Chinese Herbal Medicine Recognition Using a VCSEL-Based Time-Delay Reservoir Computing System

Dianzuo Yue 🅾, Yushuang Hou 🅾, Chunxia Hu 🅾, Cunru Zang, and Yingzhe Kou

Abstract—We numerically investigate the performance of Chinese herbal medicine (CHM) recognition using a time delay reservoir computing (TDRC) system consist of a vertical-cavity surface-emitting laser (VCSEL) subjected to optical feedback. A database of 960 images from 12 CHM categories is used to evaluate the recognition performance, most of which contain many CHM samples and have different backgrounds. The CHM image feature vector is masked and optically injected into the VCSEL, and the VCSEL output in two polarization modes are sampled as the virtual node states for readout. After testing the CHM recognition performance of TDRC under injecting the same information into two polarization modes of the VCSEL, we further test the performance of parallel processing by injecting half of the image information into each of the two polarization modes. By concatenating the virtual node states collected from two modes for readout, the recognition speed is doubled. Meanwhile, the influence of key parameters on recognition performance are investigated and the relationship between system idle state and recognition performance is also examined. The results show that, under parallel processing and orthogonal optical feedback, the TDRC achieves the minimum recognition error of 1.7% at a reservoir processing rate of 2×10^9 images per second.

Index Terms—Chinese herbal medicine recognition, reservoir computing, VCSEL.

I. INTRODUCTION

THE rapid development of neuro-inspired algorithms such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have greatly improved the information

Dianzuo Yue, Cunru Zang, and Yingzhe Kou are with the Hebei Normal University of Science and Technology, Qinhuangdao 066004, China (e-mail: ydz4158@hevttc.edu.cn; 2095344322@qq.com; 59600941@qq.com).

Yushuang Hou is with the Hebei Normal University of Science and Technology, Qinhuangdao 066004, China, and also with the Inner Mongolia University of Science and Technology, Baotou 014010, China (e-mail: hys4177@hevttc.edu.cn).

Chunxia Hu is with the Chongqing College of Electronic Engineering, Chongqing 401331, China (e-mail: 202121009@cqcet.edu.cn).
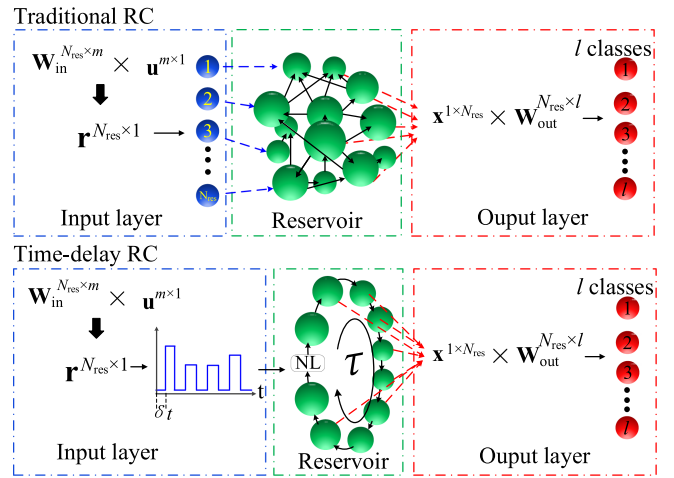
Fig. 1. Schematic diagram of traditional RC and TDRC. NL: Nonlinear node.

processing efficiency. However, when dealing with complex tasks such as face or speech recognition, these artificial neural networks (ANNs) with too many neurons or layers are usually blamed for time-consuming [1]. In the era of information explosion, it is of great significance to explore new neural network algorithms with simple structure and high training efficiency. In this case, the time-delay reservoir computing (TDRC) has received extensive attention due to its simple system structure and fast training rate in recent years [2].

The concept of TDRC is derived from traditional reservoir computing (RC) including echo state network (ESN) [3] and liquid state machine (LSM) [4]. As shown in Fig. 1, the traditional RC includes an input layer, a reservoir, and an output layer. The reservoir consists of a large number of nonlinear nodes connected to each other by randomly generated weights. The nonlinearity refers to the relationship between the input information and the output information of a nonlinear node. In theoretical studies, nonlinear nodes usually consist of sigmoid functions similar to those used in neural network [3]. In hardware implementations, nonlinear nodes can be implemented in hardware with nonlinear transformation effect on input information, such as memristor and semiconductor optical amplifier (SOA) [1], [2]. In the input layer, the feature vector $\mathbf{u} = [u_1, u_2, \ldots, u_m]^T$ needs to be multiplied by a weight matrix $\mathbf{W}_{in}$ ($\mathbf{W}_{in}$ is the multiplicand) with dimension of $N_{res} \times m$, where N_{res} is the number of nonlinear nodes in the reservoir. The resulting vector

$\mathbf{r}^{Nres\times1}$ is injected into the reservoir. Under excitation of different inputs, randomly connected reservoir nodes will generate different transient responses which are sampled as a reservoir state vector $\mathbf{x}^{1\times Nres}$. The RC output $\mathbf{y}$ can be obtained by $\mathbf{y} = \mathbf{x}\mathbf{W}_{out}$, where $\mathbf{W}_{out}$ with dimension of $N_{res} \times l$ is the trained output weight, and l is the number of classes to be identified. We can find that the key issue of RC is nonlinear mapping m-dimensional input to N_{res}-dimensional space to facilitate classification, while the procedure of training $\mathbf{W}_{out}$ is to find a linear hyperplanes to separate different classes. For TDRC, the nonlinear transformation from low-dimensional input to high-dimensional state is realized by time-multiplexing. Since TDRC has only one nonlinear node, the masked vector $\mathbf{r}$ with N_{res} elements needs to be serially injected into the reservoir. To provide a certain response time for the nonlinear node, each element of $\mathbf{r}$ is kept constant for a small time interval δ_t. In this way, the vector $\mathbf{r}$ is converted to a time series $r(t)$. Therefore, a feature vector $\mathbf{u}$ can be processed in time T, where $T = N_{res}\delta_t$. In other words, the processing speed of TDRC for a feature vector is $1/N_{res}\delta_t$. In the output layer, to obtain N_{res} reservoir states (named as virtual node states), we need to sample the state of the single nonlinear node N_{res} times within an input period T, where the sampling interval is usually set to δ_t. The collected virtual node state vector $\mathbf{x}$ is used for training and testing, which is similar to that in a traditional RC. TDRC retains the training simplicity of RC, i.e., the weights of the input layer and the reservoir are randomly generated and fixed during training [5]. Therefore, it is only necessary to train the readout weights of TDRC using simple methods of ridge regression [6] and solving pseudo-inverse [7]. TDRC replaces the large number of nonlinear nodes in traditional RC with a single nonlinear node. As a result, TDRC is much easier to implement in hardware than traditional RC. Several groups have proved that nonlinear time-delay systems such as electrical circuit system [8], optoelectronic system [9], [10], and all-optical system [11] can be used to implement TDRC. Particularly, the TDRC based on a semiconductor laser (SL) subjected to optical feedback shows the advantages of low error rate, high processing speed, low energy consumption, and parallel processing. In 2013, the SL-based TDRC implemented by Brunner et al. processed a spoken digit recognition task with a classification error of $(0.88 \pm 0.18)\%$, in which the processing speed was 3×10^6 digits per second, and the total energy consumption was about 150 W [12]. In 2020, the photonic integrated TDRC implemented by Harkhoe et al. achieved a high processing rate of 0.89 GSa/s [13]. In 2021, based on a Fabry Perot SL, Bogris et al. show the parallel processing ability of TDRC, which can achieve a potential processing speed of 8 Gsa/s [14]. The high processing rate of SL-based TDRC is due to that the response time of SL at sub-nanoseconds [2], the virtual node interval δ_t (lower than the response time of SL) is shorter than other nonlinear nodes such as modulator and nonlinear electrical circuit. Therefore, SL-based TDRC has been used for processing speech recognition [12], waveform recognition [15], nonlinear time series prediction [16], channel equalization [17], and chaotic synchronization [18].

In this work, we apply the SL-based TDRC to deal with a novel task of Chinese herbal medicine (CHM) recognition which stemming from practical applications. Since CHMs have hundreds of types with different shapes and colors, the identification and classification of CHMs usually requires well-trained operators [19]. Therefore, the automatic identification of CHM is of great significance for accelerating the development of related industries. Different from previous image recognition tasks such as handwritten digit recognition, most of the CHM images used in this task contain many pieces of CHM with a natural background, rather than hand-crafted image contains only one piece of CHM with a clean background. In [20], Xin et al. performed CHM images recognition using a CNN with 16 weight layers, which achieved an average recognition rate of 71% after 6000 iterations. After removing some special samples and increasing the number of training samples, the precision in [20] reached to 97.22%. In [19], Xu et al. achieved an accuracy over 90% for CHM images recognition using multiple attentional pyramid networks. Compared to the methods used in [19] and [20], the SL-based TDRC in this work possesses a simpler structure and less parameters, and can be implemented in hardware.

Specifically, we take a vertical-cavity surface-emitting laser (VCSEL) under optical feedback as the core nonlinear element of the time-delay reservoir. The feature vectors of CHM images are optically injected into the reservoir. The recognition performance is evaluated using a CHM data set containing 960 images in 12 categories. First, we investigate the recognition performance of the VCSEL-based TDRC under different virtual node state acquisition approaches, where the virtual node states can be obtained from different polarization modes of the VCSEL output. Then, the parallel processing ability of the TDRC is investigated, where two parts of the masked information about one image are injected into the two polarization modes, respectively. The virtual node states collected from two polarization modes are concatenated for training and testing. By adopting the parallel processing method, the recognition speed is doubled. At last, the influence of key parameters such as injection strength, feedback strength, frequency detuning, the offset phase of modulator, and the training sample size on recognition performance are investigated. The relationship between system idle state and recognition performance is also examined. The results show that under parallel processing, a minimum error rate (ER) of 1.7% can be achieved with reservoir processing speed of 2×10^9 images per second.

II. RESERVOIR COMPUTING SYSTEM MODEL

Fig. 2 shows the schematic diagram of CHM recognition system using VCSEL-based TDRC, where the system can be split into three layers of an input layer, a reservoir, and an output layer. The time-delay reservoir is composed of a VCSEL under optical injection and optical feedback. Compared with traditional edge-emitting SL, VCSEL has the characteristics of small size, low energy consumption, high response rate, and easy integration [21]. Moreover, the refractive index difference between principal axis component and orthogonal axis component in semiconductor
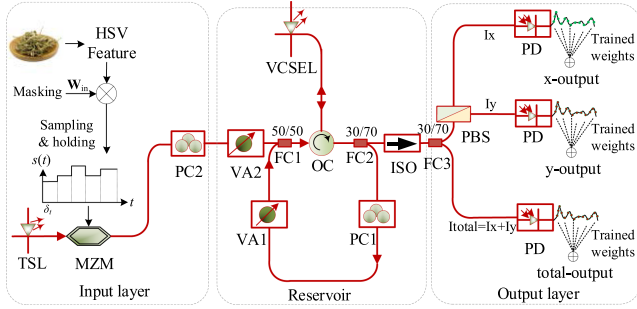
Fig. 2. Schematic diagram of CHM recognition system using the VCSEL-based TDRC. PC: Polarization controller; TSL: Tunable semiconductor laser; VA: Variable attenuator; OC: Optical circulator. MZM: Mach-Zender modulator; PBS: Polarization beam splitter; PD: Photo detector. FC: Fiber coupler.

medium of laser results in the existence of two polarization modes (XP mode and YP mode) in the output light of VCSEL [22]. It has been demonstrated that VCSEL is sensitive to optical feedback and optical injection [23], [24], [25]. For instance, in 2018, Nazhan et al. investigated the stability and sensitivity of VCSEL [23]. They found that under orthogonal optical feedback (OOF), the VCSEL shows instability in output power when the feedback level is -14 dB. Therefore, the nonlinear dynamics of VCSEL under external perturbations are usually used to generate random numbers [26], chaotic optical communication [27], and also be used in mapping low-dimensional input to high-dimensional state space for TDRC systems [28], [29]. Based on the VCSEL under parallel optical feedback (POF) or OOF, several groups have investigated the memory capability and prediction performance of the VCSEL-based TDRC system [30], [31], [32]. Previous studies have shown that to efficiently perform tasks, RC should satisfy separation and consistency properties [33]. Separation property refers to that the reservoir should exhibit sufficiently different dynamical responses under inputs belong to different levels, which requires the reservoir to be sensitive to external stimulus. Consistency means that the reservoir states should be identical or sufficiently similar under inputs belong to the same level. Under optimized parameters, VCSEL-based TDRC can make a trade-off between separation and consistency to achieve good performance. Moreover, In the case of coexistence of two polarization modes, much richer reservoir states may be achieved, which leads to a better performance. Therefore, VCSEL is considered as one of the ideal nonlinear nodes for implementing TDRC [30], [31], [32].

As shown in Fig. 2, the reservoir in this work is composed of a VCSEL and a feedback loop, where the polarization controller (PC1) and variable attenuator (VA1) in the feedback loop can control the polarization state and feedback strength, respectively. The information to be processed is injected into the reservoir by modulating the intensity of the injection light provided by a tunable semiconductor laser (TSL), where the intensity and polarization state of the injected light are controlled by VA2 and PC2, respectively. In the output layer, the intensity of XP mode (I_x), YP mode (I_y) and the total output (I_{total}) are converted into electrical signal by three photo detectors. Therefore, by sampling the intensities of XP mode and YP mode with time interval δ_t, two virtual node state matrices can be obtained for training and testing. Meanwhile, for comparison, the total

output of VCSEL is also sampled for training and testing. In the input layer, the herbal image is converted into a feature vector, and the vector is masked to input to the reservoir. For CHM recognition, the feature values of each image need to be extracted first. Considering that it is difficult to accurately extract the feature vector of CHM image using methods such as edge detection, we extract HSV feature values in this work. Compared with other color features such as RGB, HSV features can better reflect the human perception of color, where h, s, and v represent Hue, Saturation, and Value, respectively. If r, g, and b represents the three channel pixel values of a RGB image, the h, s, and v can be expressed as:

$$h = \begin{cases} 0° & \text{if max} = \text{min} \\ 60° \times \frac{g-b}{\text{max} - \text{min}} + 0°, & \text{if max} = r \text{ and } g \geq b \\ 60° \times \frac{g-b}{\text{max} - \text{min}} + 360°, & \text{if max} = r \text{ and } g < b \\ 60° \times \frac{b-r}{\text{max} - \text{min}} + 120°, & \text{if max} = g \\ 60° \times \frac{r-g}{\text{max} - \text{min}} + 240°, & \text{if max} = b \end{cases} \tag{1}$$

$$s = \begin{cases} 0, & \text{if max} = 0 \\ 1 - \frac{\text{min}}{\text{max}}, & otherwise \end{cases} \tag{2}$$

$$v = \text{max} \tag{3}$$

where max (min) represents the maximum (minimum) value among r, g, and b. The h is divided into 16 levels (0-15) with an interval of 22.5, s and v are divided into four levels (0-3) with an interval of 0.25. Then h, s, and v are weighted sum with weights of 16, 4, and 1, where the weights are determined by trial-and-error. At last, the HSV features is extracted as the histogram of the weighted sum with 256 elements. According to the operation process of TDRC, the HSV vector needs to be multiplied by a mask to enrich the reservoir state (mapping the HSV vector to a higher dimensional space) [34]. We construct a masking matrix $\mathbf{W_{in}}$ with dimension of $256 \times N_{\text{res}}$, where N_{res} is the number of virtual nodes and the values of $\mathbf{W_{in}}$ are randomly extracted from [0, 1]. As a result, we can obtain an input vector containing N_{res} elements which corresponding to N_{res} virtual nodes of the reservoir. Furthermore, considering that VCSEL needs a certain time to respond to the input, each input value of the masked information should be kept constant in a time interval δ_t to form the input information $r(t)$, where δ_t is virtual node interval.

Taking an image of Jujube as an example, Fig. 3 shows the preprocessing procedure of a CHM image. The first row of Fig. 3 shows the raw image in RGB format (Fig. 3(a)), the transferred h (Fig. 3(b)), s (Fig. 3(c)), and v (Fig. 3(d)) feature matrix, respectively. Fig. 3(e) shows the mask matrix $\mathbf{W_{in}}$ with dimension of $256 \times N_{\text{res}}$. The value of N_{res} can be adjusted in the following tests. Setting $N_{\text{res}} = 500$ in this figure is only for showing the preprocessing procedure more clearly. The masked information for each images of the training set are cascaded as a long time series $r(t)$. Before injecting into the reservoir, $r(t)$ is normalized to $[-1, 1]$ by $2(r(t)-r_{\text{min}})/(r_{\text{max}}-r_{\text{min}})-1$, where r_{min} and r_{max} refers to the minimum value and maximum value of $r(t)$, respectively. Fig. 3(f) shows the normalized series corresponding to one image, which is a part of the whole series of $r(t)$ and will be injected into the reservoir within a period T (=
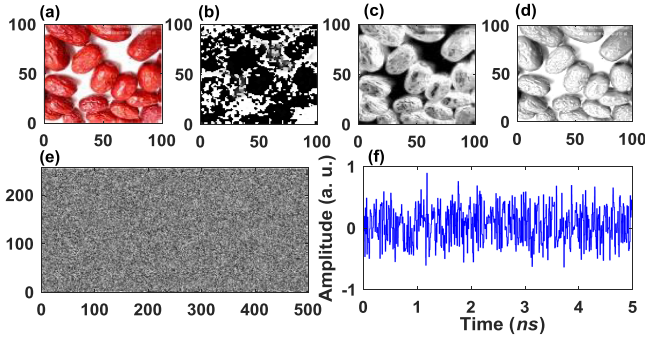
Fig. 3. Preprocessing of the CHM image in the input layer. (a) Is the raw image of Jujube, (b)–(d) is the H, S, V feature, respectively. (e) Is the mask matrix $\mathbf{W_{in}}$, (f) is the input information $r(t)$ in one input period T.

Fig. 4. Sample images of the CHM data set.

$\delta_t N_{res}$). Therefore, the reservoir processing rate for each CHM image is 1/T ($= 1/\delta_t N_{res}$), which is determined by the number of reservoir nodes N_{res} and the interval δ_t between virtual nodes.

Based on the spin-flip model (SFM), the nonlinear dynamics of the VCSEL under both optical injection and optical feedback can be expressed as [35], [36], [37], [38], [39]:

$$\frac{dE_x}{dt} = \kappa(1+i\alpha)(NE_x - E_x + inE_y) - (\gamma_a + i\gamma_p)E_x \\ + k_{inj}\varepsilon(t) + k_{fd}\Phi_x(t) + F_x(t) \tag{4}$$

$$\frac{dE_y}{dt} = \kappa(1+i\alpha)(NE_y - E_y - inE_x) + (\gamma_a + i\gamma_p)E_y \\ + k_{inj}\varepsilon(t) + k_{fd}\Phi_y(t) + F_y(t) \tag{5}$$

$$\frac{dN}{dt} = \gamma_N \left[\mu - N\left(1 + |E_x|^2 + |E_y|^2\right) + in\left(E_x E_y^* - E_y E_x^*\right) \right] \tag{6}$$

$$\frac{dn}{dt} = -\gamma_s n - \gamma_N \left[n\left(|E_x|^2 + |E_y|^2\right) + iN\left(E_y E_x^* - E_x E_y^*\right) \right] \tag{7}$$

$$\varepsilon(t) = \frac{1}{2}|\varepsilon_0|\left(1 + e^{i[r(t)+\Phi_0]}\right)e^{i2\pi\Delta f t} \tag{8}$$

$$POF: \Phi_{x,y}(t) = E_{x,y}(t-\tau)e^{-i\omega_{x,y}\tau} \tag{9}$$

$$OOF: \Phi_{x,y}(t) = E_{y,x}(t-\tau)e^{-i\omega_{y,x}\tau} \tag{10}$$

$$F_x = \sqrt{\frac{\beta}{2}}\left(\sqrt{N+n}\xi^a + \sqrt{N-n}\xi^b\right) \tag{11}$$

$$F_y = -i\sqrt{\frac{\beta}{2}}\left(\sqrt{N+n}\xi^a - \sqrt{N-n}\xi^b\right) \tag{12}$$

where E_x and E_y represent the slowly varied complex amplitude of the electric field for the XP and YP mode, respectively. N represents the total carrier inversion between conduction and valence bands, n represents the difference between carrier inversions for the spin-up and spin-down radiation channels. k is the decay rate of field. α is the linewidth enhancement factor. γ_a and γ_p represents the linear dichroism and linear birefringence, respectively. γ_s represents the spin-flip rate, γ_N represents the decay rate of total carrier population. μ refers to the normalized injection current of VCSEL, i.e., $\mu = J/J_{th}$, where J is the injection current and J_{th} is the threshold current. $\mu = 1$ implies

the injection current at the threshold current of VCSLE. k_{inj} refers to the injection strength, and k_{fd} refers to the feedback strength. Equation (8) describes the injection term, where $r(t)$ is the input information. ε_0 represents the field amplitude of the injection light. Φ_0 is the offset phase which can be adjusted by the bias voltage of the MZM. Δf is the frequency detuning between the lasing frequency of TSL and the central frequency ω_0 of the VCSEL, where $\omega_0 = 2\pi c/\lambda$ with the lasing wavelength of $\lambda = 850$ nm and light speed of c $= 3 \times 10^8$m/s. Equations (9) and (10) describes the feedback term for POF and OOF (the feedback light is rotated at 90 degree) mode, respectively. $\omega_{x,y} = \omega_0 \pm (k\gamma_a - \gamma_p)$ refers to the angular frequency of X-PC and Y-PC, respectively, and τ refers to the delay time of the feedback loop. F_x and F_y represent the Langevin noise sources added in XP mode and YP mode, respectively. ξ^a and ξ^b are complex Gaussian white noise with unit variance and zero mean, where β_{sp} is the spontaneous emission coefficient.

The output intensities of the VCSEL are sampled as virtual node states with an interval δ_t, where the intensities of XP, YP, and total intensity (XP+YP) can be expressed as $I_x = |E_x|^2$, $I_y = |E_y|^2$, and $I_{total} = I_x + I_y$, respectively. In this work, a CHM data set is used to evaluate the recognition performance of the TDRC. The data set contains 960 images for 12 types of CHMs, with each type contains 80 images. Fig. 4 shows some samples of the CHM data set. We can find that the images are not hand-crafted images which contains a single CHM sample with clean background but contains many CHM samples in each image with different backgrounds. Such kind of images is more in line with the practical application scenarios. For each type of CHM, 48 images are selected as the training set and 32 images as the testing set, i.e., 60% of the data set is used as the training set, 40% is used as the testing set. In machine learning, when the labeled sample is relatively small, the amount of training data is often chosen to be larger than that of test data since the learning performance directly depends on the training sample, and the training to test ratio is usually 6:4, 7:3, or 8:2 [40]. As a supervised learning method of machine learning, the size of training set in TDRC needs significantly bigger than the number of reservoir nodes to prevent overfitting [41]. In this work, setting 60% of the data set as the training set can effectively avoid over-fitting, while the larger training set will lead to saturation of training error. In the training procedure, the training set is injected and 12 classifiers corresponding to 12 types of CHM
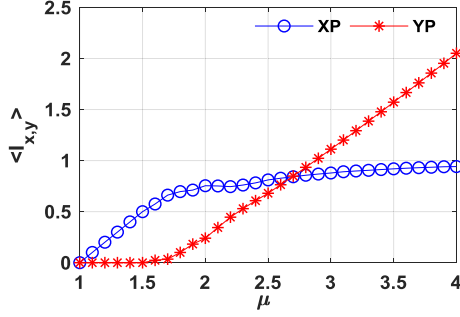
Fig. 5. XP and YP intensities of the free running VCSEL as a function of normalized current μ.

are constructed. The readout weight $\mathbf{W_{out}}$ can be obtained by using the regression algorithm of $\mathbf{W_{out}} = (\mathbf{X_{tr}^T X_{tr}})^{-1} \mathbf{X_{tr}^T y_d}$, where $\mathbf{X_{tr}}$ is the virtual node state matrix and $\mathbf{y_d}$ is the training target [42]. In the testing procedure, the testing set is injected and the virtual node state matrix for testing set $\mathbf{X_{ts}}$ is obtained. The reservoir output $\mathbf{y_t} = \mathbf{X_{ts} W_{out}}$ can be calculated. By applying the method of winner-takes-all [43], the maximum value of $\mathbf{y_t}$ is found out for determining the final recognition type of CHM. We use the ER ($= (N_{err}/N_{test}) \times 100\%$) to evaluate the recognition performance, which is the ratio of the number of wrong classified images (N_{err}) to the total number of tested images (N_{test}) [44]. Taking into account the classification performance of previous researches [19], [20], in this work, the recognition performance is considered as good when ER is lower than 10%. To accurately assess the TDRC performance, a 10-fold cross-validation method is applied [45].

III. RESULTS AND DISCUSSIONS

The states of VCSEL can be numerically obtained by solving above equations using fourth-order Runge-Kutta method with integration step of 2 ps. The parameter values are $\alpha = 3$, $k = 300$ ns^{-1}, $\gamma_a = 0.1$ ns^{-1}, $\gamma_p = 10$ ns^{-1}, $\gamma_s = 50$ ns^{-1}, $\gamma_N = 1$ ns^{-1}, $\beta_{sp} = 10^{-6}$ [27]. Since VCESL can simultaneously output XP and YP modes under suitable conditions, we first examine the output intensities of XP and YP modes under different current.

As shown in Fig. 5, for the free-running VCSEL, when the normalized injection current μ is larger than 1.0, the XP mode lasing first. YP mode starts lasing at $\mu > 1.5$ and its intensity increases to almost same as the XP mode when $\mu = 2.7$. When $\mu > 2.7$, the intensity of YP mode keeps increasing, while the XP mode is almost constant. Therefore, in the following tests we can set μ around 2.7 to take advantage of the coexistence of the two modes of VCSEL to enrich the reservoir states and improve the CHM recognition rate.

According to Fig. 5, we set $\mu = 2.7$ to investigate the CHM recognition performance of this VCSEL-based TDRC system. The virtual node interval θ is set as 10 ps, and the number of virtual nodes N_{res} is set as 100. Although small δ_t and N_{res} can lead to a high processing rate, they cannot be too small to ensure good performances. δ_t is related to the relaxation oscillation characteristic time of the VCSEL, N_{res} is related to the dimensionality of the state space. By repeated tests, we find

that setting $\delta_t = 10$ ps and $N_{res} = 100$ can make a trade-off between recognition performance and processing speed. In the input layer, we inject the same masked information into XP and YP mode simultaneously. We assume that the injection frequency detuning $\triangle f = 0$ and the injection strength k_{inj} for two modes are the same. In the output layer, we collect virtual node states from I_x, I_y, and I_{total}, respectively. Fig. 6 shows the evolution map of ER in feedback strength (k_{fd}) and injection strength (k_{inj}) parameter space. The upper panel shows the results of VCSEL under POF, and the lower panel shows the results of VCSEL under OOF. Column a, b, and c represent the test results based on virtual node states collected from I_x, I_y, and I_{total}, respectively. To highlight the color changes in the figure, the region of ER $> 60\%$ is shown in white, which makes it easier to locate the parameter region with good performance. Under POF (the first row in Fig. 6), ER values lower than 60% is mainly located at the upper left of the three figures, i.e., the region with relatively high k_{inj} and low k_{fd}. This is because relatively larger k_{fd} tends to cause system oscillations, which reduce the consistency of the reservoir's response to input information [43]. In Fig. 6(a1), (b1), and (c1), the ER lower than 10% are located at the area of 5 ns$^{-1} \leq k_{inj} \leq 15$ ns^{-1} and $k_{fd} < 2$ ns^{-1}. The lowest ERs in the three figures are 1.7%, 3%, and 3%, respectively. Compared to Fig. 6(a) and (b), the area of good performance (ER $< 10\%$) in Fig. 6(c) is slightly enlarged. Under OOF (the second row in Fig. 6), the areas of ER < 60 in three figures are similar with that under POF, and the lowest ERs of 4.6% for Fig. 6(a2), 3.4% for Fig. 6(b2), and 3.8% for Fig. 6(c2) are obtained. Moreover, the areas of ER $< 10\%$ under OOF are horizontally elongated than that under POF, which indicates the system robustness to feedback strength is improved. This phenomenon is consistent with the previous studies when SL is under OOF [30].

In the above test, we find that when the same information is injected into two modes of VCSEL, both XP and YP mode can be used for training and testing to achieve good recognition performance. Based on this result, we further exploit the coexistence property of XP and YP mode in VCSEL-based TDRC system to improve the CHM recognition rate. In this test, we split the masked sequence of one image into two parts and simultaneously inject the two parts into the XP mode and YP mode, respectively. In the readout layer, the virtual node states matrixes of $\mathbf{X^{XP}}$ (collected from XP mode) and $\mathbf{X^{YP}}$ (collected from YP mode) are concatenated as $[\mathbf{X^{XP}}; \mathbf{X^{YP}}]$ for readout. For clarity, we refer to this processing method as parallel processing [46]. To ensure the comparability of the test results, we set $N_{res} = 50$. Therefore, the number of virtual node states after concatenation of $\mathbf{X^{XP}}$ and $\mathbf{X^{YP}}$ is 100 which is same as the test in Fig. 6. Fig. 7 shows the results of parallel processing under POF and OOF. It is obviously that for parallel processing, both under POF (Fig. 7(a)) and OOF (Fig. 7(a)) the area of ER $< 60\%$ is enlarged compared with Fig. 6, but these areas are also located in the upper left of the two figures. The lowest ERs in Fig. 7(a) and (b) are 2.6% and 2.1%, respectively. Similar to the results in Fig. 6, under OOF, the robustness of the system to the feedback strength is improved, which is beneficial to the fast adjustment and stable operation of the system when implemented in hardware. Furthermore,
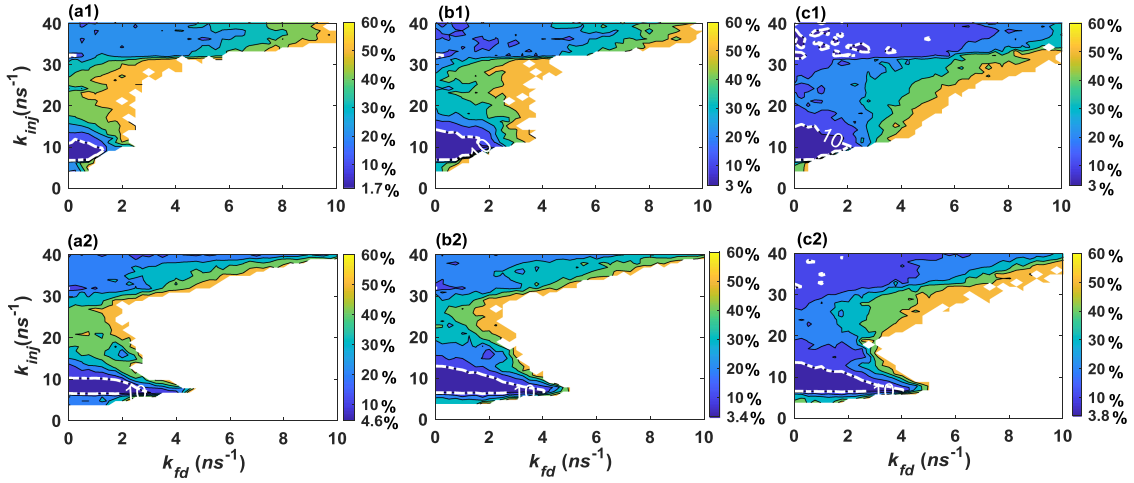
Fig. 6. Evolution map of ER in k_{fd}-k_{inj} parameter space, where the blank region is for error rate higher than 60%. The first row shows the result obtained under POF and the second row shows the results obtained under OOF. Column a, b, and c show the results based on virtual node states collected from I_x, I_y, and I_x+I_y, respectively.
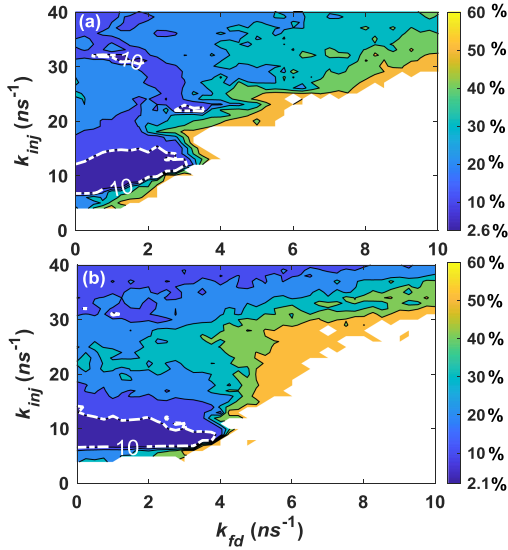

Fig. 7. Evolution map of ER in k_{fd}-k_{inj} parameter space, where the blank region is for error rate higher than 60%. (a) Is for the result obtained under POF and (b) is for the result obtained under OOF.

since N_{res} in parallel information processing method is 50, the recognition rate is doubled compared to the test in Fig. 6, i.e., a rate of 2×10^9 images per second is achieved.

In the above tests, we assumed that the frequency detuning Δf = 0 and offset phase $\Phi_0 = 0$. When implemented in hardware, Δf can be adjusted by changing the frequency of the injected light, and Φ_0 can be adjusted by changing the bias voltage of the modulator. Next, we investigate the influence of Δf and Φ_0 on CHM recognition performance under the parallel processing method. Fig. 8(a) shows the evolution map of ER in Φ_0-Δf parameter space under OOF, where $k_{inj} = 10$ ns^{-1} and $k_{fd} = 1$ ns^{-1}. In Fig. 8(a), the area of ER < 60% mainly lies between $-0.8\pi \leq \Phi_0 \leq 0.8\pi$, which is almost symmetrically distributed along $\Phi_0 = 0$. This may be due to that when Φ_0 is too large, the nonlinear transformation effect of modulator is too strong to generate appropriate virtual node states. Meanwhile, we can
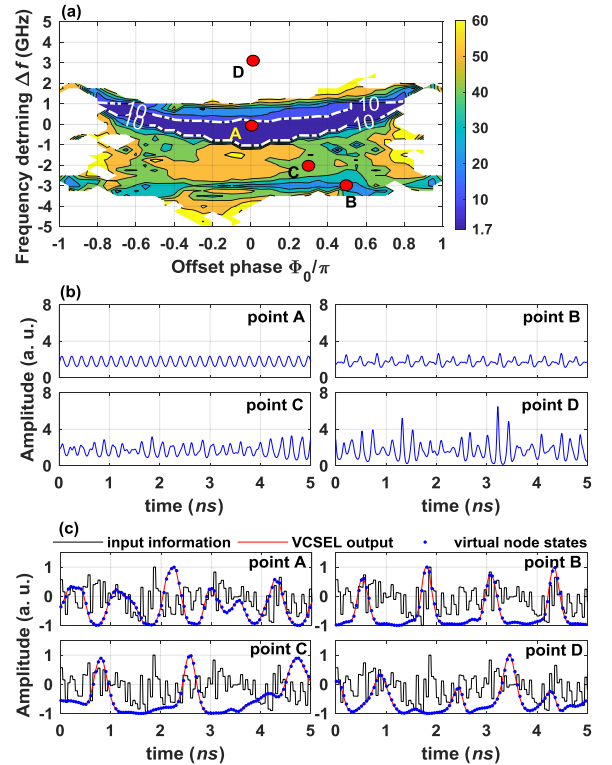

Fig. 8. (a) Evolution map of ER in Φ_0-Δf parameter space under OOF and parallel processing, where the blank region is for error rate higher than 60%. (b) Time series of point A, B, C, and D of Fig. 8(a) without information injection. (c) Input information, VCSEL output, and virtual node states of point A, B, C, and D.

find that the area of ER < 60% is mainly located between -4 ns$^{-1} \leq \Delta f$ and $\Delta f \leq 1$ns^{-1}, while the area of ER > 60% is mainly distributed at the positive detuning region. In particular, the area of ER < 10% is mainly concentrated in the range of -1 ns$^{-1} \leq \Delta f \leq 1$ns^{-1}, which is due to that the system is easier to present an injection locking state in this region to achieve good consistency [46]. The lowest ER of 1.7% is obtained at

$\Delta f = -0.2$ GHz, Φ_0 = -0.1π. In Fig. 8(a), the region of ER lower than 10% (the blue region) is relatively small, while the blank region of ER $> 60\%$ is relatively large. This is because the scanned parameter space is relatively large. In application, the small blue region in Fig. 8(a) corresponds to a relatively large parameter range. For example, the variation of offset phase Φ_0 from $-\pi$ to π corresponds to the variation of MZM bias voltage from about 3.5V to 12V with a step of 0.01V. The variation of frequency detuning Δf from -1 GHz to 1 GHz to 1GHz is also easy controlled with an adjustment step of 0.1GHz by using a TSL. GHz is also easy controlled with an adjustment step of 0.1GHz by using a TSL.

Then, we select four points of A, B, C and D marked in Fig. 8(a) as examples to further investigate the relationship between CHM recognition performance and system idle state (without information injection). Fig. 8(b) shows the time series of point A, B, C, and D, where the TDRC system is under OOF but without information injection. For point A ($\Phi_0 = 0$, $\Delta f = 0$), the time series shows a single-period oscillation state. For point B ($\Phi_0 = 0.5\pi$, Δf = -3GHz), the time series shows a multi-periods oscillation state. For point C ($\Phi_0 = 0.3\pi$, $\Delta f = $ -2GHz) and point D ($\Phi_0 = 0$, $\Delta f = 3$GHz), the system presents relatively weak and strong chaos oscillation state, respectively. Fig. 8(c) shows the injected information (black line), VCSEL output (red line), and sampled virtual node states (blue point) of point A, B, C, and D. For point A, the output of VCSEL presents six peaks under masked information injection. Based on the sampled virtual node states, an ER of 6.3% is achieved. For point B, VCSEL output four peaks, and an ER of 25.4% is achieved. For point C, the output of VCSEL presents only three peaks, and an ER of 45% is achieved. Note that the time series shown in Fig. 8(c) have been normalized for clearly show the waveform of VCSEL output. In fact, the amplitude of VCSEL output increases in order of point A, B, C, and D. For point D, three peaks are observed. Furthermore, the location of these peaks appears to be independent of the input information, which results in an ER of 79.6%. From this test we can find that in order to obtain good recognition performance, the system needs to run in an asymptotically stable idle state. In this case, the reservoir can map input to a higher-dimensional state space and make a trade-off between consistency and separation property. This correspondence between the CHM recognition performance and the system idle state is consistent with our previous experimental results [18], which provides a reference for quickly adjusting the system parameters.

At last, under optimized parameters of $k_{\text{inj}} = 10$ GHz to 1GHz is also easy controlled with an adjustment step of 0.1GHz by using a TSL. ns^{-1}, $k_{fd} = 1$ GHz to 1GHz is also easy controlled with an adjustment step of 0.1GHz by using a TSL. ns^{-1}, Δf = -0.2 GHz to 1GHz is also easy controlled with an adjustment step of 0.1GHz by using a TSL. GHz, and Φ_0 = -0.1π, we investigate the influence of the training sample size on ER, and analyze the recognition performance difference between different type of CHMs. The ERs of each CHM under different training sample sizes are listed in Table I. For each tested size, we randomly extract samples from each type of CHM as the same proportion (e.g., 30%), and run the TDRC 20 times, where the ERs in Table I is the mean value of 20 tests. For the training

TABLE I
ER OF EACH CHM UNDER DIFFERENT SIZE OF TRAINING SAMPLE

	30%	40%	50%	60%	70%
Lili	0.472	0.383	0.169	0.038	0.035
Jujube	0.475	0.342	0.092	0.029	0.023
Poria	0.505	0.424	0.112	0.016	0.018
Honeysuckle	0.253	0.185	0.016	0.002	0.002
Saffron	0.154	0.081	0.015	0.016	0.017
Locust flower	0.252	0.218	0.052	0.009	0.009
Cocklebur	0.162	0.117	0.021	0.023	0.028
Codonopsis	0.362	0.180	0.017	0.014	0.015
Wolfberry	0.707	0.653	0.248	0.035	0.041
Tangerine peel	0.292	0.154	0.003	0.002	0.002
Atractylodes	0.456	0.292	0.027	0.007	0.004
Antler	0.301	0.181	0.032	0.014	0.012

proportion of 30%, 40%, 50%, 60%, and 70%, the mean ERs are 0.37, 0.26, 0.07, 0.018 and 0.018, respectively. Obviously, a larger training sample size is conducive to decrease ER, but when the size is larger than 60%, the ER tends to saturation. However, the reduction trends of ER and the saturation ER values differ for different CHMs. For Saffron, the ER saturates to about 0.015 at 50% training sample, while for Honeysuckle, it saturates to 0.002 at 60% training sample. We speculate that these discrepancies are due to the different CHM appearances and the nature of the images taken. Although a relatively low average ER is achieved in this work, there is still room for further reduction of individual CHM ER. In our future work, we will try our best to identify the sources of the discrepancies based on TDRC, and explore better image preprocessing methods as well as feature extraction techniques to reduce the discrepancies.

IV. CONCLUSION

In conclusion, we numerically perform CHM recognition using a VCSEL-based TDRC system, where the time-delay reservoir consist of a VCSEL subjected to POF or OOF. A data set containing 960 images in 12 CHMs categories is used for evaluating the performance, where most of the images contain many pieces of CHM with different background. The HSV feature vector is masked and input to the reservoir, and the reservoir states are sampled as the virtual node states for readout. After testing the CHM recognition performance of TDRC under injecting the same information into XP and YP mode, we further tested the performance of parallel processing based on the two polarization modes of VCSEL. For parallel processing, the masked sequence corresponding to one image is split into two parts and injected into XP and YP mode, respectively. By concatenating the virtual node states collected from two modes for readout, the recognition rate is doubled. Meanwhile, the influence of key parameters such as injection strength, feedback strength, frequency detuning, and the offset phase of modulator on recognition performance are investigated. The relationship between system idle state and recognition performance are also examined. The results show that, under parallel processing and OOF, a minimum ER of 1.7% is achieved with reservoir processing rate of 2×10^9 images per second. The successfully recognition of CHM images using a VCSEL-based TDRC system is helpful to expand the application field of TDRC.

REFERENCES

[1] G. V. der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, pp. 561–576, May 2017.

[2] K. Kitayama, M. Notomi, M. Naruse, K. Inoue, S. Kawakami, and A. Uchida, "Novel frontier of photonics for data processing—Photonic accelerator," *Appl. Phys. Lett. Photon.*, vol. 4, no. 9, Sep. 2019, Art. no. 090901.

[3] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 2, pp. 78–80, Apr. 2004.

[4] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, Nov. 2002.

[5] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Netw.*, vol. 20, no. 3, pp. 391–403, Apr. 2007.

[6] N. McDonald et al., "Analysis of an ultra-short true time delay line optical reservoir computer," *J. Lightw. Technol.*, vol. 38, no. 14, pp. 3584–3591, Jul. 2020.

[7] K. Harkhoe and G. V. d. Sande, "Delay-based reservoir computing using multimode semiconductor lasers: Exploiting the rich carrier dynamics," *IEEE J. Sel. Topics Quantum Electron.*, vol. 25 no. 6, Nov./Dec. 2019, Art. no. 1502909.

[8] L. Appeltant et al., "Information processing using a single dynamical node as complex system," *Nature Commun.*, vol. 2, Sep. 2011, Art. No. 468.

[9] L. Larger et al., "Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing," *Opt. Exp.*, vol. 20, no. 3, pp. 3241–3249, Jan. 2012.

[10] Y. P. Chen et al., "Reservoir computing system with double optoelectronic feedback loops," *Opt. Exp.*, vol. 27, no. 20, pp. 27431–27440, Sep. 2019.

[11] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Exp.*, vol. 20, no. 20, pp. 22783–22795, Sep. 2012.

[12] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nature Commun.*, vol. 4, Jan. 2013, Art. no. 1364.

[13] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. V. Der Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," *Opt. Exp.*, vol. 28, no. 3, pp. 3086–3096, Feb. 2020.

[14] A. Bogris, C. Mesaritakis, S. Deligiannidis, and P. Li, "Fabry-Perot lasers as enablers for parallel reservoir computing," *IEEE J. Sel. Topics Quantum Electron.*, vol. 27, no. 2, Mar./Apr. 2021, Art. no. 7500307.

[15] D.-Z. Yue, Y.-S. Hou, Z.-M. Wu, C.-X. Hu, Z.-Z. Xiao, and G.-Q. Xia, "Experimental investigation of an optical reservoir computing system based on two parallel time-delay reservoirs," *IEEE Photon. J .*, vol. 13, no. 3, Jun. 2021, Art. no. 8500111.

[16] Y. S. Hou et al., "Prediction performance of reservoir computing system based on a semiconductor laser subject to double optical feedback and optical injection," *Opt. Exp.*, vol. 26, no. 8, pp. 10211–10219, Apr. 2018.

[17] R. M. Nguimdo, G. Verschaffelt, J. Danckaert, and S. G. Van, "Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics," *Opt. Exp.*, vol. 22, no. 7, pp. 8762–8686, Apr. 2014.

[18] D. Z. Zhong, H. Yang, J. T. Xi, N. Zeng, Z. Xu, and F. Q. Deng, "Predictive learning of multi-channel isochronal chaotic synchronization by utilizing parallel optical reservoir computers based on three laterally coupled semiconductor lasers with delay-time feedback," *Opt. Exp.*, vol. 29, no. 4, pp. 5279–5294, Feb. 2021.

[19] Y. X. Xu et al., "Multiple attentional pyramid networks for Chinese herbal recognition," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107558.

[20] X. Sun and H. N. Qian, "Chinese herbal medicine image recognition and retrieval by convolutional neural network," *PLoS One*, vol. 11, no. 6, Jun. 2016, Art. no. e0156327.

[21] F. Koyama, "Recent advances of VCSEL photonics," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4502–4511, Dec. 2006.

[22] A. Quirce et al., "Polarization switching and injection locking in vertical-cavity surface-emitting lasers subject to parallel optical injection," *Opt. Lett.*, vol. 41, no. 11, pp. 2664–2667, Jun. 2016.

[23] S. Nazhan and Z. Ghassemlloy, "Polarization output power stabilization of a vertical-cavity surface-emitting laser," *J. Opt. Soc. Amer. B: Opt. Phys.*, vol. 35, pp. 1615–1619, Jul. 2018.

[24] S. Nazhan and Z. Ghassemlooy, "Polarization switching dependence of VCSEL on variable polarization optical feedback," *IEEE J. Quantum Electron.*, vol. 53, no. 4, Aug. 2017, Art. no. 2400707.

[25] J. Paul, C. Masoller, Y. H. Hong, P. S. Spencer, and K. A. Shore, "Impact of orthogonal optical feedback on the polarization switching of vertical-cavity surface-emitting lasers," *J. Opt. Soc. Amer. B: Opt. Phys.*, vol. 24, pp. 1987–1994, Aug. 2007.

[26] A. Quirce and A. Valle, "Random polarization switching in gain-switched VCSELs for quantum random number generation," *Opt. Exp.*, vol. 30, no. 7, pp. 10513–10527, Mar. 2022.

[27] Y. Huang, P. Zhou, and N. Q. Li, "High-speed secure key distribution based on chaos synchronization in optically pumped QD spin-polarized VCSELs," *Opt. Exp.*, vol. 29, no. 13, pp. 19675–19689, Jun. 2021.

[28] Y. G. Yang, P. Zhou, P. H. Mu, and N. Q. Li, "Time-delayed reservoir computing based on an optically pumped spin VCSEL for high-speed processing," *Nonlinear Dyn.*, vol. 107, pp. 2619–2632, Feb. 2022.

[29] Y. H. Zhang, J. Robertson, S. Y. Xiang, M. Hejda, J. Bueno, and A. Hurtado, "All-optical neuromorphic binary convolution with a spiking VCSEL neuron for image gradient magnitudes," *Photon. Res.*, vol. 9, no. 5, pp. B201–B209, May 2021.

[30] J. Vatin, D. Rontani, and M. Sciamanna, "Enhanced performance of a reservoir computer using polarization dynamics in VCSELs," *Opt. Lett.*, vol. 43, no. 18, pp. 4497–4500, Sep. 2018.

[31] J. Vatin, D. Rontani, and M. Sciamanna, "Experimental reservoir computing using VCSEL polarization dynamics," *Opt. Exp.*, vol. 27, no. 13, pp. 18579–18584, Jun. 2019.

[32] X. X. Guo, S. Y. Xiang, Y. H. Zhang, L. Lin, A. J. Wen, and Y. Hao, "Four-channels reservoir computing based on polarization dynamics in mutually coupled VCSELs system," *Opt. Exp.*, vol. 27, no. 16, pp. 23293–23306, Aug. 2019.

[33] K. Hicke, M. A. Escalona-Morán, D. Brunner, M. C. Soriano, I. Fischer, and C. R. Mirasso, "Information processing using transient dynamics of semiconductor lasers subject to delayed feedback," *IEEE J. Sel. Topics Quantum Electron.*, vol. 19, no. 4, Jul./Aug. 2013, Art. no. 1501610.

[34] M. Tezuka, K. Kanno, and M. Bunsen, "Reservoir computing with a slowly modulated mask signal for preprocessing using a mutually coupled optoelectronic system," *Jpn. J. Appl. Phys.*, vol. 55, Aug. 2016, Art. no. 08RE06.

[35] T. Deng et al., "Stable propagation of inhibited spiking dynamics in vertical-cavity surface-emitting lasers for neuromorphic photonic networks," *IEEE Access*, vol. 6, pp. 67951–67958, 2018.

[36] N. Jiang, W. Pan, B. Luo, S. Xiang, and L. Yang, "Bidirectional dual-channel communication based on polarization-division-multiplexed chaos synchronization in mutually coupled VCSELs," *IEEE Photon. Technol. Lett.*, vol. 24, no. 13, pp. 1094–1096, Jul. 2012.

[37] T. Deng, J. Robertson, and A. Hurtado, "Controlled propagation of spiking dynamics in vertical-cavity surface-emitting lasers: Towards neuromorphic photonic networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 23, no. 6, Nov./Dec. 2017, Art. no. 1800408.

[38] N. Jiang et al., "Physical secure optical communication based on private chaotic spectral phase encryption/decryption," *Opt. Lett.*, vol. 44, no. 7, pp. 1536–1539, Apr. 2019.

[39] X. X. Guo, S. Y. Xiang, Y. H. Zhang, L. Lin, A. J. Wen, and Y. Hao, "Polarization multiplexing reservoir computing based on a VCSEL with polarized optical feedback," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jan./Feb. 2020, Art. no. 1700109.

[40] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.

[41] F. Köster, D. Ehlert, and K. Lüdge, "Limitations of the recall capabilities in delay-based reservoir computing systems," *Cogn. Comput.*, pp. 1–8, Oct. 2020, doi: 10.1007/s12559-020-09733-5.

[42] R. S. Zimmermann and U. Parlitz, "Observing spatio-temporal dynamics of excitable media using reservoir computing," *Chaos*, vol. 28, no. 4, Apr. 2018, Art. no. 043118.

[43] L. Larger, A. Baylon-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, "High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification," *Phys. Rev. X*, vol. 7, no. 1, Feb. 2017, Art. no. 011015.

[44] D. Z. Yue et al., "Performance optimization research of reservoir computing system based on an optical feedback semiconductor laser under electrical information injection," *Opt. Exp.*, vol. 27, no. 14, pp. 19931–19939, Jul. 2019.

[45] A. Molinaro, R. Simon, and R. Pfeiffer, "Prediction error estimation: A comparison of resampling methods," *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, Aug. 2005.

[46] J. Bueno, D. Brunner, M. C. Soriano, and I. Fischer, "Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback," *Opt. Exp.*, vol. 25, no. 3, pp. 2401–2412, Feb. 2017.