pilot signals (8) are taken as the inputs to our deep neural network (DNN). Hence, a dataset is generated via Monte-Carlo simulations for an ELAA BS with 30×30 antenna elements, resulting in 900 complex-valued features. To accommodate deep learning networks that work with real numbers, we split the complex numbers into real and imaginary parts, doubling the feature dimension to 1800 [12]. Each region that the users perceive at the ELAA-based BS is assigned a unique integer label as the output.

Two main scenarios are considered with five users ($K = 5$ and 5 VRs) and ten users ($K = 10$ and 10 VRs). Under each scenarios, two cases are considered as overlapping and non-overlapping VRs, totaling upto 4 cases. For each 5-VR dataset, we generate 50000 samples, each consisting of 1801 elements, where the initial 1800 represent features, and the last one indicates integer-labeled region output. Each VR contains 10000 samples. In the 10-VR dataset, we have 100000 samples, all sharing the same 1801-sample length, with each VR having 10000 samples. Thereby, we maintain a constant number of samples per region for all four cases. We perform data normalization, scaling all sample values within the range of -1 to 1. This normalization expedites the learning of optimal parameters in a neural network, leading to improved convergence speed [12].

In the 5-VR scenario, each region consists of 180 antenna elements, and in the 10-VR case, each region consists of 90 antenna elements. In the overlapping scenario, we introduce a 6% overlap between regions. The summary of dataset preparation is presented in Table I.

V. Deep Learning Model Architecture

Our proposed neural network architecture is tailored for time-series classification tasks, distinguishing itself from traditional machine learning algorithms, which are reported to be less efficient than deep learning frameworks [11].

Number of users	$K = 5$	$K = 10$
Number of regions	5	10
Total antenna elements	900 (30×30)	900 (30×30)
No. of antennas in each region	180	90
Feature dimension	900 (complex)	900 (complex)
Dataset sample size	50000	100000
Samples per region	10000	10000
Sample length	1801	1801
Overlapping/non-overlapping	Both	Both

TABLE I: Two scenarios with two different number of users.

The proposed model combines a 1D convolutional layer, a dense layer, a 1D max-pooling layer, a dropout layer, an LSTM layer, and a final output layer for classification tasks (see Fig. 3). As reported in the existing literature related to various architectures for time-series labeled multi-class classification problems [11], [12], [14], the LSTM-based architecture posed a computational burden due to its significantly lengthy training time in our specific context. Our experimentation with the 1D CNN architecture resulted in lower accuracy when compared to our proposed model. Further details on the comparison between LSTM, CNN, and the final proposed architecture are provided in Section VI. The final proposed architecture demonstrates its efficiency when configured with 64 filter numbers, a size of 2 in the CNN layers, and 64 units in the LSTM layer. The architectural features are summarized in Table II.

During training, our model effectively learns to predict the correct integer label (related to each VR) based on the input features. The softmax output layer consists of 5 units in the 5-VR scenario and 10 units in the 10-VR scenario, generating probability scores for each class. The class with the highest probability is ultimately predicted.

Layer name	**Layer information**	**Output dimensions**
Input	Input shape: (1800, 1)	(1800, 1)
Conv1D	Filters: 64, Filter Size: 2x1, ReLU	(1799, 64)
Dense	Units: 32, ReLU	(1799, 32)
MaxPooling1D	Pooling Size:(2)	(899, 32)
Dropout	0.3	(899, 32)
LSTM	Units: 64, Return Sequences: True	(899, 64)
Flatten	Flattened vector	(None, 57536)
Dense	Units: 10 or 5, Softmax	(10) or (5)

TABLE II: Model architecture details.

VI. Numerical Results

In this section, we present our numerical results which validate our design/analysis and illustrate the performance of our proposed deep learning framework. The path-losses of the channels between the users are modeled as $\zeta_k = \zeta_0 + 10\kappa \log(d)$, where $\kappa = 3.5$ is the path-loss exponent, $\zeta_0 = 42\,\mathrm{dB}$ is the reference path-loss, and d is the distance in

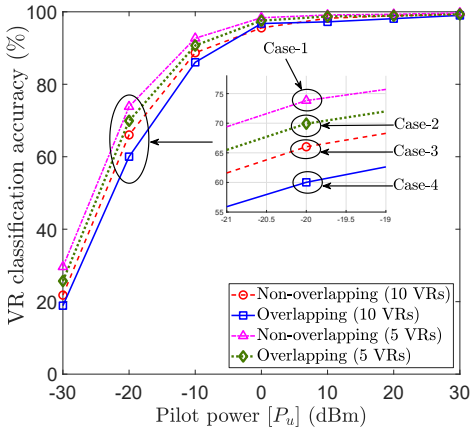

Fig. 4: The accuracy versus the pilot transmit power.

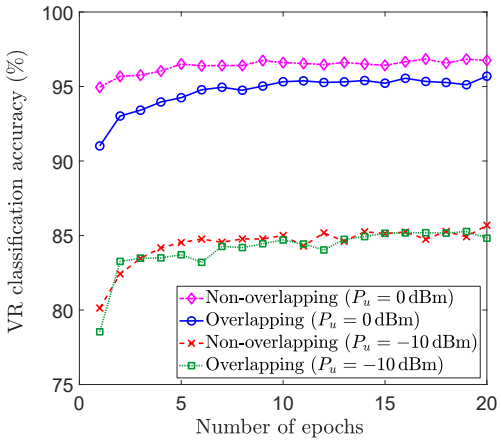

Fig. 5: The accuracy versus the epoch number with $K = 10$.

meters. The ELAA of the BS is equipped with 900 antennas ($M = 900$). The distances between the BS and users are set in between 50 m and 100 m. The noise power is modeled as $\sigma_n^2 = 10\log_{10}(N_0 B N_F)$, where the bandwidth $B = 10$ MHz, the noise figure $N_F = 7$ dB, and $N_0 = -174$ dBm/Hz. The pilot length is equal to the number of users ($\tau_u = K$), and the coherence interval consists of 196 symbols ($\tau_c = 196$) [2].

The proposed 1D-CNN-LSTM architecture was tested with batch sizes of 16, 32, and 64, and the best performance was observed with a batch size of 64. We incorporated dropout layers after 1D max-pooling layer (with a dropout rate of 0.3) and after the LSTM layer (also with a dropout rate of 0.3). The study revealed that the optimal placement for the dropout layer was after the 1D max-pooling layer with a dropout rate of 0.3. Further experiments involved testing different kernel sizes and numbers in the 1D-CNN layer and LSTM units. The best results were obtained with a kernel size of (64,2) in the CNN layer and 64 units in the LSTM layer. Other combinations suffered from generalization issues, as shown in Table III. The proposed 1D-CNN-LSTM architecture underwent training, testing, and validation on a 70-20-10 dataset split. We employed the Adam optimizer with a learning rate of 0.001.

Model	CNN (K, S)	LSTM units	Batch size	Dropout	Accuracy
1	(64,2)	64	64	0.3	High
2	(32,5)	32	32	0.3	Low
3	(32,7)	32	32	0.3	Low
4	(64,5)	64	64	0.3	Low
5	(64,7)	64	64	0.3	Low

TABLE III: Experimental configurations and results.

In Fig. 4, we investigate the VR classification accuracy versus the pilot transmit power (P_u) for four different cases. In Case-1 and -2, we assume 5 non-overlapping and overlapping VRs, respectively, while in Case-3 and -4 there are 10 non-overlapping and overlapping VRs, respectively. Fig. 4 clearly illustrates that the accuracy increases with the pilot transmit power. This is because in the presence of partial visibility, the antennas in the VR will receive higher pilot signal power, while the non-visible antennas have noise only. As expected, 5-VR cases (Case-1 and -2) perform better than the 10-VR cases (Case-3 and -4) mainly due to the fewer number

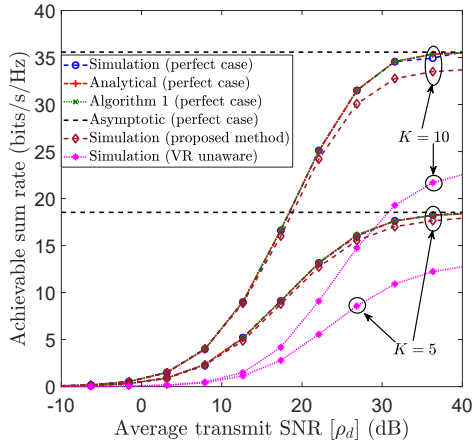

Fig. 6: The achievable sum rate versus average transmit SNR.

of confusions in 5-VR cases. It is evident that the non-overlapping cases perform better than the overlapping cases generally. As an example, with $P_u = 10$ dBm Case-1, -2, -3, and -4 provide accuracy of 98.71%, 96.75%, 96.6%, and 96.57%, respectively. In comparison, we explored LSTM and 1D CNN-based models separately. With the same pilot transmit power, the 1D CNN model achieved a maximum accuracy of 78% to 81% for all scenarios with the same learning rate and batch size, while the LSTM model reached 85% to 87%.

Fig. 5 depicts the accuracy of our proposed method against the number of epochs with $K = 10$ for two different pilot transmit powers; $P_u \in \{-10, 0\}$ dBm. Attaining the VR classification accuracy of 85.45% and 83.89% in non-overlapping and overlapping cases at a pilot transmit power of -10 dBm, and achieving accuracy of 96.82% and 95.45% for the same two cases at 0 dBm, our model exhibited rapid convergence within the initial epochs in both scenarios. The rapid convergence trend aligns with previously reported AMC models [12]. This shows that our model has desirable generalization characteristics and computational efficiency.

In Fig. 6, we compare the sum rate of the proposed VR-aware MRT precoder against the perfect case (with 100% VR classification accuracy) and a VR-unaware MRT precoder. The analytical and asymptotic rate curves are obtained via (17) and (19), respectively. Here, we have used non-overlapping VR

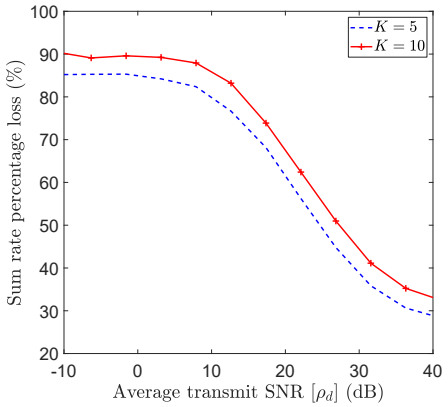

Fig. 7: The sum rate loss due to VR-unaware precoding.

cases for $K = 5$ and $K = 10$ with 98.71% and 96.6% VR classification accuracy, respectively. The VR-unaware MRT precoder is designed by assuming full visibility although the true channels exhibit VRs. At an average transmit SNR (ρ_d) of 30 dB, when compared against the sum rate provided by the perfect case with 100% VR classification accuracy (i.e., rates of 18.52 bits/s/Hz for $K = 5$ and 35.57 bits/s/Hz for $K = 10$), the sum rate achieved via the proposed method drops down to 95.46% (17.68 bits/s/Hz) and 95.08% (33.82 bits/s/Hz) for $K = 5$ and $K = 10$ cases, respectively. This drop in rate occurs due to two reasons. Firstly, in a given sample, when a VR of a single user is miss-classified (while correctly identifying the VRs of the remaining users), the downlink precoder is constructed imperfectly. This not only reduces the rate of the miss-classified user, but also reduces the rates of the remaining users by increasing the inter-user interference. Secondly, a single miss-classification will affect the transmit power allocation and thereby eventually reducing the rates of users with accurately classified VRs. Hence, Fig. 6 highlights the importance of having a higher accuracy for the VR classification in ELAAs. At $\rho_d = 30$ dB, the sum rate further drops down to 64.14% and 58.88% for VR-unaware precoder when compared against the proposed VR-aware precoder for $K = 5$ and $K = 10$, respectively. Our method only utilizes the antennas visible to a user when designing its precoder. Hence, it minimizes the inter-user interference. Since the VR-unaware precoder assumes full visibility, the inter-user interference is however higher. The VR-unaware precoder also leads to inefficient transmit power allocation.

In Fig. 7, the sum rate percentage loss caused by current state-of-the-art VR-unaware MRT precoding is studied when compared to the proposed VR-aware MRT precoding. In VR-unaware precoding, regardless of VRs, all antennas in the ELAA transmit all user signals leading to higher inter-user interference. However, in VR-aware precoding, a given antenna in ELAA transmits signals intended for the users that are visible to that antenna. Fig. 7 also reveals that the VR-unaware precoding yields higher losses when the number of users increases due to enhanced inter-user interference. In lower SNR regime, the sum rate loss is higher since the limited amount of transmit power is shared among the users when designing a VR-unaware precoder by assuming full visibility.

VII. Conclusion

We proposed 1D-CNN-LSTM deep learning architecture to accurately classify VRs in ELAA-based wireless systems. Our proposed deep ANN model, a hybrid of 1D CNN and LSTM, demonstrated higher accuracy in classifying VRs based on received pilot signals at the ELAA-based BS. The uplink channels have been estimated by using MMSE criterion. A computationally efficient transmit power allocation optimization algorithm has been proposed to prioritize user-fairness by guaranteeing a common achievable rate for all users irrespective of their different VRs and near-far effects. The effectiveness of our proposed 1D-CNN-LSTM architecture in classifying VRs in ELAAs has been validated through a rigorous set of numerical results. The robustness, accuracy, and efficiency of the proposed deep learning framework position it as a valuable tool to improve the quality of channel estimation and VR-aware precoder design for ELAAs.

References

[1] E. Bjornson *et al.*, "Massive MIMO is a reality - What is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, no. 9, pp. 3–20, Nov. 2019.
[2] T. L. Marzetta *et al.*, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
[3] X. Gao, F. Tufvesson, and O. Edfors, "Massive MIMO channels - Measurements and models," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2013, pp. 280–284.
[4] X. Li *et al.*, "Capacity Analysis for Spatially Non-Wide Sense Stationary Uplink Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 7044–7056, 2015.
[5] S. Wu *et al.*, "A Non-Stationary 3-D Wideband Twin-Cluster Model for 5G Massive MIMO Channels," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1207–1218, 2014.
[6] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 GHz," in *Proc. 6th Eur. Conf. Antennas Propag.*, 2012, pp. 433–437.
[7] E. De Carvalho *et al.*, "Non-Stationarities in Extra-Large-Scale Massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, 2020.
[8] A. Ali, E. De Carvalho, and R. W. Heath, "Linear Receivers in Non-Stationary Massive MIMO Channels With Visibility Regions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 885–888, 2019.
[9] D. Gunasinghe and G. Amarasuriya, "Achievable Rate Analysis for Extra-Large RIS-Aided Massive MIMO with Visibility Regions," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2023, pp. 1542–1547.
[10] J. Zhang *et al.*, "Average Spectral Efficiency for TDD-based non-Stationary XL-MIMO with VR Estimation," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2022, pp. 973–977.
[11] L. Dai *et al.*, "Deep Learning for Wireless Communications: An Emerging Interdisciplinary Paradigm," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 133–139, 2020.
[12] A. Jagannath and J. Jagannath, "Multi-task Learning Approach for Modulation and Wireless Signal Classification for 5G and Beyond: Edge Deployment via Model Compression," *arXiv:2203.00517*, 2022.
[13] H. Huang *et al.*, "Deep Learning for Physical-layer 5G Wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, 2019.
[14] S. Rajendran *et al.*, "Deep Learning Models for Wireless Signal Classification with Distributed Low-cost Spectrum Sensors," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, 2018.
[15] S.-H. Kim *et al.*, "A Hybrid Deep Learning Model for Automatic Modulation Classification," *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 313–317, 2022.
[16] D. Gunasinghe and G. Amarasuriya, "Statistical CSI Based Phase-Shift and Transmit Power Optimization for RIS-Aided Massive MIMO," in *Proc. IEEE Glob. Commun. Conf. (Globecom)*, 2022, pp. 4007–4012.
[17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, New York, NY, US, 2004.