



# Article Imbalanced Learning-Enhanced Beam Codebooks towards Imbalanced User Distribution in Millimeter Wave and Terahertz Massive MIMO Systems

Zhiheng Chen<sup>1</sup>, Pei Liu<sup>1,2,\*</sup> and Kehao Wang<sup>1</sup>

- <sup>1</sup> School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China; leochenzhiheng@whut.edu.cn (Z.C.); kehao.wang@whut.edu.cn (K.W.)
- <sup>2</sup> Key Lab of Broadband Wireless Communication and Sensor Network Technology (Nanjing University of Posts and Telecommunications), Ministry of Education, Nanjing 210003, China
- \* Correspondence: pei.liu@whut.edu.cn

Abstract: Millimeter wave (mmWave) and terahertz (THz) massive MIMO architectures are pivotal in the advancement of mobile communications. These systems conventionally utilize codebooks to facilitate initial connection and to manage information transmission tasks. Traditional codebooks, however, are typically composed of numerous single-lobe beams, thus incurring substantial beam training overhead. While neural network-based approaches have been proposed to mitigate the beam training load, they sometimes fail to adequately consider the minority users dispersed across various regions. The fairness of the codebook coverage relies on addressing this problem. Therefore, we propose an imbalanced learning (IL) methodology for beam codebook construction, explicitly designed for scenarios characterized by an imbalanced user distribution. Our method begins with a pre-clustering phase, where user channels are divided into subsets based on their power response to combining vectors across distinct subareas. Then, each subset is refined by a dedicated sub-model, which contributes to the global model within each IL iteration. To facilitate the information exchange among sub-models during global updates, we introduce the focal loss mechanism. Our simulation results substantiate the efficacy of our IL framework in enhancing the performance of mmWave and THz massive MIMO systems under the conditions of imperfect channel state information and imbalanced user distribution.

Keywords: beam codebook design; imbalanced learning; millimeter wave; terahertz; massive MIMO

# 1. Introduction

Millimeter wave (mmWave) and terahertz (THz) multiple-input multiple-output (MIMO) communication systems are crucial in driving the progress of the fifth generation (5G) mobile communication and future Beyond 5G. These systems incorporate large-scale antenna arrays, providing notable beamforming gains and adequate signal reception capability. Nonetheless, the implementation of fully digital transceiver architectures, necessitating an RF chain for each antenna, faces limitations due to the cost and power consumption issues linked with high-frequency mixed-signal circuits [1]. As a result, mmWave and THz systems frequently adopt either fully analog or hybrid analog/digital architectures for their combining operations [2,3]. In these configurations, transceivers utilize the networks of phase shifters. Due to the challenges associated with channel estimation and feedback, they commonly employ predefined single-lobe beamforming codebooks, such as DFT codebooks [4,5]. These classical pre-defined beamforming codebooks typically encompass a multitude of single-lobe beams, with each beam capable of steering the signal in a specific direction. These classical beamforming codebooks encounter two primary issues. Firstly, they involve extra beam training costs by covering all possible directions, which includes the redundant beams for



Citation: Chen, Z.; Liu, P.; Wang, K. Imbalanced Learning-Enhanced Beam Codebooks towards Imbalanced User Distribution in Millimeter Wave and Terahertz Massive MIMO Systems. *Electronics* 2023, 12, 4768. https://doi.org/ 10.3390/electronics12234768

Academic Editor: Da Chen

Received: 5 November 2023 Revised: 18 November 2023 Accepted: 20 November 2023 Published: 24 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). beam training. Secondly, they slightly accommodate specific scenarios such as non-line-of-sight (NLOS) conditions or imbalanced user distribution.

Recent developments have focused on enhancing the adaptability of beam codebooks to user distributions in both LOS and NLOS conditions using neural network (NN)-based approaches [6,7]. Moreover, reinforcement learning has also been utilized to adapt the beam patterns based on the surrounding environment [8]. Nevertheless, the challenges posed by imbalanced or non-uniform user distribution, crucial for 5G deployment [9,10], remain largely unexplored. Typically, in imbalanced user distribution, majority users significantly outnumber the minority users. This is related to the imbalanced long-tail data challenge in machine learning, which hampers the neural network's ability to generalize for minority classes [11]. In this case, imbalanced user distribution can lead to either overlooking or underperforming combined gain for these minority users, especially when applying NNbased methods in codebook design. Ensuring equitable codebook coverage necessitates an effective approach to address the data imbalance-related problem. Outside the realm of wireless communication, imbalanced learning (IL)-based solutions typically involve resampling the different classes in [12,13] and re-weighting featured by adjusting the loss calculated from the majority and minority classes in [14–16]. Furthermore, ensemble learning, a technique that aims to enhance the performance of a single classifier by integrating multiple complementary classifiers, has demonstrated its effectiveness in adapting to and generalizing from imbalanced data [17,18]. However, directly applying these strategies in mmWave and THz communications overlooks the richness of the geometric and physical attributes of the training data, i.e., the channels. Consequently, it is natural to propose an IL-based codebook specifically designed for imbalanced user distributions in mmWave and THz communication systems, which meets the following criteria. Firstly, the IL method aims to fully exploit the physical features of the channels. This is achieved by implementing pre-clustering, which is similar to a reverse process of DFT codebook-based combining. Secondly, the method inherits the superior feature extraction capability of NN-based approaches. This is accomplished through the utilization of multiple sub-models to deal with different classes of channels. Each of these sub-models contributes to the global model, which is similar to the essentials of ensemble learning.

In this paper, our work centers around the advancement of codebook design enhanced by the IL framework, particularly in scenarios where user distribution is notably imbalanced. The proposed approach commences with a pre-clustering of user channels based on the power responses of distinct combining vectors across various subregions, which enables a distinction between channels associated with majority and minority users. The classification creates multiple subsets. Moving forward, we introduced an innovative IL architecture to process these identified subsets. Each subset is then associated with a phase shift sub-model, which contributes to the global model after undergoing sub-forward and sub-backpropagation passes. To bolster information exchange across sub-models, we employ the focal loss function during global updating. Ultimately, iterative IL updating rounds for both global and sub-models allow the IL-based codebook to adapt to the imbalanced user distribution under imperfect channel conditions, which outperforms the NN-based baseline in the achievable rate metric.

Throughout this paper, vectors and matrices are represented by lower-case and uppercase boldface letters, respectively. The sets  $\mathbb{R}^{M \times K}$  and  $\mathbb{C}^{M \times K}$  denote the real and complex spaces of dimensions  $M \times K$ . The Hermitian transpose and the transpose of a matrix **A** are indicated by  $\mathbf{A}^{\dagger}$  and  $\mathbf{A}^{T}$ , respectively. We denote an  $M \times M$  identity matrix by  $\mathbf{I}_{M}$  and an  $M \times N$  zero matrix by  $\mathbf{0}_{M \times N}$ . A complex Gaussian random vector **x** is characterized as  $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{x}}, \boldsymbol{\Sigma})$ , where  $\bar{\mathbf{x}}$  and  $\boldsymbol{\Sigma}$  represent the mean vector and the covariance matrix, respectively. The notation  $\|\cdot\|_n$  denotes the *n*-norm of a vector.

The rest of this paper is organized as follows: Section 2 details the system model. Section 3 provides the problem description and defines the optimization objective. In Section 4, we introduce the proposed IL scheme. Section 5 describes the experimental setup and discusses the simulation results. Finally, Section 6 concludes the paper.

#### 2. System Model

## 2.1. Channel Model

This study investigates a mmWave and THz massive MIMO communication system that accommodates U single-antenna users dispersed over F subregions, which utilizes a base station (BS) equipped with a uniform linear array (ULA) comprising M antennas. The single-cluster geometric channel vector between the single antenna uth user and a M-antenna BS can be written as [19]

$$\mathbf{h}_{u} = \sqrt{\frac{M}{L}} \sum_{l=1}^{L} \alpha_{u,l} \mathbf{a}(\psi_{u,l}), \tag{1}$$

where  $\sqrt{\frac{M}{L}}$  is the normalization factor,  $\alpha_{u,l}$  denotes the complex gain of the *l*th path, and there are *L* paths in total. Also, the term  $\psi_{u,l}$  represents the angle of arrival (AOA) of the *l*th path in the angle domain  $\Psi \triangleq [0, \pi)$ . The array response vector  $\mathbf{a}(\cdot)$  is given by the ULA geometric structure of the *M*-antennas as

$$\mathbf{a}(\psi_{u,l}) = \frac{1}{\sqrt{M}} [1, \dots, e^{-j\frac{2\pi}{\lambda}d(m-1)\cos(\psi_{u,l})}, \dots, e^{-j\frac{2\pi}{\lambda}d(M-1)\cos(\psi_{u,l})}]^{\mathrm{T}} \in \mathbb{C}^{M \times 1},$$
(2)

where  $\lambda$  is the wavelength and *d* signifies the spacing between antennas. To comprehensively characterize the channel states of all *U* users, we introduce the concept of a channel set denoted as  $\mathcal{H}$ . It is formally defined as  $\mathcal{H} \triangleq \bigcup_{u=1}^{U} \mathbf{h}_{u}$  and hence  $|\mathcal{H}| = U$ .

## 2.2. User Distribution

To streamline the representation of user distribution, we divide the channel set  $\mathcal{H}$  into F subsets. This can be expressed as:  $\mathcal{H} = \bigcup_{f=1}^{F} \mathcal{H}_{f}$ . Here, each user channel group  $\mathcal{H}_{f}$  encompasses the channels of designated users belonging to class f (f = 1, ..., F). In more detail, by employing the mean AOA of the user's channel as the feature of a cluster [19], the channel subset  $\mathcal{H}_{f}$  comprises channels with the same mean AOA values that fall within the interval  $\Psi_{f} \triangleq [(f-1)\frac{\pi}{F}, f\frac{\pi}{F})$ . Here,  $\Psi$  is partitioned into F nonoverlapping subregions. Subsequently, an imbalanced user distribution is manifested by a substantial difference in the cardinality between one subset and the others, i.e.,  $\exists n \in \{1, ..., F\}$  s.t.  $\forall f \in \{1, ..., F\}$  and  $f \neq n$ ,  $|\mathcal{H}_{n}| \gg |\mathcal{H}_{f}|$ . Consequently, we can interpret  $\mathcal{H}_{n}$  as  $\mathcal{H}_{mjr}$ , representing the channel set of the majority of users. The channels not encompassed by  $\mathcal{H}_{mjr}$  are collectively categorized under  $\mathcal{H}_{mnr}$ , which denotes the channel subsets corresponding to the minority user groups.

# 2.3. Uplink Transmission

In the uplink data transmission stage, the *u*th user sends the uplink symbol  $x_u \in \mathbb{C}$  to the BS. After combining in BS, the processed signal  $z_u \in \mathbb{C}$  can be represented as

$$z_u = \mathbf{c}_{\epsilon_u}^{\dagger} \mathbf{h}_u x_u + \mathbf{c}_{\epsilon_u}^{\dagger} \mathbf{n}_u, \tag{3}$$

where the symbol  $x_u$  has a mean power limitation  $\mathbb{E}[|x_u|^2] = P_{x_u}$ . Also, the noise vector  $\mathbf{n}_u$  at the BS follows  $\mathcal{N}_{\mathbb{C}}(\mathbf{0}_{M \times 1}, \sigma^2 \mathbf{I}_M)$ . Next, the combining vector  $\mathbf{c}_{\epsilon_u}$  is the  $\epsilon_u$ th beam of the N-beam codebook  $\mathcal{C}$ , i.e.,  $\mathcal{C} \triangleq \bigcup_{n=1}^{N} {\mathbf{c}_n}, |\mathcal{C}| = N$ , and  $\epsilon_u \in {1, ..., n, ..., N}$ . In particular,  $\mathbf{c}_{\epsilon_u}$  is given as

$$\mathbf{c}_{\epsilon_{u}} = \frac{1}{\sqrt{M}} \left[ e^{j\gamma_{\epsilon_{u},1}}, ..., e^{j\gamma_{\epsilon_{u},m}}, ..., e^{j\gamma_{\epsilon_{u},M}} \right]^{\mathrm{T}}, \tag{4}$$

where  $\gamma_{\epsilon_u,m}$  represents the phase shift applied at the *m*th BS antenna. Importantly,  $\mathbf{c}_{\epsilon_u}$  is only responsible for modifying the phase of the signal and does not affect the processed signal power. To this end, it is possible to optimize the configuration of  $\mathbf{c}_{\epsilon_u}$ .

## 3. Problem Definition

In this paper, our primary goal is to devise a design for an *N*-beam codebook, referred to as C, which can adeptly accommodate the imbalanced user distribution outlined in Section 2.2. Given the substantial generalization and adaptability attributes inherent in NN architectures, we selected this approach as the backbone network to create an adaptive codebook. Hence, we aim to reframe the construction of codebook C into an optimization problem based on NNs in this section.

First of all, our aim is to formulate a beamforming codebook C with the target of optimizing the combining gain. This gain is defined as  $g_{u,\epsilon_u} \triangleq |\mathbf{c}_{\epsilon_u}^{\dagger} \mathbf{h}_u|^2$  [6]. To this end, the optimal  $\epsilon_u$  is determined by exhaustively searching through  $\mathbf{c}_n$  in the codebook, i.e.,

$$\epsilon_{u}^{*} = \arg \max_{n} xg_{u,n} = \arg \max_{n} \left| \mathbf{c}_{n}^{\dagger} \mathbf{h}_{u} \right|^{2}, \qquad (5)$$
  
s.t.  $n = 1, \dots, N,$ 

where  $\epsilon_u^*$  signifies the selection of the most suitable codeword selected from the codebook C. Following this, the optimization problem is elucidated. Specifically, the task is to maximize the combining gain across the entire dataset of user channels with the designed codebook. The expression of the superlative codebook  $C_{opt}$  can be rewritten as

$$\mathcal{C}_{\text{opt}} = \arg \max_{\mathcal{C}} \sum_{u=1}^{U} \left| \mathbf{c}_{\epsilon_{u}^{*}}^{*} \mathbf{h}_{u} \right|^{2},$$

$$\text{s.t.} \left| [\mathbf{c}_{\epsilon_{u}^{*}}]_{m} \right| = \frac{1}{\sqrt{M}}, \ m = 1, ..., M.$$
(6)

The above-mentioned restriction is introduced to ensure adherence to power limitations imposed by the combining vectors.

Furthermore, a prominent challenge inherent in optimizing this problem lies in the imbalanced distribution of users, notably the substantial imbalance between the majority and minority subsets. This disproportionality leads to a faint combining gain improvement for minority user channels when employing conventional NN-based methods. This phenomenon can be attributed to the fact that the loss values computed from the minority user channels contribute insignificantly to the overall loss during batch updates in traditional NN-based methods. Consequently, the associated combining gains for the minority users from these updates remain almost the same. In the subsequent section, we will conduct an in-depth analysis of the underlying causes for the limited contribution of the minority user losses and propose an innovative IL-based solution to effectively mitigate this issue.

#### 4. Imbalanced Learning

In this section, we employ IL to construct a more adaptive codebook in scenarios characterized by imbalanced user distribution, as compared to the NN-based approach. This adaptability arises from the independent updates of sub-models and the fair aggregation process applied to the global model. To achieve this, we introduce an initial pre-clustering method that capitalizes on the channel properties to effectively differentiate between  $\mathcal{H}_{mnr}$ and  $\mathcal{H}_{mjr}$ , subsequently assigning them to separate subsets. Following this, we propose an IL framework to attain the desired beam patterns for these diverse subsets. Finally, we present a global model structure ensuring the equitable integration of the sub-models.

#### 4.1. Model Architecture

## 4.1.1. Pre-Clustering Process

According to the user distribution elucidated in Section 2.2, our primary objective is to categorize users and their associated channels into F' distinct clusters. This process results

in distinguishing between  $\mathcal{H}_{mnr}$  and the corresponding  $\mathcal{H}_{mjr}$ . Additionally, segmenting  $\mathcal{H}$  into F' clusters can be likened to partitioning  $\Psi$  into F' discrete regions. This segmentation effectively characterizes each cluster by the median angle, denoted by  $\theta_{f'}$ , of subregion  $f'(\theta_{f'} \triangleq \frac{(2f'-1)\pi}{2F'})$  [20]. Consequently, by defining the median angle of each cluster as a distinctive feature, we can express the pre-clustered index  $f' \in \{1, ..., F'\}$  as follows

$$f' = \arg\max_{i} \left[ \mathbf{P}_{\text{PRE}} \right]_{i} = \arg\max_{i} \left[ \left| \mathbf{C}_{\text{PRE}}^{\dagger} \mathbf{h}_{u} \right|^{2} \right]_{i},$$
(7)

where the pre-clustering power vector  $\mathbf{P}_{PRE} = [\mathcal{P}_1, ..., \mathcal{P}_{f'}, ..., \mathcal{P}_{F'}]^T \in \mathbb{R}^{F' \times 1}$ , the power response to each combining vector is defined as  $\mathcal{P}_{f'} \triangleq |\mathbf{a}(\theta_{f'})^{\dagger}\mathbf{h}_u|^2$  and the pre-clustering codebook matrix  $\mathbf{C}_{PRE} = [\mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_{f'}), ..., \mathbf{a}(\theta_{F'})] \in \mathbb{C}^{M \times F'}$ . Also, the function  $\mathbf{a}(\cdot)$  signifies the array response for a given AOA. This further signifies that the channels in  $\mathcal{H}_{f'}$  share analogous power responses to the respective featured combining vector within the f'th subregion. To implement the above equation in the form of NN, the complex matrix multiplication, power computation, and argmax can be mapped into the basic NN components. These components will be introduced in the next subsection. Then, the pre-clustered channels, which belong to the subset  $\mathcal{H}_{f'}$ , will be stored in the f'th subset register waiting to be processed by the corresponding sub-model. After partitioning  $\mathcal{H}$  into F' distinct types, the next step involves deploying a model to acquire the features of  $\mathcal{H}_{f'}$ . Specifically, this paper adopts a focal loss and cross entropy loss combined IL model improved from the original NN-based self-supervised method outlined in [7]. The details of the IL forward pass architecture will be elaborated in the subsequent parts.

**Remark 1.** The pre-clustering process draws inspiration from the DFT codebook  $C_{DFT}$ , denoted by  $\mathbf{c}_n^{DFT} = \frac{1}{\sqrt{M}} \left[ 1, ..., e^{j\frac{2\pi n}{M}}, ..., e^{j\frac{2\pi n(M-1)}{M}} \right]^T$ , n = 0, 1, ..., N - 1. Given that the angle domain  $\Psi$  is defined as the range  $[0, \pi)$ , it is divided into M subareas, each with an equal interval of  $\frac{2\pi}{M}$ . Consequently, when a channel's mean AOA is congruent with a combining vector for a specific interval within  $\Psi$ , a superior power response is realized. In essence, the DFT codebook covers the entire angular space, which ensures that there is a combining vector catering to the specific channel, and thus the highest power response is attained among the pre-defined combining vectors. However, the best power response within the pre-defined combining vectors may not be optimal and requires the more adaptive NN-based method to optimize.

Informed by the exhaustive angular coverage of the DFT codebook, the initial step of our method involves calculating the power response for each channel across the subdivided angular sectors. Through this assessment, we identify the superior combining vector that results in the highest power  $\mathcal{P}_{f'}$  among the pre-defined combining vectors. This process facilitates the determination of a channel's pre-clustering class f' where the channels share similar power responses.

#### 4.1.2. Complex-Valued Fully Connected and Power Computation Layer

The layer is designed to perform complex-valued multiplicative and additive operations in the neural network backbone. In this context, the signal post-combining is represented by the resultant inner product, i.e.,

$$\mathbf{r} = \mathbf{C}_{f'}^{\dagger} \mathbf{h},\tag{8}$$

where  $\mathbf{y} = \begin{bmatrix} y_1, ..., y_n, ..., y_{N_{f'}} \end{bmatrix}^T \in \mathbb{C}^{N_{f'} \times 1}$  signifies the received signal vector after combining and  $\mathbf{h} \in \mathbb{C}^{M \times 1}$  denotes a channel from the subset  $\mathcal{H}_{f'}$ . Also, the term  $\mathbf{C}_{f'} \in \mathbb{C}^{M \times N_{f'}}$  denotes the *f*'th sub-model codebook matrix. Moreover, unlike the conventional NN approaches in computer vision, the elements of the codebook matrix do not function as parameters of the fully connected layer. Instead, the codebook matrices are generated based on the real NN weights [6]. This operation is conducted through the phase-to-complex conversion, scaled by  $\frac{1}{\sqrt{M}}$ , which is given as

$$\mathbf{C}_{f'} = \frac{1}{\sqrt{M}} \Big( \cos\left(\mathbf{\Gamma}_{f'}\right) + j * \sin\left(\mathbf{\Gamma}_{f'}\right) \Big), \tag{9}$$

where  $\Gamma_{f'} = [\gamma_1, ..., \gamma_n, ..., \gamma_{N_{f'}}] \in \mathbb{R}^{M \times N_{f'}}$  denotes the phase shift matrix and  $\gamma_n = [\gamma_{1,n}, ..., \gamma_{M,n}], n \in \{1, ..., N_{f'}\}$  represents a phase shift vector.

Following this, the power received from each combining vector is determined by evaluating the square modulus of every complex value in the received signal vector. This leads to the formulation of the received signal power  $\mathbf{p} \in \mathbb{R}^{N_{f'} \times 1}$ , which is defined as

$$\mathbf{p} = \left[p_1, \dots, p_n, \dots, p_{N_{f'}}\right]^{\mathrm{T}} = \left[|y_1|^2, \dots, |y_n|^2, \dots, |y_{N_{f'}}|^2\right]^{\mathrm{T}}.$$
 (10)

4.1.3. Softmax and Argmax Layer

This layer commences with a softmax operation to estimate the alignment 'probability' between a combining vector and the current channel **h**, which is indicated by the received power **p**. Accordingly, the softmax vector, represented by  $\mathbf{s} = [s_1, \ldots, s_n, \ldots, s_{N_{f'}}]^T \in \mathbb{R}^{N_{f'} \times 1}$ , is formulated as

$$s_n = \frac{e^{p_n}}{\sum\limits_{i=1}^{N_{f'}} e^{p_i}}.$$
(11)

On the other hand, the argmax layer is tasked with conducting the one-hot vector, denoted by  $\mathbf{o} = [o_1, \dots, o_n, \dots, o_{N_{f'}}]^T \in \mathbb{R}^{N_{f'} \times 1}$ , and it is given as

$$o_n = \begin{cases} 1, & n = \arg\max_i [\mathbf{s}]_i, \ i \in \{1, ..., N_{f'}\}, \\ 0, & \text{otherwise,} \end{cases}$$
(12)

where the arg max operation is to identify the position of the maximum value within the softmax vector. Additionally, the self-generated label vector  $\mathbf{o}$  plays a crucial role in adjusting and updating the phase shifts in the codebook. This adaptation is achieved through the utilization of the cross-entropy loss during sub-updating and the focal loss during global updating. Also, it is worth noting that a similar arg max operation is also required in the pre-clustering process to determine the class index f' associated with the optimal power response.

Transitioning to the process of learning the IL-based codebook, the subsequent subsection will delve into the details of sub-backpropagation and the comprehensive IL solution.

## 4.2. Learning the IL-Based Codebooks

After establishing the architecture for the forward pass of the sub-model, the subsequent phase entails the application of sub-backpropagation to enhance the performance of the sub-model  $\Gamma_{f'}$ . Following the sub-backpropagation, the sub-models are aggregated to assess the overall performance of the global model, which then performs the global updating through the focal loss. Ultimately, the global model is partitioned into sub-models to initiate the next round of IL.

## 4.2.1. Sub-Backpropagation

The corresponding sub-model undergoes a series of forward-backward updating iterations based on  $\mathcal{H}_{f'}$ . Within each iteration, as delineated in the preceding section, the forward processes obtain the softmax vectors **s** and one-hot labels **o** for a batch of channel vectors. In each of these iterations, the efficacy of the codebook is evaluated by

quantifying the divergence between the obtained softmax vector **s** and the target one-hot vector **o**, utilizing the cross-entropy loss function, i.e.,

$$\mathcal{L}_{B}^{CE} = \sum_{b=1}^{B} \mathcal{L}_{b}^{CE} = -\sum_{b=1}^{B} \sum_{n=1}^{N_{f'}} o_{nb} \log s_{nb},$$
(13)

where  $s_{nb}$  and  $o_{nb}$  represent the *n*th element of the *b*th tuple  $\mathbf{s}_b$ ,  $\mathbf{o}_b$  in a batch with the size of *B*. Moreover, channels for a single batch are indiscriminately chosen from both  $\mathcal{H}_{mjr}$  and  $\mathcal{H}_{mnr}$ , which constitute the comprehensive dataset  $\mathcal{H}$ . In contrast, in the IL-based batch, the batch exclusively comprises channels from either  $\mathcal{H}_{mjr}$  or  $\mathcal{H}_{mnr}$ .

Particularly, the aforementioned loss function considers the one-hot encoded label **o** as the target distribution for the model. Its objective is to minimize this value through the adjustment of phase shift vectors  $\gamma_n \in \Gamma_{f'}$  with the aim of minimizing the discrepancy between **s** and **o** to the utmost degree. Subsequently, the backpropagation gradient of phase shift vectors  $\mathbf{g}(\gamma_n)$  is computed through the chain rule, i.e., [7]

$$\mathbf{g}(\boldsymbol{\gamma}_n) = \left(\frac{\partial \mathcal{L}_B}{\partial \boldsymbol{\gamma}_n}\right)^{\mathrm{T}} = \sum_{b=1}^{B} \left(\frac{\partial \mathcal{L}_b}{\partial \mathbf{s}_b}\right)^{\mathrm{T}} \cdot \frac{\partial \mathbf{s}_b}{\partial \mathbf{p}_b} \cdot \frac{\partial \mathbf{p}_b}{\partial \mathbf{y}_b} \cdot \frac{\partial \mathbf{y}_b}{\partial \boldsymbol{\gamma}_n}.$$
 (14)

Hence, the stochastic gradient descent (SGD) based method can be employed to update the sub-model phase matrix  $\Gamma_{f'}$ , which is given as

$$\gamma_{n \text{new}} = \gamma_{n \text{current}} - \beta \cdot \mathbf{g}(\gamma_n)^{1}, \qquad (15)$$

where  $\beta$  denotes the learning rate of the SGD-based process. In conclusion, the pre-clustering process and sub-model updating are demonstrated in Figure 1. After the sub-model updates, the next stage will introduce the process of global aggregation and updating.



Figure 1. The pre-clustering process and sub-model updating.

#### 4.2.2. Global Updating

After the sub-model updating epochs, the phase shift matrices  $\Gamma_{f'}$  are fused into the global model. As previously explained,  $\Gamma_{f'}$  consists of  $N_{f'}$  distinct phase shift vectors, each corresponding to the number of combining vectors within the correlated codebook. Our objective is to evenly integrate the model across the various  $N_{f'}$  distributions, thereby addressing the challenge of imbalanced combining gain elevation in the NN-based approach.

Driven by the specified objective, the global model is constructed by concatenating the phase shift matrices of varying dimensions  $N_{f'}$  to create a unified model, which is expressed by

$$\boldsymbol{\Gamma} = \left[\boldsymbol{\Gamma}_1, ..., \boldsymbol{\Gamma}_{f'}, ..., \boldsymbol{\Gamma}_{F'}\right],\tag{16}$$

where  $\Gamma \in \mathbb{R}^{M \times N}$   $(N = \sum_{a=1}^{F'} N_a)$  represents the global phase shift matrix model and **C** denotes the corresponding global codebook. The above equation ensures the specialization of different sub-models.

Nonetheless, the process of concatenation introduces the invalid exchange of information among different sub-models, potentially sub-optimal in nature. The information exchange implies that a small proportion of the channels from a subset  $\mathcal{H}_{f'}$  tend to have better power responses on combining vectors in the adjacent sub-model after global updating, instead of its own sub-model  $\Gamma_{f'}$ . Consequently, to adapt the global user distribution within  $\mathcal{H}$ , the global model should undergo updates akin to the sub-model. However, the cross-entropy loss function is supposed to be replaced with a more adaptive focal loss function in [21]. The adaptive method is supposed to automatically rectify the imbalance in loss calculations, stemming from the imbalance of quantity between two types of channels: those seeking to change their assigned sub-models and those already well adapted to their sub-models. Specifically, the focal loss mechanism notes the lower received power (reflected on the softmax probabilities) of the exchanged channels, compared with the received power of a sub-model's originally well-adapted channels. To this end, the mechanism will introduce a more balanced loss function by improving the proportion of the losses of the low received power channels. Therefore, the focal loss function is utilized and given as [21]

$$\mathcal{L}_{B}^{\rm FL} = -\sum_{b=1}^{B} \sum_{n=1}^{N} (1 - s_{nb})^{\zeta} o_{nb} \log s_{nb}, \tag{17}$$

where the modulating factor  $(1 - s_{nb})^{\zeta}$  is added to the cross-entropy loss, with a tunable focusing parameter  $\zeta \ge 0$ . In scenarios where the received power is relatively high (i.e.,  $s_{nb} \rightarrow 1$ ), the modulating factor approaches zero. Conversely, for channels with lower received power, the modulating factor tends towards unity. As a result, when contrasted with the standard cross-entropy loss, the focal loss maintains its effectiveness for channels with lower received power while reducing the loss for those with higher received power. In essence, this adjustment amplifies the contribution of channels with lower receiver power that seeks to alter the serving sub-model within the loss function.

To initiate the subsequent IL iteration,  $\Gamma$  is separated at the columns corresponding to the prior concatenation, reverting to their pre-fusion dimensions. This step can be considered the reverse operation of Equation (16). To date, Algorithm 1 has summarized the overall IL process.

Algorithm 1 Imbalanced Learning. 1: Pre-cluster  $\mathcal{H}$  and return F' different subsets 2: Initialize *F*′ sub-models for each IL round do 3: for each subset  $\mathcal{H}_{f'}$  and the selected sub-model  $\Gamma_{f'}$  do 4: 5: for each sub-model updating epoch **do** 6: **for** each batch with size of *B* in  $\mathcal{H}_{f'}$  **do for** each channel in a batch indexed by  $b \in \{1, ..., B\}$  **do** 7: 8: Execute the sub-forward pass and obtain  $\mathbf{s}_b$ ,  $\mathbf{o}_b$ 9: Execute the sub-backpropagation pass:  $\mathcal{L}_{B}^{\text{CE}} \leftarrow \mathcal{L}_{B}^{\text{CE}} - \sum_{n=1}^{N_{f'}} o_{nb} \log s_{nb}$ end for 10: for each  $n \in \{1, ..., N_{f'}\}$  do 11: 12: Update the selected sub-model  $\Gamma_{f'}$  in the sub-backpropagation pass:  $\gamma_{n\text{new}} \leftarrow \gamma_{n\text{current}} - \beta \cdot \frac{\partial \mathcal{L}_B^{\text{CE}}}{\partial \gamma_n}$ end for 13: end for 14: 15: end for end for 16: Obtain the global model  $\mathbf{\Gamma} = [\mathbf{\Gamma}_1, ..., \mathbf{\Gamma}_{f'}, ..., \mathbf{\Gamma}_{F'}]$ 17: for each global updating epoch do 18: for each batch with size of B in  $\mathcal{H}$  do 19: **for** each channel in a batch indexed by  $b \in \{1, ..., B\}$  **do** 20: Execute the global forward pass and obtain  $\mathbf{s}_{h}$ ,  $\mathbf{o}_{h}$ 21: Execute the global backpropagation pass: 22:  $\mathcal{L}_B^{\mathrm{FL}} \leftarrow \mathcal{L}_B^{\mathrm{FL}} - \sum_{n=1}^N (1 - s_{nb})^{\zeta} o_{nb} \log s_{nb}, N = \sum_{a=1}^{F'} N_a$ end for 23: for each  $n \in \{1, ..., N\}$  do 24: Update the global model  $\Gamma$  in the global backpropagation pass: 25:  $\gamma_{n\text{new}} \leftarrow \gamma_{n\text{current}} - \beta \cdot \frac{\partial \mathcal{L}_B^{\text{FL}}}{\partial \gamma_{..}}$ end for 26: 27: end for 28: end for Execute the reverse operation of  $\Gamma = [\Gamma_1, ..., \Gamma_{f'}, ..., \Gamma_{F'}]$  and get the global-updated 29: sub-models 30: end for

**Remark 2.** While the loss function in the NN-based approach effectively quantifies the disparity between the current model and the desired response for users equally distributed, its primary limitation becomes apparent when applied in a batch-based context. Even though only the best-performing combining vectors associated with the current channel will be updated based on the one hot label, it is quite common for the same combining vector to serve both the majority and minority during a single batch. For instance, it arises when the codebook size is relatively small, and the previous batch update only contains the majority user channels and the current batch update contains minority user channels. To this end, in cases where there is an imbalance in the number of users across different regions, the loss associated with channels from  $\mathcal{H}_{mnr}$  might be unfairly averaged and consequently downplayed.

To illustrate this, it is reasonable to consider a NN-based updating process for a phase shift matrix  $\mathbf{\Gamma} \in \mathbb{R}^{M \times N}$  based on the global dataset  $\mathcal{H}$ . Let the number of channels from  $\mathcal{H}_{mjr}$  in a batch be denoted by  $B_{mjr}$  and the number of channels from  $\mathcal{H}_{mnr}$  as  $B_{mnr}$ , with  $B_{mjr} > B_{mnr}$ 

and  $B_{mjr} + B_{mnr} = B$ . In such scenarios, the loss function for NN with a batch size of B can be represented as follows

$$\mathcal{L}_{B} = -\left(\sum_{b=1}^{B_{mnr}} \sum_{n=1}^{N} o_{nb} \log s_{nb} + \sum_{b=1}^{B_{mjr}} \sum_{n=1}^{N} o_{nb} \log s_{nb}\right).$$
(18)

If  $B_{mir} \gg B_{mnr}$ , the equation above is approximated as

$$\mathcal{L}_B \approx \mathcal{L}_{B_{mjr}} = -\sum_{b=1}^{B_{mjr}} \sum_{n=1}^{N} o_{nb} \log s_{nb}.$$
(19)

This insight reveals a tendency of the NN-based method to underestimate the calculated loss in specific regions when the size of subsets in those regions is significantly smaller than the region with the predominant users. In the IL-based approaches, the shortcomings of the loss function are effectively mitigated. During the update phase of the phase shift matrix in NN, the enhancement in combining gain for minority users can be measured by the contribution of  $\mathcal{H}_{minor}$  to the overall loss function, expressed as

$$Q_{NN} \triangleq \frac{\mathcal{L}_{B_{mnr}}}{\mathcal{L}_{B}} = \frac{\sum_{b=1}^{B_{mnr}} \sum_{n=1}^{N} o_{nb} \log s_{nb}}{\sum_{b=1}^{B} \sum_{n=1}^{N} o_{nb} \log s_{nb}},$$
(20)

where Q denotes the fraction of the loss derived from  $\mathcal{H}_{mnr}$  in the cumulative loss. For a comparative perspective, by assuming the update processes of all sub-models for  $\mathcal{H}_{mnr}$  are considered, the proportional contribution  $Q_{IL}$  in the IL framework can be given as

$$Q_{IL} = \frac{\mathcal{L}_{B_{mnr}}}{\mathcal{L}_{B_{mnr}}} = \frac{\sum_{b=1}^{B_{mnr}} \sum_{n=1}^{N} o_{nb} \log s_{nb}}{\sum_{b=1}^{B_{mnr}} \sum_{n=1}^{N} o_{nb} \log s_{nb}},$$
(21)

where  $Q_{IL}$  exceeds  $Q_{NN}$  is evident in the stipulated scenario. Consequently, from Equations (20) and (21), it is inferred that minority users contribute more significantly to the loss function within the proposed framework. This ensures a more pronounced combining gain enhancement for minority users in comparison to the NN approach.

# 5. Simulation Result

In this section, we begin by providing an in-depth exposition of the dataset generation, model training, and testing configurations for the proposed algorithms. The simulation was conducted in a setting with imperfect channel conditions and a notably imbalanced user distribution.

## 5.1. Scenarios, Datasets, and Training Parameters

To assess the validity of the IL method in addressing the challenges posed by the imbalance within the channel set  $\mathcal{H}$ , we create a scenario featuring 10 non-overlapping and equally spaced subregions. The numbers of pre-clustering subsets F' in pre-clustering, BS antennas M and paths L are set as 5, 64 and 5, respectively. The AOA within each subregion follows a Laplacian distribution, all sharing the same mean angle and an angle spread of 10 degrees. More precisely, the mean AOAs in subregion n fall within the interval  $\left[\frac{\pi}{10}(n-1), ..., \frac{\pi}{10}n\right)$ . We generate a total of 10,000 realizations of  $\mathbf{h}$  in the 10th subregion, representing the majority of users. In contrast, the minority group is represented by 250 realizations of  $\mathbf{h}$ , with subregions 1,3,4,5 containing 100, 50, 50, and 50 realizations, respectively. The average achievable rate for all users is denoted by  $\frac{1}{U}\sum_{u=1}^{U} \log_2\left(1 + \frac{P_{xu}}{\sigma^2} \left|\mathbf{c}_{e_u^*}^{\dagger} \mathbf{h}_u\right|^2\right)$  and the term  $P_{xu}/\sigma^2$  is the received SNR. For assessing the performance under imperfect channel conditions,

we introduce synthetic additive white Gaussian noise to each channel's data and define  $SNR_{tr} = 20 \log_{10} \frac{[|\mathbf{h}_u|_i|^2}{\sigma'^2}$  [20]. For simulations,  $SNR_{tr}$  is set as 10 dB. To assess the efficacy of our proposed method, we perform a comparative analysis. Our method is compared with the *N*-beam DFT codebook  $\mathbf{c}_n^{\text{DFT}} = \frac{1}{\sqrt{M}} [1, ..., e^{j\frac{2\pi n}{M}}, ..., e^{j\frac{2\pi n(M-1)}{M}}]^T$ , n = 0, 1, ..., N - 1 and with the self-supervised NN solution outlined in prior work [7]. In addition, we define the upper bound equal-gain combining vector as  $\mathbf{c}_{u,egc} = \frac{1}{\sqrt{M}} \left[ e^{j \angle h_{u,1}}, e^{j \angle h_{u,2}}, ..., e^{j \angle h_{u,M}} \right]^T$ . This upper bound vector is defined based on the corresponding received power for each user, given as  $p = \left| \mathbf{c}_{u,egc}^{\dagger} \mathbf{h}_u \right|^2 = \frac{1}{M} ||\mathbf{h}_u||_1^2$  [22]. To simplify the IL model training, we maintain uniform codebook sizes across all sub-models. In a given training instance, the batch sizes during the sub-model and global updates are configured to 10 and 1000, respectively. Similarly, the epochs are set to 20 for sub-model updating and 5 for global updating. Moreover, the learning rate, the focal loss focusing parameter  $\zeta$ , and the validation rates are set as 0.001, 2, and 0.1, respectively.

# 5.2. Performance Analysis

# 5.2.1. Pre-Clustering Result

As delineated in Section 5.1, the constructed dataset comprises 10,000 realizations for region 10, while all other regions consist of 250 realizations. After pre-clustering, the channel sets  $\mathcal{H}_1$ ,  $\mathcal{H}_2$ ,  $\mathcal{H}_3$ , and  $\mathcal{H}_5$  encompass 125, 49, 95, and 9981 channels, respectively. These outcomes suggest that certain non-ideal conditions, such as constrained subregion segmentation, imperfect channel state information, and angle spacing, may impede precise clustering. However, the distinction between  $\mathcal{H}_{mir}$  and  $\mathcal{H}_{mnr}$  remains generally discernible.

#### 5.2.2. Achievable Rate and Model Convergence

Figure 2a depicts the average achievable rate against the codebook size at received SNRs of 0 dB and 5 dB. The equal-gain receiver serves as the performance upper bound. Observations from the results indicate that increased codebook size leads to augmented gains and achievable rates, which implies enhanced feature extraction during updates. Notably, saturation in the achievable rate is observed at a 64-beam codebook size, which suggests marginal gains when transitioning from 32 to 64 beams. In terms of the proposed IL method's efficacy, it surpasses the NN-based approach in the achievable rate for all cases except the 64-beam codebook. Specifically, with smaller codebooks (sizes of 8, 12, or 16 beams), the IL method demonstrates superior adaptability to  $\mathcal{H}_{mnr}$  and the imbalanced user distribution. For codebooks exceeding 16 beams, the IL method's performance is commensurate with that of the 64-beam DFT codebook, whereas the NN-based codebooks lag behind. Moreover, Figure 2b depicts the validation losses for the 16-beam IL codebook over the IL rounds, each comprising sub-model and global updating cycles. It is close to saturation in the 8th round.



**Figure 2.** The figures of achievable rate comparison utilizing different methods and the model convergence of IL-based methods: (**a**) average achievable rate versus the number of beams of the codebook utilizing different methods; and (**b**) valid loss versus the IL round.

#### 5.2.3. Beam Pattern

Figure 3 showcases the beam patterns emanating from six learned codebooks of sizes 20, 32, and 64, which demonstrates the learning outcomes for both the proposed and NN-based methods. Notably, unlike the DFT codebook, these patterns reveal that beams do not span the entire azimuth plane but are instead oriented toward user locations. However, the NN codebook tends to overlook minority channel groups in region 1, and the magnitude of beams for user channels in regions 3, 4, and 5 are not significantly enhanced after updating. This is particularly evident in the 64-beam NN codebook, where most beams cater to the majority user channels. Consequently, it fails to capture the desired beam pattern, which corroborates the analysis presented in Remark 2. In contrast, the IL-based method flexibly adapt to the minority user channels in multipath scenarios using multi-lobe beams or precisely aligns with the minority users' orientation with a magnitude exceeding 0.7.



**Figure 3.** Beam patterns for the codebook with 20, 32, 64 beams learned by the self-supervised NN and the IL solution in multipath setting: (a) 20-beam NN codebook; (b) 20-beam IL codebook; (c) 32-beam NN codebook; (d) 32-beam IL codebook; (e) 64-beam NN codebook; and (f) 64-beam IL codebook.

#### 6. Conclusions

In this paper, our focus was on the design of codebooks aided by the IL framework in the mmWave and THz massive MIMO communication systems with imbalanced user distribution. The primary contributions of our work to the field of machine learningenhanced mmWave and THz massive MIMO communication lie in revealing the negative impact of imbalanced datasets in practical communication on the overall performance of NN-based codebook design and proposing an IL-based method to handle the imbalanced samples. We began by pre-clustering  $\mathcal{H}$  based on the power responses of distinct combining vectors across various subregions, which allows us to differentiate between channels associated with majority and minority users. This classification resulted in multiple subsets. Subsequently, we introduced an innovative IL architecture to process these F' subsets. Each subset was associated with a phase shift sub-model  $\Gamma_{f'}$ , which contributed to the global model after undergoing sub-forward and sub-backpropagation passes. To enhance the information exchange across different sub-models, we utilized the focal loss function, which improved the balance during global updating. Compared to the NN-based method in previous work, the IL process ensured combining gain improvement during updates by augmenting the minority user channel's contribution to the overall loss function in each batch. Through iterative IL updating rounds for both global and sub-models, our IL-based approach effectively adapted to the imbalanced user distribution under imperfect channel conditions, surpassing the performance of the NN-based baseline in terms of achievable rate. Moreover, the depicted beam patterns further intuitively confirmed the method's

effectiveness in accommodating minority user channels. In future work, we plan to broaden the scope of our methodology to encompass hybrid beamforming techniques. This expansion will involve integrating additional modules in the backbone network for designing baseband precoders. Additionally, to test the validation of the proposed IL framework in a more practical scenario, we will consider the potential hardware impairments on the antenna array geometric that are prevalent in the real-world mmWave and THz massive MIMO systems.

**Author Contributions:** Conceptualization, Z.C. and P.L.; methodology, Z.C.; software, Z.C.; validation, Z.C.; formal analysis, Z.C.; investigation, Z.C and P.L.; writing—original draft preparation, Z.C. and P.L.; writing—review and editing, Z.C., P.L., and K.W.; supervision, P.L and K.W. All authors have read and agreed to the published version of the manuscript

**Funding:** This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB3901503, in part by the National Natural Science Foundation of China under Grant 62001336 and Grant 62371353, in part by the open research fund of Key Lab of Broadband Wireless Communication and Sensor Network Technology (Nanjing University of Posts and Telecommunications), Ministry of Education under Grant JZNY202105, and in part by the Knowledge Innovation Program of Wuhan-Shuguang Project under Grant 2023010201020316, and in part by the China Scholarship Council (CSC) under Grant 202306950052.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

#### References

- Sohrabi, F.; Yu, W. Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE J. Sel. Top. Signal Process*. 2016, 10, 501–513. [CrossRef]
- Heath, R.W.; Gonzalez-Prelcic, N.; Rangan, S.; Roh, W.; Sayeed, A.M. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J. Sel. Top. Signal Process.* 2016, 10, 436–453. [CrossRef]
- Alkhateeb, A.; Mo, J.; Gonzalez-Prelcic, N.; Heath, R.W. MIMO precoding and combining solutions for millimeter-wave systems. IEEE Commun. Mag. 2014, 52, 122–131. [CrossRef]
- 4. Hur, S.; Kim, T.; Love, D.J.; Krogmeier, J.V.; Thomas, T.A.; Ghosh, A. Millimeter wave beamforming for wireless backhaul and access in small cell networks. *IEEE Trans. Commun.* **2013**, *61*, 4391–4403. [CrossRef]
- Wang, J.; Lan, Z.; Pyo, C.w.; Baykas, T.; Sum, C.s.; Rahman, M.A.; Gao, J.; Funada, R.; Kojima, F.; Harada, H.; et al. Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems. *IEEE J. Sel. Areas Commun.* 2009, 27, 1390–1399. [CrossRef]
- Zhang, Y.; Alrabeiah, M.; Alkhateeb, A. Learning beam codebooks with neural networks: Towards environment-aware mmWave MIMO. In Proceedings of the 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 26–29 May 2020; pp. 1–5.
- 7. Alrabeiah, M.; Zhang, Y.; Alkhateeb, A. Neural networks based beam codebooks: Learning mmWave massive MIMO beams that adapt to deployment and hardware. *IEEE Trans. Commun.* **2022**, *70*, 3818–3833. [CrossRef]
- Zhang, Y.; Alrabeiah, M.; Alkhateeb, A. Reinforcement learning of beam codebooks in millimeter wave and terahertz MIMO systems. *IEEE Trans. Commun.* 2021, 70, 904–919. [CrossRef]
- Li, C.; Yongacoglu, A.; D'Amours, C. Downlink coverage probability with spatially non-uniform user distribution around social attractors. In Proceedings of the 2017 24th International Conference on Telecommunications (ICT), Limassol, Cyprus, 3–5 May 2017; pp. 1–5.
- 10. Ye, J.; Ge, X.; Mao, G.; Zhong, Y. 5G ultradense networks with nonuniform distributed users. *IEEE Trans. Veh. Technol.* 2017, 67, 2660–2670. [CrossRef]
- 11. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [CrossRef]
- Pouyanfar, S.; Tao, Y.; Mohan, A.; Tian, H.; Kaseb, A.S.; Gauen, K.; Dailey, R.; Aghajanzadeh, S.; Lu, Y.H.; Chen, S.C.; et al. Dynamic sampling in convolutional neural networks for imbalanced data classification. In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 10–12 April 2018; pp. 112–117.
- 13. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* 2009, 21, 1263–1284.
- Cui, Y.; Jia, M.; Lin, T.Y.; Song, Y.; Belongie, S. Class-balanced loss based on effective number of samples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9268–9277.
- 15. Huang, C.; Li, Y.; Loy, C.C.; Tang, X. Deep imbalanced learning for face recognition and attribute prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2781–2794. [CrossRef] [PubMed]

- 16. Cao, K.; Wei, C.; Gaidon, A.; Arechiga, N.; Ma, T. Learning imbalanced datasets with label-distribution-aware margin loss. *Adv. Neural Inf. Process. Syst.* **2019**, 32 .
- Chawla, N.V.; Lazarevic, A.; Hall, L.O.; Bowyer, K.W. SMOTEBoost: Improving prediction of the minority class in boosting. In Proceedings of the Knowledge Discovery in Databases: PKDD 2003: 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Dubrovnik, Croatia, 22–26 September 2003; Springer: Berlin/Heidelberg, Germany, 2003; pp. 107–119.
- Zhu, Z.; Wang, Z.; Li, D.; Zhu, Y.; Du, W. Geometric Structural Ensemble Learning for Imbalanced Problems. *IEEE Trans. Cybern.* 2020, 50, 1617–1629. [CrossRef] [PubMed]
- Forenza, A.; Love, D.J.; Heath, R.W. Simplified spatial correlation models for clustered MIMO channels with different array configurations. *IEEE Trans. Veh. Technol.* 2007, 56, 1924–1934. [CrossRef]
- Elbir, A.M.; Coleri, S. Federated learning for hybrid beamforming in mm-wave massive MIMO. *IEEE Commun. Lett.* 2020, 24, 2795–2799. [CrossRef]
- Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 Ocotber 2017; pp. 2980–2988.
- Love, D.J.; Heath, R.W. Equal gain transmission in multiple-input multiple-output wireless systems. *IEEE Trans. Commun.* 2003, 51, 1102–1110. [CrossRef]

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