Design of an M-Ary DLCSK Communication System Using Deep Transfer Learning

MAJID MOBINI[1], MARIJAN HERCEG[2], AND GEORGES KADDOUM[3,4] (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol 47148-71167, Iran

[2]Department of Communications, Faculty of Electrical Engineering, Computer Science and Information Technology, University of Osijek, 31000 Osijek, Croatia

[3]Department of Electrical Engineering, University of Québec, École de technologie supérieure, Montreal, QC H3C 1K3, Canada

[4]Cyber Security Systems and Applied AI Research Center, Lebanese American University, Beirut 13-5053, Lebanon

CORRESPONDING AUTHOR: M. MOBINI (e-mail: mobini2002@gmail.com)

ABSTRACT Conventional coherent chaos-based communication systems require synchronization of chaotic signals, which is still practically unattainable in a noisy environment. Moreover, in non-coherent schemes, a part of the bit duration is spent sending non-information-bearing reference samples, which deteriorates the Bit Error Rate performance (BER) of these systems. To tackle these problems, this paper designs an M-ary Deep Learning Chaos Shift Keying (M-ary DLCSK) system. The proposed receiver uses a Convolutional Neural Network (CNN)-based classifier that recovers M-ary modulated data. The trained NN model grasps different chaotic maps, estimates channels, and classifies the received signals effectively. Moreover, we consider a Transfer Learning (TL) framework that enhances the noise performance and classification results. Due to the generalization capabilities of TL, the trained NN can work in different Signal-to-Noise Ratio (SNR) conditions without the need for re-training. We compare the BER performance, complexity, and bandwidth efficiency of the M-ary DLCSK receiver with existing receivers. The results demonstrate that the M-ary DLCSK receiver is the first practical system that achieves the theoretical BER performance of the coherent CSK systems under Rayleigh fading channels. Moreover, the proposed system provides a considerable performance advantage compared to the existing DL-based receivers under Rayleigh fading channels. For example, the BER performance of 8-ary DLCSK shows a gain of 0.1 over the Long Short-Term Memory (LSTM)-aided DNN systems at the target $E_b/N_0 = 14dB$. These features make M-ary DLCSK an attractive candidate for several applications, such as Massive Multiple-Input Multiple Output (MIMO), Vehicle-to-everything (V2X), Quantum, and optical communication systems.

INDEX TERMS Chaos-based communications, computational complexity, constellation diagram, deep transfer learning.

I. INTRODUCTION

THE OBJECTIVE of any digital modulation technique is to obtain a reliable Bit Error Rate (BER) with minimum energy per bit. The second important objective of digital modulations is to achieve a suitable bandwidth efficiency (BE), which can be calculated by dividing the data rate by the channel bandwidth. Deep Learning Chaos Shift Keying (DLCSK) system [1], was shown to achieve a remarkable BER performance improvement compared to the conventional Differential Chaos Shift Keying (DCSK)

scheme. However, in the primary form of the DLCSK system (i.e., binary DLCSK), only one bit is delivered in each transmission. In this paper, in order to enhance the BE of the basic DLCSK system, an M-ary DLCSK scheme is designed based on the conventional coherent CSK modulation scheme. The proposed receiver employs a Convolutional Neural Network (CNN)-based classifier that recovers the M-ary modulated data. The trained Neural Network (NN) model can effectively classify the received signals from harsh wireless channels. Moreover, we provide an approach

based on the fundamentals of Deep Transfer Learning (DTL) to enhances the noise performance of the M-ary DLCSK system. The proposed DTL technique allows us to use the trained receiver in different noise conditions without the need for re-training.

In digital communication systems, data are transmitted by mapping data bits to symbols, and symbols to sample functions of analog waveforms. Conventional chaos-based modulation schemes are commonly categorized into two classes, i.e., coherent and non-coherent [2]. A symbol may be retrieved by coherent detection, where the basis function can be regenerated using chaotic synchronization. In a coherent system, all sample functions are known. Using non-coherent detection, sample functions are unknown at the receiver, and the reference signals should be transmitted for the basis function recovery. In a coherent detection scheme, such as Chaos Shift Keying (CSK) [3], [4], data is transmitted as a combination of basis functions chosen from chaotic waveforms. In coherent CSK, synchronization of the chaotic sequences is commonly required in order to enable basis functions recovery. A coherent CSK scheme with one basis function (i.e., antipodal CSK [3]) theoretically achieves the BER performance of a Binary Phase Shift Keying (BPSK) system under Additive White Gaussian Noise (AWGN) channels. However, owing to the cross-correlation between chaotic signals and the problem of basis function recovery, this performance cannot be achieved practically. With no available solution to the problem of basis function recovery, chaotic switching CSK modulation has been proposed using two basis functions [4]. The chaotic switching CSK can theoretically reach the BER performance of the Frequency Shift Keying (FSK) modulation scheme under AWGN channels [5]. This level of BER performance can be achieved only if the cross-correlation problem is solved and the basis functions can be recovered successfully. However, since complete synchronization is still complicated to attain in realistic environments, this theoretical performance is not practically attainable [6]. If it is not possible to recover the basis functions, the DCSK scheme [7] may provide better performance.

Non-coherent detection schemes, such as DCSK, rely on a reference transmission, where the "reference" and "information-bearing" signals are transmitted in successive time slots. In fact, DCSK is a variant of CSK, where the basis functions are transmitted over the channel (Instead of using synchronization for the basis function recovery), and the information is recovered from the correlation between the two aforementioned signals. With non-coherent DCSK schemes, the Channel State Information (CSI) and synchronization of chaotic sequences are not needed on the receiver side. However, the main weakness of DCSK systems stems from repetition of the chaotic signals in the time domain. This repetition leads to a degradation of the data rate and energy efficiency, as half of the bit time interval is used for delivering non-information-bearing samples [8].

Most of the existing chaos-based communication systems are based on the DCSK scheme and transmission of the reference signals. To improve the performances of DCSK systems, many research works have been performed to utilize the resources in the time, frequency, space, and code domain to enhance the data rate, remove the delay lines and further enhance the reliability. For example, in Multicarrier DCSK (MC-DCSK) systems [9], the reference signal is transmitted over a predefined subcarrier frequency, and multiple data-bearing signals sharing the same reference are transmitted over the remaining subcarriers. Since multiple data-bearing signals share the same reference signal, both spectrum and energy efficiencies are improved. However, transmission errors in the reference chaotic signal would induce error propagations when directly used for demodulations. To address this issue, an algorithm has been proposed in [10] to update the reference signals iteratively. In the iterative receiver, the correlation coefficients between the reference signal and the information-bearing signal are evaluated. The renewed reference chaotic signal is fed back and acts as the inputs in the next iteration. The reference signal can be updated until the iteration-stopping criterion is reached. The iterative receiver improves the Signal-to-Noise Ratio (SNR) of the reference signals to compensate for the destructive effects of the channels. However, iteration at the receiver induces additional complexity and latency. In [11], a low-rank approximation of matrices (LRAM) detection method is proposed. Instead of directly using the received reference chaotic signal for demodulation, the authors proposed to apply the LRAM method to jointly estimate the reference signal and the information-bearing signal. They used the singular value decomposition and generalized LRAM methods to minimize the distances between the estimates and transmitted data. Although the SNR can be improved by the methods presented in [10] and [11], these receivers have poor performances under practical conditions, such as Rayleigh fading channels.

In this paper, we would like to design a coherent CSK receiver, without the need for chaotic synchronization or reference transmission for basis function recovery. To this aim, we train the NN with the goal of the chaotic basis function recovery. Indeed, the NN learns the chaotic basis functions, adjusts its hyperparameters, and used them in order to detection of M-ary modulated data. By taking advantage of DTL, the M-ary DLCSK system can improve the performance of existing chaos-based modulation benchmarks.

In a communication channel, the received signal is usually distorted by the destructive effects of the channel. Channel estimation plays a critical role in eliminating the impacts of wireless channels. Several pilot-aided and pilotless channel estimation methods have been proposed using DL [12]. The pilot-aided channel estimation achieves high estimation accuracy. However, the transmitted pilot sequence occupies the valuable time and frequency resources. In this paper, we adopt a pilot-less approach to implicitly extract the channel features and recover data symbols. In the proposed design,

the received signals act as input to the NN model. The received signals are filtered by the channel and carry the channel features. Thus, the proposed receiver can learn the channel features from the received signals, and needs no additional pilot for channel estimation. The above-mentioned design not only allows the NN to grasp the dynamics of various chaotic maps but also enables it to learn the channel features at the same time.

In order to design a lightweight receiver, a BER-versus-complexity study is also performed. The BER performance of the CNN model is compared to existing NN models, such as Bidirectional Long Short-Term Memories (BiLSTMs) [39] under the same computational complexity conditions. The presented complexity analysis enables us to study the dependency of BER on the computational complexity and to find the best NN model for a certain complexity level. Furthermore, simulation results are compared with a theoretical lower bound to provide insightful conclusions.

A. RELATED WORKS AND MOTIVATIONS

One well-known challenge of coherent chaos-based modulations is the synchronization of the chaotic sequences in harsh environments, where many researches have been done to achieve a robust synchronization. The robustness of Deep Learning (DL)-based synchronization systems was evaluated in [6], where a chaos synchronization system has been designed based on the trainable NN model. The results demonstrated that DL reduces the synchronization error compared to traditional synchronization systems. However, an error remains between the recovered signal and initial signal, due to stochastic phenomena or harsh wireless channels. This synchronization error implies that future works should test other ways to recover the chaotic basis functions.

Two recent papers have united chaos-based communications with DL methods [14], [15]. In [14], a smart Orthogonal Frequency Division Multiplexing (OFDM) DCSK demodulator using a Long Short-Term Memory-aided Deep Neural Network (LSTM-aided DNN) is suggested. By exploiting the optimization and classification capabilities of NNs, the LSTM-aided DNN receiver attains improved BER performances. Similarly, in [15], the authors used the LSTM-aided DNN architecture to provide a demodulator for Multilevel Code-shifted Differential Chaos Shift Keying (MCS-MDCSK) system. However, both [14] and [15] are based on conventional DCSK modulation and reference signal transmission. The main drawback of DCSK-based modulations is that the reference transmission increases the overhead, errors, and complexity of the system.

In our recent paper [1], to overcome the reference transmission problems, we proposed the DLCSK system that employs DTL for the basis function recovery. Specifically, the main difference between the DLCSK and LSTM-aided DNN systems [15] is that, in a DLCSK system, the transmitter does not transmit the reference signals, whereas an LSTM-aided DNN uses the reference signals for basis functions recovery. The first version of the DLCSK system (i.e.,

binary DLCSK) only uses two chaotic maps and delivers one bit in each transmission. Moreover, the cross-correlation between the chaotic maps and the complexity aspects are not considered in [1]. In this paper, our focus is on the design of an M-ary DLCSK system using CNNs that can transmit more than one bit in each transmission. The proposed design enhances BE while limiting computational complexity. The mentioned features make M-ary DLCSK an attractive candidate for several applications, such as Quantum classifiers [16], optical communications [17], secure data transmission [18], vehicular communications [19], [20], Internet of Things (IoT) [21], underwater communications [22], and massive Multiple-Input Multiple Output (MIMO) systems [23].

B. INNOVATIVE ASPECTS

Innovative aspects of this paper are outlined below:

- A new M-ary DTL-based receiver is designed that allows us to increase the BE and BER performance of conventional M-ary chaos-based communication schemes. Different from the existing LSTM-aided DNN receivers [14], [15], which use DCSK-based schemes, our receiver is designed based on the basics of the coherent CSK schemes. Thus, the proposed M-ary DLCSK system does not transmit the reference signals, which reduces the overhead, errors, and complexity of the system.

- The proposed design employs TL technique to enhance its generalization capabilities. With a slightly increased complexity, the trained receiver shows an outstanding robustness against noise. Moreover, we present general expressions for the computational complexity of the designed NN-based receivers. This enables us to study the dependency of the BER on the complexity and to find a light-weight NN-based receiver.

- A geometrical approach based on the Bloch sphere is considered to find *maximally orthogonal* points (or quasi-orthogonal signals) on the constellation. In particular, in order to reduce the negative effects of the cross-correlation problem, we introduce a map selection algorithm to provide a set of quasi-orthogonal chaotic maps.

- We derive a lower bound for the BER of M-ary DLCSK systems based on the presented constellation. The difference between this bound and simulated results leads to insightful conclusions.

The rest of this paper is organized as follows: In Section II, the characteristics of existing chaos-based systems are described with an emphasis on CSK. In Section III, the general architecture of the proposed M-ary DLCSK system, the role of DTL, dataset generation, channel estimation, and map selection are presented. We also introduce a theoretical lower bound for the BER of M-ary DLCSK systems. In Section IV, we analyze the data rate and BE of the proposed M-ary DLCSK scheme and provide a comparison with other benchmarks. We further describe how to determine the complexity of the NN models in this Section. In

Section V, simulation results and discussions are presented, while Section VI draws the concluding remarks.

II. CSK SYSTEM AND WEAKNESS POINTS

The suggested transmitter is based on the conventional coherent CSK transmitter, presented in [5]. In this section, the structure of an M ary CSK transmitter, as well as existing correlation receivers, are discussed. Several derivations of the conventional CSK modulation scheme will be used in this paper as a comparative to illustrate the performance improvements obtained from the contributions of this paper.

A. M-ARY CSK TRANSMITTER

Consider the conventional M-ary CSK communication system with two (or more) basis functions, where the basis functions can be derived from the chaotic waveforms [5]. In an M-ary CSK modulation scheme, we use continuous chaotic signals $g_j(t)$, $j \in \{1, \ldots, M\}$, with a bit duration T_b. Assume that $T_b = \beta T_c$, where T_c is the time between each chaotic sample (chip), and β is the spreading factor. Hence, the continuous-time spreading signal $g_j(t)$ is given by

$$g_j(t) = \sum_{k=1}^{\beta} g_{j,k} h(t - kT_c), \quad (1)$$

where $g_{j,k}$, $j \in \{1, \ldots, M\}$ denotes k^{th} sample of the discrete-time chaotic sequences, presented in Table 1. Moreover, $h(t)$ is the square-root-raised-cosine filter, which is commonly used as the pulse-shaping filter in communication systems.

Let $s_i(t)$, $i \in \{1, 2, \ldots, M\}$ stand for the *signal set*, which is a linear combination of chaotic basis functions $g_j(t)$, $j \in \{1, 2, \ldots, M\}$. The following equation indicates the modulation process, i.e., mapping of symbols to analog waveforms

$$s_i^{(z)}(t) = \sum_{j=1}^{M} w_{i,j}^{(z)} g_j^{(z)}(t), \begin{cases} 0 < t < T_b, \\ i, j \in 1, 2, \ldots, M, \\ z = 1, 2, \ldots, Z, \end{cases} \quad (2)$$

where $s_i^{(z)}(t)$ is the zth modulated signal being transmitted, Z is the total number of data symbols, $g_j(t)$, $j \in \{1, 2, \ldots, M\}$ are chaotic basis functions, and the weights $w_{i,j}^{(z)}$ are the elements of the *signal vector*. In other words, based on the current data symbol $i^{(z)}$, some weights $w_{i,j}^{(z)}$ are multiplied by continuous chaotic signals $g_j(t)$, $j \in \{1, 2, \ldots, M\}$. The vector representation of the modulated signals with energy per symbol $\sqrt{E}$ can be written as

$$s_1 = \left(\sqrt{E}, 0, \ldots, 0 \right)$$
$$s_2 = \left(0, \sqrt{E}, \ldots, 0 \right)$$
$$\vdots = \vdots$$
$$s_M = \left(0, 0, \ldots, \sqrt{E} \right). \quad (3)$$

This paper designs an M-ary DLCSK receiver based on the constellation diagram of the classical M-ary CSK communication system. The signal-space diagram for 2-ary DLCSK
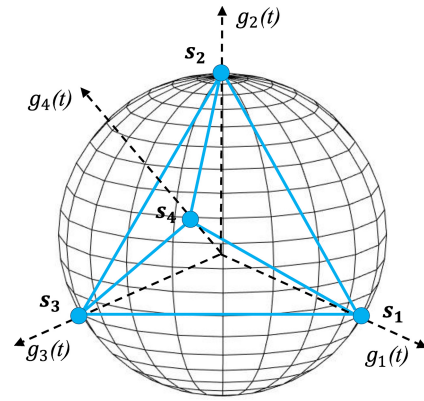

FIGURE 1. Signal-space diagram of 2-ary DLCSK (solid points on $g_1(t)$ and $g_2(t)$), and its extension to 4-ary DLCSK (solid points on $g_3(t)$ and $g_4(t)$).

(solid points on $g_1(t)$ and $g_2(t)$), and its extension to 4-ary DLCSK using solid points on $g_3(t)$ and $g_4(t)$ are shown in Figure 1. Unlike the LSTM-aided DNN [15], in which the user data bits are transmitted using MCS-MDCSK modulation [44] (which is a DCSK-based scheme), the M-ary DLCSK receiver is based on the coherent CSK modulation scheme [5]. In such a system, a necessary condition for maximum noise performance is that the chaotic sample functions should have constant energy. Therefore, in the remainder of this paper, we normalize the chaotic basis functions such that $\int_0^{T_b} g_j(t)dt = 0$ and $\int_0^{T_b} g_j^2(t)dt = 1$.

B. EXISTING CORRELATION RECEIVERS

Existing M-ary CSK receivers should be constructed with at least M correlators. For example, considering a chaotic switching CSK system with $M = 2$, where the signal set is $s_i^{(z)}(t) = w_{i,1}^{(z)} g_1^{(z)}(t) + w_{i,2}^{(z)} g_2^{(z)}(t)$, the message is detected by forming the decision variables $D_{i,1}^{(z)}$ and $D_{i,2}^{(z)}$, i.e., correlating the received signal with two regenerated chaotic basis functions $g_1'(t)$ and $g_2'(t)$. The correlation receiver estimates the elements $w_{i,j}$ of the signal vector. In the noise-free case, with perfect regeneration of the chaotic basis functions, we have $g_i'(t) = g_i(t)$. With a constant E, and orthonormal basis functions in the interval $[0, T_b]$, the outputs of two correlators are given by

$$D_{1,1}^{(z)} = w_{1,1}^{(z)} \int_0^{T_b} g_1^{2(z)}(t) \, dt = \sqrt{E},$$
$$D_{1,2}^{(z)} = w_{1,2}^{(z)} \int_0^{T_b} g_1^{(z)}(t) g_2^{(z)}(t) \, dt = 0, \quad (4)$$

when symbol "1" is transmitted, and

$$D_{2,1}^{(z)} = w_{2,1}^{(z)} \int_0^{T_b} g_1^{(z)}(t) g_2^{(z)}(t) \, dt = 0,$$
$$D_{2,2}^{(z)} = w_{2,2}^{(z)} \int_0^{T_b} g_2^{2(z)}(t) \, dt = \sqrt{E}, \quad (5)$$

when symbol "2" is transmitted. In particular, if $D_{i,1}^{(z)} > D_{i,2}^{(z)}$, the decision is "1," and if $D_{i,1}^{(z)} < D_{i,2}^{(z)}$, the decision is "2."

Consider the output of the coherent correlation receiver under the AWGN channel. In this case, the chaotic basis functions $g'_j(t)$, $j = 1, \ldots, M$ are regenerated from the noisy received signal $s_i^{(z)}(t) + \boldsymbol{\varepsilon}^{(z)}(t)$, where $\boldsymbol{\varepsilon}^{(z)}(t)$ is assumed to be independent AWGN. If we assume that the symbol "1" has been sent, the decision variable $D_{i,1}^{(z)}$, at the output of the first correlator, can be written as

$$
\begin{aligned}
D_{i,1}^{(z)} &= \int_{T_{sync}}^{T_b} \left[\boldsymbol{s}_i^{(z)}(t) + \boldsymbol{\varepsilon}^{(z)}(t) \right] \boldsymbol{g'}_1^{(z)}(t)\, dt \\
&= \int_{T_{sync}}^{T_b} \left[w_{i,1}^{(z)} \boldsymbol{g}_1^{(z)}(t) + w_{i,2}^{(z)} \boldsymbol{g}_2^{(z)}(t) + \boldsymbol{\varepsilon}^{(z)}(t) \right] \boldsymbol{g'}_1^{(z)}(t)\, dt \\
&= w_{i,1}^{(z)} \underbrace{\int_{T_{sync}}^{T_b} \boldsymbol{g}_1^{(z)}(t) \boldsymbol{g'}_1^{(z)}(t)\, dt}_{A} + w_{i,2}^{(z)} \underbrace{\int_{T_{sync}}^{T_b} \boldsymbol{g}_2^{(z)}(t) \boldsymbol{g'}_1^{(z)}(t)\, dt}_{B} \\
&\quad + \underbrace{\int_{T_{sync}}^{T_b} \boldsymbol{\varepsilon}^{(z)}(t) \boldsymbol{g'}_1^{(z)}(t)\, dt}_{C},
\end{aligned}
\tag{6}
$$

where T_{sync} is the synchronization transient time for each symbol. If chaotic synchronization is retained, $T_{sync} = 0$. Using orthonormal basis functions, the terms A and B are equal to 1 and 0, respectively. Thus, an estimate of $w_{i,1}^{(z)}$ is provided by

$$
D_{i,1}^{(z)} = w_{i,1}^{(z)} + \int_0^{T_b} \boldsymbol{\varepsilon}^{(z)}(t) \boldsymbol{g}_1^{(z)}(t)\, dt
\tag{7}
$$

The following outlines the problems of existing correlator-based receivers:

• *Synchronization problem:* The decision variable is a random variable whose mean value depends on the quality of the basis functions recovery (see the terms A and B in (6)). If perfect chaotic synchronization is not sustained, we have $T_{sync} \neq 0$ and $\boldsymbol{g'}_1^{(z)}(t) \neq \boldsymbol{g}_1^{(z)}(t)$. Using the DLCSK structure mitigates this problem.

• *Autocorrelation problem:* To obtain the best noise performance, basis functions must be orthonormal such that $\int_0^{T_b} g_1^2(t)dt = 1$ (See the term A in (6)). Chaotic basis functions are at best orthonormal over a symbol duration only in the mean, i.e., $E[\int_0^{T_b} g_1^2(t)dt] = 1$, where $E[\,\cdot\,]$ denotes the expectation operator [5]. The variance of the basis functions leads to an estimation error. This error can be decreased by increasing the length of the chaotic signals, or by normalizing the basis functions such that the transmitted energy E is kept constant.

• *Cross-correlation problem:* In the general case, $\int_0^{T_b} g_2(t)g_1'(t)dt \neq 1$ (see the term B in (6)). Some geometrical approaches, such as the Bloch sphere [25], can be employed to find *maximally orthogonal* points (or quasi-orthogonal signals [26]) on the constellation. However, signal classification becomes a complex problem when the number of classes increases [27]. The goal of the proposed map selection algorithm is to find M points on the sphere from a larger set of points.

• *Noise problem:* The term C in (6) shows that the mean value of the decision variable can be affected by noise and channel distortions. The DLCSK approach can reduce this negative effect using DTL [1]. If the receiver is well trained, DTL could extract the characteristics of the signals and channels, which contributes to reducing this term.

III. *M*-ARY DLCSK SYSTEM MODEL

Figure 2 depicts the general architecture of the *M*-ary DLCSK system. In this design, the proposed scheme contains two main phases, namely offline training, and online deployment. In the offline training phase, the chaotic maps are transmitted under different channel conditions. When the offline training phase is over, the online deployment (test) phase initializes the modulated data transmission and data detection. Note that the same structure is utilized for the map selection process. However, we should select the appropriate chaotic maps before the start of the training and testing phases. Table 1 introduces the used discrete-time chaotic sequences g_k and their generator functions used in this paper. In Table 1, a and b stand for the bifurcation parameters. For example, the Logistic map generates non-periodic and non-converging signals with $3.57 \leq a \leq 4$. By choosing a set of proper bifurcation parameters and initial values, the generated signals show chaotic properties which can be used for several scenarios, such as secure chaos-based communication systems [1].

The proposed *M*-ary DLCSK receiver has two main differences compared to its counterparts, such as LSTM-aided OFDM-DCSK [14] and LSTM-aided DNN [15]. The main difference is the mechanism for recovering the chaotic basis functions. A major weakness of the existing DCSK-based systems is the repeated transmissions of the reference signals, which lead to a degradation of data rate, energy efficiency, security, and reliability. The proposed modulation scheme drops out the reference transmissions, provides a coherent scheme, and achieves an outstanding performance among the chaos-based modulations. Owing to the difficulties of attaining complete chaotic synchronization in the conventional CSK systems, we propose a receiver based on the NNs. The capabilities of the NNs enable us to achieve a CSK receiver without the need for reference transmission or chaotic synchronization.

Another characteristic of the *M*-ary DLCSK system is that the training of the NN model is performed under SNR variations. In particular, in each transmission, the training SNR (σ_{tr}) is a random variable in a limited SNR range as $[\sigma_{tr,min}, \sigma_{tr,max}]$dB. In this paper, $\sigma_{tr,min}$ and $\sigma_{tr,max}$ is selected in the limited range $[11, 23]$dB and the NN model is tested over a wide range of $E_b/N_0 \geq 0$dB.

A. THE PROPOSED TRAINING STRATEGY USING DTL

We first give the definitions of a *domain* and a *task*. A domain $\mathcal{D}$ consists of two components: a feature space ∇ and a marginal probability distribution $P(r)$, where $r = \{r^{(1)}, \ldots, r^{(n)}\} \in \mathcal{R}$. Given a specific *domain* $\mathcal{D} = \{\mathcal{R}, P(r)\}$,
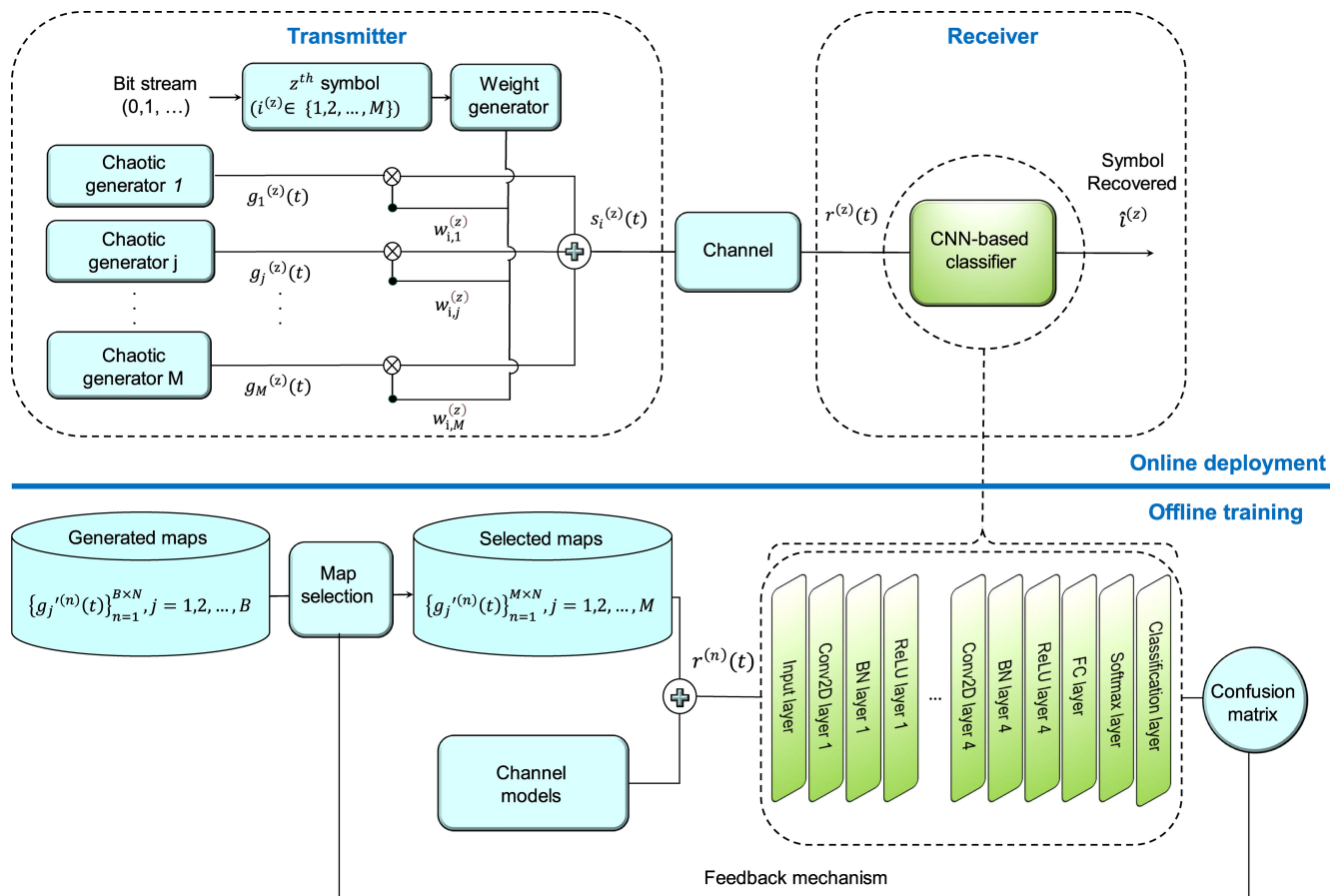
FIGURE 2. *M*-ary DLCSK communications system.

a *task* $\mathcal{T}$ consists of two components: a label space $\mathcal{J}$ and a function $f(\cdot)$ (i.e., $\mathcal{T} = \{\mathcal{J}, f(\cdot)\}$). The function $f(\cdot)$ is not observed but can be learned from the training data, which consists of pairs $\{r^{(n)}, j^{(n)}\}$, where $r^{(n)} \in \mathcal{R}$ and $j^{(n)} \in \mathcal{J}$. The function $f(\cdot)$ can be used to predict the corresponding label, $f(r)$, of a new instance r. From a probabilistic viewpoint, $f(r)$ can be written as $P(j|r)$.

In this paper, we consider one source domain (i.e., $\mathcal{D}_S$), and one target domain (i.e., $\mathcal{D}_T$). We denote the source domain data as $\mathcal{D}_S = \{r_S^{(n)}, j_S^{(n)}\}_{n=1}^N$, where $r_S^{(n)} \in \mathcal{R}_S$ is the data instance and $j_S^{(n)} \in \mathcal{J}_S$ is the corresponding class label. In our signal classification problem, $\mathcal{D}_S$ is a set of received chaotic signals with their associated labels. Similarly, we denote the target domain data as $\mathcal{D}_T = \{r_T^{(n)}, j_T^{(n)}\}_{n=1}^N$, where the input $r_T^{(n)} \in \mathcal{R}_T$ and $j_T^{(n)} \in \mathcal{J}_T$ is the corresponding output.

TL can be defined as [13]:

Definition 1: Given the source tasks $\mathcal{T}_S$, the source domain $\mathcal{D}_S$, the target task $\mathcal{T}_T$, and the target domain $\mathcal{D}_T$, the aim of transfer learning is to improve the performance of the target task $\mathcal{T}_T$ using the knowledge from $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$.

In the above definition, a domain is a pair $D = \{\mathcal{R}, P(r)\}$. Thus, the condition $\mathcal{D}_S \neq \mathcal{D}_T$ implies that either $\mathcal{R}_S \neq \mathcal{R}_T$ or $P(r_S) \neq P(r_T)$. In our design, the feature spaces between the domains are similar, i.e., $\mathcal{R}_S = \mathcal{R}_T$. The channel coefficients of $\mathcal{D}_S$ and $\mathcal{D}_T$ are also drawn from similar distributions. The only feature which we use for TL is the SNR. In particular, in each transmission, the training SNR is a random variable in a limited SNR range as $[\sigma_{tr,\ min}, \sigma_{tr,\ max}]$dB. Since the training is performed using stochastic SNRs and testing is performed for a fixed SNR, the marginal distributions are different between the two domains, i.e., $P(r_S) \neq P(r_T)$. In other words, the feature spaces between the domains are the same, but the marginal probability distributions between the two domains are different. Therefore, the condition $P(r_S) \neq P(r_T)$ is satisfied, and the target region signal classification can be formulated as a TL problem. Using this framework, the classification function $f(\cdot)$ can be learned from the training data and then used to classify the target signals. In our problem, $\mathcal{D}_S$ is a set of noisy signals $r_S^{(n)}$ with their associated labels, i.e., $j_S^{(n)}$. The received signals $r_S^{(n)}$ are filtered by the channel and carry the channel features. Thus, the NN learns the dynamics of various chaotic maps and the channel features at the same time. The trained NN can classify the target signals $r_T^{(z)}, z \in \{1, \ldots, Z\}$ and determine the corresponding labels $j_T^{(z)}$, where Z shows the number of target signals.

TABLE 1. The used chaotic maps.

Chaotic map	Generator Function
Chebyshev map [28]	$g_{k+1} = 1 - g_k^2$
Logistic map [29]	$g_{k+1} = ag_k(1 - g_k), \ a = 3.6$
Cubic map [29]	$g_{k+1} = 4g_k^3 - 3g_k$
Tent-like map [30]	$g_{k+1} = a(1 - \|1 - 2g_k\|), a = 0.9$
Circle map [30]	$g_{k+1} = mod(g_k + b - (\frac{a}{2\pi})sin(2\pi g_k), 1)$ $a = 0.5, b = 0.9$
Iterative map [30]	$g_{k+1} = sin(\frac{a\pi}{g_k}), a = 0.7$
Bernoulli shift map [30]	$\begin{cases} g_{k+1} = 1.2g_k - 1, \ if \ g_k > 0 \\ g_{k+1} = 1.2g_k + 1, \ if \ g_k \leq 0 \end{cases}$
Tent map [30]	$\begin{cases} g_{k+1} = \frac{g_k}{0.7}, \ if \ g_k < 0.7 \\ g_{k+1} = \frac{10}{3}(1 - g_k), \ if \ g_k \geq 0.7 \end{cases}$
Hénon map [31]	$\begin{cases} g_{k+1} = 1 + y_k - 1.4g_k^2 \\ y_{k+1} = 0.3g_k \end{cases}$
PAM map [32]	$\begin{cases} g_{k+1} = 1.5(1 - \|g_k\|) \ if \ \|g_k\| < 1 \\ g_{k+1} = 1.5(\|g_k\| - 2) \ if \ \|g_k\| \geq 1 \end{cases}$
Modified PAM map [32]	$\begin{cases} g_{k+1} = -1.5(\|g_k\|) \ if \ \|g_k\| < 1 \\ g_{k+1} = -1.5(\|g_k\| - 2) \ if \ \|g_k\| \geq 1 \end{cases}$

We should emphasize that in the proposed design, only feature which we have used for TL is the stochastic SNRs. If we train the NN with fixed SNRs, the problem reduces to a regular DL problem. Training with random SNRs improves the robustness and generalization capabilities of the proposed system in noisy environments. This is especially useful when we want to test the NN in different SNR conditions without the need for re-training. This TL problem is also known as *transductive TL* in the literature. In the transductive TL setting, the source and target tasks are the same, while the source and target domains are different [13]. In this situation, no labeled data in the target domain are available, while a lot of labeled data in the source domain are available. When $P(r_S) \neq P(r_T)$, the transductive TL is related to *domain adaptation*, *sample selection bias*, and *co-variate shift* methods whose assumptions are similar [13]. Note that the word "transductive" is used with several meanings. In the traditional machine learning setting, transductive learning [62] refers to the situation where all test data are required to be seen at training time, and that the learned model cannot be reused for future data. In this paper, we use the definition of transductive learning presented in [13]. Unlike [62], in the categorization presented in [13], the authors use the term transductive to emphasize the concept that the tasks must be the same and there must be some unlabeled data available in the target domain.

Table 2 provides a comparison between the *M*-ary DLCSK and its DL-based counterparts, i.e., LSTM-aided DNN [15]

and LSTM-aided OFDM-DCSK [14], from the viewpoint of the used features. As shown in Table 2, existing DL-based demodulators are based on the DCSK modulation scheme. In the LSTM-aided DNN receiver, the NN model is trained with the MCS-MDCSK modulated signals and tested with similar modulated signals. In fact, the basis function is recovered by the transmission of the reference signals. In such a case, a corrupted reference affects the demodulation of one or more data-bearing signals. In other words, in the DCSK-based modulations, transmission errors in the reference chaotic signal would induce propagation errors. The *M*-ary DLCSK system is based on the coherent CSK scheme. In the *M*-ary DLCSK system, the basis function recovery is performed by training the NN. The NN learns the characteristics of each chaotic map, where each map represents one class. The NN adjusts its hyperparameters and uses them in the deployment phase. Therefore, the characteristics of each basis function are transferred from the source domain to the target domain by reusing the hyperparameters, and we can recover the chaotic basis signals at the receiver without the need for chaotic synchronization. Consequently, in the testing phase, we can assume that the receiver knows the references and demodulates the unlabeled *M*-ary CSK modulated signals. Although the sample basis signals are known at the receiver side of a DLCSK system in the testing phase, this knowledge is transferred from the training phase. According to the above discussion, since the proposed model is based on coherent modulation, the performance of the proposed system is improved.

It is worth mentioning that in all systems presented in Table 2, the feature spaces between the domains are similar, i.e., $\mathcal{R}_S = \mathcal{R}_T$. However, in the *M*-ary DLCSK, since the training is performed using stochastic SNRs and testing is performed for a fixed SNR, the marginal distributions are different between the two domains, i.e., $P(r_S) \neq P(r_T)$. In other words, the feature spaces between the domains are the same, but the marginal probability distributions between the two domains are different. Therefore, according to *Definition 1*, the condition $P(r_S) \neq P(r_T)$ is only satisfied for the *M*-ary DLCSK system. Thus, only the *M*-ary DLCSK system is trained under a TL framework.

B. DATASET GENERATION FOR TRAINING AND TESTING

In the training phase, the transmitter generates M sets of the chaotic signals, where each of these sets includes N chaotic signals, i.e., $\{g'^{(n)}_j(t)\}^N_{n=1}$, $j = 1, 2, \ldots, M$, where n denotes the nth signal transmission. In other words, we use M sets of the training signals, each including $S = N\beta$ samples, where N is the number of signals, and β is the spreading factor. All chaotic signals $g'^{(n)}_j(t)$ have equivalent class labels as $j_S^{(n)} \in \{1, 2, \ldots, M\}$. The received signals $r_S^{(n)}(t)$ act as input to the NN in the source domain

$$r_S^{(n)}(t) = \sum_{l=1}^{L} \alpha_l^{(n)} g'^{(n)}_j(t - \tau_l) + \varepsilon^{(n)}(t), \qquad (8)$$

TABLE 2. Example features of the *M*-ary DLCSK and LSTM-aided DNN systems.

System	Source features $\mathcal{D}_S^{(n)}$	Target features $\mathcal{D}_T^{(n)}$
M-ary DLCSK	M-ary CSK modulated signals with known labels Variable SNRs ($\sigma_{tr}^{(n)} \in [\sigma_{tr,min}, \sigma_{tr,max}]$) Multi-path fading	Unlabeled *M*-ary CSK modulated signals Fixed SNRs ($\sigma_{tr}^{(n)} = \sigma_{tr}$) Multi-path fading
LSTM-aided DNN [15]	MCS-MDCSK modulated signals with known labels Fixed SNRs ($\sigma_{tr}^{(n)} = \sigma_{tr}$) Multi-path fading	Unlabeled MCS-MDCSK modulated signals Fixed SNRs ($\sigma_{tr}^{(n)} = \sigma_{tr}$) Multi-path fading
LSTM-aided OFDM-DCSK [14]	OFDM-DCSK modulated signals with known labels Fixed SNRs ($\sigma_{tr}^{(n)} = \sigma_{tr}$) Multi-path fading	Unlabeled OFDM-DCSK modulated signals Fixed SNRs ($\sigma_{tr}^{(n)} = \sigma_{tr}$) Multi-path fading

where α_l^n and τ_l are the channel gain and delay of the lth path, respectively. L is the number of paths, and $\varepsilon^{(n)}(t)$ is independent noise, which is assumed to be AWGN with zero mean and variance $N_0/2$, where N_0 is defined as the Power Spectral Density (PSD) of the AWGN. After $M \times N$ transmissions, we have a set of training vectors $\{r_S^{(n)}(t)\}_{n=1}^{M \times N}$, and corresponding labels, $\{j_S^{(n)}\}_{n=1}^{M \times N}$. Thus, the training set of the *M*-ary DLCSK receiver can be expressed as

$$\mathcal{D}_S = \left\{ r_S^{(n)}(t), j_S^{(n)} \right\}_{n=1}^{M \times N}, \ j_S = 1, 2, \ldots, M, \qquad (9)$$

where $M \times N$ is the number of signals in the training set. Assuming a Rayleigh fading channel, $r_S^{(n)}(t)$ can be modeled as a complex-valued random variable, i.e., $r_S(t) = \Re(r_S(t)) + \Im(r_S(t))$, where $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imaginary parts of a complex number, respectively. Therefore, the vector r_S can be separated into two vectors, i.e., $r_S = [\Re(r_S), \Im(r_S)]$, before being fed into the classifier, and the training set can be rewritten as

$$\mathcal{D}_S = \left\{ [\Re(r_S), \Im(r_S)]^{(n)}, j_S^{(n)} \right\}_{n=1}^{M \times N}. \qquad (10)$$

After the NN training, the trained classifier can be used for online demodulation.

To generate the test signals for the evaluation of the proposed design, the *M*-ary modulator maps a symbol $i^{(z)} \in \{1, \ldots, M\}$, ($1 \leq z \leq Z$), to a chaotic waveform $s_i^{(z)}(t)$ using (2). A CNN can learn the channel features and dynamics of the chaotic signals under the given environment. Thus, the test dataset can be written as

$$\mathcal{D}_T = \left\{ r_T^{(z)}(t), j_T^{(z)} \right\}_{z=1}^{Z}, \ j_T = 1, 2, \ldots, M, \qquad (11)$$

where Z shows the number of testing chaotic signals or target signals.

C. CHANNEL ESTIMATION
A multipath fading channel is assumed such that the channel coefficients are constant over a symbol duration and vary from one symbol to another. To achieve reliable communication, channel estimation plays a key role in eliminating the impact of fading channels. In [12], the communication system is considered as a black-box, and an end-to-end DL architecture is used for signal detection. Using the black-box approach mentioned above, encoding, decoding, channel estimation, and all other functionalities of a communication link are implicitly embedded in the NN model. This method is not able to explicitly find the channel time-frequency response and is not effective for applications that need to have the complete channel response. Inspired by [12], to eliminate the requirement of channel estimation, we adopt a DL-based approach to implicitly extract the channels' features and recover the data symbols. In other words, the NN in the source domain adjusts itself to be utilized in the target domain. Thus, the proposed receiver needs no additional pilot transmission for channel estimation. Note that, the channel coefficients of $\mathcal{D}_S$ and $\mathcal{D}_T$ are drawn from similar distributions. In this design, the received signals (the chaotic signals filtered by the channel) act as input to the NN model. Since the channel influences the transmitted signals, each of the received signals includes the channel features. Thus, the NN can learn the dynamics of various chaotic maps, where each of the received signals carries the channel features.

The Probability Density Function (PDF) of the channel coefficient α in the multipath Rayleigh fading channels can be written as [34]

$$q(\alpha | \delta) = \frac{\alpha}{\delta^2} e^{\frac{-\alpha^2}{2\delta^2}}, \qquad (12)$$

where $\delta > 0$ is the scale parameter of the distribution representing the root-mean-square value of the received voltage signal. Based on the above-mentioned assumptions the NN should be trained with complex-valued input vectors r_S. At the receiver side, these complex-valued input vectors are split into real and imaginary parts, i.e., $r_S = [\Re(r_S), \Im(r_S)]$. Thus, under the fading channels, we have two input feature vectors. A Softmax layer estimates two probability vectors $p_{n,j}$ from the above-mentioned input distribution, where j_S shows each of the possible classes (i.e., $j_S \in \{1, 2, \ldots, M\}$), in (13). In the test phase, we utilize this trained model to demodulate test inputs.
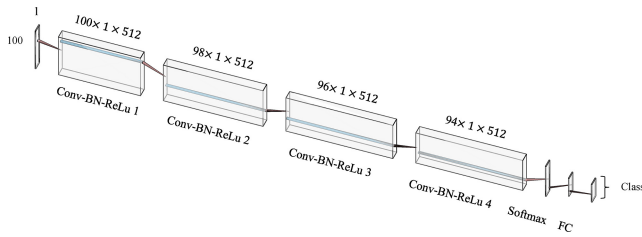
FIGURE 3. The structure of the proposed CNN-based classifier.

In this paper, we assume that the training SNR $\sigma_{tr}^{(n)}$ changes after the nth channel realization (or transmission). In particular, $\sigma_{tr}^{(n)}$ is a Gaussian random variable, such that $\sigma_{tr}^{(n)} \in [\sigma_{tr,min}, \sigma_{tr,max}]$, where $\sigma_{tr,min}$ and $\sigma_{tr,max}$ are optional SNR bounds. In this paper, $\sigma_{tr,\,min}$ and $\sigma_{tr,\,max}$ are selected in the limited range $[11, 23]$dB, and tested over a wide ranger of $E_b/N_0 \geq 0$dB. Note that in the current form of the DLCSK systems, we only consider one source and one target domain including a number of chaotic signals that are transmitted over the channels with a given distribution. Therefore, using state-of-the-art training methods, such as meta-transfer learning, that consider multiple source features will be effective. Furthermore, training on wider range of SNRs and other features of practical channels can be comprehensively investigated in future works.

D. CNN-BASED RECEIVER

Research efforts have demonstrated that CNNs are effective in studies and applications involving chaotic signals, such as chaotic biomedical signal analysis, speech processing, and chaos identification systems [48], [49], [50]. Moreover, the selection of CNNs allows us to gain useful theoretical insight into the proposed system. We will see later that we can calculate a lower bound approximation for the BER of a correlation-based M-ary CSK receiver. Since correlation and convolution operations are almost identical, the output of a correlator can be considered an approximation of the output of a convolutional layer. Consequently, this lower bound can also be considered as a lower bound for the M-ary DLCSK system. These theoretical analyses can be useful for different purposes, such as validation of simulation results or designing power allocation strategies.

The receiver classifies each of the received test signals r_T to the corresponding labels $j_T \in \{1, 2, \ldots, M\}$. As shown in Figure 3, the proposed CNN-based classifier has 16 base layers: one sequence input layer, four 2-D convolution layers (Conv2D), four Batch Normalization (BN) layers, and four ReLU layers. The Fully Connected (FC), Softmax, and classification layers are located at the end of the process. The sequence input layer is only used to fetch sequential input values. Since each of the received vectors is separated into two vectors, i.e., real and imaginary vectors, input values into the NN are of size $2 \times \beta$.

In our work, four Conv2D layers are employed to simultaneously produce an acceptable BER performance and a

reasonable complexity level. By changing the number of Conv2D layers between 1 and 10, we found that four Conv2D layers result in an acceptable BER, and increasing the number of Conv2D layers has no significant effect on the BER performance. Each Conv2D layer includes n_f ($1 \leq n_f \leq 1024$) sliding filters. We specify a stride size equal to 1 for step size, with padding size 0. The BN layer normalizes a mini-batch of data across all transmissions. Several BN layers are operated between the convolutional layers. In the BN layer, the input vector is normalized to have zero mean and unit variance. In this way, the value of input vectors is adapted to the range wherein the activation functions have high gradients. Thus, the vanishing and exploding gradient problems are solved and the convergence process is accelerated. The ReLU layer handles the nonlinearity in the model and realizes a threshold operation to each segment of the input, where any value less than zero is set to zero. The FC layer enhances stability by performing more non-linear operations and specifies the output size. For M-ary DLCSK with M output classes, there is an FC layer with M output neurons. We will see later that the number of input neurons of an FC layer depends on the size of the input sequence $n_s = \beta$, filter size n_k, and the number of Conv2D layers. The Softmax layer is an activation function that delivers results to the classification layer by computing a probability for each input sequence $U(t)$. The output values show the probability that the sequence $U(t)$ belongs to the class j. Therefore, the Softmax layer contains j nodes, which is equal to the number of classes. The utilized Softmax function can be expressed as [35],

$$\gamma(U(t))_j = \frac{e^{(U(t))_j}}{\sum_{k \in \mathcal{K}} e^{(U(t))_k}}, \qquad (13)$$

where $\gamma(U(t))_j = p_{n,j}$ is the probability that vector $U(t)$ belongs to the jth class ($j \in \mathcal{K}$), and $\mathcal{K} = \{1, \ldots, M\}$ shows all possible classes. For example, in a 2-ary DLCSK system, $p_{n,0}$ and $p_{n,1}$, ($1 \leq n \leq N$), shows the probability that the transmitted symbol is "1" or "2," respectively.

The main objective of the training process is to minimize the categorical cross-entropy cost function [36]. The cross-entropy is a widely used cost function for classification tasks. This function penalties the NN model for wrong decisions and gives a larger gradient value that leads to faster convergence. The calculated cross-entropy loss function is given by

$$\mathrm{L}(\boldsymbol{\theta}) = -\frac{1}{a_t} \sum_{t=1}^{a_t} \sum_{j=1}^{M} p'_{n,j} \log(p_{n,j}), \qquad (14)$$

where a_t is the mini-batch size, $\boldsymbol{\theta}$ indicates the set of network parameters, $p_{n,j}$ is the Softmax layer output for class j and observation n, and $p'_{n,j} \in \{0, 1\}$ is a binary indicator that specifies the correctness of the class label j for observation n. The Stochastic Gradient Descent (SGD) method is widely used an iterative algorithm to solve this optimization problem [36]. To find the optimal network parameters, the
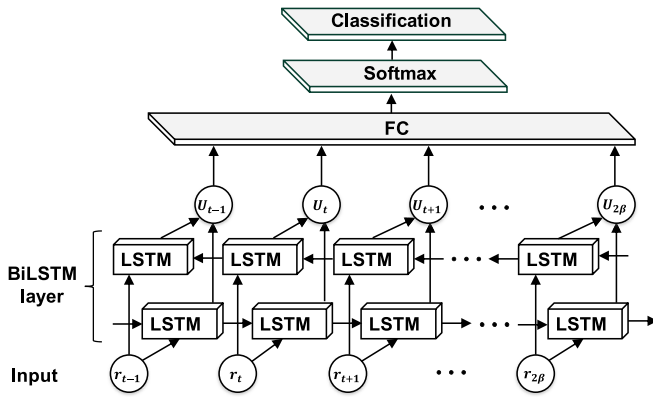
FIGURE 4. The structure of the proposed BiLSTM-based classifier.

SGD algorithm starts with a random initial value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and repeatedly updates it. $\tilde{L}(\boldsymbol{\theta})$ gives an approximation of the cost function at each iteration, which is computed for a random mini-batch of training samples of size a_t. At the end of the training process, a network with optimized weights and biases that can be used for online demodulation is obtained.

E. BILSTM-BASED RECEIVER
In addition to the CNN-based DLCSK, we recommend a classification model based on the Bidirectional Long Short-Term Memory (BiLSTM) networks for comparative analysis. Different types of Recurrent Neural Networks (RNNs) are often used to learn complex dynamics from the input data. However, classic RNNs suffer from some issues, such as vanishing gradients and the tendency to take into account only short-term dependencies. LSTMs are able to hold previous information for a longer period. LSTMs have demonstrated outstanding performance in practical chaotic signal processing, classification of chaotic physiological signals, and prediction of non-linear dynamics [51], [52], [53]. As presented in Figure 4, based on the excellent performance of LSTMs, we chose Bidirectional LSTMs to design an NN-based receiver that classifies the received chaotic signals. A BiLSTM can process input data from front to back and back to front. The designed classifier has a well-known structure including five base layers: a sequence input, a BiLSTM, an FC, a Softmax, and a classification layer. Theoretical tasks of an LSTM cell can be found in [37]. The FC, Softmax, and classification layers are similar in both the BiLSTM and CNN-based classifiers, as previously described in (13) and (14).

F. MAP SELECTION
The chaotic basis signals used in the M-ary DLCSK system can be generated by different chaotic map generators or by a single generator with various initials or bifurcation parameters. In all the cases mentioned above, the generated chaotic basis signals have a level of cross-correlation that deteriorates the BER performance of the system. In

this paper, M different chaotic map generators are used to produce the number of M chaotic signals. In each transmission, based on the current data symbol, the transmitter chooses one of these maps and transmits a segment (with the length of β) of the produced long chaotic signals. To reduce the negative effects of the cross-correlation, we use relatively longer segments (i.e., $\beta = 50$) for all experiments.

Furthermore, among all possible solutions, we focus on the map selection algorithm and demonstrate that it can effectively reduce the destructive effects of the cross-correlation by choosing a quasi-orthogonal signal set. Future studies could examine other algorithms for improving the statistical characteristics of the chaotic basis signals or providing strictly orthogonal basis signals. For example, the proposed scheme can be combined with the Walsh codes [44]. In this paper, we applied the Gram-Schmidt algorithm [24] to the initial long signals for convenience. Alternatively, the Gram-Schmidt algorithm can be applied to the transmitted segments of chaotic signals. The following focuses on the proposed map selection algorithm.

Most existing map selection algorithms analyze the cross-correlation properties of the sequences made by the chaotic maps [38]. In addition to the cross-correlation of the transmitted signals, the BER performance of the proposed system also depends on the selected classifier and the channel conditions. If we adopt the output performance as the decision criterion, we can consider the end-to-end effects of our selected set of maps, whereas cross-correlation optimization may only have a limited effect on the performance (refer to the (6)). Indeed, we would like to choose M proper maps using the confusion matrix [39] through a simple feedback mechanism. The confusion matrix contains valuable information about the number of correctly and incorrectly classified symbols. These correct and incorrect classifications are on the diagonal and non-diagonal elements of the confusion matrix, respectively. Thus, we consider the *total number of incorrect classifications* which is the sum of the numbers on the non-diagonal elements.

We would like to derive an optimal policy for selecting a sub-set $\mathcal{D}_{sel} = \{g_j^{(n)}(t)\}_{n=1}^N$, $j = 1, 2, \ldots, M$, that contains M chaotic maps, from a larger set $\mathcal{D}_{all} = \{g_m^{(n)}(t)\}_{n=1}^N$, $m = 1, 2, \ldots, B$, where $B > M$. To select M maps from the total B maps, $A = \binom{B}{M} = \frac{B!}{M!(B-M)!}$ different cases (different sets of maps) are possible. We realize all the possible cases, record all related confusion matrices, and use a metric to select the case that provides the best performance. The set that leads to a lower number of incorrectly classified symbols is selected. We define $\phi_a = Sum\{Incorrect\ classifications\}$, $a = 1, 2, \ldots, A$, as the selection criterion, where A shows the number of possible cases (possible sets of maps). Calculating ϕ_a for all possible cases helps us choose the case with the best result. ϕ_a can be calculated from the non-diagonal elements of the confusion matrix. In other words, the total number of incorrect classifications can be obtained by calculating the

---

Algorithm 1: Selecting M Maps From B Maps

Data: A set of chaotic signals using B maps, i.e.,
$\mathcal{D}_{all} = \{g_m^{(n)}(t)\}_{n=1}^N$, $m = 1, 2, ...B$. Each map includes N chaotic signals $g_m(t)$. The selection criterion is ϕ_a, $a = 1, 2, ..., A$, where $A = \binom{B}{M} = \frac{B!}{M!(B-M)!}$ shows the number of possible cases.

Result: Selection of M chaotic maps that lead to the lowest ϕ_a

begin

    for $a = 1 : A$ **do**

        Calculate all possible ϕ_a using the number of A confusion matrices:

        $\phi_a = Sum\{Incorrect\ classifications\}$, $a = 1, 2, ..., A$.

    Sort all the obtained ϕ_a;

    Return M maps corresponding to the lowest ϕ_a;

    End

Algorithm 2: Training of an M-ary DLCSK System Using DTL

Data: A set of chaotic signals using M maps, i.e.,
$\mathcal{D}_{sel} = \{g_j^{(n)}(t)\}_{n=1}^N$, $j = 1, 2, ...M$. For each map, N chaotic signals are generated as $g_j^{(n)}(t)$, $n = 1, 2, ..., N$. In addition, $[\sigma_{tr,min}, \sigma_{tr,max}]$ determines the training SNR region and L shows the number of paths;

Result: A trained M-ary DLCSK receiver;

begin

    for $i = 1 : M$ **do**

        for $n = 1 : N$ **do**

            Generate $\sigma_{tr}^{(n)} \in [\sigma_{tr,min}, \sigma_{tr,max}]$;

            Generate $\alpha_l^{(n)}$, $\tau_l^{(n)}$;

            Transmit n^{th} signal, i.e., $g_i^{(n)}(t)$;

    Receive all signals $\{r_S^{(n)}(t)\}_{n=1}^{M \times N}$, with known labels $\{j_S^{(n)}\}_{n=1}^{M \times N}$;

    Separate $r_S^{(n)}(t)$ into real and imaginary parts to form the final training set $\mathcal{D}_{sel} = \{[\Re\{r_S\}, \Im\{r_S\}]^{(n)}, j_S^{(n)}\}_{n=1}^{M \times N}$;

    Define a CNN classifier and train it using $\mathcal{D}_{sel}$;

    End

---

sum of the numbers on the non-diagonal elements. Once the best case is determined, we can extract the *selected maps* to use them for training. The proposed map selection process is summarized in Algorithm 1. After finding the proper maps, the desired dataset can be produced. The proposed training algorithm is also described in Algorithm 2.

G. LOWER BOUND APPROXIMATION FOR THE BER OF M-ARY DLCSK

Obtaining complete chaotic synchronization and recovery of the chaotic basis functions is difficult under a harsh wireless environment. Failure to solve this problem has impeded the theoretical analysis of coherent CSK systems. Few articles have only analyzed the theoretical performance of the binary CSK. Since this paper is the first attempt to design a coherent M-ary DLCSK system, there is no theoretical benchmark to compare and validate the simulation results. Therefore, we intend to obtain the performance of a correlation-based M-ary CSK receiver and perform a comparative investigation.

Based on the given constellation, we can calculate the BER of a correlator receiver. It is worth mentioning that correlation and convolution operations are almost identical. Thus, the output of a correlator can be considered as an approximation of the output of a convolutional layer. In other words, since M-ary DLCSK works with conventional CSK transmitters, we can use the properties of CSK transmitters to calculate a lower bound for the BER of M-ary DLCSK systems under AWGN channels. However, correlation receivers use synchronization for reference regeneration, whereas M-ary DLCSK systems utilize DTL for training and regeneration of the reference signals. Indeed, the convolutional layer captures the features well and reproduces the reference signals. Since the distance between any pair of the constellation points is equal to $\sqrt{2E}$, the BER of

the optimum detector is independent of the transmitted signal. Thus, to compute the BER performance of the system, we can suppose that the test signal $s_1 = (\sqrt{E}, 0, \ldots, 0)$ is transmitted. With this assumption, the received vector is $r = (\sqrt{E} + \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)$, where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M$ are zero-mean, mutually independent Gaussian random variables with variance $N_0/2$.

Using the method presented in [46], the decision variables in orthogonal signaling can be defined as $D_{i,j} = r \cdot s_i'$, $1 \leq i, j \leq M$, where s_i', $1 \leq i \leq M$ is a regenerated version of $s_i, 1 \leq i \leq M$ at the receiver side. Since we assumed that s_1 has been transmitted, the decision variables can be written as

$$D_{i,1} = E + \sqrt{E}\varepsilon_1$$
$$D_{i,j} = \sqrt{E}\varepsilon_i, \ 2 \leq j \leq M. \tag{15}$$

The coherent correlation receiver estimates the elements $w_{i,j}$ of the signal vector. In the noise-free case, with perfect regeneration of the basis functions, $s_i' = s_i$. Since we assumed that s_1 has been transmitted, if $D_{i,1} > D_{i,j}$, $j = 2, 3, \ldots, M$, the decision is "1" (i.e., the decision is correct). Thus, the probability of a correct decision can be written as

$$P_c = P[D_{i,1} \geq D_{i,2}, D_{i,1} \geq D_{i,3}, \ldots, D_{i,1} \geq D_{i,M}|s_1]$$
$$= P[\sqrt{E} + \varepsilon_1 > \varepsilon_2, \sqrt{E} + \varepsilon_1 > \varepsilon_3, \ldots, \sqrt{E} + \varepsilon_1 > \varepsilon_M]. \tag{16}$$

Events $\sqrt{E} + \varepsilon_1 > \varepsilon_2, \sqrt{E} + \varepsilon_1 > \varepsilon_3, \ldots, \sqrt{E} + \varepsilon_1 > \varepsilon_M$ are dependent due to the existence of the random variable ε_1. Conditioning on ε_1, these events are independent. With

independent and identically distributed (iid) random variables ε_i, P_c can be expressed as

$$P_c = \int_{-\infty}^{\infty} \left(P\left[\varepsilon_2 < \varepsilon + \sqrt{E}|s_1, \varepsilon_1 = \varepsilon \right] \right)^{M-1} p_{\varepsilon_1}(\varepsilon)d\varepsilon, \tag{17}$$

We have

$$P\left[\varepsilon_2 < \varepsilon + \sqrt{E}|s_1, \varepsilon_1 = \varepsilon \right] = 1 - Q\left(\frac{\varepsilon + \sqrt{E}}{\sqrt{N_0/2}} \right). \tag{18}$$

Hence,

$$P_c = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi N_0}} \left[1 - Q\left(\frac{\varepsilon + \sqrt{E}}{\sqrt{N_0/2}} \right) \right]^{M-1} e^{-\frac{\varepsilon^2}{N_0}} d\varepsilon. \tag{19}$$

and

$$P_e = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[1 - (1 - Q(x))^{M-1} \right] e^{-\frac{\left(x - \sqrt{\frac{2E}{N_0}} \right)^2}{2}} dx, \tag{20}$$

where $x = \frac{\varepsilon + \sqrt{E}}{\sqrt{N_0/2}}$. P_e shows the probability of error of the considered constellation and provides a lower bound for the simulated BER. Due to the symmetry of the constellation, the probabilities of receiving any of the messages $i = 2, 3, \ldots, M$, when s_1 is transmitted, are equal. Therefore, the probabilities of receiving any of the messages s_i when s_1 is transmitted can be written as $P[s_i|s_1] = \frac{P_e}{2^\varphi - 1}$. Assume that s_1 corresponds to a data sequence of length φ with "0″ as the first bit. The error probability at this bit is the probability of detecting a s_i with a "1″ at the first bit. Since there are $2^{\varphi-1}$ sequences, the probability of error can be written as

$$P_b = 2^{\varphi-1} \frac{P_e}{2^\varphi - 1} = \frac{2^{\varphi-1}}{2^\varphi - 1} P_e. \tag{21}$$

The theoretical probability of error can be approximated using (20) and (21).

IV. COMPLEXITY, DATARATE, AND BANDWIDTH EFFICIENCY ANALYSIS

In this section, the computational complexity, data rate, and BE of the proposed M-ary DLCSK scheme are analyzed. Furthermore, a set of comparative analyses with chaos-based communication benchmarks is presented.

A. COMPLEXITY ANALYSIS

The following presents a complexity study on the proposed M-ary DLCSK receiver based on the method introduced in [40]. The advantage of this method over other complexity calculation methods is that it allows us to access the details of layers, such as the number of convolutional filters or the number of hidden units. Since the number of additions is often ignored in complexity evaluations, we calculate the complexity in terms of Real Multiplications per Symbol (RMpS) for the suggested NN architectures. As discussed in [40] and [41], considering an offline training

process, we only evaluate the complexity at the deployment stage, where the receiver demodulates the transmitted data. Moreover, the complexity of nonlinear activation functions is not considered in this paper, due to the fact that typically their operations are based on an approximation approach, rather than direct multiplication. For example, in the classical lookup tables-based approximation methods, the tasks can be carried out with a few computations [45]. The number of multiplications will be zero for the input layer since the sequence input layer only receives and groups the input data.

We first consider the computational complexity of the BiLSTM-based M-ary DLCSK receiver. Let β be the size of the input sequence and n_i be the number of feature vectors, which is one and two for AWGN and Rayleigh channels, respectively. With these values, the dimension of the sequence input layer can be written as βn_i. The complexity of a BiLSTM layer with n_h hidden units can be calculated as [40]

$$C_{BiLSTM} = 2\beta n_i n_h (4n_i + 4n_h + 3 + n_o), \tag{22}$$

where n_o is the number of outputs per symbol. The complexity of the FC layer, with the output dimension of $\mathcal{L}_{fc}$, can be computed as $\beta n_i \mathcal{L}_{fc}$. Similarly, the complexity of the classification layer with the output dimension of $\mathcal{L}_{class}$ can be calculated as $\mathcal{L}_{fc} \mathcal{L}_{class}$). The total computational complexity of the BiLSTM-based M-ary DLCSK receiver is the sum of the computational complexities of the above-mentioned layers. As shown in Table 3, the computational complexity per symbol of the proposed BiLSTM-based M-ary DLCSK receiver is dependent on the values of n_h, which also affects the BER performance.

Now, we consider the computational complexity of the CNN-based M-ary DLCSK receiver. The computational complexity of the first 1-D convolutional layer can be written as [40]

$$C_{CNN} = n_i n_f n_k (\beta - n_k + 1), \tag{23}$$

where n_f shows the number of convolutional filters and n_k stands for the kernel (or filter) size. The output dimension of the first convolutional layer can be calculated as $\mathcal{L}_{c,1} = (\beta - n_k + 1)$, which can be considered as the input dimension of the next layer. In other words, the input dimension of each of the next convolutional layers, i.e., $\mathcal{I}_{c,\theta}$, can be calculated as $\mathcal{I}_{c,\theta} = (\mathcal{L}_{c,\theta-1} - n_k + 1)$, where $2 \leq \theta \leq 4$. The computational complexity of the FC and classification layers can be calculated in the way that is already described for BiLSTM-based receivers. Finally, the total computational complexity of the CNN-based M-ary DLCSK receiver can be calculated by summing the calculated complexities for all layers.

From the viewpoint of flexibility, since the total computational complexity can be adjusted based on the user requirements, the NN-based receivers can provide more flexible designs. For instance, the results of this study suggest that increasing the number of convolutional filters (n_f) or

TABLE 3. Computational complexity comparisons.

System	Complexity (RMpS)
M-ary DLCSK (BiLSTM)	$O(\beta n_i^2 n_h + \beta n_i n_h^2 + \beta n_i n_h + \beta n_i \mathcal{L}_{fc,1} + \mathcal{L}_{fc,1} \mathcal{L}_{class})$
M-ary DLCSK (CNN)	$O(n_i n_f n_k \mathcal{L}_{c,1} + n_i n_f n_k \mathcal{L}_{c,2} + n_i n_f n_k \mathcal{L}_{c,3} + n_i n_f n_k \mathcal{L}_{c,4} + \mathcal{L}_{c,4} \mathcal{L}_{fc,1} + \mathcal{L}_{fc,1} \mathcal{L}_{class})$
LSTM-aided DNN [15]	$O(\beta \mathcal{N} \mathcal{H} \mathcal{L}_{fc,1} + \beta \mathcal{N} \mathcal{L}_{fc,1} \mathcal{L}_{fc,2} + \beta \mathcal{N} \mathcal{L}_{fc,2} \mathcal{L}_L + \mathcal{N} \mathcal{L}_L \mathcal{L}_{fc,3} + \mathcal{N} M \mathcal{L}_{fc,3})$
Iterative receiver [10]	$O(\mathcal{N} \mathcal{H} \beta + \mathcal{N} \beta \Phi + \mathcal{N} M)$
LRAM-based detector [11]	$O(\mathcal{N} \mathcal{H} \beta + \mathcal{N} \beta^2 + \mathcal{N} M)$

equivalently increasing the complexity up to 10^6 RMpS, can lead to an enhancement in the BER performance under AWGN channels. However, the NNs with large RMpS (typically when RMpS$\geq 10^8$) are too complex for hardware implementation. It is worth mentioning that all architectures proposed in this paper have complexity values in the range from 10^3 to 10^7 RMpS.

Table 3 provides a comparison between the complexity of the M-ary DLCSK system and benchmark schemes. In Table 3, the computational complexity of the LSTM-aided DNN detector [15] is presented, where $\mathcal{L}_L$ and $\mathcal{L}_{fc,\theta}$ denote the output dimensions of the LSTM layer and θ^{th} FC layer, respectively. In addition, Table 3 presents the computational complexity of the general iterative receiver [10] and the low-rank approximation of matrices (LRAM)-based detector [11], where Φ stands for the number of iterations.

B. DATARATE AND BANDWIDTH EFFICIENCY ANALYSIS

In the following, we study the data rate and BE of the proposed system and provide a comparison with other benchmarks, namely, M-ary DCSK [42], Multilevel Code-shifted Differential Chaos Shift Keying (MCS-DCSK) [43], Multilevel Code-shifted Differential Chaos Shift Keying with M-ary modulation (MCS-MDCSK) [44], and LSTM-aided DNN systems [15]. In this paper, the data rate (R) is defined as the number of transmitted bits per symbol duration.

In order to enhance the data rate of the binary DCSK, an M-ary DCSK is proposed in [42]. In the M-ary DCSK system, the basis functions are generated as a product of Walsh functions and chaotic waveforms. The Walsh functions assure the orthogonality of basis functions. In the M-ary DCSK scheme, $M \cdot \beta$ samples are needed to transmit $log_2 M$ bits of data. Therefore, the data rate and BE of M-ary DCSK system can be written as $R_1 = log_2 M$ and $BE_1 = log_2 M/(M \cdot \beta)$, respectively. The M-ary DCSK system offers a good BER performance under AWGN channels. However, the way that

the M-ary DCSK system utilizes the Walsh codes results in a relatively low BE. Moreover, M-ary DCSK requires a number of delay elements at the transmitter and receiver that increases exponentially with the number of bits per symbol.

In [43], the MCS-DCSK modulation has been proposed to carry multiple bits. Unlike the M-ary DCSK system that uses Walsh codes to separate information-bearing signals, the MCS-DCSK system utilizes one of the Walsh functions for the transmission of the reference signal and uses the remaining Walsh functions for the transmission of the reference signal and uses the remaining Walsh functions for the information-bearing signals. Since the reference signal and a number of information-bearing signals are transmitted in the same slot, the BE of the MCS-MDCSK system is enhanced considerably. However, the receiver of the MCS-DCSK system requires a number of delay elements that reduce the BER performance of this system. According to [43], for a given $\mathcal{H}$-order Walsh code matrix, the maximum data rate and BE of the MCS-DCSK can be written as $R_2 = \mathcal{H} - 2$ and $BE_2 = (\mathcal{H} - 2)/(M \cdot \beta)$, respectively.

In [44], the authors proposed the MCS-MDCSK system by combining MCS-DCSK with M-ary modulation. In particular, the MCS-MDCSK system uses the orthogonal Walsh functions to carry in-phase and quadrature components of the M-ary constellation symbols. Since the reference signal and numerous information-bearing signals are transmitted in the same slot, the BE of the MCS-MDCSK system is enhanced. In the MCS-MDCSK system, $\mathcal{N}$ represents the number of M-ary constellation symbols transmitted in a symbol duration and its maximum value is $\mathcal{N}_{max} = \frac{\mathcal{H}-2}{2}$. According to [44], the maximum data rate and BE of MCS-MDCSK system can be calculated as $R_3 = (\frac{\mathcal{H}-2}{2}) log_2 M$ and $BE_3 = (\frac{\mathcal{H}-2}{2}) log_2 M/(M \cdot \beta)$, respectively. The data rate of MCS-MDCSK can be increased in a linear fashion with $\mathcal{H}$ and in a logarithmic way with the modulation order M. The LSTM-aided DNN system [15], proposes a DL-based detector for MCS-MDCSK modulation and does not change the transmitter structure of the benchmark MCS-MDCSK system. Thus, the data rate and BE remain the same as the MCS-MDCSK system.

In the proposed M-ary DLCSK scheme, the number of $log_2 M$ can be transmitted in each symbol duration. Thus, its data rate is similar to the M-ary DCSK system. The current form of the M-ary DLCSK does not use delay lines. Moreover, since only one M-ary symbol is transmitted in each time slot, it needs only β samples for each symbol transmission. Note that the M-ary DCSK system transmits $M \cdot \beta$ samples for each M-ary symbol. Thus, the BE of the M-ary DLCSK is M times larger than the BE of an M-ary DCSK system. Based on the above discussion, the data rate and BE of the M-ary DLCSK system can be written as $R_4 = log_2 M$ and $BE_4 = (log_2 M)/\beta$, respectively. The data rate and BE of M-ary DLCSK system increase with the increment of M. The BE comparisons are presented in Table 4, where all systems have the same spread factor β.

TABLE 4. Data rate and bandwidth efficiency of the *M*-ary DCSK, MCS-DCSK, MCS-MDCSK, LSTM-aided DNN, and *M*-ary DLCSK systems.

System	Data Rate	Bandwidth Efficiency
M-ary DCSK [42]	$log_2 M$	$\frac{log_2 M}{M \cdot \beta}$
MCS-DCSK [43]	$\mathcal{H} - 2$	$\frac{\mathcal{H} - 2}{M \cdot \beta}$
MCS-MDCSK [44]	$(\frac{\mathcal{H}-2}{2})log_2 M$	$(\frac{\mathcal{H}-2}{2})\frac{log_2 M}{M \cdot \beta}$
LSTM-aided DNN [15]	$(\frac{\mathcal{H}-2}{2})log_2 M$	$(\frac{\mathcal{H}-2}{2})\frac{log_2 M}{M \cdot \beta}$
M-ary DLCSK	$log_2 M$	$\frac{log_2 M}{\beta}$

V. SIMULATION RESULTS

In this section, we perform a comparison between the BER performance of the *M*-ary DLCSK and chaos-based modulation benchmarks under AWGN and multi-path fading channels. We also compute the complexity costs of the suggested NN models, i.e., the BiLSTM and CNN models.

A. SIMULATION SETUP

For all the used chaotic maps, presented in Table 1, the bifurcation parameters and initial values are chosen in such a way that the generated signals show chaotic properties. Thus, the generated signals can be used in secure chaos-based communication systems. To generate signals by the Chebyshev map, Initial state is set to $g_1 = 0.35$. For other maps, initial state is set to $g_1 = 0.9$. In order to generate different training and testing datasets, the total number of 2×10^6 chaotic samples (4×10^4 signals of length $\beta = 50$) are first generated using the chaotic generator functions, presented in Table 1. In all experiments, we use fixed-length signals, i.e., $\beta = 50$. The generated dataset is then partitioned into the training set $\mathcal{D}_S$ (50% of the total samples) and the test set $\mathcal{D}_T$ (50% of the total samples). Pre-defined functions of the MATLAB Neural Network Toolbox are utilized to implement the classifiers and define training options. The learning rate, the number of epochs, and the mini-batch size are set to $\eta = 0.01$, $E_p = 12$, and $a_t = 50$, respectively. In this paper, we consider the system's performance under the AWGN, single-path ($L = 1$), and two-path ($L = 2$) Rayleigh fading channels. A single-path fading model with zero delay and an average path gain of 0 dB is considered. In addition, a two-path channel, with identical average power gains, is considered. The average power gain in each path is assumed to be 0.5, (i.e., $E(\alpha_1^2) = E(\alpha_2^2) = 0.5$), with $\tau_1 = 0$, and $\tau_2 = 2$. We assume that the channel coefficients (α_l) and training SNR (σ_{tr}) change after each transmission.
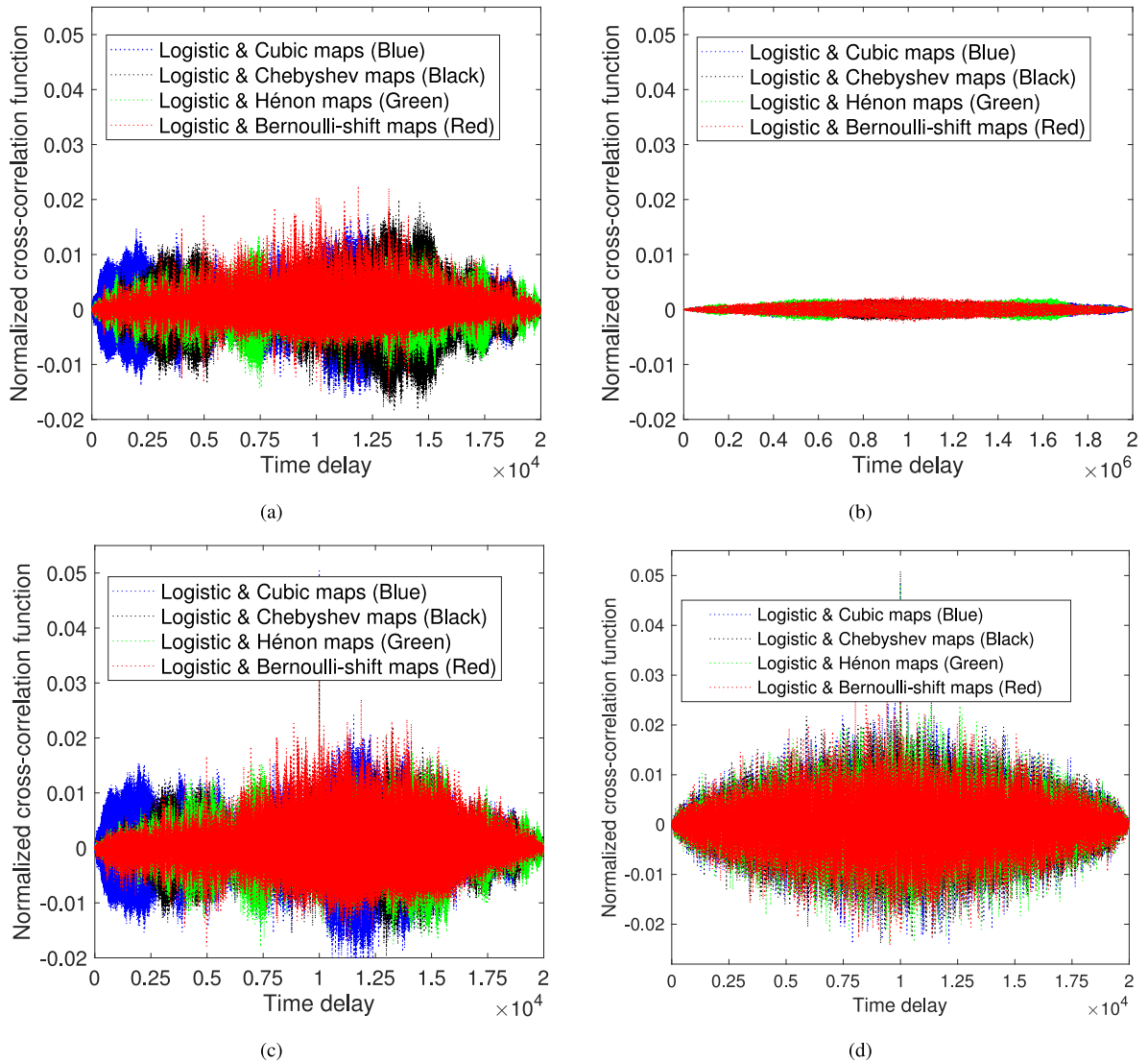
B. CROSS-CORRELATION PROBLEM

Consider the transmitter of a 2-ary DLCSK system. There are two long chaotic signals with the length of 10^4 samples generated by two different chaotic generators. In each transmission, the transmitter selects one of these maps based

on the current data symbol and transmits a chaotic signal with the length of $\beta = 50$ samples. In other words, if the data symbol "1" is to be sent, a signal generated by the first generator, and if the symbol "2" is to be sent, a signal generated by the second generator is transmitted. Since the investigation of short chaotic signals involves checking a huge number of possible combinations between the chaotic signals, we consider the correlation properties of the long chaotic signals. Figure 5 (a) indicates the normalized cross-correlation between a signal generated by the Logistic map and four chaotic signals obtained by different chaotic generators, i.e., the Chebyshev, Bernoulli shift, Cubic, and Hénon maps. All chaotic signals have the length of 10^4 samples. The cross-correlation function measures the similarity between a sequence and shifted copies of another sequence as a function of the time delay (shift). Since the cross-correlation value is highly sensitive to the length of the chaotic signals, the question that arises is how many samples of the chaotic signals should be considered when we want to select a set of proper maps. For example, as shown in Figure 5 (b), when the length of the signals increases from 10^4 to 10^6 samples, the maximum absolute value of the cross-correlation between the chaotic signals increases considerably. Figure 5 (c) presents the normalized cross-correlation of the noisy versions of the chaotic signals. When $SNR = 10$dB, the cross-correlation values are different with the free noise case. We can conclude that selecting the proper chaotic maps based on the cross-correlation values, without considering the channel conditions, may lead to considerable performance degradation. The BER performance investigation of the 2-ary DLCSK system, presented in Figure 6, verifies that adopting improper chaotic maps results in remarkable performance degradation. In Figure 6, we use different pairs of chaotic maps to inspect the effects of chaotic map selection on the BER performance of the 2-ary DLCSK system. When $E_b/N_0 > 16$dB, selecting the Logistic and Cubic maps results in a better BER performance compared to the case where we use the Logistic and Chebyshev maps. On the other hand, when $E_b/N_0 < 16$dB, selecting the Logistic and Cubic maps leads to better performance compared to the Logistic and Chebyshev maps. Similarly, ignoring the effects of practical channels, such as fading correlation, may lead to considerable performance degradation. Thus, instead of considering cross-correlation values, we introduce a map selection algorithm that chooses the chaotic maps based on the end-to-end performance of the *M*-ary DLCSK system.

C. SELECTED MAPS

In the first example, we consider the map selection problem for a typical 2-ary DLCSK system. The data set is created using two long chaotic signals with the length of 10^4 samples. Based on the current data symbol, the transmitter chooses one of these maps and transmits a signal with the length of β. Although the Gram-Schmidt method is used in this paper to orthogonalize these long signals, the results show that a level of cross-correlation remains between the

FIGURE 5. Normalized cross-correlation values between the Logistic map and the Chebyshev, Bernoulli shift, Cubic, and Hénon maps using (a) $S = 10^4$ free-noise samples, (b) $S = 10^6$ free-noise samples, (c) $S = 10^4$ noisy samples ($SNR = 10$dB), (d) $S = 10^4$ noisy samples ($SNR = 10$dB) degraded by a single-path Rayleigh fading channel.

transmitted signals. Our map selection algorithm is based on the end-to-end performance of the system, which does not take the cross-correlation values into account. The results show that by evaluating the merits of different chaotic maps, we can obtain a set of quasi-orthogonal signals. In the first example, we would like to select two proper maps ($M = 2$) from a larger set consisting of five chaotic maps ($B = 5$). In other words, $\mathcal{D}_{sel} = \{g_i^{(n)}(t)\}_{n=1}^N$, $i \in \{1, 2\}$, and $\mathcal{D}_{all} = \{g_m^{(n)}(t)\}_{n=1}^N$, $m \in \{1, 2, \ldots, 5\}$. The Logistic, Chebyshev, Bernoulli shift, Cubic, and Hénon maps are considered as $\mathcal{D}_{all}$. Thus, the number of possible pairs is $\binom{5}{2} = 10$.

Figure 6 depicts the BER curves of four pairs from all ten possible pairs, i.e., a case in which the Logistic map is utilized along with the Chebyshev, Bernoulli shift, Cubic, and Hénon maps. The obtained results indicate that choosing the Bernoulli shift and Logistic functions leads to

better end-to-end BER performance under AWGN channels. Figure 6 also shows the theoretical lower bounds for the BER performances of M-ary DLCSK (20), for $M = 2$, 4, 8. As shown, the BER of the chaotic switching CSK and the calculated lower bound (for M=2) are almost identical. This verifies that this lower bound is an acceptable approximation for an M-ary DLCSK system with a single convolutional layer. Compared to the Chaotic switching CSK, the BER performance of 2-ary DLCSK degrades under the AWGN channels. The main reason lies in that the BER performance of the chaotic switching CSK is obtained assuming perfect synchronization, whereas the performance obtained by the 2-ary DLCSK only uses a limited set of training samples for the basis function recovery.

In Figure 7, we study the role of the confusion matrix in the proposed map selection algorithm. In this example, to construct a 4-ary DLCSK system, we select $M = 4$
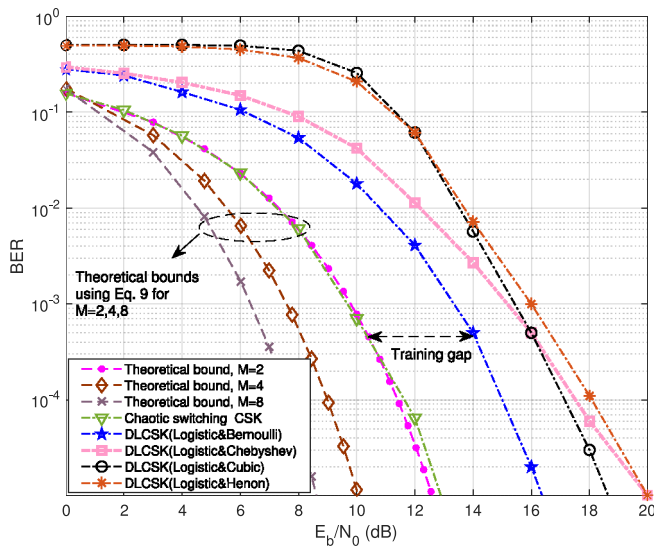
FIGURE 6. BER performance of 2-ary DLCSK system using four different pairs of chaotic maps for $\sigma_{tr}^{(n)} \in [15, 19]$dB.
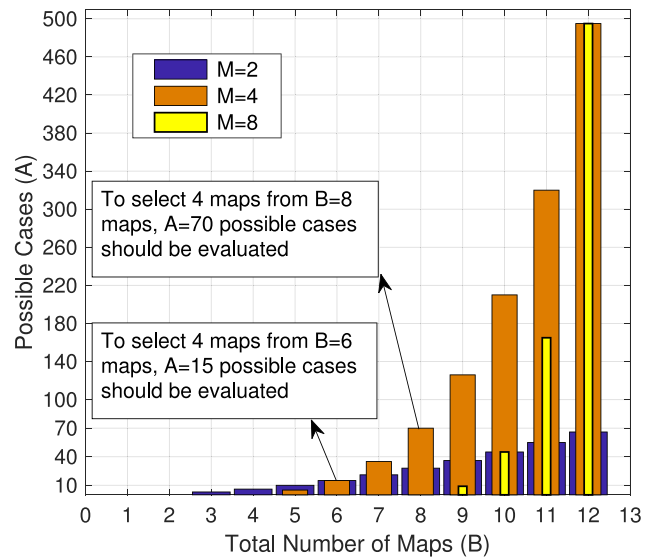

FIGURE 8. The effects of increasing the size of set $\mathcal{D}_{all}$ (i.e., B) on the number of possible cases A, for $M = 2, 4, 8$.

FIGURE 7. An example of the confusion matrix of a 4-ary DLCSK system, tested using 2×10^4 symbols, for $E_b/N_0 = 18$dB.

maps from the total $B = 6$ maps. To select four maps from the six maps, $A = \binom{6}{4} = 15$ different cases (different sets of maps) are possible. We obtained all 15 different confusion matrices and recorded the number of incorrectly classified symbols ϕ_a, $a = \{1, 2, \ldots, 15\}$, for these 15 confusion matrices. Finally, the case which results in a lower ϕ_a, $a = \{1, 2, \ldots, 15\}$, is selected. Figure 7 depicts one of 15 possible confusion matrices of the 4-ary DLCSK system, for test $E_b/N_0 = 18$dB. The columns display the predicted maps/symbols, and the rows stand for the true maps/symbols. In this example, using 2×10^4 test symbols (10^6 samples), the calculated metric is $\phi_1 = 23$ which is equal to the sum of incorrectly classified symbols. The confusion matrix also helps in inspecting the merits of chaotic maps and making

design decisions. Using the proposed map selection algorithm, the Logistic, Chebyshev, Bernoulli shift, and PAM maps are selected for the transmission of a symbol set for $M = 4$.

Generally, when the modulation order (M) is given, the probability of finding a fully orthogonal signal set increases by choosing a larger B. Consequently, with a larger B, the probability of obtaining an improved BER performance increases. However, when a larger B is chosen, the number of possible cases (A) to be evaluated also increases exponentially. As shown in Figure 8, by choosing a larger B, the number of possible cases and the complexity of the map selection algorithm augments.

D. THE EFFECTS OF TRAINING SNR

For both the training and deployment of the receiver, the training SNR (σ_{tr}) is a key parameter to determine the performances of the NN-based receiver. In Figure 9, we study the effects of training on a relatively wide range of SNR values. As shown in Figure 9, training on relatively lower SNRs ($\sigma_{tr}^{(n)} \in [11, 15]$dB), leads to a lower BER for the low-SNR conditions (i.e., when $E_b/N_0 < 14$dB). Since the training process is performed under different channel conditions, NN can indirectly estimate noise distribution. However, when $\sigma_{tr}^{(n)} \in [11, 15]$dB, the BER is high for $E_b/N_0 \geq 14$dB. When $\sigma_{tr}^{(n)} \in [15, 19]$dB, M-ary DLCSK shows a more robust behaviour for $E_b/N_0 \geq 14$dB. When the values of $\sigma_{tr}^{(n)}$ are relatively high, the NN can grasp the clean signals and performs well at higher test SNRs. Summarily, in order to obtain better BER performances in low-noise conditions, the receiver should be trained at higher SNRs. The training options of the proposed scheme enable us to design a receiver with a flexible data rate depending upon specific channel conditions.
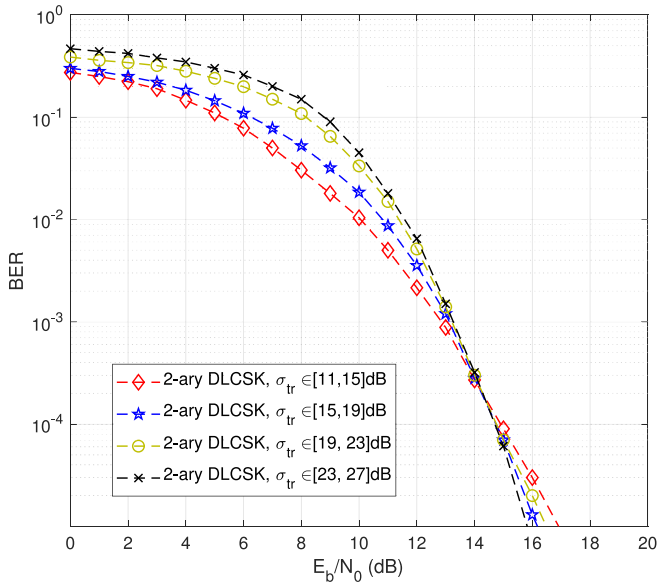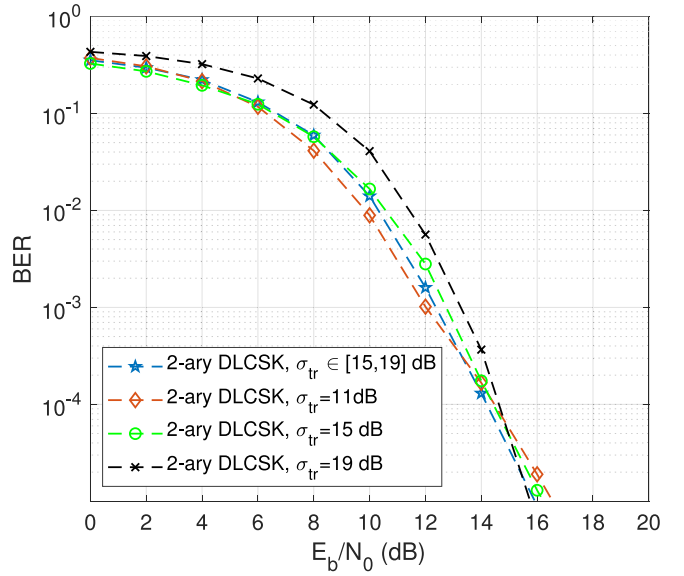
FIGURE 9. The effects of different training SNR ranges on the performance of 2-ary DLCSK system under AWGN channels.

**FIGURE 10.** The effects of training on a wide range of SNRs.

In Figure 10, we address the effect of using the TL technique. In the proposed scheme, we train the NN with stochastic SNRs to achieve optimal BER performance over a wide range of SNRs. If we train NN with fixed SNRs, the problem reduces to a regular DL problem. The proposed TL technique improves the generalization capabilities of the system. Thus, we can employ the trained NN in different SNR conditions without the need for re-training. In Figure 10, in order to further investigate the effect of training SNR, we consider two different scenarios. In the first scenario, both the training and testing processes are performed at fixed SNRs (i.e., $\sigma_{tr}^{(n)} = \sigma_{tr}$). In particular, we set $\sigma_{tr} = 11, 15$, and 19dB. In the second scenario, we assume that the training SNR changes in each transmission. In other words, in this case, the $\sigma_{tr}^{(n)}$ is a random variable such that $\sigma_{tr}^{(n)} \in [15, 19]$dB. For both scenarios, the NN model is tested over a wide range of energy per bit to noise power spectral ratios, i.e., $E_b/N_0 \geq 0$dB. Training on a lower SNR value (consider the case of $\sigma_{tr} = 11$dB) gives a good BER performance for $E_b/N_0 \leq 12$dB. However, in this case, the BER is high for $E_b/N_0 \geq 12$dB. Generally, the results indicate that the NN trained over a relatively wide range of SNRs is suitable for a large range of test SNRs. For example, when $\sigma_{tr}^{(n)} \in [15, 19]$dB, there is up to 2dB improvement compared to the case where the NN is trained at the fixed SNR $\sigma_{tr} = 19$dB.

E. PERFORMANCE-VERSUS-COMPLEXITY ANALYSIS

Figure 11 depicts the effects of the n_k (filter size or kernel size) on the BER performance of the CNN-based 2-ary DLCSK under AWGN channels. In this experiment the NN is trained on $\sigma_{tr}^{(n)} \in [15, 19]$dB and tested for $E_b/N_0 = \{6, 8, 10, 12\}$dB. Assuming a certain number of
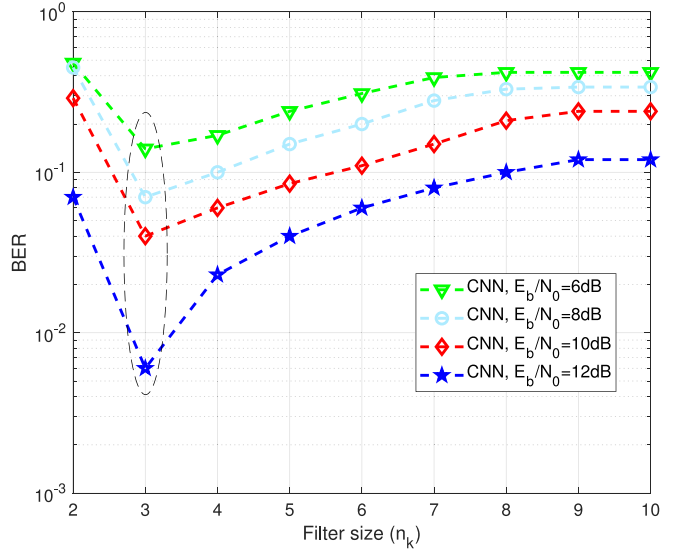

FIGURE 11. The effect of the filter size (n_k) on the BER performance of the CNN-based 2-ary DLCSK under AWGN channels, for $\sigma_{tr}^{(n)} \in [15, 19]$dB.

convolutional layers (four layers), and a certain number of filters (i.e., $n_f = 256$), we tested different values of n_k ($2 \leq n_k \leq 10$) and found that $n_k = 3$ results in the best BER performance. Therefore, in the rest of this paper, we assume a constant filter size of $n_k = 3$ and only investigate the effects of the number of filters (n_f) on the complexity and BER performance of the system. Table 5 presents an example of the calculated complexity values (C) and output dimensions (output dim.) for $n_i = \{1, 2\}$ and $n_s = 50$. In this example, we consider $n_f = 512$ and $n_h = 25$ for the BiLSTM and CNN models, respectively. With $n_i = 1$, $n_f = 512$, and $n_h = 25$, both the BiLSTM and CNN models have the same computational complexity order, i.e., 2.7×10^5 RMpS. In this way, we can manage the computational complexity of the used NNs.

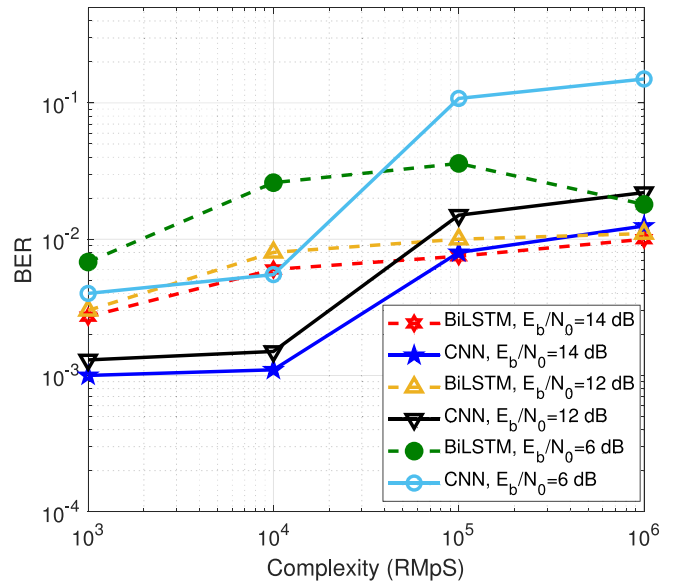**TABLE 5.** Calculated computational complexity with $n_i = \{1, 2\}$, $n_s = 50$, $n_f = 512$, and $n_h = 25$.

| NN	Layers	$n_i = 1$		$n_i = 2$	
		output dim.	C	output dim.	C
BiLSTM	Sequence input	50	0	100	0
	BiLSTM	50	270000	100	285000
	FC	2	100	2	100
	Classification	2	4	2	4
CNN	Sequence input	50	0	100	0
	1st conv2d	48	73728	98	301056
	2nd conv2d	46	70656	96	294912
	3rd conv2d	44	67584	94	288768
	4th conv2d	42	64512	92	282624
	FC	2	84	2	184
	Classification	2	4	2	4
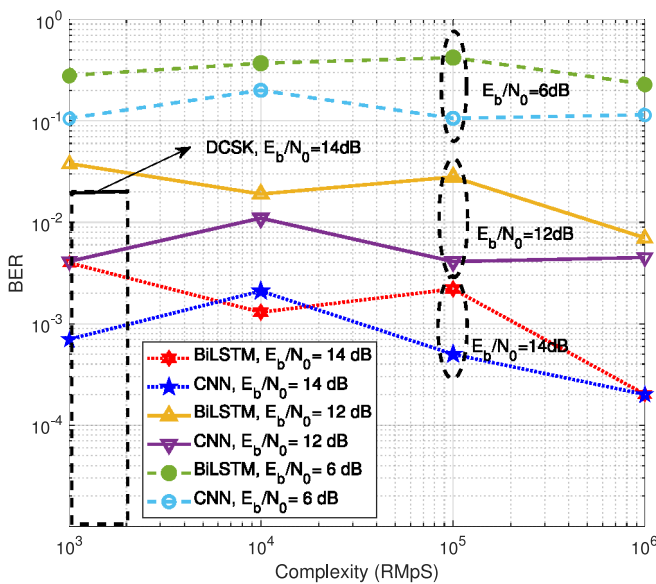
FIGURE 12. BER performances of the 2-ary DLCSK system under AWGN channels, for complexity values in the $[10^3, 10^7]$ RMpS range.
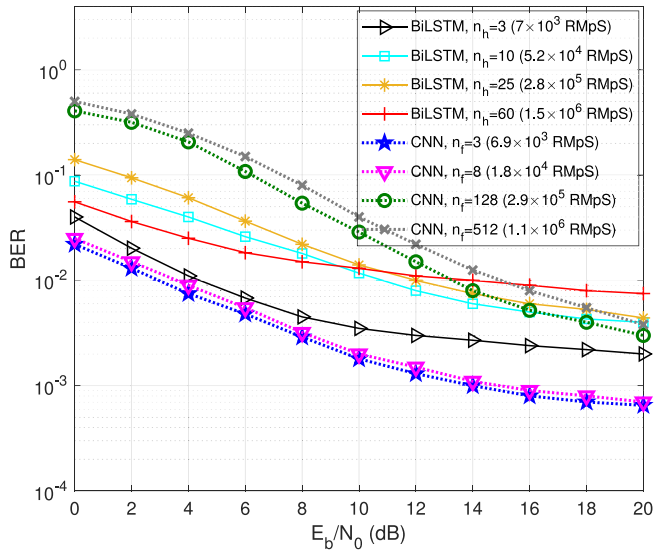

FIGURE 13. BER performances of the 2-ary DLCSK system under two-path Rayleigh fading channels, for complexity values in the $[10^3, 10^7]$ RMpS range.
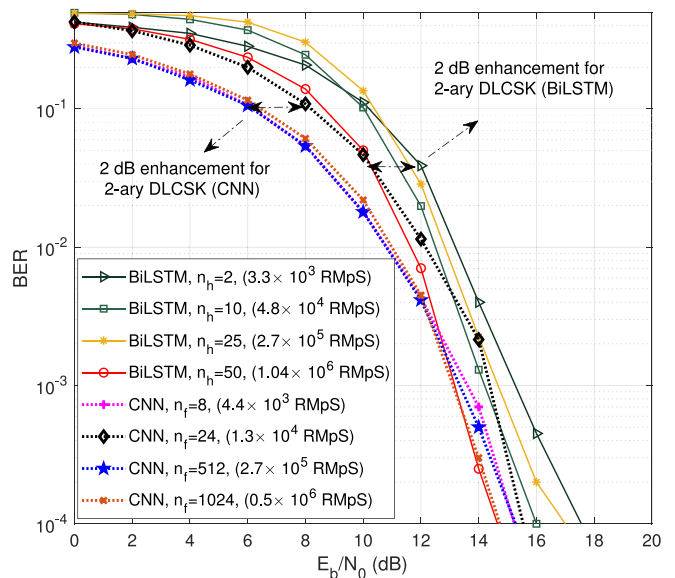
FIGURE 14. BER performances of the 2-ary DLCSK system under AWGN channels, for different NN models and different complexity levels.

Figure 12 compares the BER performances of the suggested NN models under similar computational complexity conditions. In this simulation, the 2-ary DLCSK system is trained with $\sigma_{tr}^{(n)} \in [15, 19]$dB, and tested at the target $E_b/N_0 = 6, 12, 14$dB. This simulation investigates which NN model rovides better performance under AWGN channels if we restrict the complexity to a certain level. We consider the complexity values in the $[10^3, 10^6]$ RMpS range. This range guarantees that the hardware implementation of the receivers is feasible using Field-Programmable Gate Array (FPGA) devices [40]. The results show that the BER performance changes with the changes of n_h and n_f. Although these changes are not exactly predictable, it is observable that the CNN model leads to a better BER performance than the BiLSTM model when the computational complexity is less than 10^6 RMpS.

Figure 13 performs a BER performances-versus-complexity study for the suggested NNs two-path Rayleigh

fading channels. In this experiment, the 2-ary DLCSK system is trained with $\sigma_{tr}^{(n)} \in [15, 19]$dB, and tested at the target $E_b/N_0 = 6, 12, 14$dB. Unlike the case of the AWGN channels, the performance of the CNN model deteriorates when n_f increases. Moreover, the CNN model performs better than the BiLSTM model when the computational complexity is less than 10^5 RMpS.

Figure 14 compares the performances of the suggested NN models for different complexity conditions under AWGN channels. We consider the complexity values in the $[10^3, 10^6]$ RMpS range. The BER performance obviously improves by increasing n_h and n_f. When $E_b/N_0 < 12$dB, the best
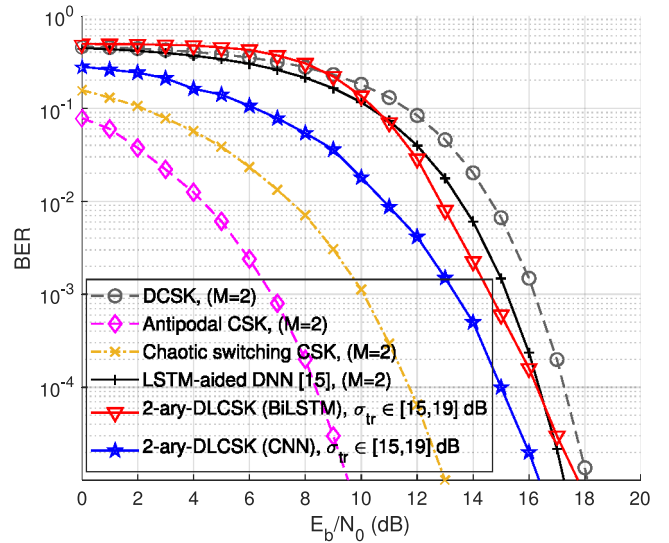
FIGURE 15. BER performances of the 2-ary DLCSK system under two-path Rayleigh fading channels, for different NN models, and different complexity levels.

FIGURE 16. BER curves of the DCSK, antipodal CSK, chaotic switching CSK, LSTM-aided DNN (with $M=2$, $\mathcal{N}=1$), and 2-ary DLCSK (for $\sigma_{tr}^{(n)} \in [15, 19]$dB) under AWGN channels.

performance for 2-ary DLCSK is provided by the CNN model, with a computational complexity of about 10^5 RMpS (i.e., $n_f = 512$). Generally, increasing the complexity level from 10^3 to 10^6 RMpS leads to about 2dB BER performance improvement. When $E_b/N_0 > 12$dB, increasing the number of filters up to $n_f = 1024$ leads to the best BER performance.

Figure 15 analyzes the BER performances of the 2-ary DLCSK system for different complexity conditions under two-path Rayleigh fading channels. Unlike the case of AWGN channels, the performance of 2-ary DLCSK system degrades by increasing n_h and n_f. The best performance for 2-ary DLCSK is provided by the CNN model, with a computational complexity of about 10^3 RMpS (i.e., $n_f = 3$). However, we found that for higher modulation orders, i.e., $M \geq 4$, the best performance for M-ary DLCSK is provided by the CNN model with $n_f = 8$.

F. BER PERFORMANCE OF M-ARY DLCSK IN AWGN CHANNELS

Figure 16 compares the BER performances of the non-coherent DCSK, antipodal CSK, chaotic switching CSK, LSTM-aided DNN receiver [15], and 2-ary DLCSK systems under AWGN channels, for $\sigma_{tr}^{(n)} \in [15, 19]$dB. An antipodal CSK scheme only uses one chaotic basis function, and can theoretically reach the noise performance of BPSK. However, owing to the cross-correlation between chaotic signals and the problem of basis function recovery, this performance cannot be achieved practically. The chaotic switching CSK is a coherent CSK scheme with two basis functions that can theoretically reach the BER performance of the Frequency Shift Keying (FSK) modulation scheme under AWGN channels. This level of BER performance can be achieved only if the cross-correlation problem is solved and the basis functions can be recovered at the demodulator successfully. If it is not possible to recover the basis functions, the DCSK

scheme may provide better performance. However, the BER performance of non-coherent receivers depends on the channel bandwidth. Further, the transmission of the reference signals increases the overhead, errors, and complexity of the system.

Note that the simulated antipodal CSK and chaotic switching CSK systems assume exact synchronization of the chaotic sequences, and are only presented as a benchmark for comparison purposes. We tested different training SNR intervals as $[\sigma_{tr,min}, \sigma_{tr,max}]$dB, and found that the system shows a relatively better BER performance when $\sigma_{tr}^{(n)} \in [15, 19]$dB. From the energy efficiency point of view, when the receiver is previously trained using chaotic references, we do not need to transmit (or regenerate) the reference signals, which results in lower energy consumption. By exploiting the optimization and classification capabilities of NNs, the 2-ary DLCSK system attains an improved BER performance compared to the DCSK system. As shown, the proposed receiver obtains a BER gain of 0.2 compared to the conventional non-coherent DCSK system at $E_b/N_0 = 16$dB.

Figure 16 also compares the simulated BER performances of the 2-ary DLCSK and LSTM-aided DNN systems [15]. Note that in this experiment, the computational complexity of the CNN and LSTM-based DLCSK receivers is 2.70×10^5 RMpS, and the computational complexity of the LSTM-aided DNN receiver is 2×10^5 RMpS. In the LSTM-aided DNN system, the orthogonality is satisfied via the Walsh codes to construct an orthogonal signal set, whereas M-ary DLCSK produces a quasi-orthogonal signal set through the map selection algorithm. In the LSTM-aided DNN system, the number of M-PSK symbols transmitted in a symbol duration is shown by $\mathcal{N}$. For $M = 2$, the number of the Walsh-coded signals should be greater or equal to $2\mathcal{N} + 1 = 3$, i.e., one Walsh-coded signal is used as the reference signal, and two Walsh codes are used for
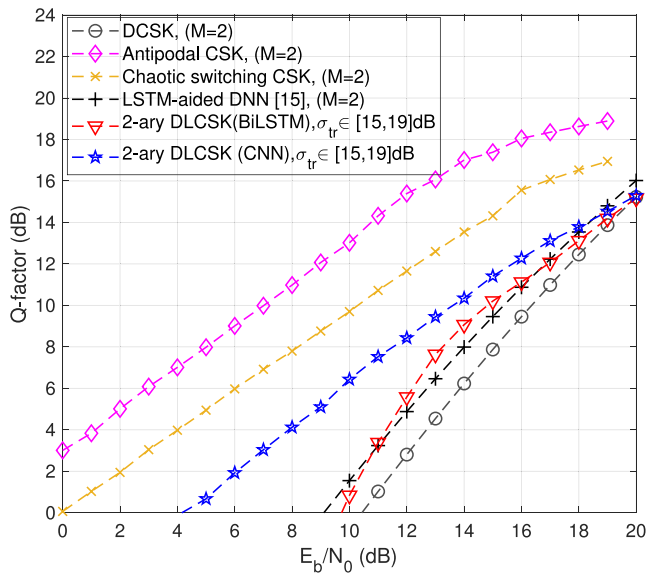
FIGURE 17. Q-factor levels of the DCSK, antipodal CSK, chaotic switching CSK, LSTM-aided DNN (with M=2, $\mathcal{N} = 1$), and 2-ary DLCSK (for $\sigma_{tr}^{(n)} \in$ [15, 19]dB) under AWGN channels.

FIGURE 18. BER performances of M-ary DLCSK (with $\sigma_{tr}^{(n)} \in$ [15, 19]dB) and LSTM-aided DNN [15] systems under AWGN channels, for $M = 2, 4, 8$.

spreading the real and imaginary parts of the PSK symbols. For a fair comparison, it is assumed that both systems transmit one symbol in each transmission (i.e., $\mathcal{N} = 1$). As indicated, the 2-ary DLCSK system provides a considerable performance advantage compared to the LSTM-aided DNN receiver. For example, the performance of 2-ary DLCSK shows a gain of 0.1 over the LSTM-aided DNN system at $E_b/N_0 = 16$dB. The main reason lies in that the 2-ary DLCSK system uses the previously trained NN for the chaotic basis functions recovery, whereas the LSTM-aided DNN system transmits the references over the AWGN channel. Based on the results obtained in Figure 16, the 2-ary DLCSK system offers the best performance among the simulated modulation schemes in terms of the E_b/N_0B. For example, at the target BER 10^{-3}, the CNN-based receiver shows a 3dB E_b/N_0 improvement compared to the conventional DCSK, 2dB compared to the LSTM-aided DNN, and 1.5dB compared to the BiLSTM-based receiver.

In Figure 17, assuming an AWGN channel, the performances of DCSK, antipodal CSK, chaotic switching CSK, LSTM-aided DNN (with M=2, $\mathcal{N} = 1$), and 2-ary DLCSK (for $\sigma_{tr}^{(n)} \in$ [15, 19]dB) are compared in terms of the Q-factor level. The Q-factor can be calculated by $Q = 20log_{10}(\sqrt{2}\ erfc^{-1}(2 \times BER))$. At the target $E_b/N_0 = 14$dB, the CNN-based receiver shows a 4dB Q-factor gain compared to the conventional DCSK, 2dB compared to the LSTM-aided DNN, and 1dB compared to the BiLSTM-based receiver.

In Figure 18, we present a performance comparison between M-ary DLCSK and LSTM-aided DNN systems under AWGN channels, for $M = 2, 4, 8$. In the LSTM-aided DNN system, $\mathcal{N}$ represents the number of M-PSK symbols transmitted in a symbol duration and its maximum value is $\mathcal{N}_{max} = \frac{\mathcal{H}-2}{2}$, where $\mathcal{H}$ is the number of Walsh
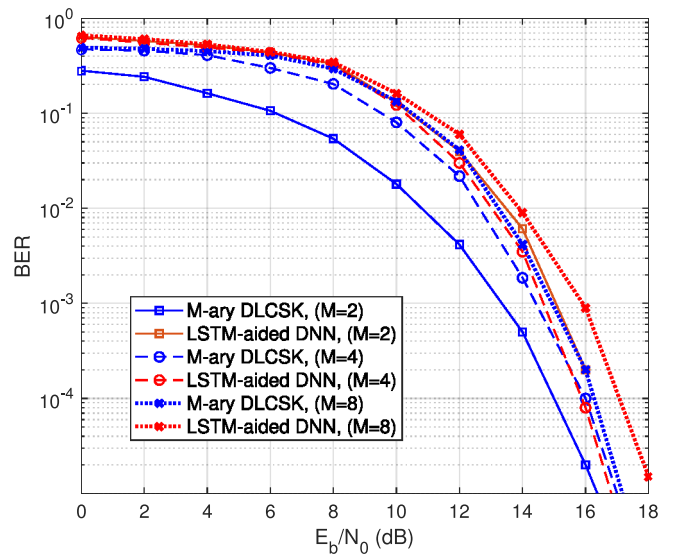
code sequences. For example, when $M = 4$, using five Walsh-coded signals, the maximum number of transmitted PSK symbols is $\mathcal{N}_{max} = 1$. Thus, for a fair comparison, we assume that both the LSTM-aided DNN and 4-ary DLCSK systems transmit one symbol in each symbol duration. First, consider the BER curves of the LSTM-aided DNN systems. When $M = 4$, the LSTM-aided DNN achieves better BER performance than the case of $M = 2$. This is due to the fact that with $M = 4$, this system can be considered as a combination of two 2-ary modulation (i.e., DCSK). However, unlike the case of DCSK, the LSTM-aided DNN system only transmits one reference signal for the demodulation of two information-bearing signals. In this way, the overhead per transmitted bit reduces and the BER performance improves. The BER performance of the LSTM-aided DNN system, except in the above special case, deteriorates when M increases. The main reason lies in that the decision boundaries of M-ary constellation are decreasing when M increases. In the proposed M-ary DLCSK system, the BER performance always deteriorates when M increase. Compared to LSTM-aided DNN, the M-ary DLCSK shows better performance for all values of M. The performance of the M-ary DLCSK system deteriorates with the increase of M. Notably, since the proposed system does not transmit any reference, the value of the overhead per transmitted bit is fixed, and it has no role in performance deterioration. A level of cross-correlation remains between the chaotic signals, while higher-order modulation reduces the Euclidean distance between adjacent symbols, thereby leading to performance degradation. In this experiment, using the proposed map selection algorithm, the Logistic, M-PAM, Tent, Bernoulli shift, Circle, Iterative, Tent-like, and PAM maps are selected for the transmission of a symbol set for $M = 8$. As previously discussed, choosing proper chaotic maps has a considerable impact on the value of this performance degradation.
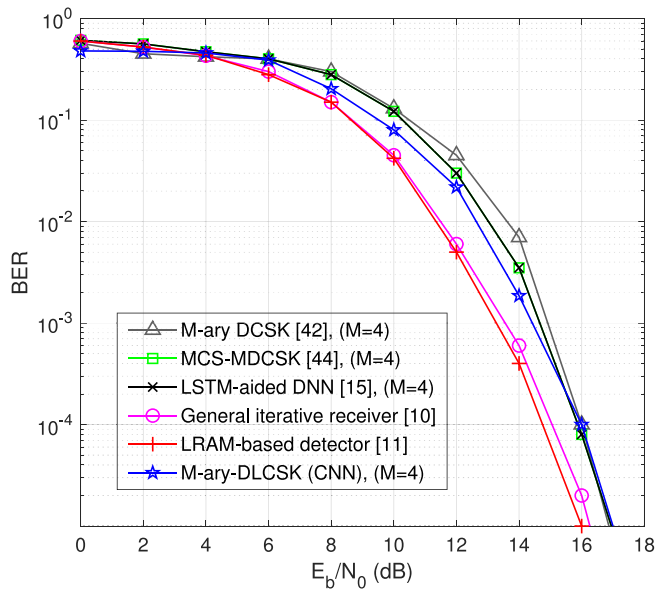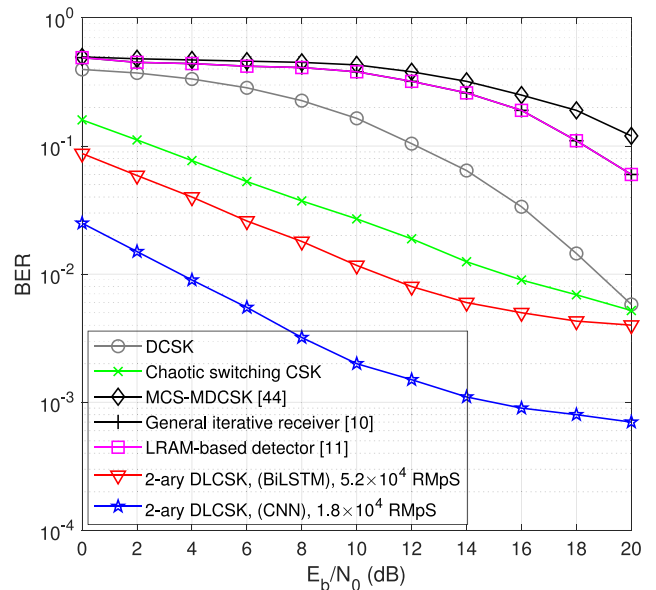
FIGURE 19. BER performances of 4-ary DLCSK (with $\sigma_{tr}^{(n)} \in [15, 19]$dB) and chaos-based modulation benchmarks under AWGN channels.

FIGURE 20. BER performances of 2-ary DLCSK (with $\sigma_{tr}^{(n)} \in [15, 19]$dB) and chaos-based modulation benchmarks under single-path Rayleigh fading channels.

Figure 19 depicts the BER performances of 4-ary DLCSK, 4-ary DCSK [42], LSTM-aided DNN receivers (with $M = 4$ and $\mathcal{N} = 1$) [15], general iterative receiver [10], and LRAM-based detector [11] under AWGN channels. As shown, the 4-ary DLCSK system achieves a slightly better BER performance than the LSTM-aided DNN system under AWGN channels. The main reason is that the reference signals at the M-ary DLCSK are obtained by the training, instead of the transmission over the channel. Moreover, since the M-ary DLCSK is trained over a wider range of SNRs, it shows a more robust behavior compared to its DL-based counterpart. We can notice that the BER performances of the general iterative receiver [10] and the LRAM-based detector [11] are better than the proposed system under AWGN channels. The reason is that these systems use additional noise reduction mechanisms. For example, an iterative receiver evaluates the reference signals and updates them for the next iteration. In this way, the SNR of reference signals and obtained performance can be improved in the iterative receiver. On the other hand, the complexity of LRAM-based detectors and iterative receivers is very high since additional noise reduction operations are employed for these systems. In this simulation, the number of iterations in the general iterative receiver is set to 100 to ensure the convergence of the iteration.

G. BER PERFORMANCE OF M-ARY DLCSK IN MULTI-PATH FADING CHANNELS

Figure 20 presents the BER performances of the 2-ary DLCSK, non-coherent DCSK, chaotic switching CSK, MCS-MDCSK [44], general iterative receiver [10], and LRAM-based detector [11] under single-path Rayleigh fading channels. Unlike the case of AWGN channels, 2-ary DLCSK shows a considerable BER enhancement under

fading channels. The main reason lies in that the M-ary DLCSK receiver can demodulate received signals without the transmission of reference signals through a destructive environment. Consider the MCS-MDCSK system, where the reference signals are transmitted over the channel with the goal of basis function recovery. Since the transmitted reference is corrupted by noise and fading, the receiver cannot properly demodulate information-bearing signals. As shown, since the destructive effects of the fading channels, such as fading correlation, are more than the destructive effects of the AWGN channels, the decrease in performance is also more obvious. Unlike the case of AWGN channels, the BER performances of the LRAM-based detectors and iterative receivers degrade dramatically under fading channels. For example, the BER performance of 2-ary DLCSK shows a gain of 0.2 over the general iterative receiver and LRAM-based detector at $E_b/N_0 = 18dB$. In addition, Figure 20 shows that the 2-ary DLCSK system provides a substantial performance advantage compared to the theoretical performance of the chaotic switching CSK [5]. The reason is that the 2-ary DLCSK design has powerful learning and feature extraction capabilities. Moreover, since the receiver is trained under channel variations and different SNR conditions, it shows robust behavior against the destructive effects of the channels.

In Figure 21, we study the effects of training SNR on the system's BER performance under the single-path ($L = 1$) and two-path ($L = 2$) Rayleigh fading channels. We assumed that the channel coefficients $\alpha_l^{(n)}$ and training SNR ($\sigma_{tr}^{(n)}$) change after each transmission. In the case of two-path channels, the best BER performance for M-ary DLCSK is provided by $\sigma_{tr}^{(n)} \in [15, 19]$dB. In the case of two-path channels, when the receiver is trained at ($\sigma_{tr}^{(n)} \in [11, 15]$dB), M-ary DLCSK
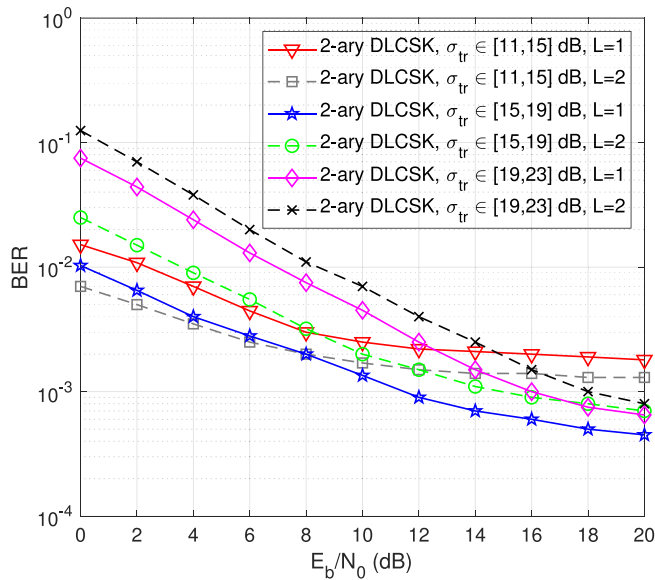
FIGURE 21. BER performances of 2-ary DLCSK system under Rayleigh fading channels, for different training SNR ranges and different number of paths (L = 1, 2).
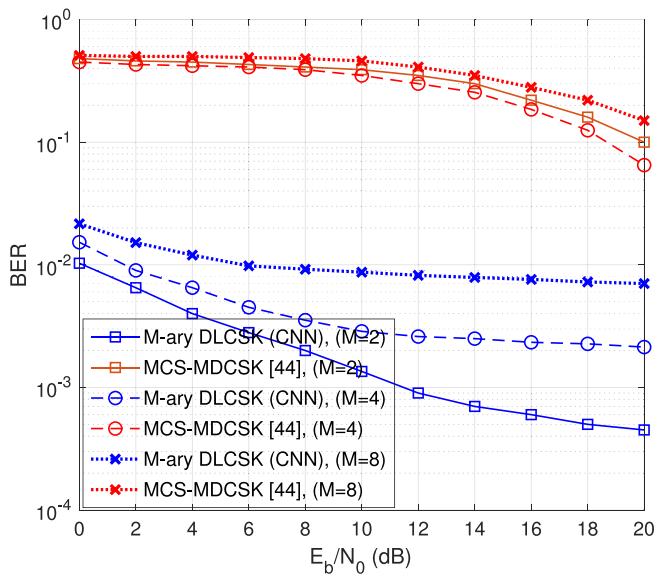

FIGURE 22. BER performances of *M*-ary DLCSK and MCS-MDCSK systems under single-path Rayleigh fading channels, for *M* = 2, 4, 8.

reaches to the best results for $E_b/N_0 < 12$dB. However, for $E_b/N_0 > 12$dB, selecting $\sigma_{tr}^{(n)} \in [15, 19]$dB results in a more robust behavior.

In Figure 22, we make a performance comparison between *M*-ary DLCSK and MCS-MDCSK systems over single-path Rayleigh fading channels, for *M* = 2, 4, 8. When *M* = 2, the MCS-MDCSK system transmits one reference signal for the demodulation of one information-bearing signal. When *M* = 4, since the MCS-MDCSK system transmits one reference signal for the demodulation of all information-bearing signals, the overhead per transmitted bit is reduced. The MCS-MDCSK system reaches its best performance with *M* = 4. For *M* > 4, the BER performance of the MCS-MDCSK system degrades since the decision boundaries of
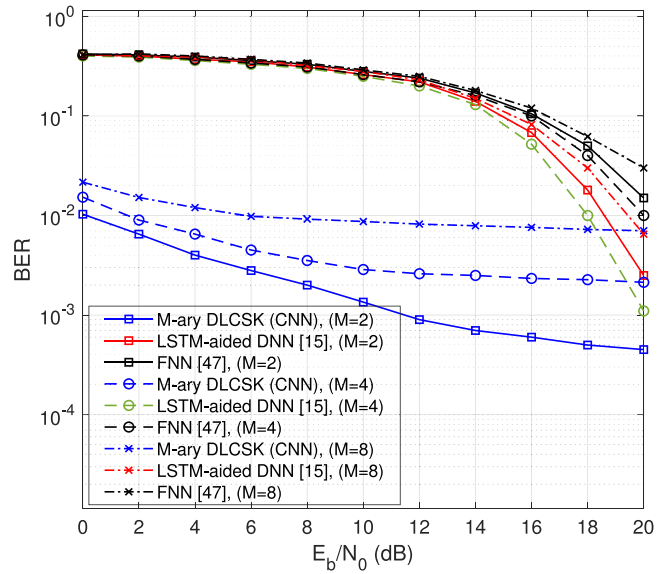

FIGURE 23. BER performances of *M*-ary DLCSK, LSTM-aided DNN [15], and FNN [47] systems under sinle-path Rayleigh fading channels, for *M* = 2, 4, 8.

TABLE 6. Training settings of the system used in Figure 23 (for *M* = 4).

Settings	FNN	LSTM-aided DNN	*M*-ary DLCSK
Transfer learning	No	No	Yes
Complexity	1.75×10^4	1.76×10^4	1.9×10^4
Training time (s)	198	272	321

the constellation are decreasing. The preferable performance of the MCS-MDCSK system under multi-path fading channels occurs in medium modulation orders (i.e., *M* = 4 or 8). Compared to the MCS-MDCSK system, the *M*-ary DLCSK can exploit the capabilities of the NNs to obtain an improved BER performance. However, since higher-order modulation shortens the Euclidean distance between adjacent symbols, the BER performance worsens when *M* increases. Moreover, choosing improper chaotic maps may lead to considerable performance degradation in the *M*-ary DLCSK systems.

In Figure 23, we make a performance comparison between *M*-ary DLCSK system and its DL-based counterparts, i.e., the LSTM-aided DNN, and FNN receiver [47], under single-path Rayleigh fading channels, for *M* = 2, 4, 8. The LSTM-aided DNN and FNN systems are almost similar, except that the LSTM layer is replaced with an FC layer in the FNN system. Each input vector of the FC layer in the FNN receiver is a concatenation of the reference signal and one of the information-bearing signals. As indicated in Table 6 for the case of *M* = 4, to make a fair comparison, the parameters, such as the number of convolutional filters and hidden units, are adjusted to make the complexity and training time of the NNs as similar as possible.

Summarily, the *M*-ary DLCSK system obtains an outstanding performance compared to the existing chaos-based benchmarks, such as the LSTM-aided DNN systems. The main reason lies in that the *M*-ary DLCSK system uses the

trained NN for the basis function recovery, instead of the reference transmissions. Note that in the LSTM-aided DNN receiver, a received reference is used for the demodulation of several data-bearing signals. Thus, when a reference signal is degraded, it affects the detection of several data bits. The value of the BER performance improvement in scenarios where the effect of the channel is more impressive, such as the fading channels, is more than the AWGN channels. For example, the BER performance of 8-ary DLCSK shows a gain of 0.1 over the LSTM-aided DNN (with $M = 8$) at the target $E_b/N_0 = 14dB$. Note that the M-ary DLCSK shows a level of error that does not disappear with increasing the SNR. The reason is that a level of cross-correlation remains between the used chaotic signals. Thus, designing new architectures by considering Walsh codes and using other orthogonalization methods can be a promising direction for future research.

Generally, the BER performance and complexity of the algorithms are two fundamental issues for the practical implementation of the NN-based communication systems [40]. The M-ary DLCSK scheme has the potential to be used in practical systems since it provides reasonable BER performances and complexity. Moreover, the M-ary DLCSK system does not require delay lines. Note that most of the traditional DCSK-based schemes, such as M-ary DCSK and MCS-DCSK schemes, require delay lines in the transmitter and receiver, which can be challenging to implement in practical circuits. From the viewpoint of time complexity, the time-consumption of the proposed design is lower than the general iterative receivers [10]. Although the training of NNs is time-consuming, the received signals can be recovered in real-time during the deployment phase [15]. Moreover, several parallelization approaches can be employed to save time costs [54]. Since the computational complexity of the M-ary DLCSK can be adjusted based on the user demands, the proposed receiver can provide a low time consumption. The proposed receiver does not need to change the transmitter structure of the existing M-ary CSK systems. Thus, several important aspects, such as the security aspect, remain the same as the benchmark CSK systems. This feature makes M-ary DLCSK a suitable candidate for secure communication scenarios, such as biomedical signal transmission [55], Optical fiber and Free-Space Optical (FSO) systems [56], [57], [58], Vehicular Ad-hoc Networks (VANETs) [59], Non-Orthogonal Multiple Access (NOMA) [60], and Multiple-Input Multiple-Output (MIMO) schemes [61].

VI. CONCLUSION

This paper designed a novel M-ary DLCSK system with flexible reliability, BE, and complexity. Unlike existing DL-based receivers, which transmit the reference signals over the channel for basis functions recovery, our NN-based receiver recovers the basis functions by exploiting the power of DTL. This paper can be considered one of the first papers that use TL techniques in chaos-based communications.

The presented scheme can be the beginning of different challenges and studies in different fields as outlined below:

• The proposed scheme can be considered for many communication systems and practical applications, such as Quantum classifiers, V2X communications, Massive MIMO systems, Multi-user (MU) scenarios, Multiple-Access (MA) structures, etc., where the proposed technique can be effective. In addition, in order to the performance improvement, various techniques, such as Forward Error Correction (FEC) codes, Multi-carrier (MC) techniques, and power allocation strategies, can be considered.

• The problem of basis function recovery has impeded the theoretical studies and practical applications of coherent chaos-based modulations. The theoretical analysis of the coherent CSK systems is a promising idea for future research. The results of these analyses can be used for different purposes, such as designing resource allocation strategies.

• In this paper, we only considered a given environment, including a number of limited features. Therefore, using state-of-the-art training methods, such as meta-transfer learning, can be effective. Moreover, we assumed that the training SNR is a random variable in a limited SNR range. Considering other scenarios and training on wider ranges of SNR can be investigated in future works.

• Hyper-parameter optimization is crucial to the learning quality, complexity, and convergence rate. Using efficient hyper-parameter optimization methods, such as the Bayesian method or simulated annealing algorithm can yield more reliable results.

• In this paper, we focused on the map selection algorithm to find a pseudo-orthogonal signal set. Future studies could examine several state-of-the-art orthogonalization methods to find strictly orthogonal signal sets. For example, combining with the Walsh codes and using the Gram-Schmidt algorithm may be effective.

• In this paper, in order to reduce the negative effects of the cross-correlation between the chaotic signals, we used a fixed and relatively large spreading factor $\beta = 50$. Further investigation must be carried out to find the optimal value of β. Moreover, a large β may have destructive effects on other criteria, such as the computational complexity. Future research should consider interactions between beta and other qualitative and quantitative measures.

• Using state-of-the-art NN-based approaches, such as attention mechanisms, Multi-Layer Perceptron (MLP), CNN+MLP, ConvLSTMs, and CNN+BiLSTM may improve the results as was demonstrated in the case of other applications, such as equalizers.

REFERENCES

[1] M. Mobini, G. Kaddoum, and M. Herceg, "Design of a SIMO deep learning-based chaos shift keying (DLCSK) communication system," *Sensors*, vol. 22, no. 1, p. 333, 2022.

[2] Y. Xia, C. K. Tse, and F. C. M. Lau, "Performance of differential chaos-shift-keying digital communication systems over a multipath fading channel with delay spread," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 51, no. 12, pp. 680–684, Dec. 2004.

[3] H. Dedieu, M. P. Kennedy, and M. Hasler," "Chaos shift keying: Modulation and demodulation of a chaotic carrier using self-synchronizing Chua's circuits," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 40, no. 10, pp. 634–642, Oct. 1993.

[4] T. Yang, "Recovery of digital signals from chaotic switching," *Int. J. Circuit Theory Appl.*, vol. 23, no. 6, pp. 611–615, 1995.

[5] G. Kolumban, M. P. Kennedy, Z. Jako, and G. Kis, "Chaotic communications with correlator receivers: Theory and performance limits," *Proc. IEEE*, vol. 90, no. 5, pp. 711–732, May 2002.

[6] M. Mobini and G. Kaddoum, "Deep chaos synchronization," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1571–1582, 2020.

[7] G. Kolumbán, B. Vizvári, W. Schwarz, and A. Abel, "Differential chaos shift keying: A robust coding for chaotic communication," in *Proc. NDES*, Seville, Spain, 1996, pp. 87–92.

[8] Z. Galias and G. M. Maggio, "Quadrature chaos shift keying: Theory and performance analysis," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 12, pp. 1510–1519, Dec. 2021.

[9] G. Kaddoum, F.-D. Richardson, and F. Gagnon, "Design and analysis of a multi-carrier differential chaos shift keying communication system," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3281–3291, Aug. 2013.

[10] B. Chen, L. Zhang, and Z. Wu, "General iterative receiver design for enhanced reliability in multi-carrier differential chaos shift keying systems," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7824–7839, Nov. 2019.

[11] L. Zhang, J. Zheng, B. Chen, and Z. Wu, "Reliable low- rank approximation of matrices detection aided multicarrier DCSK receiver design," *IEEE Syst. J.*, vol. 15, no. 4, pp. 5277–5288, Dec. 2021.

[12] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[14] L. Zhang, H. Zhang, Y. Jiang, and Z. Wu, "Intelligent and reliable deep learning LSTM neural networks-based OFDM-DCSK demodulation design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16163–16167, Dec. 2020.

[15] H. Zhang, L. Zhang, J. Cheng, Y. Jiang, and Z. Wu, "An intelligent detection based on deep learning for multilevel code shifted differential chaos shift keying system with M-ary modulation," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 1, pp. 155–169, Mar. 2022.

[16] E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," 2018, *arXiv:1802.06002*.

[17] H. Chen, P. Chen, S. Wang, S. Lai, and R. Chen, "Reference-modulated PI-DCSK: A new efficient chaotic permutation index modulation scheme," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9663–9673, Sep. 2022.

[18] M. Herceg, G. Kaddoum, D. Vranješ, and E. Soujeri, "Permutation index DCSK modulation technique for secure multiuser high-data-rate communication systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 2997–3011, Apr. 2018.

[19] H. Ye, G. Y. Li, and B. H. Fred Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.

[20] M. Rice et al., "Physical-layer security for vehicle-to-everything networks: Increasing security while maintaining reliable communications," *IEEE Trans. Veh. Technol.*, vol. 15, no. 3, pp. 68–76, Sep. 2020.

[21] Y. Fang, J. Zhuo, H. Ma, S. Mumtaz, and Y. Li, "Design and analysis of a new index-modulation-aided dcsk system with frequency-and-time resources," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7411–7425, Jun. 2023.

[22] Y. Yuting, Z. Gangqiang, and L. Junkai, "Design of FM-DCSK waveform for underwater covert communication system," in *Proc. OES China Ocean Acoust. (COA)*, 2021, pp. 651–654.

[23] K. Kim, J. Lee, and J. Choi, "Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO system," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 828–831, Apr. 2018.

[24] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997.

[25] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, Feb. 2020.

[26] M. S. Willsey, K. M. Cuomo, and A. V. Oppenheim, "Quasi-orthogonal wideband radar waveforms based on chaotic systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1974–1984, Jul. 2011.

[27] C. Miles et al., "Correlator convolutional neural networks as an interpretable architecture for image-like quantum matter data," *Nature Commun.*, vol. 12, no. 1, pp. 1–7, 2021.

[28] L. Teng, H. H. Iu, X. Wang, and X. Wang, "Novel chaotic behavior in the Muthuswamy–Chua system using Chebyshev polynomials," *Int. J. Numer. Model. Electron. Netw. Devices Fields*, vol. 28, no. 3, pp. 275–286, 2015.

[29] S. Saremi and S. Mirjalili, "Integrating chaos to biogeography based optimization algorithm," *Int. J. Comput. Commun. Eng.*, vol. 2, no. 6, pp. 655–658, 2013.

[30] S. Mirjalili. "Chaos theory and meta-heuristics, MATLAB Central File Exchange." 2023. Accessed: Jul. 10, 2023. [Online]. Available: (https://www.mathworks.com/matlabcentral/fileexchange/47215-chaos -theory-and-meta-heuristics)

[31] S. Amoh, X. Zhang, G. Chen, and T. Ueta, "Bifurcation analysis of a class of generalized Hénon maps with hidden dynamics," *IEEJ Trans. Elect. Electron. Eng.*, vol. 16, no. 11, pp. 1456–1462, 2021.

[32] M.P. Kennedy, R. Rovatti, and G. Setti Eds., *Chaotic Electronics in Telecommunications*. Boca Raton, FL, USA: *CRC Press*, 2000.

[33] C. Liu, Z. Wei, D. W. K. Ng, J. Yuan, and Y. C. Liang, "Deep transfer learning for signal detection in ambient backscatter communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1624–1638, Mar. 2021.

[34] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Hoboken, NJ, USA: Prentice-Hall, 1996.

[35] T. Erpek, T.J. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep learning for wireless communications," *Development and Analysis of Deep Learning Architectures*. Cham, Switzerland: Springer, 2020, pp. 223–266.

[36] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Physical layer deep learning of encodings for the MIMO fading channel," in *Proc. 55th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, 2017, pp. 76–80.

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

[38] A. Litvinenko and A. Aboltins, "Use of cross-correlation minimization for performance enhancement of chaotic spreading sequence based asynchronous DS-CDMA system," *Proc. IEEE 4th Workshop Adv. Informat. Electron. Elect. Eng. (AIEEE)*, 2016, pp. 1–6.

[39] W. Zhang and Z. Zhang, "Belief function based decision fusion for decentralized target classification in wireless sensor networks," *Sensors*, vol. 15, no. 8, pp. 20524–20540, 2015.

[40] P.J. Freire et al., "Performance versus complexity study of neural network equalizers in coherent optical systems," *J. Lightw. Technol.*, vol. 39, no. 19, pp. 6085–6096, Oct. 2021.

[41] T. Osogami, H. Kajino, and T. Sekiyama, "Bidirectional learning for time-series models with hidden units," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2711–2720.

[42] G. Kis, "Performance analysis of chaotic communications systems," Ph.D. dissertation, Dept. Meas. Inf. Syst., Budapest Univ. Technol. Econom., Budapest, Hungary, 2005.

[43] T. Huang, L. Wang, W. Xu, and F. C. M. Lau, "Multilevel code-shifted differential-chaos-shift-keying system," *IET Commun.*, vol. 10, no. 10, pp. 1189–1195, 2016.

[44] X. Cai, W. Xu, R. Zhang, and L. Wang, "A multilevel code shifted differential chaos shift keying system with M-ary modulation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 8, pp. 1451–1455, Aug. 2019.

[45] Y. Xie, A. N. Joseph Raj, Z. Hu, S. Huang, Z. Fan, and M. Joler, "A twofold lookup table architecture for efficient approximation of activation functions," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 28, no. 12, pp. 2540–2550, Dec. 2020.

[46] J.G. Proakis and M. Salehi, *Digital Communications*, New York, NY, USA: McGraw-Hill, pp. 203–208, 2008.

[47] M. Gao, Y. Li, O. A. Dobre, and N. Al-Dhahir, "Joint blind identification of the number of transmit antennas and MIMO schemes using Gerschgorin radii and FNN," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 373–387, Jan. 2019.

[48] A. B. Abdul Qayyum, A. Arefeen, and C. Shahnaz, "Convolutional neural network (CNN) based speech-emotion recognition," in *Proc. IEEE Int. Conf. Signal Process. Inf. Commun. Syst. (SPICSCON)*, 2019, pp. 122–125.

[49] J. Nam and J. Kang, "Classification of chaotic signals of the recurrence matrix using a convolutional neural network and verification through the lyapunov exponent," *Appl. Sci.*, vol. 11, no. 1, p. 77, 2020.

[50] J. Nam and J. Kang, "Classification of chaotic squeak and rattle vibrations by CNN using recurrence pattern," *Sensors*, vol. 21, no. 23, p. 8054, 2021.

[51] A. Foroutannia and M. Ghasemi, "Predicting cortical oscillations with bidirectional LSTM network: A simulation study," *Nonlinear Dyn.*, vol. 111, no. 9, pp. 8713–8736, 2023.

[52] A. Garg, A. Kapoor, A. K. Bedi, and R. K. Sunkaria, "Merged LSTM Model for emotion classification using EEG signals," in *Proc. Int. Conf. Data Sci. Eng. (ICDSE)*, 2019, pp. 139–143.

[53] C. Bhattacharya and A. Ray, "Thresholdless Classification of chaotic dynamics and combustion instability via probabilistic finite state automata," *Mech. Syst. Signal Process.*, vol. 164, Feb. 2022, Art. no. 108213.

[54] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun. "ShuffleNet V2: Practical guidelines for efficient CNN architecture design." 2018. [Online]. Available: arXiv:1807.11164.

[55] R. Shahzadi, S. M. Anwar, F. Qamar, M. Ali, J. J. Rodrigues, and M. Alnowami, "Secure EEG signal transmission for remote health monitoring using optical chaos," *IEEE Access*, vol. 7, pp. 57769–57778, 2019.

[56] E. E. Elsayed and B. B. Yousif, "Performance enhancement of the average spectral efficiency using an aperture averaging and spatial-coherence diversity based on the modified-PPM modulation for MISO FSO links," *Opt. Commun.*, vol. 463, May 2020, Art. no. 125463.

[57] B. B. Yousif, E. E. Elsayed, and M. M. Alzalabani, "Atmospheric turbulence mitigation using spatial mode multiplexing and modified pulse position modulation in hybrid RF/FSO orbital-angular-momentum multiplexed based on MIMO wireless communications system," *Opt. Commun.*, vol. 436 pp. 197–208, Apr. 2019.

[58] B. B. Yousif and E. E. Elsayed, "Performance enhancement of an orbital-angular-momentum-multiplexed free-space optical link under atmospheric turbulence effects using spatial-mode multiplexing and hybrid diversity based on adaptive MIMO equalization," *IEEE Access*, vol. 7, pp. 84401–84412, 2019.

[59] M. Mobini, and M.R. Zahabi, "An approach to security in VANETs using chaos-based cryptography," in *Proc. 15th Int. Conf. Traffic Transp. Eng.*, 2016, pp. 1–10.

[60] I. M. Al-Musawi, W. A. Al-Hussaibi, and F. H. Ali, "Efficient secure NOMA schemes based on chaotic physical layer security for wireless networks," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 2425–2443, 2022.

[61] N. Ishikawa, J. M. Hamamreh, E. Okamoto, C. Xu, and L. Xiao, "Artificially time-varying differential MIMO for achieving practical physical layer security," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2180–2194, 2021.

[62] T. Joachims, "Transductive inference for text classication using support vector machines," in *Proc. 16th Int. Conf. Mach. Learn.*, San Francisco, CA, USA, 1999, pp. 200–209.

MARIJAN HERCEG received the B.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering, Computer Science, and Information Technology, University of Osijek, Osijek, Croatia, in 2002 and 2010, respectively, where he is currently a Full Professor with the Department of Communications, Faculty of Electrical Engineering, Computer Science and Information Technology. Primarily his research work includes advanced modulation techniques for impulse radio ultra-wideband systems, chaos-based communication systems, multiple access techniques, and sigma-delta A/D converters. He also reviews IEEE journals and is a respectful member of IEEE Communications Society.

GEORGES KADDOUM (Senior Member, IEEE) received the bachelor's degree in electrical engineering from the École Nationale Supérieure de Techniques Avancées, Brest, France, the M.S. degree in telecommunications and signal processing from the Université de Bretagne Occidentale and Telecom Bretagne, Brest, in 2005, and the Ph.D. degree (Hons.) in signal processing and telecommunications from the National Institute of Applied Sciences, University of Toulouse, Toulouse, France, in 2009. He is currently an Associate Professor and a Tier 2 Canada Research Chair with the École de Technologie Supérieure (ÉTS) and the Université du Québec, Montréal, Canada, and also a Faculty Fellow with the Cyber Security Systems and Applied AI Research Center, Lebanese American University, Beirut, Lebanon. Since 2010, he has been working as a Scientific Consultant in the field of space and wireless telecommunications for several U.S. and Canadian companies. Over 200 journals and numerous conference papers has been published, following that two pending patents are there as well. His recent research activities cover mobile communication systems, modulations, security, space communications, and navigation. In 2014, he was awarded for his contribution in the ÉTS Research Chair in physical-layer security for wireless networks. He received the Best Papers Awards at the 2014 IEEE International Conference on Wireless and Mobile Computing, Networking, Communications, and the 2017 IEEE International Symposium on Personal Indoor and Mobile Radio Communications, with three and four co-authors, respectively. Moreover, he received IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer Award for the years 2015, 2017, and 2019. In addition, he received the Research Excellence Award from the Université du Québec in 2018. In 2019, he was honored with the Research Excellence Award from the ÉTS in recognition for his outstanding research outcomes. He is also serving as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and IEEE COMMUNICATIONS LETTERS.

MAJID MOBINI received the M.Sc. degree (holding first rank) in communications engineering from the Department of Electrical and Computer Engineering, Tehran Polytechnic University, Iran, and the Ph.D. degree in communications engineering from the Babol Noshirvani University of Technology in 2019. His current field of work includes wireless, mobile, optical, RF, cellular, MIMO, semantic, and satellite communications, convex optimization, game theory, signal processing, geometric problems, chaos and complex networks, machine learning, deep learning, virtual agents, reinforcement learning, natural language processing, ethics, human-level AI, robotic process automation, smart cities, data mining, decision making, predictive analysis, quantum devices, cybersecurity, cryptography, remote sensing, hyper-spectral image processing, radar systems, UAVs, electronic warfare, bio-medical instruments, medical diagnostics, radiology, ghost imaging, robot control, neuro-feedback, remote surgery, smart hospitals, telehealth systems, brain-computer interfaces, environment control, genome analysis, and interdisciplinary research.