Machine Learning Based Tuner for Frequency-Selective PAPR Reduction

Selahattin Gökceli, Taneli Riihonen, *Senior Member, IEEE*, Toni Levanen, and Mikko Valkama, *Fellow, IEEE*

***Abstract*—Frequency-selective peak-to-average power ratio (PAPR) reduction is essential in networks such as 5G New Radio (NR) that support frequency-domain multiplexing of users and services. However, stemming from the frequency-selective shaping of the involved clipping noise, the relation between the intended PAPR target and the actually realized PAPR is known to be heavily nonlinear, which complicates the PAPR reduction. In this article, a novel machine learning (ML)-based solution, called *PAPRer*, is proposed to automatically and accurately tune the optimal PAPR target for frequency-selective PAPR reduction. This is achieved by utilizing the features related to the used clipping noise filter and minimization of the defined loss function, through supervised learning, which quantifies the PAPR target estimation accuracy. An analytical clipping noise power-based method is also devised for reference purposes. Extensive numerical evaluations in 5G NR context are provided and analyzed, showing that PAPRer can very accurately predict and tune the optimal PAPR target. These results, together with the provided complexity assessment, demonstrate that the proposed PAPRer offers a favorable performance-complexity tradeoff in choosing the optimal PAPR target for frequency-selective PAPR reduction.**

***Index Terms*—5G NR, clipping and filtering, machine learning, PAPR, power-efficiency, supervised learning.**

I. Introduction

Fifth-generation New Radio (5G NR) is providing large improvements in achievable wireless data rates, reliability, latency, and energy consumption [1], [2]. Cyclic prefix (CP) orthogonal frequency-division multiplexing (OFDM) is the main physical-layer waveform of 5G NR, having multiple benefits against other physical-layer solutions; however, CP-OFDM has also an inherent problem with high peak-to-average power ratio (PAPR) [3]. To this end, iterative clipping and filtering (ICF) [4], partial transmit sequence [5], selected mapping [6], tone injection [7], and tone reservation (TR) [8] are among the prominent PAPR reduction solutions in the literature. In general, the ICF is among favorable methods as it offers a good balance between PAPR reduction performance and processing complexity [4].

The support for heterogeneous quality-of-service (QoS) requirements is one modern feature of 5G networks [1], [2]. However, conventional PAPR reduction methods do not support such feature. Specifically, ordinary ICF allocates clipping noise uniformly over the active processing band, thus distorting the physical resource block (PRB)-specific mean-squared error (MSE) or error vector magnitude (EVM) requirements [4], [9]. Instead, clipping noise should be controlled at PRB level to allow for efficient frequency-domain multiplexing of services with different MSE/EVM limits.

Manuscript received 16 May 2022; revised 16 August 2022 and 27 September 2022; accepted 10 November 2022. Date of publication 18 November 2022; date of current version 18 April 2023. This work was supported by the Academy of Finland under Grants 315858, 319994, 338224, 332361, and 341489. The review of this article was coordinated by Prof. Zhu Han. *(Corresponding Author: Mikko Valkama.)*

Selahattin Gökceli, Taneli Riihonen, and Mikko Valkama are with the Tampere University, 33720 Tampere, Finland (e-mail: selahattin.gokceli@tut.fi; taneli.riihonen@tuni.fi; mikko.valkama@tuni.fi).

Toni Levanen is with the Nokia Mobile Networks, 33100 Tampere, Finland (e-mail: toni.a.levanen@nokia.com).

Digital Object Identifier 10.1109/TVT.2022.3223157

Inspired by the above, the *frequency-selective PAPR reduction* studied in [10] shapes the clipping noise based on the PRB-specific MSE limits. The iterative clipping and weighted error filtering (ICWEF) approach proposed in [9] is one effective example of such frequency-selective PAPR reduction methods. However, the reported methods shape the clipping noise in the frequency domain, which leads to degradation in PAPR performance for certain PAPR targets [10]. Thus, unlike in conventional methods such as ICF, there is a nonlinear relation between the intended PAPR target and the actually realized PAPR. This is a new challenging problem that is associated with frequency-selective PAPR reduction, and the optimal PAPR target that leads to the lowest PAPR value must be known to maximize the performance. The trivial solution is to run the algorithm exhaustively for various different PAPR targets and then pick the optimal PAPR target therefrom. However, this is not a practical solution due to the involved high complexity.

This article proposes a machine learning (ML)-based solution, called *PAPRer*, that tunes automatically the optimal PAPR target for frequency-selective PAPR reduction, while considering ICWEF as an example application. In general, there are huge number of possible ICWEF masks [9], corresponding to applicable modulations and the associated MSE/EVM requirements at different PRBs. Thus, as the PAPR performance differs for each mask, the PAPR target used in ICWEF algorithm should be dynamically adapted. Accordingly, the PAPRer finds the nonlinear relation between the frequency-selective ICWEF mask and the optimal PAPR target, utilizing supervised learning. Additionally, the PAPRer builds only on a few data features and is thus suitable for multiple different ML algorithms. The PAPRer is also applicable with any frequency-selective PAPR reduction method — not only the considered ICWEF — as it exploits only the modulation and coding scheme (MCS) configuration and the interval of all possible PAPR targets, neither of which are intrinsic to ICWEF.

Furthermore, the PAPRer not only finds the optimal PAPR target, but also predicts the achieved PAPR value quite accurately for any given PAPR target. This is another crucial benefit which can be exploited for optimal MCS configuration. Conventionally, PAPR performance is not considered as criterion for feasible MCS values as it is not easy to estimate the PAPR performance. However, as will be shown, the PAPRer is efficient and accurate in estimating the optimal PAPR target, and can thus be used to quantify the suitability of different MCSs. Additionally, a reference method that is based on the analytical clipping noise power estimation is also introduced and compared against the PAPRer. By following the 5G NR specifications and guidelines, different numerical evaluations are conducted and the performance-complexity tradeoffs offered by the proposed PAPRer are assessed and shown.

Finally, for the purpose of clarity and presentation completeness, it is noted that in [11], the importance of utilizing the optimal PAPR target was also discussed — however, in more ordinary PAPR reduction context without frequency-selective clipping noise shaping. Moreover, multiple ML-based PAPR reduction methods, as such, have been proposed in the recent literature. To this end, a model-driven TR method was proposed in [12], involving an efficient training procedure. Training efficiency was pursued also in [13], where an extreme learning TR method was proposed. A PAPR reduction network called PRnet was, in turn, proposed in [14], where good PAPR performance was obtained by utilizing an autoencoder structure. Furthermore, an ML-based PAPR reduction method was also proposed in [15], which targets a balance between good bit error rate (BER) and PAPR performance. In [16], a deep unfolding model was proposed to reduce the PAPR under transmit

power constraints. An ML-based hyperparameter optimization method was proposed in [17] to optimize efficiently the various parameters of selected PAPR reduction schemes, with focus on massive multiple-input multiple-output (MIMO) systems. However, none of these works consider the frequency-selective PAPR reduction and the associated optimal target PAPR selection problem which are the technical focus of this article.

The rest of this article is organized as follows. In Section II, the ICWEF processing is shortly reviewed, followed by the proposed clipping noise power-based and ML-based PAPR target section methods. Numerical results are provided and analyzed in Section III, together with the PAPRer complexity assessment. Finally, conclusions are drawn in Section IV.

II. System Model and Proposed Solutions

A. ICWEF Processing Principle

The ICWEF [9] starts with the OFDM waveform processing, with N_{act} and $N_{\text{OFDM}} > N_{\text{act}}$ denoting the number of active subcarriers and the nominal OFDM transform size, respectively. The active subcarriers are here assumed to be allocated symmetrically around the direct-current (DC) bin by considering double-sided bin indexing. Oversampling factor of N_{ov} is also utilized to obtain accurate waveform characteristics [18] and the size of the oversampled inverse discrete Fourier Transfrom (IDFT) reads $N = N_{\text{OFDM}} N_{\text{ov}}$.

At the IDFT output, the discrete-time OFDM symbol samples that formally correspond to the 0$^{\text{th}}$ iteration of ICWEF read

$$x^{(0)}[n] = \frac{1}{\sqrt{N_{\text{act}}}} \sum_{k=-N_{\text{act}}/2}^{N_{\text{act}}/2-1} X^{(0)}[k] e^{j2\pi kn/N}, \tag{1}$$

where $n = 0, \ldots, N-1$ is the relative sample index inside the OFDM symbol and $k \in \{-N_{\text{act}}/2, \ldots, N_{\text{act}}/2 - 1\}$ denotes the active subcarrier index, with $X^{(0)}[k]$ denoting the data symbol at subcarrier k while other IDFT bins are set to zero. The sample-wise PAPR of the time-domain signal $x^{(l)}[n]$, with l denoting the ICWEF iteration index, is computed as

$$\text{PAPR}\left(x^{(l)}[n]\right) = \frac{|x^{(l)}[n]|^2}{\frac{1}{N}\sum_{n=0}^{N-1} |x^{(l)}[n]|^2}. \tag{2}$$

The PAPR target is represented by λ_{target} and, if $\max_{n=0,1,\ldots,N-1} \text{PAPR}(x^{(l)}[n]) > \lambda_{\text{target}}$, the soft limiter-based clipping is applied as

$$\bar{x}^{(l)}[n] = \begin{cases} A^{(l-1)} e^{j\angle x^{(l-1)}[n]}, & \text{if } |x^{(l-1)}[n]| > A^{(l-1)}, \\ x^{(l-1)}[n], & \text{otherwise}, \end{cases} \tag{3}$$

where $\angle x$ and $|x|$ represent the phase angle and modulus of a complex number x, respectively. Moreover, the clipped version of $x^{(l-1)}[n]$ is expressed by $\bar{x}^{(l)}[n]$, while $A^{(l-1)}$ that denotes the amplitude threshold value is computed as

$$A^{(l-1)} = \sqrt{\lambda_{\text{target}} \cdot \frac{1}{N} \sum_{n=0}^{N-1} |x^{(l)}[n]|^2}. \tag{4}$$

It is assumed that clipping function in (3) starts the next iteration and to reflect that, the iteration index is increased by one in this step.

The clipped time-domain signal is converted to frequency domain through discrete Fourier Transfrom (DFT) as

$$\bar{X}^{(l)}[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \bar{x}^{(l)}[n] e^{-j2\pi kn/N}, \tag{5}$$

while the prevailing clipping noise can be extracted as

$$C^{(l)}[k] = \bar{X}^{(l)}[k] - X^{(0)}[k]. \tag{6}$$

Then the ICWEF mask to be used for out-of-band emission and frequency-selective clipping noise filtering, is computed for the ith considered modulation or MCS as [9]

$$H^{(l)}_{\text{ICWEF},i}[k] = \begin{cases} \frac{E_i}{|C^{(l)}[k]|}, & \text{if } k \in \mathcal{K}_{\mathcal{M}_i} \wedge E_i < |C^{(l)}[k]|, \\ 1, & \text{if } k \in \mathcal{K}_{\mathcal{M}_i} \wedge E_i \geq |C^{(l)}[k]|, \\ 0, & \text{if } k \notin \mathcal{K}_{\mathcal{M}_i}, \end{cases} \tag{7}$$

where $\mathcal{K}_{\mathcal{M}_i}$ denotes the index set that includes the subcarriers that are modulated with ith modulation of the modulation set $\mathcal{K}_{\mathcal{M}}$. Moreover, E_i is the level that the clipping noise sample $C^{(l)}[k]$ is filtered to, stemming from the EVM threshold of the ith considered modulation, viz. $E_i = \frac{\text{EVM}_i\,[\%]}{100\%}$.

However, in the ICWEF method, the set $\mathcal{K}_{\text{act}}$ does not only include $\mathcal{K}_{\mathcal{M}_i}$, but also the clipping noise-free subcarriers. This is expressed as

$$\mathcal{K}_{\text{act}} = \mathcal{K}_{\mathcal{F}} \cup \bigcup_{i=1}^{\text{card}(\mathcal{K}_{\mathcal{M}})} \mathcal{K}_{\mathcal{M}_i}, \tag{8}$$

where $\mathcal{K}_{\mathcal{F}}$ contains the clipping noise-free subcarriers. The full ICWEF mask that is applied in lth iteration reads then

$$H^{(l)}_{\text{ICWEF}}[k] = \sum_{i=1}^{\text{card}(\mathcal{K}_{\mathcal{M}})} H^{(l)}_{\text{ICWEF},i}[k], \tag{9}$$

where $\text{card}(\cdot)$ is the cardinality of the argument set. Next, the ICWEF filtering is applied and the filtered clipping noise is added to original frequency-domain data signal, which reads

$$X^{(l)}[k] = X^{(0)}[k] + H^{(l)}_{\text{ICWEF}}[k] C^{(l)}[k]. \tag{10}$$

This is then followed by IDFT to obtain $x^{(l)}[n]$. When the PAPR target is reached or L iterations are completed, OFDM waveform processing is finalized with CP addition and weighted overlap-and-add (WOLA) operations, and $x^{(L)}[n]$ is obtained. Algorithm 1 in [9] summarizes the operations in the considered baseline ICWEF method.

Importantly, in (7), the ICWEF mask controls clipping noise level inside the passband and the optimal PAPR target, $\lambda_{\text{target,opt}}$, should be used in (3) and (4) to get the best PAPR performance under these limitations. Hence, optimally tuning λ_{target} for any given ICWEF mask is the main goal of this study.

B. Clipping Noise Power-Based PAPR Target Selection

As the first but secondary approach, we consider an analytical approximation to the PAPR target selection problem and use this scheme as a reference for the ML-based PAPRer. The derivations build on [19] for ordinary ICF while are here extended to the frequency-selective ICWEF method.

Following the Bussgang theorem [20], by targeting $\lambda_{\text{target}} \in \mathcal{K}_T$ (the set of the utilized PAPR targets), the time-domain signal obtained after first clipping operation becomes

$$\bar{x}^{(1)}[n] = \alpha x^{(0)}[n] + d^{(1)}[n], \tag{11}$$

where $d^{(1)}[n]$ is the nth sample of the uncorrelated clipping noise and α is the attenuation factor, which is defined as

$$\alpha = 1 - e^{-\lambda_{\text{target}}} + \frac{\sqrt{\pi \lambda_{\text{target}}}}{2} \text{erfc}\left(\sqrt{\lambda_{\text{target}}}\right). \tag{12}$$

Similar to [19], the time-domain clipping noise after first clipping operation can then be expressed using (11) as

$$c^{(1)}[n] = \bar{x}^{(1)}[n] - x^{(0)}[n] = (\alpha - 1)x^{(0)}[n] + d^{(1)}[n]. \quad (13)$$

The corresponding correlation function reads

$$R_{c^{(1)}}[u] = (\alpha - 1)^2 R_{\mathbf{x}^{(0)}}[u] + R_{d^{(1)}}[u], \quad (14)$$

where $R_{\mathbf{x}^{(0)}}[u]$ is the uth sample of the autocorrelation of the original OFDM signal and $R_{d^{(1)}}[u]$ is the uth sample of the correlation function of the uncorrelated clipping noise. Then, the power spectral density (PSD) of the clipping noise arising after the first clipping operation reads

$$\rho_{c^{(1)}}[k] = \frac{1}{\sqrt{N}} \sum_{u=0}^{N-1} R_{c^{(1)}}[u] e^{-j2\pi ku/N}. \quad (15)$$

Then, to express the clipping noise power after L iterations, the noise enhancement factor can be defined as [19]

$$\beta^{(L)}[k] = \frac{E\left[\tilde{C}^{(L)}[k]\left(\tilde{C}^{(1)}[k]\right)^*\right]}{E\left[\tilde{C}^{(1)}[k]\left(\tilde{C}^{(1)}[k]\right)^*\right]}, \quad (16)$$

where now $\tilde{C}^{(l)}[k] = H^{(l)}_{\text{ICWEF}}[k]C^{(l)}[k]$. Here, to configure the mask, no modulation limits are assumed for $H^{(l)}_{\text{ICWEF}}[k]$. The total clipping noise power obtained after L iterations reads then

$$P_{c^{(L)}}(\lambda_{\text{target}}) = \sum_{k=-N_{\text{act}}/2}^{N_{\text{act}}/2-1} \beta^{(L)}[k]\rho_{c^{(1)}}[k]. \quad (17)$$

The maximum noise power that a given ICWEF mask can support reads now

$$P_{\text{ICWEF}} = \sum_{i=1}^{\text{card}(\mathcal{K}_{\mathcal{M}})} E_i^2 \text{card}\left(\mathcal{K}_{\mathcal{M}_i}\right). \quad (18)$$

Finally, the optimal PAPR target can be computed by comparing (18) to the analytically obtained total clipping noise power levels in (17), expressed as

$$\hat{\lambda}_{\text{target,opt}} = \arg \min_{\lambda_{\text{target}} \in \mathcal{K}_T} |P_{c^{(L)}}(\lambda_{\text{target}}) - P_{\text{ICWEF}}|. \quad (19)$$

C. PAPRer: Proposed ML-Based PAPR Target Tuner

As established in the previous section, the maximum clipping noise power that an ICWEF mask can support provides information about the optimal PAPR target. The PAPRer utilizes a supervised learning approach, and stemming from (18), it builds on the modulation set $\mathcal{K}_{\mathcal{M}}$, the sizes of the sets $\mathcal{K}_{\mathcal{M}_i}$, and the set of used PAPR targets $\mathcal{K}_T$ as the features. Hence, the PAPRer can be mathematically expressed as

$$\hat{\lambda}_{\text{target,opt}} = F\left(\mathcal{K}_{\mathcal{M}_1}, \mathcal{K}_{\mathcal{M}_2}, \ldots, \mathcal{K}_{\mathcal{M}_{\text{card}(\mathcal{K}_{\mathcal{M}})}}, \mathcal{K}_{\mathcal{M}}, \mathcal{K}_T\right). \quad (20)$$

The ML algorithm that is used with PAPRer finds the nonlinear relation $F(\cdot)$ between the features and the actual optimal PAPR target, while then obtains $\hat{\lambda}_{\text{target,opt}}$ using the detected pattern. In our numerical 5G NR evaluations, four modulations are considered with the PAPRer, i.e., $\text{card}(\mathcal{K}_{\mathcal{M}}) = 4$. However, the proposed PAPRer supports conceptually any set of modulations/MSE limits, while $\text{card}(\mathcal{K}_{\mathcal{M}}) = 4$ is just an example configuration for the numerical evaluations.The overall processing can be divided into the offline and online phases, which can be summarized as follows.

TABLE I
FEATURES AND OUTPUT VARIABLE USED IN MODEL TRAINING

Features			Output
$\mathcal{K}_{\mathcal{M}_1}, \mathcal{K}_{\mathcal{M}_2}, \ldots, \mathcal{K}_{\mathcal{M}_{\text{card}(\mathcal{K}_{\mathcal{M}})}}$	$\mathcal{K}_{\mathcal{M}}$	$\mathcal{K}_T$	PAPR (CCDF@10^{-4})

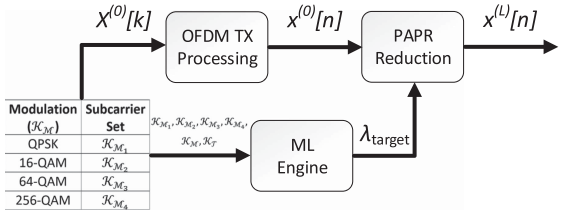

Fig. 1. Conceptual illustration of the PAPRer online processing for auto-tuning of the PAPR target. The PAPRer is trained offline.

Offline/Training Phase: For creating the learning dataset, the $\mathcal{K}_T$ is set first as $\mathcal{K}_T = \{3\text{ dB}, 4\text{ dB}, \ldots, 8\text{ dB}\}$ and a high number of random ICWEF masks is generated as in (9) by randomly varying $\mathcal{K}_{\mathcal{M}_i}$. Then, L iterations of the ICWEF algorithm are run for each mask and the PAPR target, and the obtained PAPR values corresponding to complementary cumulative distribution function (CCDF) level of 10^{-4}, are collected. The training of the PAPRer builds on this dataset, illustrated conceptually in Table I.

Online Phase: The online processing is illustrated conceptually in Fig. 1. Based on the MCS configuration, the features are given to the trained PAPRer. Then, PAPRer decides the $\hat{\lambda}_{\text{target,opt}}$, and PAPR reduction is performed accordingly.

Next, the considered ML algorithms and the associated loss function will be detailed. We note that conceptually, the PAPRer supports not only deep neural networks (DNNs), but also decision tree-based ensemble learning algorithms.

1) XGBoost Regressor: As a decision tree-based ensemble learning method, the eXtreme Gradient Boosting (XGBoost) is an effective algorithm offering high algorithm speed and high prediction accuracy [21]. In addition, it also provides an analysis on the feature importance and in this study, it is also used to determine the important data features. In the evaluations, two different parametrizations are considered for XGBoost. These are named as XGB#1 and XGB#2, and are configured with 50 estimators/maximum depth of 10 and, 80 estimators/maximum depth of 20, respectively. Moreover, in both regression models, learning rate of 0.1 is used.

2) Deep Neural Networks (DNNs): In this study, in addition to input and output layers, multiple hidden layers are used in the considered DNN architecture. Accordingly, two different parametrizations are considered also for DNN algorithm. The first case, DNN#1, contains three hidden layers and one output layer, with the number of nodes in each layer being 10, 6, 4, and 1, respectively. Similarly, the second DNN case, DNN#2, is implemented with four hidden layers and one output layer, with number of nodes in each layer being 40, 20, 12, 8, and 1, respectively. The DNN solutions are implemented using the Keras library, running on top of TensorFlow framework. In the implementation, the network weights are initialized using the normal distribution and rectified linear activation function (ReLU) is used as the activation function. In addition, stochastic gradient descent method is used as the optimizer. In training, 150 iterations are run and batch size of 10 is utilized.

It is further noted that the hyperparameters for both XGBoost and DNN are determined based on the prediction performance. Utilizing low number of hyperparameters also brings some reasonable prediction

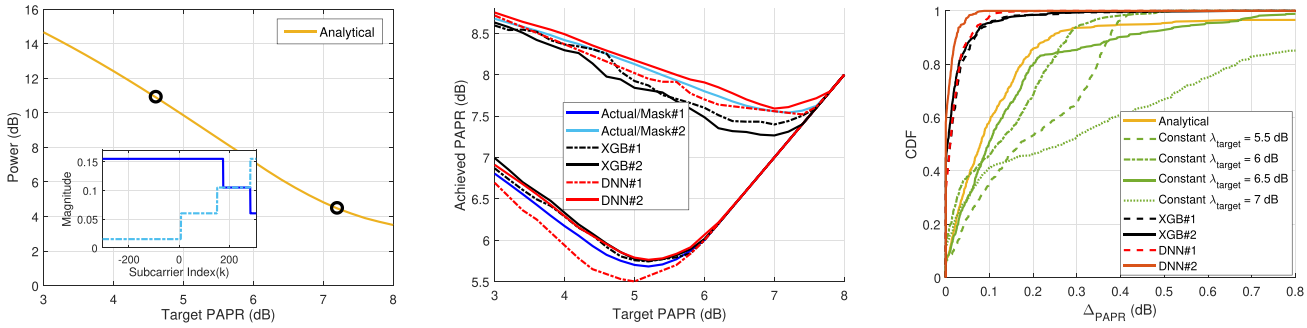

Fig. 2. Results for 20 MHz 5G NR channel with 30 kHz SCS and 51 PRBs. Two example ICWEF masks and the total clipping noise powers obtained after $L = 20$ iterations of ICWEF are shown in (a). Black circles represent the PAPR targets that are selected by the analytical reference method. In (b), the actual and predicted PAPR performance of the ML methods are shown at CCDF level of 10^{-4}, for PAPR targets varying from 3 dB to 8 dB. In (c), the results for Δ_{PAPR} metric are shown in terms of CDF for the analytical reference method, the ML-based methods and the four example constant λ_{target} cases.

performance, but we target to achieve a good PAPR target prediction accuracy and thus configure a rather high number of hyperparameters in this work.

3) Loss Function: As the loss function, we utilize the well-known modeling MSE, which for our problem scenario reads

$$\mathcal{L}\left(\lambda_{\text{target,opt}}, \hat{\lambda}_{\text{target,opt}}\right) = \frac{1}{N_{\text{TS}}} \sum_{i=1}^{N_{\text{TS}}} \left(\lambda_{\text{target,opt}}(i) - \hat{\lambda}_{\text{target,opt}}(i)\right)^2, \quad (21)$$

where $\lambda_{\text{target,opt}}(i)$ and $\hat{\lambda}_{\text{target,opt}}(i)$ denote the actual and the predicted optimal PAPR targets for the ith training sample, and N_{TS} is the number of training samples. In addition, the coefficient of determination (i.e., R^2 score) is also utilized to quantify the performance of the ML algorithms, expressed as

$$\text{R}^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}, \quad (22)$$

where $\text{SS}_{\text{res}} = \sum_i (\lambda_{\text{target,opt}}(i) - \hat{\lambda}_{\text{target,opt}}(i))^2$ and $\text{SS}_{\text{tot}} = \sum_i (\lambda_{\text{target,opt}}(i) - \overline{\lambda}_{\text{target}})^2$ are the residual sum of squares and the total sum of squares while $\overline{\lambda}_{\text{target}} = \frac{1}{N_{\text{TS}}} \sum_{i=1}^{N_{\text{TS}}} \lambda_{\text{target,opt}}(i)$.

III. Numerical Results and Analysis

The PAPRer is next evaluated in terms of PAPR, MSE and R^2 score. For reference, also the analytical method from Section II-B building on clipping noise power calculations is considered. Two different radio interface numerologies, namely a 10 MHz 5G NR channel with 60 kHz subcarrier spacing (SCS) and 11 PRBs, and a 20 MHz 5G NR channel with 30 kHz SCS and 51 PRBs, are considered based on the 5G NR specifications [22]. The CCDF is used to quantify the PAPR performance. Specifically, the CCDF probability of 10^{-4} and PAPR targets from 3 dB to 8 dB are considered. With ML-based approaches, a training dataset that contains features and the actual PAPR targets is created by running simulations for 1000 random ICWEF masks and the trained models are validated with another 500 random ICWEF masks.

A. PAPR Performance

The proposed solutions are quantified in terms of Δ_{target}, which is defined as the difference between the actual optimal PAPR target, $\lambda_{\text{target,opt}}$, and the predicted PAPR target, $\hat{\lambda}_{\text{target,opt}}$. When calculating $\lambda_{\text{target,opt}}$, PAPR values corresponding to CCDF probability level of 10^{-4} are considered as noted above. Furthermore, Δ_{PAPR} that is the difference between the PAPR values corresponding to CCDF level of 10^{-4} that are obtained with $\lambda_{\text{target,opt}}$ and $\hat{\lambda}_{\text{target,opt}}$, respectively, are also quantified for the 1000 different ICWEF masks.

TABLE II
Mutual Comparison of the Considered Solutions in Terms of Δ_{target}, Where Both the CCDF Probability Level of 10^{-4} and the 95% CI are Shown

Case	R^2		MSE		Δ_{target} (dB)	
	Train	Test	Train	Test	@10^{-4}	CI (95%)
Analytical	N.A.	N.A.	N.A.	N.A.	0.44	0.42–0.46
$\lambda_{\text{target}} = 6$ dB	N.A.	N.A.	N.A.	N.A.	0.63	0.58–0.67
$\lambda_{\text{target}} = 6.5$ dB	N.A.	N.A.	N.A.	N.A.	0.58	0.55–0.62
XGB#1	99%	93%	-23 dB	-14 dB	0.19	0.17–0.21
XGB#2	99%	95%	-23 dB	-15 dB	0.17	0.15–0.19
DNN#1	96%	95%	-16 dB	-15 dB	0.17	0.16–0.18
DNN#2	99%	98%	-23 dB	-19 dB	0.10	0.09–0.12

20 MHz 5G NR channel case with 30 kHz SCS and 51 PRBs.

The 20 MHz 5G NR channel is used as the main evaluation case. First, to evaluate and illustrate the analytical reference method, an initial evaluation is conducted for two distinctive ICWEF masks which are shown in Fig. 2(a). Here, $L = 20$ iterations are considered with both ICWEF and ICF. In Fig. 2(a), clipping noise power levels with respect to several PAPR targets are given for ICF, which is analytically obtained as in (17) and shown with the yellow line. Here, maximum noise power levels that can be supported by the ICWEF masks are 11 dB and 4.2 dB, respectively. These masks cannot reach particular PAPR targets because they cannot support the required noise power levels. The PAPR targets that are selected by the analytical reference method are shown with black circles, the numerical values being 4.6 dB and 7.4 dB for the first and second masks, respectively. However, the actual optimal PAPR targets for these two masks are 5.2 dB and 7.2 dB, respectively, and thus the average prediction error for these two cases is 0.4 dB, which can be considered as a relatively high error level.

Next, the ML algorithms are evaluated by considering the training and validation phases, and compared against the analytical reference method as well as selected constant λ_{target} cases. The results are presented in Table II. Here in the last two columns, Δ_{target} results are given for both the 10^{-4} CCDF level and the 95% confidence interval (CI). As can be seen, the prediction performance with the reference method and the two different constant λ_{target} cases is limited, resulting in high errors of 0.44, 0.63 and 0.58 dB, respectively. In contrast, the XGBoost ML approach provides good improvement over the reference method as the average values of Δ_{target} are equal to 0.19 dB and 0.17 dB, respectively. Here, both XGB#1 and XGB#2 perform similarly. Considering the

TABLE III
MUTUAL COMPARISON OF THE CONSIDERED SOLUTIONS IN TERMS OF Δ_{TARGET}, WHERE BOTH THE CCDF PROBABILITY LEVEL OF 10^{-4} AND THE 95% CI ARE SHOWN

Case	R^2		MSE		Δ_{target} (dB)	
	Train	Test	Train	Test	@10^{-4}	CI (95%)
Analytical	N.A.	N.A.	N.A.	N.A.	0.80	0.75–0.85
$\lambda_{\text{target}} = 6$ dB	N.A.	N.A.	N.A.	N.A.	1.01	0.95–1.06
$\lambda_{\text{target}} = 6.5$ dB	N.A.	N.A.	N.A.	N.A.	0.73	0.68–0.77
XGB#1	99%	93%	-23 dB	-14 dB	0.35	0.31–0.38
XGB#2	99%	95%	-23 dB	-15 dB	0.26	0.24–0.28
DNN#1	96%	95%	-16 dB	-15 dB	0.27	0.25–0.29
DNN#2	99%	98%	-23 dB	-19 dB	0.18	0.16–0.20

10 MHz 5G NR channel case with 11 PRBs and 60 kHz SCS.

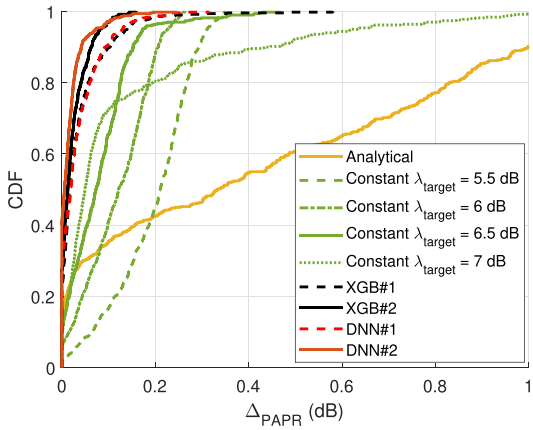

Fig. 3. The results in terms of the Δ_{PAPR} metric for the analytical reference method, the ML-based methods and the four example constant λ_{target} cases. 10 MHz 5G NR channel with 11 PRBs and 60 kHz SCS.

complexity advantages of XGB#1 (cf. Section III-B), it can be considered as the preferred choice. The DNN methods also provide huge performance gains over the reference method, with the average Δ_{target} values being 0.17 dB and 0.10 dB, respectively. Here, DNN#2 has significantly better prediction performance than the other methods. It also outperforms the analytical solution dramatically, the performance gap being around 0.35 dB.

Next, the prediction performance is evaluated for the two masks shown in Fig. 2(a), and the obtained results are shown in Fig. 2(b). The PAPR result curves shown in this figure indicate the importance of utilizing optimal PAPR target. Here, $\lambda_{\text{target,opt}}$ for the first mask is 5.2 dB which leads to minimum realized PAPR value, and it is predicted correctly by DNN#2 and both XGB solutions. With DNN#1 and the analytical solution, $\Delta_{\text{target}} = 0.2$ dB and $\Delta_{\text{target}} = 0.6$ dB, respectively. For the second mask, $\lambda_{\text{target,opt}}$ is equal to 7.2 dB and all methods provide $\Delta_{\text{target}} = 0.2$ dB. Hence, the ML methods are robust and provide high performance in both cases.

Next, Δ_{PAPR} is assessed with respect to cumulative distribution function (CDF) for the considered ML methods, the reference method and the constant λ_{target} cases. While PAPR targets from 3 dB to 8 dB were considered in other evaluations, we focus here on four different constant PAPR targets that provide the best Δ_{PAPR} performance. The evaluation of Δ_{PAPR} is essential as the achieved PAPR performance is of highest interest while also evidencing the importance of robust PAPR target estimation. The results are presented in Fig. 2(c), where the distributions of Δ_{PAPR} are provided for the 1000 random ICWEF masks. As can be observed, DNN solutions perform very well, with Δ_{PAPR} being mostly around 0–0.05 dB. The XGB methods facilitate also very good performance up to CDF levels of around 0.9–0.95 with Δ_{PAPR} being less than 0.1 dB. However, at very low probability levels in the CDF tail, they show lower reliability with Δ_{PAPR} values ranging up to 0.5–0.7 dB. On the other hand, the reference method and the constant λ_{target} cases are clearly inferior to the ML-based approaches.

To further assess the performance of the methods, the evaluations for Δ_{target} and Δ_{PAPR} are next carried out also for the 10 MHz 5G NR channel case, with the obtained results being shown in Table III. Again, the performance is limited for the analytical reference method and the two constant λ_{target} cases, the average prediction error being 0.80, 1.01 and 0.73 dB, respectively. Furthermore, the XGBoost algorithm again outperforms the reference cases significantly with the average values of Δ_{target} being equal to 0.35 dB and 0.26 dB, for XGB#1 and XGB#2, respectively. In this case, XGB#2 performs clearly better than XGB#1. Additionally, the DNN methods provide again clear performance advantages, with the DNN#2 being able to provide an average prediction error of 0.18 dB, while the performance gap with respect to the analytical reference solution is already around 0.60 dB. Thus, based on the results for both the 10 MHz and 20 MHz 5G NR channel cases, the DNN#2 can be seen as the preferred choice to obtain near-optimal PAPR performance and thereon to maximize the transmission power at the power amplifier (PA) output. Considering its relatively low computational complexity (cf. Section III-B), it is also feasible for practical implementations with great complexity benefits compared to brute force search based optimal PAPR target selection.

Finally, the Δ_{PAPR} metric is quantified for the 10 MHz 5G NR channel case by considering the setup used for generating the results in Fig. 2(c). The obtained results are presented in Fig. 3. In this case, the DNN and XGB methods provide similar performance results, and they again significantly outperform the analytical reference cases with the Δ_{PAPR} being less than 0.1 dB up to CDF levels of around 0.8–0.90. The reference method and the constant λ_{target} cases perform clearly worse, with Δ_{PAPR} being usually higher than 0.1 dB. Specifically, the Δ_{PAPR} can reach up to 1 dB for the analytical solution.

B. Complexity

Complexity is an essential figure of merit for any processing solution. As the PAPR reduction parameters are static, we focus here on the complexity of the ML model online execution and omit the complexity of the training phase. The XGBoost approach is computationally efficient as it only uses conditional operators to find the optimal branches for the given input. However, the cost of these is noticeable in the corresponding time complexity which is thus next addressed. The time complexity of the XGBoost algorithm can be denoted as $O(N_{\text{est}} N_{\text{d}})$, where N_{est} and N_{d} are the number of trees and the maximum depth of the trees [21]. The DNN solution, in turn, is a feedforward neural network (NN) with multiple hidden layers. Hence, the complexity is equal to $O(N_{\text{feat}} N_{\text{node},1} + \sum_i^{N_{\text{lay}}-1} N_{\text{node},i} N_{\text{node},i+1})$ with N_{feat}, $N_{\text{node},i}$ and N_{lay} denoting the number of features, number of nodes that layer i contains and the total number of layers, respectively. For reference, the complexity of the original CP-OFDM processing is equal to the complexity of one fast Fourier transform (FFT), i.e., $O(N \log_2 N)$.

For the considered modulations and assuming the 20 MHz channel case, $N_{\text{OFDM}} = 1024$ and with an oversampling factor of four, $N = 4096$. On the other hand, since $N_{\text{est}} = \{50, 80\}$ and $N_{\text{d}} = \{10, 20\}$, both XGB cases lead only to a small increase in complexity when compared to basic CP-OFDM processing. Similarly, in the two DNN cases, $N_{\text{feat}} = 9$ while $N_{\text{node}} = \{10, 6, 4, 1\}$ and $N_{\text{lay}} = 4$ or $N_{\text{node}} = \{40, 20, 12, 8, 1\}$ and $N_{\text{lay}} = 5$, thus the resulting complexity increase is again very minor. Concretely, when evaluating the execution time of XGB#2 and DNN#2 on a state-of-the-art central processing unit (CPU), the proposed methods result only in some 25% complexity

increase, with respect to the CP-OFDM waveform processing, highlighting the efficiency of the ML models.

IV. Conclusion

A novel solution was presented to solve the optimal PAPR target selection problem in frequency-selective PAPR reduction, focusing on the ICWEF as a concrete example. The proposed solution is an ML-based method where certain features related to ICWEF mask are exploited to accurately learn and predict the optimal PAPR target through supervised learning. Moreover, an alternative analytical clipping noise power based method was also described for reference. Specifically, this analytical method computes the maximum clipping noise power that a given ICWEF mask can support and then finds the applicable PAPR target. As presented through 5G NR standard compliant numerical evaluations, the analytical solution and the considered ML-based methods are able to predict the optical PAPR targets with accuracies in the order of $\simeq$ 0.4–0.5 dB and $\simeq$ 0–0.2 dB, respectively. Moreover, the ML models were shown to be computationally feasible and robust, making them attractive candidates for optimizing the PAPR performance in 5G NR and future 6G networks.

References

[1] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR : The Next Generation Wireless Access Technology*. Cambridge, MA, USA: Academic Press, 2018.

[2] 3GPP TS 38.300 V15.4.0, " New Radio (NR); Overall description; Stage-2," Tech. Spec. Group Radio Access Network, Rel. 15, Dec. 2018.

[3] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1567–1592, Oct.-Dec. 2013.

[4] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *Electron. Lett.*, vol. 38, no. 5, pp. 246–247, Feb. 2002.

[5] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *Electron. Lett.*, vol. 33, no. 5, pp. 368–369, Feb. 1997.

[6] R. W. Bauml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electron. Lett.*, vol. 32, no. 22, pp. 2056–2057, Oct. 1996.

[7] M. Hu, W. Wang, W. Cheng, and H. Zhang, "Initial probability adaptation enhanced cross-entropy-based tone injection scheme for PAPR reduction in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6674–6683, Jul. 2021.

[8] J. Hou, J. Ge, and F. Gong, "Tone reservation technique based on peak-windowing residual noise for PAPR reduction in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5373–5378, Nov. 2015.

[9] S. Gökceli et al., "Novel iterative clipping and error filtering methods for efficient PAPR reduction in 5G and beyond," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 48–66, 2021.

[10] S. Gökceli, T. Levanen, T. Riihonen, M. Renfors, and M. Valkama, "Frequency-selective PAPR reduction for OFDM," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6167–6171, Jun. 2019.

[11] T. Kageyama, O. Muta, and H. Gacanin, "Performance analysis of OFDM with peak cancellation under EVM and ACLR restrictions," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6230–6241, Jun. 2020.

[12] X. Wang, N. Jin, and J. Wei, "A model-driven DL algorithm for PAPR reduction in OFDM system," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2270–2274, Jul. 2021.

[13] Z. Li, N. Jin, X. Wang, and J. Wei, "Extreme learning machine-based tone reservation scheme for OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 30–33, Jan. 2021.

[14] M. Kim, W. Lee, and D.-H. Cho, "A novel PAPR reduction scheme for OFDM system based on deep learning," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 510–513, Mar. 2018.

[15] Z. Liu et al., "Low-complexity PAPR reduction method for OFDM systems based on real-valued neural networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1840–1844, Nov. 2020.

[16] M.-T. Nguyen, G. Kaddoum, B. Selim, K.V. Srinivas, and P. F. De Araujo-Filho, "Deep unfolding network for PAPR reduction in multi-carrier OFDM systems," *IEEE Commun. Lett.*, vol. 26, no. 11, pp. 2616–2620, Nov. 2022.

[17] A. Kalinov, R. Bychkov, A. Ivanov, A. Osinsky, and D. Yarotsky, "Machine learning-assisted PAPR reduction in massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 537–541, Mar. 2021.

[18] H. Ochiai and H. Imai, "Performance analysis of deliberately clipped OFDM signals," *IEEE Trans. Commun.*, vol. 50, pp. 89–101, Jan. 2002.

[19] K. Bae, J. G. Andrews, and E. J. Powers, "Quantifying an iterative clipping and filtering technique for reducing PAR in OFDM," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1558–1563, May 2010.

[20] H. E. Rowe, "Memoryless nonlinearities with gaussian inputs: Elementary results," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1519–1525, Sep. 1982.

[21] T. Chen and C. Guestrin, "XGBoost: A. scalable tree boosting system," in *Proc. Of KDD '16*, New York, NY, USA:ACM, 2016, pp. 785–794.

[22] 3GPP TS 38.104 V15.4.0, " New Radio (NR); Base station (BS) radio transmission and reception," Tech. Spec. Group Radio Access Network, Rel. 15, Dec. 2018.