

Article

Design of a SIMO Deep Learning-Based Chaos Shift Keying (DLCSK) Communication System

Majid Mobini ^{1,*} , Georges Kaddoum ²  and Marijan Herceg ³ 

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

² Département de Génie Électrique, University of Québec, École de Technologie Supérieure, Montréal, QC H3C1K3, Canada; georges.kaddoum@etsmtl.ca

³ Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, 31000 Osijek, Croatia; marijan.herceg@ferit.hr

* Correspondence: Mobini2002@gmail.com

Abstract: This paper brings forward a Deep Learning (DL)-based Chaos Shift Keying (DLCSK) demodulation scheme to promote the capabilities of existing chaos-based wireless communication systems. In coherent Chaos Shift Keying (CSK) schemes, we need synchronization of chaotic sequences, which is still practically impossible in a disturbing environment. Moreover, the conventional Differential Chaos Shift Keying (DCSK) scheme has a drawback, that for each bit, half of the bit duration is spent sending non-information bearing reference samples. To deal with this drawback, a Long Short-Term Memory (LSTM)-based receiver is trained offline, using chaotic maps through a finite number of channel realizations, and then used for classifying online modulated signals. We presented that the proposed receiver can learn different chaotic maps and estimate channels implicitly, and then retrieves the transmitted messages without any need for chaos synchronization or reference signal transmissions. Simulation results for both the AWGN and Rayleigh fading channels show a remarkable BER performance improvement compared to the conventional DCSK scheme. The proposed DLCSK system will provide opportunities for a new class of receivers by leveraging the advantages of DL, such as effective serial and parallel connectivity. A Single Input Multiple Output (SIMO) architecture of the DLCSK receiver with excellent reliability is introduced to show its capabilities. The SIMO DLCSK benefits from a DL-based channel estimation approach, which makes this architecture simpler and more efficient for applications where channel estimation is problematic, such as massive MIMO, mmWave, and cloud-based communication systems.

Keywords: chaos shift keying; deep learning; LSTM; multi-antenna



Citation: Mobini, M.; Kaddoum, G.; Herceg, M. Design of a SIMO Deep Learning-Based Chaos Shift Keying (DLCSK) Communication System. *Sensors* **2022**, *22*, 333. <https://doi.org/10.3390/s22010333>

Academic Editor: Leopoldo Angrisani

Received: 11 November 2021

Accepted: 17 December 2021

Published: 3 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chaotic signals are wide-band noise-like signals with robust and reproducible statistical features [1,2]. Thus, digital modulation using chaotic signals presents an inherently simple solution for robust and secure communications over multi-path fading channels [3,4]. Chaos-based modulations are generally classified into two main classes [5]. The first is coherent detection schemes, such as the Chaos Shift Keying (CSK) scheme [6], in which data are transmitted in a combination of basis functions obtained from chaotic waveforms. Chaotic synchronization is commonly used in the recovery of basis functions on the receiver side [7–9]. Theoretically, CSK modulation can achieve the Bit Error Rate (BER) performance of Binary Phase Shift Keying (BPSK) [10]. In application, this performance is not achievable as some problems need to be solved, namely the recovery of basis functions and the estimation [11–16]. Since chaotic synchronization is still practically impossible in a turbulent environment [17], coherent schemes can not work properly in utilization.

The Differential Chaos Shift Keying (DCSK) system [18] is a variant of CSK where the basis functions have a special arrangement and the information can be revealed from the correlation between the parts of the basis functions. In DCSK modulation, each bit duration

is divided into two equal slots, where the first slot is allocated to a reference signal whereas the second slot is used to send information-bearing signal. At the receiver side, these two signals are correlated to detect the transmitted bit information. Non-coherent schemes, such as DCSK, do not require Channel State Information (CSI) or chaotic synchronization on the receiver side [19–21]. The low data rate and energy efficiency of this scheme are significant drawbacks, as half of the bit time interval is used for sending non-information-bearing samples [22–25]. Moreover, sending the same signal twice negatively affects security [26].

The problem of basis function recovery from the received signals, independent of modulation, has impeded the progress of coherent chaos-based communications. To tackle this problem, we propose a Deep Learning (DL)-based receiver that enables us to implicitly estimate wireless channels and retrieve the transmitted message. We showed that using a trainable receiver can mitigate the chaotic synchronization problem in the existing correlator receivers.

Accepting the definitions of coherent and non-coherent chaos-based communications systems introduced in [8], the symbol may be recovered by means of coherent detection, where all possible sample functions are known, or by non-coherent detection, where one or more characteristics of the sample functions are estimated. Although in the proposed receiver, the sample functions are available, but the available information is previously estimated from the transmitted signals. Regarding the above definitions, though the proposed receiver has similarities to the coherent and non-coherent receivers, it does not exactly fit into the conventional categories. Hence, we use the term Deep Learning-based Chaos Shift Keying (“DLCSK”) to introduce a new category of chaos-based communications systems. Note that, a trained DLCSK receiver can easily work with existing CSK transmitters.

The proposed scheme shows a BER gain compared to conventional DCSK. This advantage is obtained at the cost of adding an offline training step. In other words, we train the receiver by chaotic reference signals, instead of reference signal transmission in the test phase. When the training process is performed under different channel conditions, NN grasps different chaotic maps and implicitly estimates channel distribution. As a result, DLCSK consumes fewer resources (energy and time) to transmit one bit and shows a more robust behavior.

1.1. Background

DL is a powerful tool that can be implemented for solving problems that are complicated to describe through mathematical models [27]. There are two general methods for designing DL-based architectures, namely DL-based receiver design and joint DL-based transceiver design [28]. A DL-based joint transceiver design optimizes the system as an end-to-end auto-encoder [29]. In contrast, a DL-based receiver optimizes one or more blocks in the receiver [30]. As an example of the latter, in [31], the authors suggested a DL-based receiver to indirectly estimate wireless channels and retrieve the signals. They demonstrated that DL-based receivers have the ability to learn the features of wireless channels, including non-linear distortions. Practical wireless channels change over time with a large dynamic range. In this case, NNs can hardly perform well in the detection tasks under different channel coefficients [32]. To overcome this challenge, Deep Transfer Learning (DTL) adopts a NN to extract the time-varying features with a few online training data by transferring knowledge from a source domain to a target domain [33–35]. Motivated by this approach, in this paper, we try to propose a framework for detection of the chaotic-modulated signals.

In this study, our focus is on the design of a DL-based receiver that can be easily harmonized with existing CSK transmitters. Recently, DL have attracted a lot of attention because of its effectiveness in data-driven analysis of chaotic dynamics [36]. Some studies have used DL in chaos-based communications [37–40], however, their methods need further explorations. In [39], the authors proposed an algorithm to demodulate the reference chaotic signals iteratively. In contrast to traditional demodulation methods, utilizing time or frequency resources to improve reliability, the iterative receiver addresses the feature extraction capability of neural networks (NNs). In [40], an intelligent OFDM-

DCSK demodulator is proposed using an LSTM-aided NN that withdraw the correlations between chaotic modulated OFDM-DCSK signals in order to retrieve the transmitted information. However, the LSTM-aided receiver transmits reference signals that do not carry useful information. A similar problem can be observed in the other proposed DL-based modulations (e.g., in [38,39]).

In this paper, we present an innovative Long Short-Term Memory (LSTM)-based receiver that does not require any reference signal transmissions or chaotic synchronization circuits for data detection. The proposed DL-based Chaos Shift Keying (DLCSK) system benefits from other advantages of NNs, such as generalization and fast training. The multi-antenna design of the DLCSK receiver that, uses an LSTM for each antenna is also presented. Many articles have also combined Multi-Antenna technology with Chaos-based communications [41–53]. For instance, a MIMO STBC-DCSK system does not require any complicated channel estimation, carrier synchronization, or rake reception [54]. However, these conventional techniques transmit the reference signals over the channel, which increases the redundancy and complexity of the system. Previous research in the context of neural network ensembles shows that combining network architectures is frequently more accurate than using single networks [55–57]. The proposed Single Input Multi Output (SIMO) DLCSK focuses on a fusing method to achieve a diversity gain, where the outputs of all LSTMs are combined using a majority voting strategy. This design maintains the advantages of traditional Multi-Antenna and Chaos-based systems, i.e., it does not require rake receivers, CSI, and chaotic synchronization. DLCSK is an appealing candidate for secure data transmission in cloud-based systems [58–60], vehicular communications [61–63], and massive MIMO systems [64–67] due to the aforementioned characteristics.

1.2. Contributions

The objective of this work is to develop a DL-based receiver that benefits from the inherent security of chaotic signals and the merits of NNs. This design will robust the capabilities of the existing chaos-based schemes, such as reliability and energy efficiency. Innovative aspects of this paper are briefly defined below:

- We train a LSTM-based classifier that enables online classification of the received chaotic signals. Implementing this method can mitigate the chaotic synchronization problem in the existing CSK receivers;
- The DLCSK modulation/demodulation scheme does not need any reference transmission for basis functions recovery, unlike reference-based non-coherent modulations. According to the above advantages, the BER performance of a single antenna DLCSK is close to the performance of the antipodal CSK under a Rayleigh fading channel that shows an outstanding BER performance among all chaotic modulations [10].
- We have proposed a Multi-Antenna architecture of the DLCSK system. Multiple LSTMs are used at the receiver end to obtain a diversity gain. We numerically simulate the SIMO DLCSK structure and state the advantages of fusing the hard outputs of the LSTMs to come to a decision.

The rest of this paper is organized in following sections:

- In Section 2, a statistical study on the existing CSK systems and correlation receivers is presented;
- In Section 3, the structure of the proposed SIMO DLCSK system and the basics of the LSTM-based classifier are described;
- In Section 4, simulation results and discussions are explained;
- In Section 5, the conclusions are explained.

2. Traditional Correlation Receivers

This section presents a statistical study on a typical CSK system equipped with the correlation receivers for a deeper identification of its unsolved problems. In contrast with regular communications systems in which the basis functions are periodic and orthogonal (e.g., sinusoidal functions), in a chaotic communications system, the basis functions are not

orthogonal necessarily and differ from symbol to another. Consider now a coherent CSK system using two basis functions [10]. The elements of the signal set are given by

$$\mathbf{s}_b = s_{b1}\tilde{\mathbf{x}} + s_{b2}\tilde{\mathbf{x}}, \quad (1)$$

where b is the index of the current transmission bit, the weights s_{b1} and s_{b2} are the elements of the signal vector, and the basis functions $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ are discrete-time chaotic signals with β samples, i.e., $\tilde{\mathbf{x}} = \{\tilde{x}_1, \dots, \tilde{x}_\beta\}$ and $\tilde{\mathbf{x}} = \{\tilde{x}_1, \dots, \tilde{x}_\beta\}$. A binary bit data symbol is spread by a chaotic signal with the bit duration T_b . Thus, we have $T_b = \beta T_c$, where T_c is the time between each chaotic sample (chip). The transmitted sample functions are $s_1 = \sqrt{E_b} \tilde{\mathbf{x}}$ and $s_2 = \sqrt{E_b} \tilde{\mathbf{x}}$, displaying symbols "0" and "1", respectively. The corresponding signal vectors are $(s_{11} \ s_{12}) = (\sqrt{E_b} \ 0)$ and $(s_{21} \ s_{22}) = (0 \ \sqrt{E_b})$, where E_b stands for the average energy per bit. With two basis functions, the receiver should be configured with at least two correlators. The message is detected by correlating the received signal with two reference signals $\tilde{\mathbf{x}}'$ and $\tilde{\mathbf{x}}'$, and forming the corresponding observation signals (so-called decision variables) D_{b1} and D_{b2} . If $D_{b1} > D_{b2}$, then the decision is "1", and if $D_{b1} < D_{b2}$, then the decision is "0". Consider now the output of a correlator under a noisy channel. In a coherent CSK system, the reference signals $\tilde{\mathbf{x}}'$ and $\tilde{\mathbf{x}}'$ are derived from the noisy received signal $(\mathbf{s}_b + \mathbf{n})$. The decision variable D_{b1} , at the output of the correlator, can be written as

$$\begin{aligned} D_{b1} &= \int_{T_s}^T [\mathbf{s}_b + \mathbf{n}] \tilde{\mathbf{x}}' dt \\ &= \int_{T_s}^T [s_{b1}\tilde{\mathbf{x}} + s_{b2}\tilde{\mathbf{x}} + \mathbf{n}] \tilde{\mathbf{x}}' dt \\ &= s_{b1} \int_{T_s}^T \tilde{\mathbf{x}}\tilde{\mathbf{x}}' dt + s_{b2} \int_{T_s}^T \tilde{\mathbf{x}}\tilde{\mathbf{x}}' dt \\ &\quad + \int_{T_s}^T \mathbf{n}\tilde{\mathbf{x}}' dt, \end{aligned} \quad (2)$$

where T_s is the synchronization transient time for each bit duration. Note that D_{b1} is a random variable, whose mean value depends on the E_b and the quality of the regenerated reference signal $\tilde{\mathbf{x}}'$. If a perfect synchronization is maintained throughout the transmission, we have $T_s = 0$, $\tilde{\mathbf{x}}' = \tilde{\mathbf{x}}$, and $\tilde{\mathbf{x}}' = \tilde{\mathbf{x}}$. In this case

$$\begin{aligned} D_{b1} &= s_{b1} \int_0^{T_b} \tilde{\mathbf{x}}\tilde{\mathbf{x}} dt + s_{b2} \int_0^{T_b} \tilde{\mathbf{x}}\tilde{\mathbf{x}} dt \\ &\quad + \int_0^{T_b} \mathbf{n}\tilde{\mathbf{x}} dt. \end{aligned} \quad (3)$$

The chaotic basis function $\tilde{\mathbf{x}}$ is different in every transmission, and the variance of $\int_0^{T_b} \tilde{\mathbf{x}}^2 dt$ causes detection error. This variance can be reduced by increasing the bit duration T_b . Alternatively, it can be zeroed by modifying the generated basis functions, such that the transmitted energy E_b is kept constant. A constant E_b is achievable by normalizing the basis functions before each symbol transmission, such that $\int_0^{T_b} \tilde{\mathbf{x}}^2 dt = 1$.

The second term on the right-hand side of Equation (3) results in the cross-correlation estimation problem. It can also be reduced by selecting long chaotic signals or using Walsh functions [9]. By choosing orthonormal basis functions, the estimated symbol s_{b1} can be written as

$$D_{b1} = s_{b1} + \int_0^T \mathbf{n}\tilde{\mathbf{x}} dt. \quad (4)$$

In contrast with the traditional demodulation methods that often use synchronization (coherent schemes) or delayed reference transmission (non-coherent schemes) for reference recovery, the proposed DLCK utilizes DL for training and recovery of the reference signals \tilde{x}' and \tilde{x}' . By using DL, it not only mitigates the need for chaotic synchronization during transmissions, but also can reduce the effect of the last term in Equation (4) using an indirect channel estimation method. It should be noted that our main focus is not on the auto and cross-correlation problems, or selecting proper chaotic maps. Therefore, we have chosen the Chebyshev map and logistic map to make chaotic sequences simpler in terms of generating.

3. SIMO DLCSK System Model

For the general architecture of the SIMO DLCSK system, refer to Figure 1. The proposed design involves two phases. In the first phase, the chaotic sequences are transmitted through an offline training phase under different channel conditions, and then the deployment/test phase initializes the modulated data transmission.

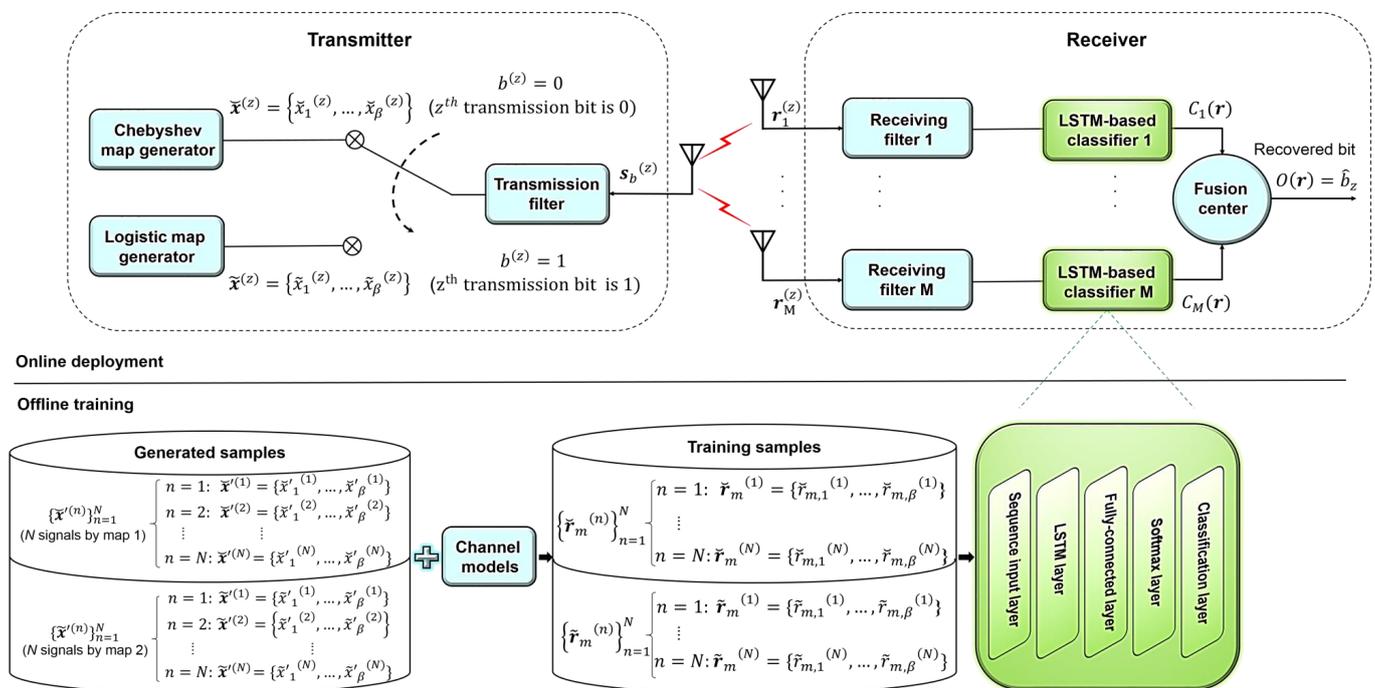


Figure 1. The SIMO DLCSK digital communication system.

3.1. Chaotic Signals

We consider discrete-time chaotic signals generated by the Chebyshev map and the Logistic map for “0” and “1”, respectively. The chaotic Chebyshev map generated by the second order Chebyshev polynomial function is used, which can be written as $\tilde{x}_{t+1} = 1 - \tilde{x}_t^2$ [68]. Another chaotic map is the Logistic map, which is generated using a recursive function, $\tilde{x}_{t+1} = \rho \tilde{x}_t (1 - \tilde{x}_t)$ [69]. The parameter ρ is called the bifurcation parameter, and the values of interest for ρ are in the interval $[1, 4]$. For $3.57 \leq \rho \leq 4$, the generated sequence is non-periodic and non-converging [70].

In this study, we evaluate the effect of the changes in ρ on the classifier’s performance. The parameter changes can be intentional (to achieve higher accuracy), or unintentional (due to environmental conditions) [71]. When $\rho < 3.57$, the generated signals by Logistic map show a periodic behavior and a classifier can easily separate it from a chaotic signal, (e.g., a signal generated by the Chebyshev map). There is a trade-off between the classification accuracy and the security. Since a periodic waveform can bring negative effect on the security, we select $\rho > 3.6$ for a case that definitely guarantees the chaotic

dynamics. The chaotic maps have zero mean and unit variance, i.e., $E(\check{x}_t) = E(\tilde{x}_t) = 0$ and $E(\check{x}_t^2) = E(\tilde{x}_t^2) = 1$, where $E(\cdot)$ denotes the probabilistic expectation operator.

3.2. Transmitter

The transmitter structure is similar to conventional coherent CSK transmitters. Thus, the proposed DLCSK demodulator can easily work with existing CSK transmitters. In the training phase, the transmitter generates two sets of the chaotic signals, $\{\check{x}'^{(n)}\}_{n=1}^N$ and $\{\tilde{x}'^{(n)}\}_{n=1}^N$ using the Chebyshev and the Logistic maps, respectively. Each one of these sets includes the number of N chaotic signals with length β samples, i.e., $\check{x}'^{(n)} = \{\check{x}'_1^{(n)}, \dots, \check{x}'_\beta^{(n)}\}$ and $\tilde{x}'^{(n)} = \{\tilde{x}'_1^{(n)}, \dots, \tilde{x}'_\beta^{(n)}\}$.

In other words, we use two sets of the chaotic signals, each with length $S = N\beta$ samples, where N and β are the number and the length of the chaotic signals, respectively. All of the chaotic signals $\check{x}'^{(n)}$ and $\tilde{x}'^{(n)}$ with known binary labels $j^{(n)} \in \{0, 1\}$ are transmitted repeatedly to train the receivers. All the M antennas obtain the corresponding altered signals, i.e., $\check{r}_m^{(n)} = \{\check{r}_{m,1}^{(n)}, \dots, \check{r}_{m,\beta}^{(n)}\}$ and $\tilde{r}_m^{(n)} = \{\tilde{r}_{m,1}^{(n)}, \dots, \tilde{r}_{m,\beta}^{(n)}\}$, for input to the NN. After the number of $2N$ transmissions, we have a set of received vectors $\{\mathbf{r}_m^{(n)}\}_{n=1}^{2N} = \{\check{r}_m^{(n)}, \tilde{r}_m^{(n)}\}_{n=1}^N$ with known class labels $\{j^{(n)}\}_{n=1}^{2N} = \{0^{(n)}, 1^{(n)}\}_{n=1}^N$ to train the NN through the supervised learning framework. Therefore, training set of the m th DLCSK receiver can be expressed as

$$\mathcal{D}_m = \{\mathbf{r}_m^{(n)}, j^{(n)}\}_{n=1}^{2N}, \quad (5)$$

where $2N$ is the number of chaotic signals in the training set. Assuming a Rayleigh fading channel, the $r_{m,t}$ (t th sample of the vector \mathbf{r}_m) can be modeled as a complex-valued random variable, i.e., $r_{m,t} = \Re(r_{m,t}) + \Im(r_{m,t})$, where the operators $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imaginary parts of a complex number, respectively. Therefore, the vector \mathbf{r}_m can be separated into two vectors, i.e., $\mathbf{r}_m = [\Re(\mathbf{r}_m), \Im(\mathbf{r}_m)]$, before being fed into the classifier, and the training set can be rewritten as

$$\mathcal{D}_m = \{[\Re(\mathbf{r}_m), \Im(\mathbf{r}_m)]^{(n)}, j^{(n)}\}_{n=1}^{2N}. \quad (6)$$

Once the DL-based receiver is trained, it can be used for online demodulation and data detection. In the test/deployment phase, the DLCSK modulator maps a transmission bit $b^{(z)} \in \{0, 1\}$, ($1 \leq z \leq Z$), to a chaos waveform, where $b^{(z)}$ denotes z^{th} transmission bit and Z is the number of data bits. Thus, in the test phase, the transmitter generates two sets of the chaotic signals, i.e., $\{\check{x}^{(z)}\}_{z=1}^Z$ and $\{\tilde{x}^{(z)}\}_{z=1}^Z$ using the Chebyshev and the Logistic maps, respectively. Each of these sets includes the number of Z chaotic signals with length β samples, i.e., $\check{x}^{(z)} = \{\check{x}_1^{(z)}, \dots, \check{x}_\beta^{(z)}\}$ and $\tilde{x}^{(z)} = \{\tilde{x}_1^{(z)}, \dots, \tilde{x}_\beta^{(z)}\}$. According to the Equation (1), and depending on the current bit $b^{(z)}$, the signal $\mathbf{s}_b^{(z)}$ is transmitted, i.e.,

$$\mathbf{s}_b^{(z)} = \begin{cases} \sqrt{E_b} \check{x}^{(z)} & \text{if } b^{(z)} = 0 \\ \sqrt{E_b} \tilde{x}^{(z)} & \text{if } b^{(z)} = 1. \end{cases} \quad (7)$$

Since in every practical communication system, the chaos generator circuits may operate under different environmental conditions, it is essential to consider a parameter mismatch between the training and testing phases. To evaluate the system's generalization capability and robustness, we use different parameter settings to achieve $\{\check{x}'^{(n)}\}_{n=1}^N \neq \{\check{x}^{(z)}\}_{z=1}^Z$ and $\{\tilde{x}'^{(n)}\}_{n=1}^N \neq \{\tilde{x}^{(z)}\}_{z=1}^Z$. Hence, to generate Logistic maps, different bifurcation parameters are chosen as ρ_{train} and ρ_{test} . In addition, to generate Chebyshev maps, different initial states are chosen for the training and the deployment phases, i.e., $\check{x}'_1 \neq \check{x}_1$.

The transmission filter (or pulse shaping filter) is commonly used in communication systems. This filter can take different forms, such as the Gaussian filter or raised-cosine filter [72]. In this paper, we consider a rectangular pulse of unit amplitude on $[0, T_c]$, where

T_c represents the chip time. The noise power can be restricted by a receiving filter at the receiver side. Note that, these filters are not our main focus.

3.3. Channel Model and Estimation

In an AWGN channel, a one-dimensional noisy version of the transmitted signal s_b can be observed at the receiver side. The m th receiver obtains the altered signals r_m and makes a decision about the transmitted bit. In addition to the Gaussian noise effect, many other stochastic phenomena may occur in a practical communication channel. Conventional channel estimation approaches are sensitive in terms of the quality of the pilot signals. One way of augmenting DL models is through the use of learning channel variations [73]. If the LSTM-based classifier is trained using a dataset that contains signals transmitted under different channel conditions, the classifier will be resilient to channel changes, eliminating the need for instantaneous channel estimation [74,75]. In this paper, we assume that the channel changes from one signal transmission to another, and the transmitted signal acts as a pilot that carries channel information. Thus, the NN can simultaneously obtain different chaotic maps and estimate the statistical distribution of the fading.

Consider a multi-path fading channel, with L independent paths, where the channel coefficients follow a Rayleigh distribution. Therefore, the probability density function (pdf) of the channel coefficient α can be given as [76],

$$q(\alpha|\delta) = \frac{\alpha}{\delta^2} e^{-\frac{\alpha^2}{2\delta^2}}, \quad (8)$$

where $\delta > 0$ is the scale parameter of the distribution representing the root mean square value of the received voltage signal. Considering a multi-antenna system with the number of M antennas, if the t th transmitted sample is shown by $s_{b,t}$, the t th received sample at the m th antenna can be modeled as

$$r_{m,t} = \sum_{l=1}^L \alpha_{m,l} s_{b,t-\tau_{m,l}} + n_{m,t}, \quad m = 1, 2, \dots, M, \quad (9)$$

where $\alpha_{m,l}$ represents the channel gain of l th path between the transmit antenna and m th receive antenna, and $\tau_{m,l}$ is the delay of the l th path. L is the number of paths, and $n_{m,t}$ is independent noise at each antenna, which is assumed to be additive white Gaussian noise with zero mean and variance $N_0/2$.

Since we assume that the channel changes from one signal transmission to another, the channel gain $\alpha_{m,l}^{(n)}$ and training SNR $\sigma_{tr}^{(n)}$ changes after the n th channel realization. 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 values. The goal of training process is to train a NN with a complex-valued vector r_m . The input complex values are split into real and imaginary parts, i.e., $r_m = [\Re(r_m), \Im(r_m)]$, before being fed into the classifier. Therefore, we have two feature vectors containing the channel coefficients and stochastic phases. Later, a Softmax layer estimates probability vectors $p_{n,j}$ from input distribution, for the n th observation, where j shows the possible classes (i.e., $j \in \{1, 2\}$), and optimizes the cross-entropy cost function in Equation (13). In the test phase, we use this trained network to detect unknown inputs.

3.4. Receiver

We consider multiple antennas and LSTMs at the receiver end to establish a SIMO design and to obtain a diversity gain. This architecture relies on an ensemble method to fuse several classifiers with the goal of increasing the classification accuracy. In the following, we introduce the structure of a single LSTM-based classifier. The proposed classifier has five base layers: Sequence input layer, LSTM/BiLSTM, fully connected layer (size 2), Softmax, and Classification layer. The sequence input layer is only used to fetch sequential input values of length 2β . The adopted LSTM cell (unit) is shown in Figure 2.

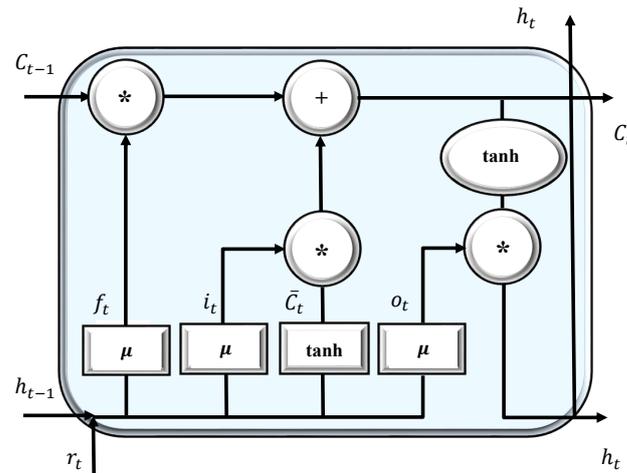


Figure 2. The LSTM cell structure.

The forget gate f_t determines how much of the current cell state should be forgotten, and the output gate o_t controls which part of the information should be sent to the output. C_{t-1} and C_t , respectively, show the state value of the memory unit at the previous step and current step. Then, h_{t-1} and h_t indicate the output of the previous and the current states, respectively, whereas r_t and μ represents the current input and sigmoid function, respectively. Equation (10) illustrates the LSTM cell calculation process [77]. The forget gate decides which information will be remembered or forgotten based on the last hidden layer output h_{t-1} and the current input r_t . The memory cell value C_t is determined by the current value C_t , its own state C_{t-1} , input, and forget gates. The operator (*) represents element-wise matrix multiplication, while (\cdot) denotes point multiplication. w represents the weight and b is the bias parameter.

$$i_t = \mu(w_i \cdot [h_{t-1}, r_t] + b_i), \quad (10a)$$

$$f_t = \mu(w_f \cdot [h_{t-1}, r_t] + b_f), \quad (10b)$$

$$C_t = \tanh(w_c \cdot [h_{t-1}, r_t] + b_c), \quad (10c)$$

$$C_t = f_t * C_{t-1} + i_t * C_t, \quad (10d)$$

$$o_t = \mu(w_o \cdot [h_{t-1}, r_t] + b_o), \quad (10e)$$

$$h_t = o_t * \tanh(C_t). \quad (10f)$$

In this work, motivated by the features of LSTM, a Bidirectional (BiLSTM) arrangement is implemented for classification tasks. The hidden state of BiLSTM at times t can be calculated by the weighted sum of the forward hidden state \vec{h}_t and the backward hidden state \overleftarrow{h}_t as follows:

$$\vec{h}_t = \text{LSTM}(r_t, \vec{h}_{t-1}), \quad (11a)$$

$$\overleftarrow{h}_t = \text{LSTM}(r_t, \overleftarrow{h}_{t-1}), \quad (11b)$$

$$U_t = w_t \vec{h}_t + v_t \overleftarrow{h}_t + b_t, \quad (11c)$$

where w_t and v_t denote the weights corresponding to \vec{h}_t and \overleftarrow{h}_t , respectively. The number of hidden units indicates the number of BiLSTM units that need to be placed in the hidden layer of the network.

The fully connected layer receives the output of the BiLSTM layer in order to increase the stability of the output by performing more non-linear operations. There are two fully connected layers for two output classes. The Softmax layer is an activation function that

calculates a probability for each sequence and sends results to the next layer. This layer contains two nodes, which is the same as the number of output classes. The utilized Softmax function can be written as [78],

$$\gamma(U_t)_j = \frac{e^{(U_t)_j}}{\sum_{i \in I} e^{(U_t)_i}}, \quad (12)$$

where $\gamma(U_t)_j = p_{n,j}$ is the probability that vector U_t is a member of j th class ($j \in \{0, 1\}$), and $I = \{0, 1\}$ is a set of all possible classes. In other words, $p_{n,0}$ and $p_{n,1}$, ($1 \leq n \leq N$), represents the probability that the transmitted chaotic signal is "0" or "1", respectively.

The goal of the training process is to minimize the categorical cross-entropy loss function,

$$L(\theta) = -\frac{1}{a_t} \sum_{t=1}^{a_t} \sum_{j=0}^1 p'_{n,j} \log(p_{n,j}), \quad (13)$$

where a_t is the mini-batch size, θ indicates set of network parameters corresponding to the different layers, $p_{n,j}$ is the Softmax's layer output probability for output class j and observation n , and $p'_{n,j} \in \{0, 1\}$ represents the binary indicator if class label j is the correct classification for observation n . A popular algorithm to obtain θ is the Stochastic Gradient Descent (SGD) method [79], which starts with a random initial value $\theta = \theta_0$, and iteratively updates θ as

$$\theta_{k+1} = \theta_k - \eta \nabla \tilde{L}(\theta_k), \quad (14)$$

where $\eta > 0$ is the learning rate, and $\tilde{L}(\theta_k)$ is an approximation of the loss function which is computed for a random mini-batch of training examples of size a_t at each iteration. Through offline training, a network with optimized weights and biases that can be used for online signal demodulation is formed. The proposed training algorithm is summarized in Algorithm 1.

Algorithm 1: Training of m th receiver.

> At the Transmitter side:

- 1: Input parameters: Number of generated signals ($2N$), Spread factor (β).
- 2: Generate two sets of the chaotic signals, each with length N , using the Chebyshev and the Logistic maps, i.e., $\{\tilde{x}^{(n)}\}_{n=1}^N$ and $\{\tilde{x}'^{(n)}\}_{n=1}^N$.
- 3: Normalize chaotic signals.
- 4: Transmit all generated signals over the channel:
 - for* $n = 1 : 2N$
 - Generate a random $\sigma_{tr}^{(n)} \in [\sigma_{tr,min}, \sigma_{tr,max}]$;
 - Generate a random $\alpha_{m,l}^{(n)}$;
 - Transmit n th symbol;

End for.

> At the receiver side:

- 5: The m th antenna receives altered signals $\{\mathbf{r}_m^{(n)}\}_{n=1}^{2N} = \{\tilde{\mathbf{r}}_m^{(n)}, \tilde{\mathbf{r}}_m'^{(n)}\}_{n=1}^N$ with known labels $\{j^{(n)}\}_{n=1}^{2N} = \{0^{(n)}, 1^{(n)}\}_{n=1}^N$.
 - 6: Separate $\mathbf{r}_m^{(n)}$ into real and imaginary parts to form the training set: $\mathcal{D}_m = \{[\Re\{\mathbf{r}_m\}, \Im\{\mathbf{r}_m\}]^{(n)}, j^{(n)}\}_{n=1}^{2N}$.
 - 7: Train m th NN-based receiver including:
 - Sequence input layer;
 - LSTM/BiLSTM layer;
 - Fully connected layer;
 - Softmax layer;
 - Classification layer;
 - 8: End of Training.
-

3.5. Decision Combining Rule

Reliable communication over multi-path channels highly depends on the condition of the paths, and the probability of deep fade. Spatial diversity is used in conventional coherent communication systems for combating the destructive effects of small-scale fading, and thereby for improving reliability. However, these receivers are very sensitive to the accuracy of the channel estimation process. Since the proposed classifier is trained under different channel conditions, the receiver does not require complex channel estimation techniques or soft data combining methods, such as Equal Gain Combining (EGC) and Maximal Ratio Combining (MRC) [80], for data detection. We can combine hard outputs of several classifiers to achieve a diversity gain through a simple decision rule.

Based on the received vector $r_{t,m}$, each of the LSTMs can produce a local decision and report this decision to a Fusion Center (FC), which makes the final decision. There are several methods of fusing the decisions of the classifiers, such as majority voting and ensemble averaging [81]. In this paper, having the binary-valued decisions, the FC applies the majority voting fusion rule to generate the ultimate decision. The class with the highest overall output is selected as the ultimate decision. Mathematically, the decision class $O(\mathbf{r})$ can be calculated as [82]

$$O(\mathbf{r}) = \arg \max_j \sum_{m=1}^M \mathcal{O}_{j,m} (C_m(\mathbf{r}) = j), \quad (15)$$

where M is the number of classifiers, $j \in \{0, 1\}$ denotes the j th class, $C_m(\mathbf{r})$ represents the output of the m th classifier for the received vector \mathbf{r} , and $\mathcal{O}_{j,m}$ is a binary characteristic function that can be defined as

$$\mathcal{O}_{j,m} = \begin{cases} 1 & \text{if } C_m(\mathbf{r}) = j \\ 0 & \text{if } C_m(\mathbf{r}) \neq j. \end{cases} \quad (16)$$

4. Simulation Results

In this section, we first provide a comparison between the BER performance of the Single-input Single-output (SISO) DLCSK and the conventional DCSK over AWGN and multi-path Rayleigh fading channels. Then, evaluation of various parameters such as bifurcation parameters and the number of antennas on the system performance takes place. In all simulations the classifiers are trained only at a limited SNR range, i.e., $\sigma_{tr}^{(n)} \in [\sigma_{tr,min}, \sigma_{tr,max}]$ dB, while tested over a wide range of E_b/N_0 values greater than 0 dB.

In order to generate chaotic signals, we choose two discrete time recursive functions, i.e., Chebyshev and Logistic maps, which have been used extensively in practical communication systems [40,83]. The transmitter creates two sets of chaotic signals using these two maps during the training phase, according to the values in Table 1. For example, in case 1, each of these sets includes the number of $N = 2000$ chaotic signals, each with length $\beta = 50$ samples. In other words, we use two sets of the training chaotic signals, each with total length $S = 50 \times 2000 = 10^5$ samples. All of the chaotic signals have equivalent class labels as $j \in \{0, 1\}$. The receiver obtains the corresponding altered signals and forms the final training set as input of the NN. Therefore, after the number of $2 \times 2000 = 4000$ signal transmissions, we will obtain a training set, consisting of 4000 received signals along with their corresponding class labels j . In the test phase, a binary "0" will be sent, transmitting a Chebyshev map, and if "1" is to be sent, a Logistic map is transmitted.

In our simulations, to evaluate the system's generalization capability and robustness against parameter mismatch, we use different parameter settings for the training and testing phases. Therefore, to generate Logistic maps, the bifurcation parameters are, respectively, chosen as $\rho_{train} = 3.6$ and $\rho_{test} = 3.3$ for the training and the deployment phases, with an initial state as $\check{x}_1 = 0.3$. To generate Chebyshev maps, initial states are chosen as $\check{x}_1 = 0.3535$ and $\check{x}_1 = 0.3$ for the training and the deployment phases, respectively.

Predefined functions of the MATLAB Neural Network Toolbox can be used to define an LSTM network and specify training options, including Learning Rate (η), and the number of hidden layers (H). In all experiments, the learning rate is set to $\eta = 0.01$. The other DL parameters are selected based on the SIMO DLCSK parameters. Although we consider the AWGN and Rayleigh fading channels, SIMO DLCSK can also be applied to any other channel model as well. The considered network parameters are listed in Table 1.

Table 1. List of the parameters.

Variable	Description	Case 1	Case 2
S	Number of training samples (for each class)	10^5	10^4
N	Number of training signals (for each class)	2000	200
β	Length of chaotic signals	50	50
w_v	Number of validation samples (for each class)	3×10^3	10^3
Z	Number of test signals	10^4	10^4
a_t	Mini-batch size	200	200
H	Number of hidden units	100	10
E	Number of epochs	20	500
η	Learning Rate	10^{-2}	10^{-2}

4.1. Training Convergence

Figure 3, illustrates the training metrics of a single classifier under the AWGN channel and how the training is performed for a random training SNR value $\sigma_{tr}^{(n)} \in [19, 23]$ dB. Each epoch is a full pass through the dataset, and each iteration is an update of the network parameters. Figure 3 also shows validation metrics, which are recorded each time the program validates the network. For example, the validation accuracy presents the classification accuracy on the validation set, which gives an idea of the generalization of the model. Final validation metrics are labeled “Final” in the plot.

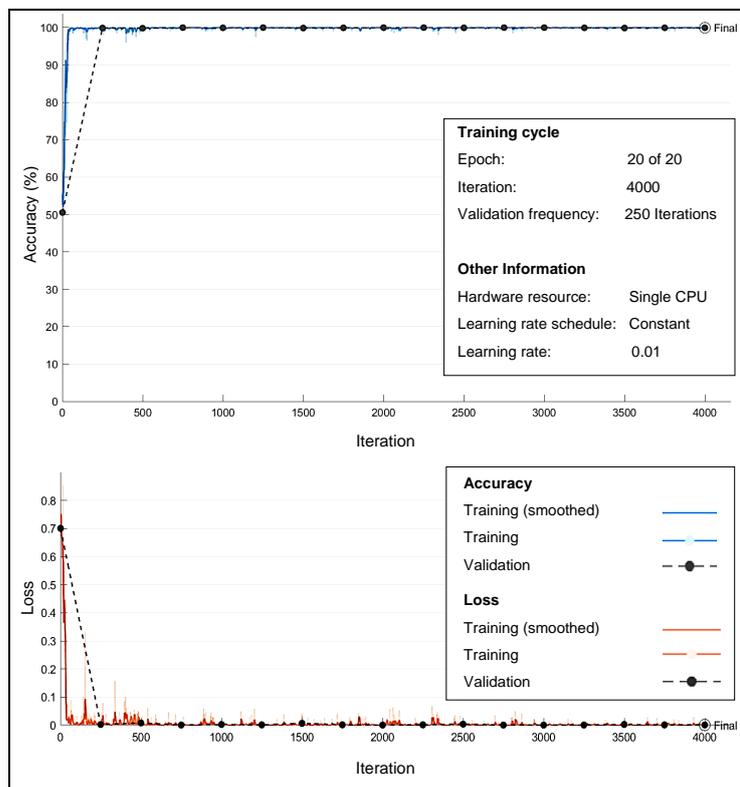


Figure 3. Training metrics when the training SNR $\sigma_{tr}^{(n)} \in [19, 23]$ dB.

The training accuracy represents the classification accuracy of each mini-batch during the training process. The smoothed training accuracy is less noisy than the training accuracy and makes it easier to observe trends. The training loss shows the value of the loss function, i.e., the categorical cross-entropy, on each mini-batch. The results show that the loss function converges rapidly within about 100 iterations.

When the DLCSK system is trained at relatively lower SNRs (For instance, when $\sigma_{tr}^{(n)} \in [11, 15]$ dB), training accuracy converges more slowly. Since, the chaotic signals contain more stochastic features, the fluctuations of the curve and difficulty of training increases. However, for $\sigma_{tr}^{(n)} \in [11, 15]$ dB, the model can achieve a relatively good accuracy rate, and converge to a small final error.

4.2. BER Performance under AWGN Channel

Figure 4 shows the simulated results of the BER performance of the DLCSK, non-coherent DCSK, chaotic switching CSK, and antipodal CSK over the AWGN channel. It is noteworthy that the simulated coherent CSK systems are plotted assuming a perfect chaotic synchronization and only provide benchmark data for evaluation. Antipodal CSK is a chaotic modulation scheme with one basis function that can theoretically achieve the BER performance of binary phase shift keying (BPSK) under AWGN channels. In our simulations, the Logistic map is used for antipodal CSK, such that if “0” is to be sent, a chaotic signal \check{x} is transmitted, and if “1” is to be sent, $-\check{x}$ is transmitted. The results are also compared with those of chaotic switching, a special case of CSK with two basis functions, in which the transmitted samples are obtained from the Chebyshev and the Logistic maps. The BER curve of the chaotic switching scheme is also simulated assuming a correlation receiver with perfect chaotic synchronization.

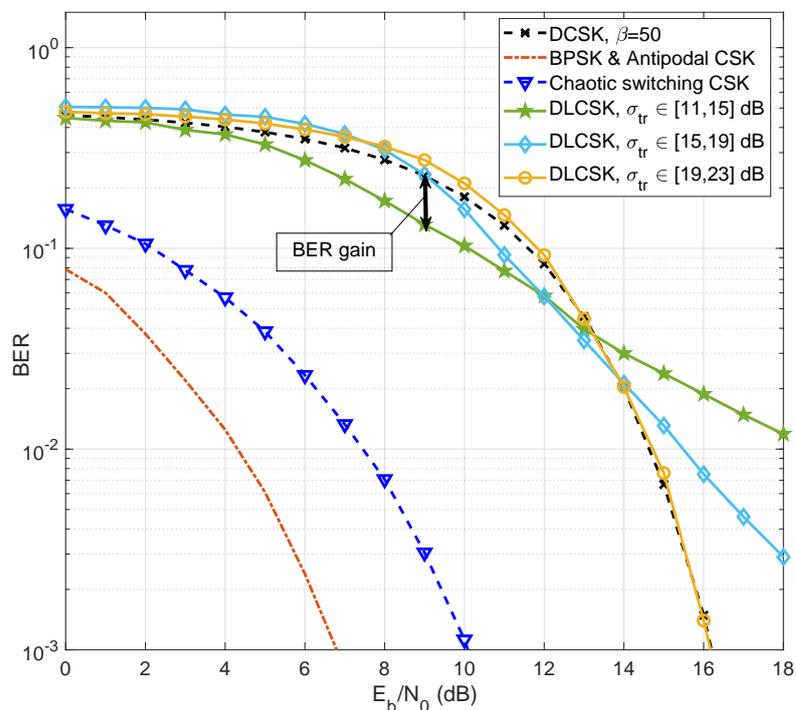


Figure 4. BER curve of the SISO DLCSK under AWGN channel for different $[\sigma_{tr,min}, \sigma_{tr,max}]$, $\beta = 50$, in comparison with the benchmark modulations.

The proposed SISO DLCSK scheme shows a gain compared to conventional DCSK in the lower SNRs, when the DLCSK system is trained at relatively lower SNR values (Training SNR $\sigma_{tr}^{(n)} \in [11, 15]$ dB). Since the auto and cross correlation properties of the chaotic signals are similar in this experiment, according to Equation (3), this gain comes from reducing the

third term, i.e., $\int_0^T n\tilde{x} dt$. When the training process is performed under different channel conditions, NN can indirectly estimate noise distribution, for use in iterative minimization of the cross-entropy cost function. As an important result, DLCSK shows a more robust behaviour in the test phase.

When σ_{tr} is relatively high, the LSTM can only grasp the clean signal. For proper training, the SNR value must help the LSTM learn both clear and noisy samples. The results show that under the AWGN channels, when the $\sigma_{tr}^{(n)} \in [19, 23]$ dB, the performance of the SISO DLCSK is close to the conventional DCSK system. Therefore, for better BER performance in high-SNR conditions, the receiver should be trained at higher SNRs. These results seem trustworthy because the above-mentioned training options allow us to design a receiver with a flexible data rate depending upon specific channel conditions.

Figure 5 depicts Monte Carlo simulations of the BER performances obtained from the SISO DLCSK system for $\beta = 50$ under an AWGN channel. The DLCSK system is trained over a limited SNR range, i.e., $[19, 23]$ dB, and tested over the whole SNR range. The other parameters for both case 1 and 2 are selected according to Table 1. The results show that the BER performance of DLCSK has a low sensitivity to changes in the hyper-parameters, such as the number of training samples S and the number of hidden units H . There is a trade-off between security and classification accuracy. When a system is trained with $\rho_{train} = 3.6$, using $\rho_{test} = 3.6$ results in more non-periodic behavior. However, reducing the ρ_{test} may result in a better BER performance.

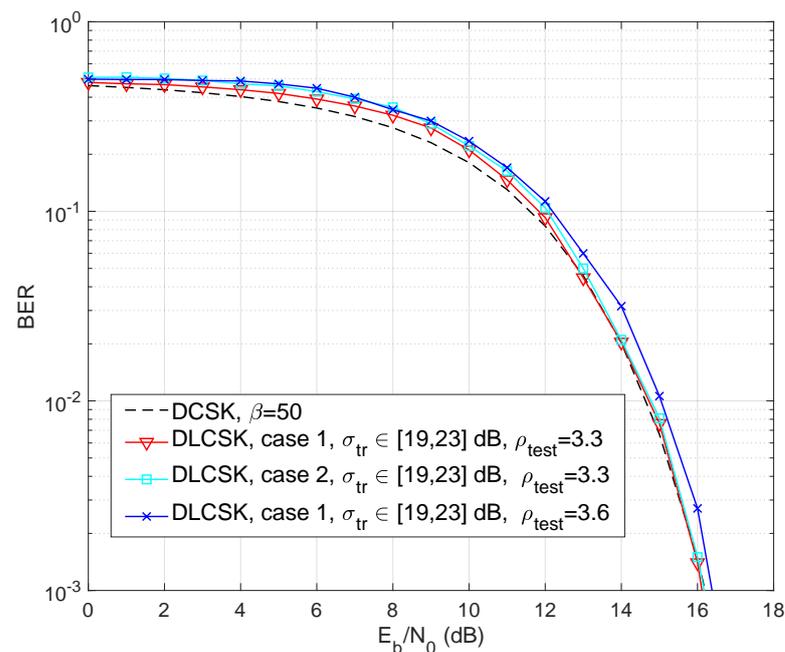


Figure 5. BER curve of the SISO DLCSK under AWGN channel, trained at a random SNR $\in [19, 23]$ dB, and $\beta = 50$, for case 1, and case 2. There is also the results for $\rho_{test} \in \{3.3, 3.6\}$.

4.3. Confusion Matrix, Sensitivity, and Specificity

We provide the confusion matrix of the proposed method, which helps in analyzing the performance of our classification algorithm. Figure 6 depicts an example of two confusion matrices with 10,000 chaotic symbols for AWGN channels for $\beta = 50$, $\sigma_{tr}^{(n)} \in [19, 23]$ dB, and test SNRs = $\{16, 23\}$ dB.

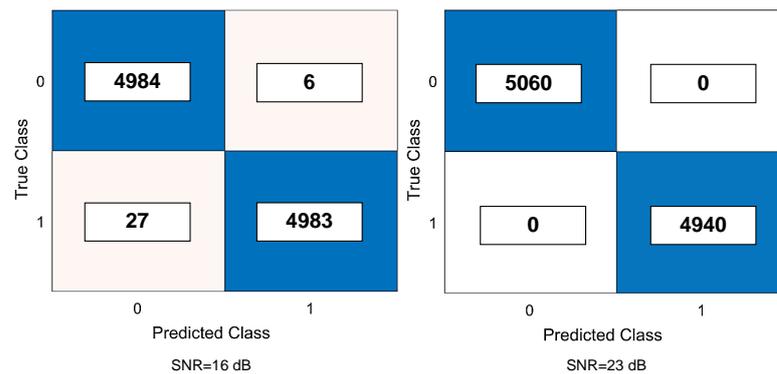


Figure 6. Confusion chart of the LSTM-based classifier under AWGN channel for $\beta = 50$, $\sigma_{tr}^{(n)} \in [19, 23]$ dB, and test SNRs = $\{16, 23\}$ dB.

In Figure 7, the sensitivity and specificity measures extracted from the confusion matrix are introduced. Both are statistical measures for the performance of a binary classification test that are widely used in the literature [84,85]. The sensitivity or True Positive Rate (TPR) measures the proportion of Logistic maps that are correctly identified. The specificity or True Negative Rate (TNR) measures the proportion of Chebyshev maps that are correctly identified. The terms "True Positive (TP)", "False Positive (FP)", "True Negative (TN)", and "False Negative" address the correctness of a classification test. For example, if the condition is sending the signals generated by the Logistic map, "TP" means "correctly predicted as Logistic map", "FP" means "incorrectly predicted as Logistic map", "TN" means "correctly predicted as Chebyshev map", and "FN" means "incorrectly predicted as Chebyshev map".

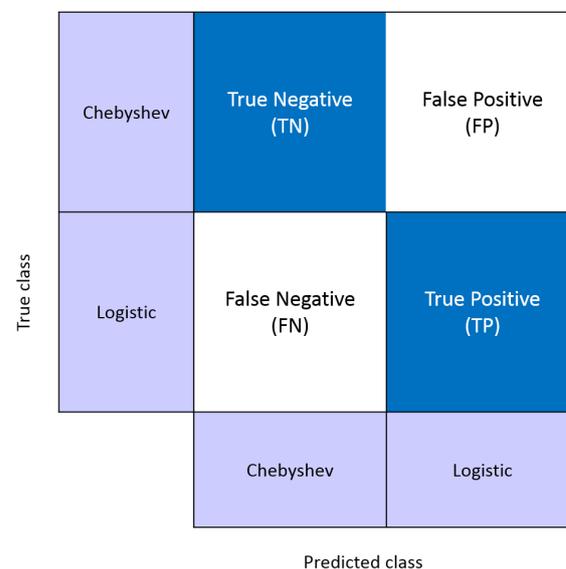


Figure 7. Calculation of sensitivity and specificity from the confusion matrix.

Figure 8 shows the sensitivity and specificity of different bifurcation parameters. The sensitivity curve in this test shows how capable LSTM can classify samples that are generated by the Logistic maps. Sensitivity can also be referred to as the recall or hit rate [86]. It is the percentage of true Logistic maps out of all samples generated by the Logistic maps (i.e., $TP/(TP + FN)$). The specificity is a measure of how well the test can identify true Chebyshev maps, which can be expressed as $TN/(TN + FP)$. For high SNRs (i.e., $SNR > 16$ dB), even using a single LSTM-based classifier results in sensitivity and specificity rates $> 95\%$. However, there is a trade-off between sensitivity and specificity for low SNRs, where the curves show low sensitivity and high specificity. For example, when

SNR = 12 dB, $\rho_{train} = 3.6$, and $\rho_{test} = 3.3$, the sensitivity of a single LSTM-based classifier is 65.69% and its specificity is 99.96%, its false negatives and false positives rates are 34.31% and 0.04%, respectively.

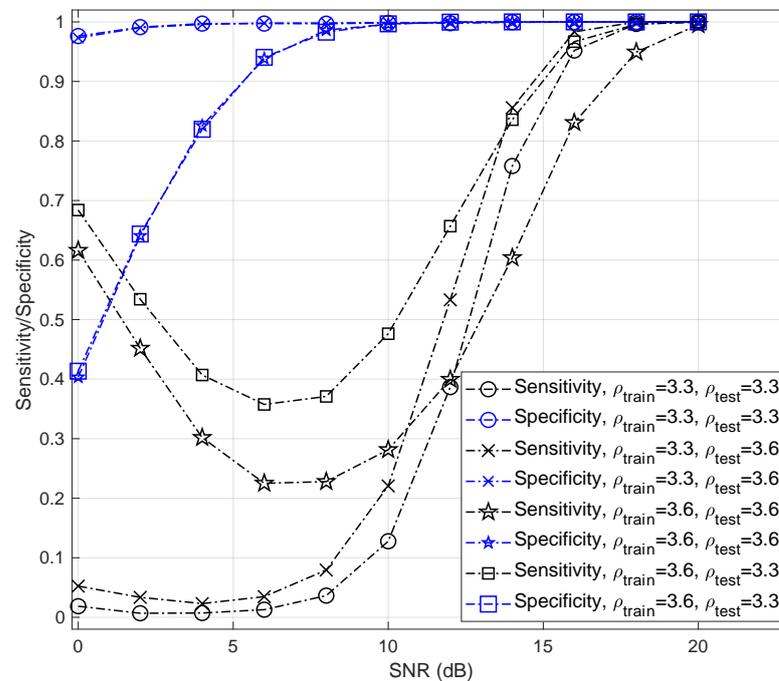


Figure 8. Sensitivity and specificity of LSTM-based classifier under AWGN channel for different values of bifurcation parameters, $\beta = 50$, and $\sigma_{tr}^{(n)} \in [19, 23]$ dB.

Sensitivity values contain useful information about the receiver's performance that can be utilized for different goals. Here, we evaluate the effect of the changes in bifurcation parameter on the classifier's performance. For example, $\rho_{train} = 3.6$ and $\rho_{test} = 3.3$ leads to a better classification accuracy than other parameter settings for SNR < 14 dB. However, with $\rho_{test} = 3.3$, the Logistic map shows a periodic behavior and this change can bring negative effect on security.

Figure 9 presents the performance obtained for case 1 using different numbers of training epochs (E). The system achieves its peak when $E = 20$. This Figure also measures the sensitivity to the number of epochs. It can be observed that the SISO DLCSK system is relatively robust to the number of epochs. For example, for a target BER performance of 10^{-2} , there is a 1 dB gap between the worst-case and the best-case scenarios.

4.4. BER Performance under Multi-Path Rayleigh Fading Channels

Figure 10 illustrates a comparison between the BER performance of the SISO DLCSK and conventional DCSK over multi-path Rayleigh fading channels, for $\beta = 50$ and $\sigma_{tr}^{(n)} \in [19, 23]$ dB. We consider two cases corresponding to different path gain ratios and different path delays. In the first case, a two-path channel is considered ($L = 2$) in which the two paths have similar average power gain. In this case, the average power gain in each path is 0.5, (i.e., $E(\alpha_1^2) = E(\alpha_2^2) = 0.5$), with $\tau_1 = 0$, and $\tau_2 = 2$. In the second case, three paths ($L = 3$) are considered with different average power gains. The average power gains are $E(\alpha_1^2) = 1/7$, $E(\alpha_2^2) = 2/7$, and $E(\alpha_3^2) = 4/7$ with $\tau_1 = 0$, $\tau_2 = 3$, and $\tau_3 = 6$. The average power gain of the third path is 3 dB below the second path, and for the second path it is 3 dB below the first path. When the receiver is already accustomed to chaotic signals, we do not need to transmit a reference signal, it means less energy will be used to transmit one bit. As shown in Figure 10, the SISO DLCSK overcomes the BER inefficiency of the existing DCSK under Rayleigh fading channels. For further benchmark comparisons, we consider the antipodal CSK system. The BER performance of DLCSK is close to antipodal

CSK when $E_b/N_0 \leq 10$ dB, and both systems have similar performance when E_b/N_0 is more than 10 dB.

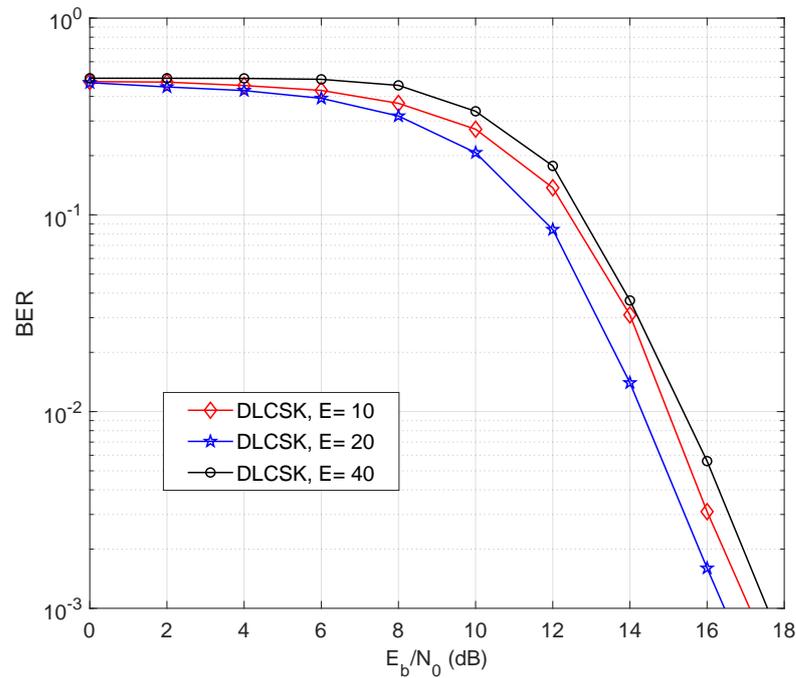


Figure 9. BER of SISO DLCSK under AWGN channel, for $\beta = 50$ and different numbers of training epochs (E).

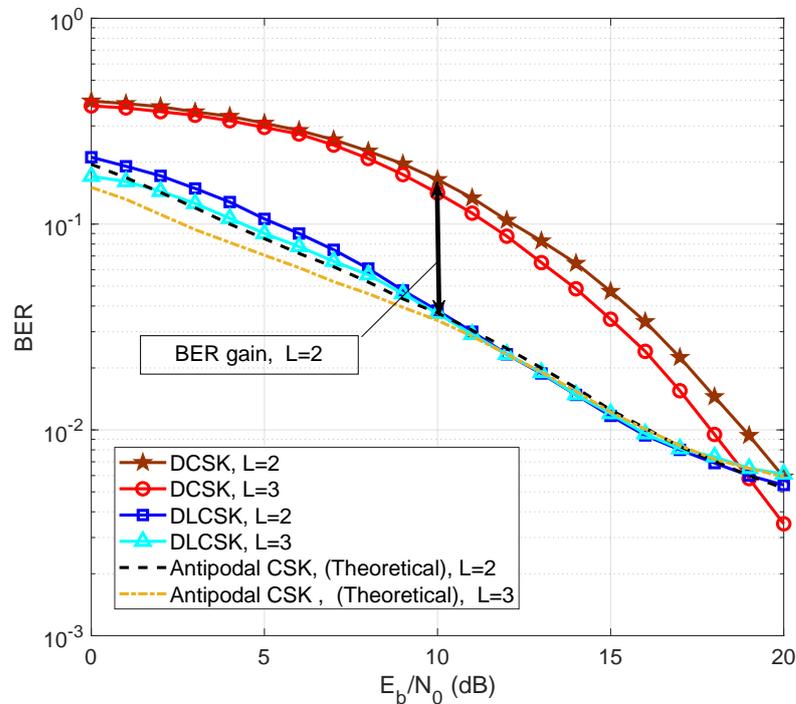


Figure 10. BER of SISO DLCSK for two-path and three-path Rayleigh fading channels, $\sigma_{tr}^{(n)} \in [19, 23]$ dB and $\beta = 50$, in comparison with the benchmark modulations.

Figure 11 compares the BER performance of the SIMO DLCSK with the SISO-DLCSK under Rayleigh fading channels. The number of receive antennas is $M = \{1, 3, 5\}$, and the spreading factor is $\beta = 50$. The results show that the SIMO DLCSK attains a better BER performance due to the diversity gain. For example, when $M = 1$, to achieve the target

$BER = 10^{-2}$, the required SNR will be about 20 dB. If we increase the number of antennas, i.e., $M = 5$, the required SNR to achieve the $BER = 10^{-2}$ will only be about 15 dB. Figure 11 also compares the BER performance of the SIMO DLCSK with a recently published LSTM-aided OFDM DCSK in [40]. The LSTM-aided NN calculates the correlations between chaotic modulated OFDM-DCSK signals in order to retrieve the data. For a fair comparison, we consider a SIMO DLCSK with $M = 5$ and an LSTM aided OFDM DCSK with $K = 5$, where K indicates the number of independent sub-channels in such a system. The SIMO DLCSK with $M = 5$ independent sub-channels can be fairly compared to LSTM-aided OFDM DCSK with $K = 5$ sub-channels. The LSTM-aided receiver sends reference signals that do not carry useful information. The proposed DLCSK scheme shows an expected gain compared to the LSTM-aided OFDM DCSK because of using indirect channel estimation and reducing the burden of the additional reference sample transmission.

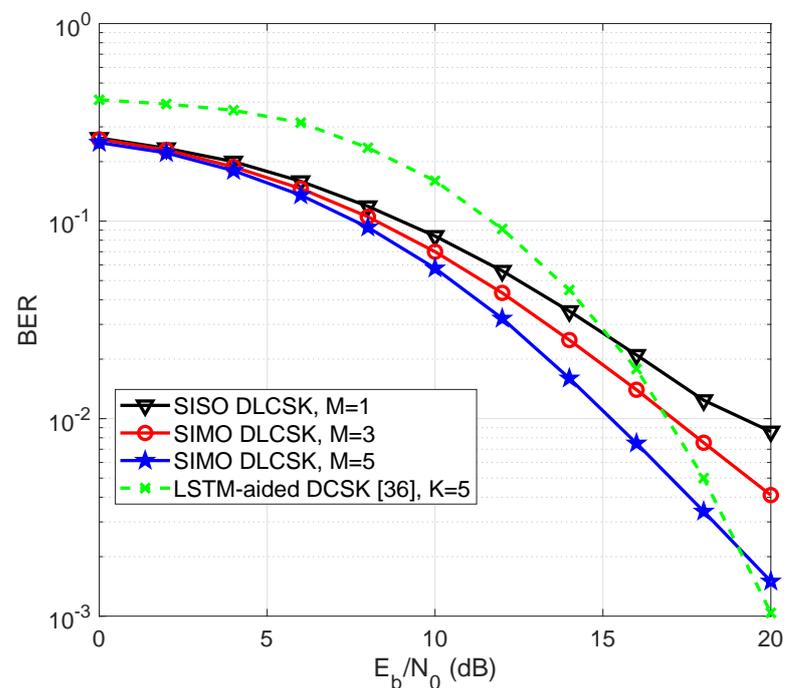


Figure 11. BER curve of the SIMO DLCSK under Rayleigh fading channel, for $\beta = 50$, in comparison with the LSTM-aided OFDM DCSK [40], for $K = 5$.

5. Conclusions

This paper represents an attempt to develop a new generation of chaos-based communication systems based on DL. We have introduced a trainable DLCSK receiver, which does not need any reference signal transmission or chaotic synchronization. Thus, the DLCSK is more practical, in terms of reliability and compatibility with modern communication infrastructure. A multi-antenna design is presented to achieve a diversity gain. The main objective of the multi-antenna receiver is to improve the classification accuracy of the individual classifiers. Simulation results verify that the SIMO DLCSK system provides an excellent BER performance without channel estimation and complex combining modules at the receiver side. The proposed architecture is an appealing candidate for next generation wireless communications, such as massive MIMO systems and cloud/edge-based communications.

Author Contributions: Conceptualization, G.K. and M.H.; methodology, G.K., M.H. and M.M.; software, M.H. and M.M.; validation, M.H. and G.K.; formal analysis, G.K. and M.H.; investigation, G.K. and M.H.; resources, G.K.; data curation, G.K.; writing—original draft preparation, M.M.; writing—review and editing, G.K., M.H.; visualization, M.M.; supervision, G.K. and M.H.; project administration, G.K. and M.H.; funding acquisition, G.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Tier2 Canada research chair entitled ‘Towards a Novel and Intelligent Framework for the Next generations of IoT Networks.’

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Harikrishnan, N.B.; Nagaraj, N. When Noise meets Chaos: Stochastic Resonance in Neurochaos Learning. *Neural Netw.* **2021**, *143*, 425–435.
2. Vuong, D.P.; Le, D.K.; Nguyen, K.K.; Van Nguyen, B. Correlation receiver with nonlinearity blanking for DCSK systems under pulse jamming attack. *IEEE Access* **2019**, *7*, 25037–25045. [[CrossRef](#)]
3. Cai, G.; Song, Y. Closed-form BER expressions of M-ary DCSK systems over multipath Rayleigh fading channels. *IEEE Commun. Lett.* **2020**, *24*, 1192–1196. [[CrossRef](#)]
4. Chen, P.; Fang, Y.; Su, K.; Chen, G. Design of a capacity-approaching chaos-based multi-access transmission system. *IEEE Trans. Veh. Technol.* **2017**, *66*, 10806–10816. [[CrossRef](#)]
5. Mazzini, G.; Setti, G.; Rovatti, R. Chaotic complex spreading sequences for asynchronous DS-CDMA—Part I: System modeling and results. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **1997**, *44*, 937–947. [[CrossRef](#)]
6. Dedieu, H.; Kennedy, M.P.; Hasler, M. 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.* **1993**, *40*, 634–642. [[CrossRef](#)]
7. Parlitz, U.; Chua, L.O.; Kocarev, L.; Halle, K.S.; Shang A. Transmission of digital signals by chaotic synchronization. *Int. J. Bifurc. Chaos* **1992**, *2*, 973–977. [[CrossRef](#)]
8. Kolumbán, G.; Kennedy, M.P.; Chua, L.O. The role of synchronization in digital communications using chaos. I. Fundamentals of digital communications. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **1997**, *44*, 927–936. [[CrossRef](#)]
9. Kennedy, M.P.; Kolumbán, G.; Kis, G.; Jákó, Z. Performance evaluation of FM-DCSK modulation in multipath environments. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2000**, *47*, 1702–1711. [[CrossRef](#)]
10. Kennedy, M.P.; Kolumbán, G.; Jákó, Z.; Kis, G. Chaotic communications with correlator receivers: Theory and performance limits. *Proc. IEEE* **2002**, *90*, 711–732.
11. Schweizer, J.; Schimming, T. Symbolic dynamics for processing chaotic signal—II. Communication and coding. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2001**, *48*, 1283–1295. [[CrossRef](#)]
12. Mazzini, G.; Rovatti, R.; Setti, G. Sequence synchronization in chaos-based DS-CDMA systems. *Proc. IEEE Int. Symp. Circuits Syst. (IsCAS)* **1998**, *44*, 937–947.
13. Jovic, B.; Unsworth, C.P.; Sandhu, G.S.; Berber, S.M. A robust sequence synchronization unit for multi-user DS-CDMA chaos-based communication systems. *Signal Process.* **2007**, *87*, 1692–1708. [[CrossRef](#)]
14. Kaddoum, G.; Roviras, D.; Chargé, P.; Fournier-Prunaret, D. Robust synchronization for asynchronous multi-user chaos-based DS-CDMA. *Signal Process.* **2009**, *89*, 807–818. [[CrossRef](#)]
15. Yeo, K.; Melnyk, I. Deep learning algorithm for data-driven simulation of noisy nonlinear dynamical system. *J. Comput. Phys.* **2019**, *376*, 1212. [[CrossRef](#)]
16. Burel, G.; Boudier, C. Blind estimation of the pseudo-random sequence of a direct sequence spread spectrum signal. In Proceedings of the 21st Century Military Century Military Communication (MILCOM), Los Angeles, CA, USA, 22–25 October 2000; Volume 2, pp. 967–970.
17. Tam, W.M.; Lau, F.C.M.; Tse, C.K.; Lawrance, A.J. Exact analytical bit error rates for multiple access chaos-based communication systems. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2004**, *376*, 473–481.
18. Kolumbán, G.; Vizvári, G.K.; Schwarz, W.; Abel, A. Differential chaos shift keying: A robust coding for chaos communication. In Proceedings of the Workshop Non-Linear Dynamics Electronic Systems (NDES), Seville, Spain, 27–28 June 1996; pp. 92–97.
19. Kis, G.; Jákó, Z.; Kennedy, M.P.; Kolumbán, G. Chaotic communications without synchronization. In Proceedings of the 6th IEE Conference on Telecommunications, 29 March–1 April 1998; pp. 49–53.
20. Kennedy, M.P.; Rovatti, R.; Setti, G. *Chaotic Electronics in Telecommunications*; RC Press: London, UK, 2000.
21. Mazzini, G.; Rovatti, R.; Setti, G. Interference minimisation by autocorrelation shaping in asynchronous DS-CDMA systems: Chaos-based spreading is nearly optimal. *Electron. Lett.* **1999**, *35*, 1054–1055. [[CrossRef](#)]
22. Galias, Z.; Maggio, G.M. Quadrature chaos shift keying: Theory and performance analysis. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2001**, *48*, 1510–1519. [[CrossRef](#)]

23. Yang, H.; Jiang, G.-P. High-efficiency differential-chaos-shift-keying scheme for chaos-based noncoherent communication. *IEEE Trans. Circuits Syst. II Exp. Briefs* **2012**, *59*, 312–316. [[CrossRef](#)]
24. Kaddoum, G.; Shokraneh, F. High-efficiency Analog network coding for multiuser multi-carrier differential chaos shift keying communication system. *IEEE Trans. Wireless Commun.* **2015**, *14*, 1492–1505. [[CrossRef](#)]
25. Kaddoum, G. Design and performance analysis of a multi-user OFDM based differential chaos shift keying communication system. *IEEE Trans. Commun.* **2015**, *64*, 1–12.
26. Taleb, F.; Bendimerad, F.T.; Roviras, D. Very high efficiency differential chaos shift keying system. *IET Commun.* **2016**, *10*, 2300–2307. [[CrossRef](#)]
27. Cao, X.; Liu, L.; Cheng, Y.; Shen, X. Towards energy-efficient wireless networking in the big data era: A survey. *IEEE Commun. Surveys Tuts.* **2018**, *20*, 303–332. [[CrossRef](#)]
28. Dai, L.; Jiao, R.; Adachi, F.; Vincent Poor, H.; Hanzo, L. Deep learning for wireless communications: An emerging interdisciplinary paradigm. *IEEE Wirel. Commun.* **2020**, *27*, 133–139. [[CrossRef](#)]
29. Merima, K.; Kazaz, T.; Moerman, I.; De Poorter, E. End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications. *IEEE Access* **2018**, *6*, 18484–18501.
30. Nariman, F.; Goldsmith, A. Neural network detection of data sequences in communication systems. *IEEE Trans. Signal Process.* **2018**, *66*, 5663–5678.
31. Ye, H.; Li, G.Y.; Juang, B.H. Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 114–117. [[CrossRef](#)]
32. Zha, X.; Peng, H.; Qin, X.; Li, G.; Yang, S. A deep learning framework for signal detection and modulation classification. *Sensors* **2019**, *19*, 4042. [[CrossRef](#)]
33. Liu, C.; Wei, Z.; Ng, D.W.K.; Yuan, J.; Liang, Y.C. Deep Transfer Learning for Signal Detection in Ambient Backscatter Communications. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 1624–1638. [[CrossRef](#)]
34. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1–40. [[CrossRef](#)]
35. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In Proceedings of the International Conference on Artificial Neural Networks (ICANN), Rhodes, Greece, 4–7 October 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 270–279.
36. Pathak, J.; Hunt, B.; Girvan, M.; Lu, Z.; Ott, E. Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach. *Phys. Rev. Lett.* **2018**, *120*, 024102. [[CrossRef](#)]
37. Mobini, M.; Kaddoum, G. Deep Chaos Synchronization. *IEEE Open J. Commun. Soc.* **2020**, *1*, 1571–1582. [[CrossRef](#)]
38. Zhang, H.; Zhang, L.; Cheng, J.; Jiang, Y.; Wu, Z. An Intelligent Detection Based on Deep Learning for Multilevel Code Shifted Differential Chaos Shift Keying System With M-ary Modulation *IEEE Trans. Cogn. Commun.* **2021**, *66*, 1451–1455. [[CrossRef](#)]
39. Chen, B.; Zhang, L.; Wu, Z. General iterative receiver design for enhanced reliability in multi-carrier differential chaos shift keying systems. *IEEE Trans. Commun.* **2020**, *67*, 7824–7839. [[CrossRef](#)]
40. Zhang, L.; Zhang, C.; Jiang, Y.; Wu, Z. Intelligent and Reliable Deep Learning LSTM Neural Networks-based OFDM-DCSK Demodulation Design. *IEEE Trans. Veh. Technol.* **2020**, *69*, 16163–16167. [[CrossRef](#)]
41. Kaddoum, G.; Soujeri, E.; Arcila, C.; Eshteiwi, K. I-DCSK: An improved non-coherent communication system architecture. *IEEE Trans. Circuits Syst. II Exp. Briefs* **2015**, *62*, 1–5.
42. Wang, S.; Lu, S.; Zhang, E. MIMO-DCSK communication scheme and its performance analysis over multipath fading channels. *J. Syst. Eng. Electron.* **2013**, *24*, 729–733. [[CrossRef](#)]
43. Fang, Y.; Xu, J.; Wang, L.; Chen, G. Performance of MIMO relay DCSK-CD systems over Nakagami fading channels. *IEEE Trans. Circuits Syst. I Reg. Papers* **2013**, *3*, 757–767. [[CrossRef](#)]
44. Kumar, A.; Sahu, P.R. Performance analysis of DCSK-SR systems based on best relay selection in multiple MIMO relay environment. *AEU Int. J. Electron. Commun.* **2016**, *70*, 18–24. [[CrossRef](#)]
45. Zhang, G.; Zhao, C.; Zhang, T. Performance analysis of MISO-MU-OHE-DCSK system over Rayleigh fading channels. *AEU Int. J. Electron. Commun.* **2020**, *115*, 153048. [[CrossRef](#)]
46. Ahmadinejad, A.; Talebi, S. Performance evaluation of chaotic spreading sequences in a multi-user MIMO-OFDM system. *Phys Commun.* **2016**, *19*, 7–11. [[CrossRef](#)]
47. Kaddoum, G.; Vu, M.; Gagnon, F. Performance analysis of differential chaotic shift keying communications in MIMO systems. In Proceedings of the 2011 IEEE International Symposium of Circuits and Systems (ISCAS), Rio de Janeiro, Brazil, 15–18 May 2011; pp. 1580–1583.
48. Ma, M.A.; Kan, H.B. Space-time coding and processing with differential chaos shift keying scheme. In Proceedings of the IEEE International Conference on Communications (ICC), Dresden, Germany, 14–18 June 2009; pp. 1–5.
49. Wang, L.; Min, X.; Chen, G. Performance of SIMO FM-DCSK UWB system based on chaotic pulse cluster signals. *IEEE Trans. Circuits Syst. I Reg. Papers* **2011**, *58*, 2259–2268. [[CrossRef](#)]
50. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM fully convolutional networks for time series classification. *IEEE Access* **2017**, *6*, 1662–1669. [[CrossRef](#)]
51. Ma, H.; Cai, G.; Fang, Y.; Wen, J.; Chen, P.; Akhtar, S. A new enhanced energy-detector-based FM-DCSK UWB system for tactile Internet. *IEEE Trans. Ind. Inform.* **2019**, *15*, 3028–3039. [[CrossRef](#)]
52. Chen, P.; Shi, L.; Fang, Y.; Cai, G.; Wang, L.; Chen, G. A coded DCSK modulation system over Rayleigh fading channels. *IEEE Trans. Commun.* **2018**, *66*, 3930–3942. [[CrossRef](#)]

53. Xu, W.; Tan, Y.; Lau, F.C.M.; Kolumbán, G. Design and optimization of differential chaos shift keying scheme with code index modulation. *IEEE Trans. Commun.* **2018**, *66*, 1970–1980. [[CrossRef](#)]
54. Chen, P.; Wang, L.; Lau, F.C.M. One Analog STBC-DCSK Transmission Scheme not Requiring Channel State Information. *IEEE Trans. Circuits Syst. I Reg. Papers* **2013**, *60*, 1027–1037. [[CrossRef](#)]
55. Rodriguez, J.J.; Kuncheva, L.I.; Alonso, C.J. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630. [[CrossRef](#)] [[PubMed](#)]
56. Ke, R.; Li, Z.; Tang, J.; Pan, Z.; Wang, Y. Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 54–64. [[CrossRef](#)]
57. Saqlain, M.; Jargalsaikhan, B.; Lee, J.Y. A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 171–182. [[CrossRef](#)]
58. Sadeghi, M.; Larsson, E.G.; Lee, J.Y. Physical adversarial attacks against end-to-end autoencoder communication systems. *IEEE Commun. Lett.* **2019**, *23*, 847–850. [[CrossRef](#)]
59. Mobini, M.; Zahabi, M.R. Masking communication using Sprott94 case A chaotic system in AWGN channel. *J. World's Electr. Eng. Technol.* **2018**, *7*, 9–16.
60. Fu, J.S.; Liu, Y.; Chao, H.C.; Bhargava, B.K.; Zhang, Z.J. Secure data storage and searching for industrial IoT by integrating fog computing and cloud computing. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4519–4528. [[CrossRef](#)]
61. Ye, H.; Li, G.Y.; Juang, B.H.F. Deep reinforcement learning based resource allocation for V2V communications. *IEEE Trans. Ind. Inform.* **2019**, *68*, 3163–3173. [[CrossRef](#)]
62. Liang, L.; Peng, H.; Li, G.Y.; Shen, X. Vehicular communications: A physical layer perspective. *IEEE Trans. Veh. Technol.* **2017**, *66*, 10647–10659. [[CrossRef](#)]
63. Feng, D.; Lu, L.; Yuan-Wu, Y.; Li, G.Y.; Li, S.; Feng, G. Device to-device communications in cellular networks. *IEEE Commun. Mag.* **2014**, *52*, 49–55. [[CrossRef](#)]
64. Kim, K.; Lee, J.; Choi, J. Deep learning based pilot allocation scheme (DL-PAS) for 5G massive MIMO system. *IEEE Commun. Lett.* **2018**, *22*, 828–831. [[CrossRef](#)]
65. Liu, Z.; Zhang, L.; Wu, Z.; Bian, J. A secure and robust frequency and time diversity aided OFDM-DCSK modulation system not requiring channel state information. *IEEE Trans. Commun.* **2019**, *68*, 1684–1697. [[CrossRef](#)]
66. Sanguinetti, L.; Björnson, E.; Hoydis, J. Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination. *IEEE Trans. Commun.* **2019**, *68*, 232–257. [[CrossRef](#)]
67. Chen-Hu, K.; Liu, Y.; Armada, Y. Toward Non-Coherent Massive MIMO-OFDM Down-Link based on Differential Modulation. *IEEE Trans. Veh. Technol.* **2020**, *69*, 11281–11294. [[CrossRef](#)]
68. Teng, L.; Iu, H.H.; Wang, X.; Wang, X. Novel chaotic behavior in the Muthuswamy–Chua system using Chebyshev polynomials. *Int. J. Numer. Model. Electron. Netw. Devices Fields* **2015**, *28*, 275–286. [[CrossRef](#)]
69. Wang, Y.; Liu, Z.; Ma, J.; He, H. A pseudorandom number generator based on piecewise logistic map. *Nonlinear Dyn.* **2016**, *83*, 2373–2391. [[CrossRef](#)]
70. Heidari-bateni, G.; McGillem, C.D. Chaotic directsequence spread-spectrum communication system. *IEEE Trans. Commun.* **1994**, *42*, 1524–1527. [[CrossRef](#)]
71. Miliou, A.N.; Antoniadis, I.P.; Stavrinos, S.G.; Anagnostopoulos, A.N. Secure communication by chaotic synchronization: Robustness under noisy conditions. *Nonlinear Anal. Real World Appl.* **2007**, *8*, 1003–1012. [[CrossRef](#)]
72. Herceg, M.; Kaddoum, G.; Vranješ, D.; Soujeri, E. Permutation index DCSK modulation technique for secure multiuser high-data-rate communication systems. *IEEE Trans. Veh. Technol.* **2017**, *67*, 2997–3011. [[CrossRef](#)]
73. Wang, T.; Wen, C.K.; Wang, H.; Gao, F.; Jiang, T.; Jin, S. Deep learning for wireless physical layer: Opportunities and challenges. *China Commun.* **2017**, *14*, 92–111. [[CrossRef](#)]
74. Liu, C.; Liu, X.; Ng, D.W.K.; Yuan, J. Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications. *IEEE Trans. Wirel. Commun.* **2021**. [[CrossRef](#)]
75. Soltani, M.; Pourahmadi, V.; Mirzaei, A.; Sheikhzadeh, H.; Jin, S. Deep Learning-Based Channel Estimation *IEEE Commun. Lett.* **2019**, *23*, 652–655. [[CrossRef](#)]
76. Rappaport, T.S. *Wireless Communications: Principles and Practice*; IEEE Press: Piscataway, NJ, USA, 1996; pp. 210–214.
77. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
78. Erpek, T.; O'Shea, T.J.; Sagduyu, Y.E.; Shi, Y.; Clancy, T.C. Deep learning for wireless communications. In *Development and Analysis of Deep Learning Architectures*; Springer: Cham, Switzerland, 2020; pp. 223–266.
79. O'Shea, T.J.; Erpek, T.; Clancy, T.C. "Physical layer deep learning of encodings for the MIMO fading channel. In Proceedings of the 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 3–6 October 2017; pp. 76–80.
80. Chen, Z.; Yuan, J.; Vucetic, B. Analysis of transmit antenna selection/maximal-ratio combining in Rayleigh fading channels. *IEEE Trans. Veh. Technol.* **2005**, *54*, 1312–1321. [[CrossRef](#)]
81. Webb, A.R.; Copsey, K.D.; Vucetic, B. Ensemble methods. *Stat. Pattern Recognit.* **2011**, 361–403. [[CrossRef](#)]
82. Mohandes, M.; Deriche, M.; Aliyu, S.O. Classifiers combination techniques: A comprehensive review. *IEEE Access* **2018**, *6*, 19626–19639. [[CrossRef](#)]
83. Kaddoum, G.; Wireless chaos-based communication systems: A comprehensive survey. *IEEE Access* **2016**, *4*, 2621–2648. [[CrossRef](#)]

84. Zhang, W.; Zhang, Z. Belief function based decision fusion for decentralized target classification in wireless sensor networks. *Sensors* **2015**, *15*, 20524–20540. [[CrossRef](#)] [[PubMed](#)]
85. Ekbiyik, K.; Akbunar, Ö.; Ekti, A.R.; Kurt, G.K.; Görçin, A. On the investigation of wireless signal identification using spectral correlation function and SVMs. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–5.
86. Whang, J.; Whang, Y.; Li, W.; Gui, G.; Gacanin, H.; Adachi, F. Automatic Modulation Recognition Method for Multiple Antenna System Based on Convolutional Neural Network. In Proceedings of the 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Victoria, BC, Canada, 4–7 October 2020; pp. 1–5.