A Review of Designs and Applications of Echo State Networks

Chenxi Sun[1,2], Moxian Song[1,2], Shenda Hong[3,4], and Hongyan Li [*,2]

[1]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China.
[2]School of Electronics Engineering and Computer Science, Peking University, Beijing, China.
[3]National Institute of Health Data Science, Peking University, Beijing, China.
[4]Institute of Medical Technology, Health Science Center, Peking University, Beijing, China.

December 8, 2020

ABSTRACT

Recurrent Neural Networks (RNNs) have demonstrated their outstanding ability in sequence tasks and have achieved state-of-the-art in wide range of applications, such as industrial, medical, economic and linguistic. Echo State Network (ESN) is simple type of RNNs and has emerged in the last decade as an alternative to gradient descent training based RNNs. ESN, with a strong theoretical ground, is practical, conceptually simple, easy to implement. It avoids non-converging and computationally expensive in the gradient descent methods. Since ESN was put forward in 2002, abundant existing works have promoted the progress of ESN, and the recently introduced Deep ESN model opened the way to uniting the merits of deep learning and ESNs. Besides, the combinations of ESNs with other machine learning models have also overperformed baselines in some applications. However, the apparent simplicity of ESNs can sometimes be deceptive and successfully applying ESNs needs some experience. Thus, in this paper, we categorize the ESN-based methods to basic ESNs, DeepESNs and combinations, then analyze them from the perspective of theoretical studies, network designs and specific applications. Finally, we discuss the challenges and opportunities of ESNs by summarizing the open questions and proposing possible future works.

1 Introduction

In the last decades, Recurrent Neural Networks (RNNs) based methods demonstrated their outstanding ability in time-series prediction tasks and have become very attractive for their potentially wide range of applications, such as biological, medical, linguistic, social and economic. RNNs represent a very powerful generic tool, integrating both large dynamical memory and highly adaptable computational capabilities. They are the Machine Learning (ML) model most closely resembling biological brains, the substrate of natural intelligence. Meanwhile, deep RNN are able to develop in their internal states a multiple time-scales representation of the temporal information, a much desired feature.

In the context of deep learning, error backpropagation (BP) [1] is to this date one of the most important achievements in RNNs training, but only with a partial success. One of the limitations is that bifurcations can make training non-converging, computationally expensive and poor local minima [2]. Moreover, the exploding gradient problem often occurs in the training process, where the stable prediction performance cannot be easily ensured [3]. Some developments in BP for RNNs [4] perform better on problems which require long memory, but it's hard for BP RNN training, unless networks are specifically designed to deal with them [5]. Long short-memory (LSTM) and gated recurrent unit (GRU) are advanced designs to mitigate shortcomings of RNN. But when the length of input sequence exceeds a certain limit, the gradient will still disappear. Meanwhile, each LSTM cell have four full connection layers, if the time span of LSTM is large and the network is very deep, the calculation will be very heavy and time-consuming. Further, too many parameters will lead to over fitting risk.

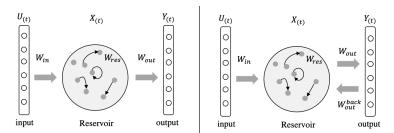*Corresponding author: leehy@pku.edu.cn

Figure 1: Basic echo state networks

Reservoir computing (RC) has emerged in the last decade as an alternative to gradient descent methods for training RNNs with a strong theoretical ground [6, 7, 8, 9, 10]. Echo State Network (ESN) [6] is one of the key RC. ESNs are practical, conceptually simple, and easy to implement. ESNs employ the multiple high-dimensional projection in the large number of states of the reservoir with strong nonlinear mapping capabilities, to capture the dynamics of the input. This basic idea was first clearly spelled out in a neuroscientific model of the corticostriatal processing loop [3]. ESNs enjoy, under mild conditions, the so-called echo state property [6], that ensures that the effect of the initial condition vanishes after a finite transient. The inputs with more similar short-term history will evoke closer echo states, which ensure the dynamical stability of the reservoir.

Recently, the introduction of the Deep Echo State Network (DeepESN) model [11, 12] allowed to study the properties of layered RNN architectures separately from the learning aspects. Remarkably, such studies pointed out that the structured state space organization with multiple time-scales dynamics in deep RNNs is intrinsic to the nature of compositionality of recurrent neural modules. The interest in the study of the DeepESN model is hence twofold. On the one hand, it allows to shed light on the intrinsic properties of state dynamics of layered RNN architectures. On the other hand it enables the design of extremely efficiently trained deep neural networks for temporal data.

Meanwhile, with the development of deep learning, a variety of neural network structures have been proposed, like auto-encoder (AE) [13], generative adversarial networks (GAN) [14], convolutional neural networks (CNNs) [15] and restricted Boltzmann machine (RBM) [16]. ESNs have combined these different structures and achieved state-of-the-art performance in specific tasks and different applications, such as industrial [17], medical [18], financial [19] and robotics with reinforcement learning [20, 21].

ESNs from their beginning proved to be a highly practical approach to RNN training. It is conceptually simple and computationally inexpensive. It reinvigorated interest in RNNs, by making them accessible to wider audiences. However, the apparent simplicity of ESNs can sometimes be deceptive. Successfully applying ESNs needs some experience. Thus, we summarize the advancements in the development, analysis and applications of ESNs to summarize the experience of existing work.

The rest of the paper is organized as follows. Section 2 analyzes the basic ESNs from the basic definitions to specific designs. Section 3 introduces the DeepESNs with its different structures. Section 4 describes the existing combinations of ESN and other machine learning methods. Section 5 gives the benchmarks, the data-driven and the real-ward applications of ESNs. Section 6 and 7 raise the challenges, opportunities and make conclusions.

2 Basic echo state networks

2.1 Preliminaries

Definition 1 (Echo State Networks (ESNs)) *Echo state networks are a type of fast and efficient recurrent neural network. A typical ESN consists of an input layer, a recurrent layer, called reservoir, with a large number of sparsely connected neurons, and an output layer. The connection weights of the input layer and the reservoir layer are fixed after initialization, and the output weights are trainable and obtained by solving a linear regression problem.*

In an ESN, $u(t) \in R^{D \times 1}$ denotes the input value at time t of the time series, $x(t) \in R^{N \times 1}$ denotes the state of the reservoir at time t. $y(t) \in R^{M \times 1}$ denotes the output value. $W_{in} \in R^{N \times D}$ represents the connection weights between the input layer and the hidden layer, $W_{res} \in R^{N \times N}$ notes the connection weights in hidden layer. The state transition equation is in Equation 1, where f is a nonlinear function such as $\tan$ and sigmoid.

$$x(t) = f(W_{in}u(t) + W_{res}x(t-1))$$
$$y(t) = W_{out}x(t)$$

(1)

Only parameters of readout weights W_{out} are subject to training. Which can obtain a closed-form solution by extremely fast algorithms such as ridge regression. Assuming the internal states and desired outputs are stored in X and Y. By training to find a solution to the least squares problem in Equation 2, the calculation formula of the readout weights W_{out} is got in Equation 3.

$$\min_{W_{out}} ||W_{out}X - Y||_2^2 \tag{2}$$

$$W_{out} = Y \cdot X^{-1} \tag{3}$$

Before training phase, three are three main hyper-parameters of ESNs need to initialize - w^{in}, α and $\rho(W_{res})$.

- w^{in} is an input-scaling parameter, the elements in W_{in} are commonly randomly initialized from a uniform distribution in $[-w^{in}, w^{in}]$.

- α is the sparsity parameter of W_{res}, denoting the proportion of non-zero elements in matrix.

- $\rho(W_{res})$ is the spectral radius parameter (the largest eigenvalue in absolute value) of W_{res}. W_{res} initialize from a matrix W, where the elements of W are generated randomly in $[-1, 1]$ and $\lambda_{max}(W)$ is the largest eigenvalue.

$$W_{res} = \rho(W_{res}) \cdot \frac{W}{\lambda_{max}(W)} \tag{4}$$

The extreme training efficiency of the ESN approach derives from the fact that only the readout weights are trained, while the weights of input and reservoir are initialized under certain conditions and then are left untrained.

After training phase, ESNs can give predictions $\hat{Y}$ of new data $\hat{U}$ with the trained W_{out} by Equation 5.

$$\begin{aligned} \hat{X}_t &= f(W_{in}\hat{U}_t + W_{res}\hat{X}_{t-1}) \\ \hat{Y}_t &= W_{out}\hat{X}_t \end{aligned} \tag{5}$$

Assuming that the size of the reservoir are fixed by N, The input time series data has T-length D-dimension. For the typical reservoir computing Equation 1, its complexity is computed in Equation 6.

$$C_{typical} = \mathcal{O}(\alpha TN^2 + TND) \tag{6}$$

A pivotal property of ESNs is Echo State Property (ESP), which characterizes valid ESN dynamics. ESP essentially describes that the states of reservoir should asymptotically depend only on the driving input signal, which means the state is an echo of the input, meanwhile, and the influence of initial conditions should progressively vanish with time. Briefly, according to Jeager's paper [6], the ESN is that every echo state vector x is uniquely determined for every input sequence u. It implies that the nearby echo states have the similar input histories.

Definition 2 (Echo State Property (ESP)) *A network with state transition equation F with has the echo state property if for each input sequence $U = [u(1), u(2), ...u(N)]$ and all couples of initial states x, x', it could hold the condition in Equation 7.*

$$||F(U, x) - F(U, x')|| \to 0 \quad as \quad N \to \infty \tag{7}$$

The existence of ESP are verified by a necessary condition and a sufficient condition with spectral radius $\rho(W_{res})$ and the largest singular value $\overline{\sigma}(W_{res}) = ||W||_2$:

$$\begin{aligned} echo\ state\ property &\Rightarrow \rho(W_{res}) < 1 \\ \overline{\sigma}(W_{res}) < 1 &\Rightarrow echo\ state\ property \end{aligned} \tag{8}$$

Two conditions are commonly used in ESN literature for reservoir initialization [22]. In recent years, the conditions for the ESP have been further investigated and refined in successive contributions [23, 24, 10, 25, 26]. For example, basd on the two basic conditions, Buehner et al. [23] provided a rigorous bound for guaranteeing asymptotic stability of ESN, which was required in the applications where stability is required. Authors extended L2-norm $||W_{res}||_2$ to the D-norm

3

$||W_{res}||_D = \overline{\sigma}(DW_{res}D^{-1})$. They connected two conditions of echo state property by $||W_{res}||_D = \rho(W_{res}) + \epsilon$ and the sufficient condition changes to Equation 9.

$$\overline{\sigma}D(W_{res}D^{-1}) < 1 \Rightarrow echo\ state\ property \tag{9}$$

Short-Term Memory (STM) denotes the memory effects connected with the transient activation dynamics of networks. Memory capacity (MC) is a quantitative measure of STM. MC is calculated by the determination coefficient of W_{out}^k, $d[W_{out}^k](u(n-k), y_k(n)) = \frac{cov^2(u(n-k), y_k(n))}{\sigma^2(u(n))\sigma^2(y_k(n))}$ when training the ESN to generate at its output units delayed versions $u(n-k)$. The STM capacity of the network according to MC is described in Equation 10

$$MC = \sum_{k=1}^{\infty} MC_k$$
$$MC_k = \max_{W_{out}^k} d[W_{out}^k](u(n-k), y_k(n)) \tag{10}$$

2.2 Designs

2.2.1 Basic models

In 2002, Jeager [6] proposed and named the echo state networks (ESNs) at the first time. He introduced ESN as a computationally efficient and easy learning method based on artificial recurrent neural networks (RNNs). The schema of ESN is that we described in section 2.1. The ESN were suitable for process chaotic time series, abounded nonlinear dynamical systems of the sciences and engineering. In 2004, Jeager et al. [27] applied ESN for predicting a chaotic time series and the accuracy were improved by a factor of 2400 over previous techniques.

In 2007, Jeager et al. [28] proposed another important structure, leaky-ESN, formalized in Equation 11, where $a \in [0,1]$ is the reservoir neuron's leaking rate and γ is the compound gain. Leaky-ESN's reservoir used leaky integrator neurons and the leaking rate was determined and fixed after model selection. By the effect analysis and experiments in slow dynamic systems, noisy time series and time-warped dynamic patterns, leaky-ESN outperformed than ESN and could incorporate a time constant to act as a low-pass filter and the dynamics can be slowed down.

$$x(t) = (1 - a\gamma)x(t-1) + \gamma f(W_{in}u(t) + W_{res}x(t-1)) \tag{11}$$

ESNs with output feedbacks was also emerging. The hidden state is not only affected by the input of this time and the hidden state of the previous time, but also affected by the output of the previous time, shown as Figure 1 right. The state transition equation changed to Equation 12.

$$x(t) = (1 - a\gamma)x(t-1) + \gamma f(W_{in}u(t) + W_{res}x(t-1) + W_{out}^{back}y(t-1)) \tag{12}$$

Based on these original versions, many deformations have been proposed.

2.2.2 Designs of reservoir

- **Dynamic weights of reservoir**.

In the basic versions of ESNs, the weights in reservoir are fixed after initialization. However, the randomly initialized reservoir could not be always the most suitable one for the specific data and tasks, and it brings about complications with choosing starting parameters. Thus, the dynamic reservoir changing when processing data is a main focus when designing reservoir.

Mayer et al. [29] changed the random weights to adapted weights by a method of adapting the recurrent layer dynamics and self-prediction. The weights of reservoir changed to $\hat{W}_{res}$ by the output weight through output-prediction $W_{out}^1 x_t = y(t)$ and self-prediction $W_{out}^2 x_t = x_{t+1}$ in Equation 13, where α is constant parameter. By adding self-prediction, the modified ESN could moderate the noise sensitivity.

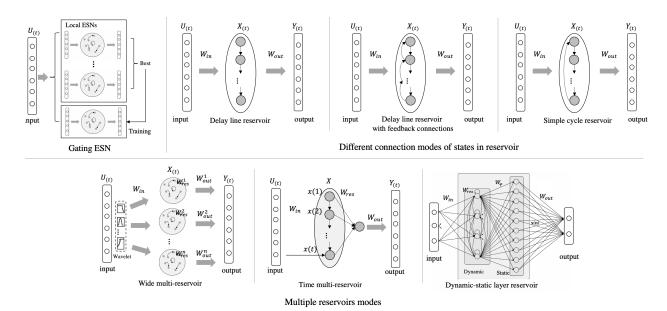$$\hat{W}_{res} = (1 - \alpha)W_{res} + \alpha W_{out}^2 \tag{13}$$

Figure 2: Designs of reservoir

Hajnal et al.[30] introduced a notion of critical echo state networks (CESN), re-setting W_{res} to $\hat{W}_{res}$ by Equation 14. In this case, $\hat{W}_{res}$ were able to approximate non-periodic dynamical systems, because the hidden layer were embedded by the input matrix W_{in} and output matrix W_{out}, and they can modify finite cycles.

$$\hat{W}_{res} = W_{res} + W_{in}W_{out} \tag{14}$$

Fan et al. [31] proposed principle neuron reinforcement (PNR) algorithm to alter the strength of internal synapses within the reservoir by modifying the reservoir connections towards the goal of optimizing the neuronal dynamics. The PNR altered the connection according to the output weights of a pre-training stage, shown in Equation 15.

$$W_{res}(t) = (1 + \rho(W_{res}))W_{res}(t - 1), \quad if \quad W_{out} > \xi \tag{15}$$

Babinec et al. [32] optimized ESN by the updating of W_{res} with Anti-Oja's learning (AO) by Equation 16, where η is a small positive constant. AO changed the weights by decreasing correlation between neurons. It increased the diversity between hidden neurons and the internal state dynamics became much richer.

$$W_{res}(n + 1) = W_{res}(n) - \Delta W_{res}(n)$$
$$\Delta W_{res}(n) = \eta y(n)(x(n) - y(n)W_{res}(n)) \tag{16}$$

Boedecker et al. [33] derived a new unsupervised learning rule based on intrinsic plasticity (IP) [34] to design weights of reservoir adapting dependency online. Then, Equation 1 changed to Equation 17. Where a is the gain vector and c is the bias vector. And the objective in Equation 2 changed to Kullback-Leibler divergence $D_{KL} = \sum p(y)log\frac{p(y)}{\overline{p}(y)}$. Where $p(y)$ denotes the sampled output distribution of a reservoir neuron and $\overline{p}(y)$ is the desired output distribution with Laplace distribution $\overline{p}(y) = \frac{1}{2\sigma}e^{(-\frac{|x-\mu|}{\sigma})}$.

$$y(t) = f(diag(a)W_{res}x(t - 1) + diag(a)W_{in}u(t) + c) \tag{17}$$

There are also many other designs about the dynamic connections. For example, Qiao et al. [35] proposed an adaptive lasso ESN (ALESN) to solve the problem of collinearity in randomly and sparsely connected reservoir. Wang et al. [36] used local plasticity rules, that allow different neurons to use different parameters to learn local features, to replace the same type of plasticity rule used in the unmodified ESN, that makes the connections between neurons remain to be global. Babinec et al. [37] designed gating ESNs by combining some local expert ESNs with different spectral radius and used the them of local ESN with best prediction result to train gating ESN, shown in the first structure of Figure 2.

- **Different connection modes of reservoir**.

There are three basic connections of states: 1) delay line reservoir (DLR), 2) delay line reservoir with feedback connections (DLRB) and 3) simple cycle reservoir (SCR) shown in Figure 2. In [38], authors tested the the minimal complexity of reservoir construction for obtaining competitive models and the memory capacity of such simplified reservoirs. They finally concluded that the memory capacity of SCR can be made arbitrarily close to the proved optimal value.

Besides the basic connections, Sun et al. [39, 40] proposed the adjacent-feedback loop reservoir (ALR) based on SCR. ALR had a deterministic reservoir structure, in which the reservoir units were connected orderly in the loop manner and were also connected to the preceding one by adjacent feedback. In [41], authors also designed a circle reservoir structure with wavelet-neurons. To evaluate the different connections of reservoirs,

For designing the size and topology of multiple reservoirs automatically, Growing ESN (GESN) [42, 43] was proposed. It owned a growing reservoir with multiple sub-reservoirs by adding hidden units to the network group by group based on block matrix theory. And in [44], authors introduced using PSO and SVD algorithms to pre-train a growing ESN with multiple sub-reservoirs by optimizing singular values.

- **Modes of multiple reservoirs**.

The above structures were built on the single reservoir, in practice, ESN can be extended with more reservoirs to achieve higher accuracy.

Chen et al. [45] designed shared reservoir modular ESN (SRMESN) to first divided the neural state space into several modules of subspaces with independent output weight and then put the data belonging to each subspace into the same reservoir. Which means the ESNs own n reservoirs rather than only one, shown in Figure 2. Many designs were proposed based on this. In [46], the inputs of this structure were utilized different wavelets of sequence; In [47], authors built a stacking ESN by stacking ensemble of reservoir computing learners based that structure; In [48], variational mode decomposition (VMD) was used to pick data as the input of different reservoirs.

Meanwhile, another structure is shown in 2, the output $y(t)$ could also calculated from the multiple hidden states from time 1 to t [49]. However, the linear nature of the output layer may limit the capability of exploring the available information, since higher-order statistics of the signals are not taken into account. There are also other deigns of the output layer. In [50], authors used the nonlinear readout to represent the output probabilities; In [51], authors presented a Volterra filter structure and used principal component analysis to reduce the number of effective signals transmitted to the output layer.

Gallicchio and Micheli [8] pointed out that mapping the states of reservoir of ESN to the high-dimensional feature space was beneficial to both memory and good nonlinear mapping capabilities of the network, and proposed to extend the ESN with a random static nonlinear hidden layer. This type of structural variant is collectively referred to as φ-ESN, shown in Figure 2. In [52], An orthogonal least squares φ-ESN (OLS-φ-ESN) was proposed for generating the nodes of the extended static nonlinear hidden layer by applying incremental randomized learning method.

There are also other designs based on multiple reservoirs. In [53], authors proposed Broad-ESN through the unsupervised learning algorithm of restricted Boltzmann machine (RBM) to determine the number of reservoirs and aimed to fully reflect dynamic characteristics of a class of multivariate time series. Xue et al. [54] proposed decoupled ESN (DESN), used lateral inhibition unit to decouple the representation of state before input into multiple reservoirs. This structure have better robustness with respect to the random generation of reservoir weight matrix and feedback connections.

- **Analysis of short-term memory in reservoirs**.

One of the most important properties of ESN is short-term memory, first introduced in [55]. And all of the basic ESN including the above reservoir designs worked by modeling it.

The topological structure of the reservoir relates to STM. Ma et al. [56] first tested it. They concluded that the reservoir topology can be constructed according to the given memory span r, r means the network can recover an input signal r time steps ago. They also designed a model to convert r to W_{res}.

Different inputs also influence the STM. Barancok et al. [57] calculated memory capacity (MC) of the networks for various input data, both random and structured, and showed that for uniformly distributed input data, the higher interval shift lead to higher MC; for structured data, depending on data properties.

Meanwhile, the hyper-parameters w^{in}, $\rho(W_{res})$, α can also affect STM. Farkas et al. [58] systematically analyzed the MC from the perspective of that three hyper-parameters and their relations. They concluded that the optimal w^{in} value

moves with the reservoir size N, while $\rho(W_{res})$ and α approach for large N. And the reservoir size increases the MC sublinearly.

Besides, Koryakin et al. [59] proofed that the reservoir size and the output feedback strength crucially co-determines the likelihood of generating an effective ESN by experiments - somewhat smaller networks can yield better performance; the output feedback strength drives the dynamic reservoir but it can also block suitable reservoir dynamics.

Further, Stockdill et al. [60] determine how leaking, loops, cycles and discrete time to influence the suitability of the reservoir. The results were that the single most important feature of a reservoir neural network is the discrete time step nature of input propagation; The loops and cycles can replicate each other; The potential limitation of energy conservation is equivalent to limiting the spectral radius.

2.2.3 Optimization of hyper-parameters

Some parameters in ESNs can be adjusted in advance - three hyper-parameters w^{in}, α and $\rho(W_{res})$, which are initialized before training in basic ESN, leaky rate a, active function f, training readout methods and regularization parameters. Different initialization results in different performance.

Venayagamoorthy et al. [61] studied effects of spectral radius $\rho(W_{res})$ and settling time ST, the number of iterations of reservoir between input and output. The experiments showed that $\rho(W_{res}) = 0.8$ gives best result and the increase in ST adversely affects the ESN performance. Which showed that ESNs prefer short-term memory and desire no long-term echoing arrangement.

The standard practice is the random initialization of the hyper-parameters, subject to few constraints. Although this results in a simple-to-solve optimization problem, it is in general suboptimal, and several additional criteria have been devised to improve its design. Rather than tuning by hand, more and more works focus on evolutionary optimization to optimize the parameters.

The basic algorithms for searching optimal solution can apply in this context. Cai et al. [62] have tested genetic algorithm (GA) and cross-validation (CV) method when optimizing the network parameters. The simulation showed that GA optimized ESN had better prediction accuracy while CV had better optimizing efficiency. Luo et al. [63] applied PSO and made up two interacting aspects of regularization of W_{out}. Venayagamoorthy et al. [61] optimized relevant parameters of leaky-ESN by PSO. Martin et al. [64] combined self-assembly (SA) and PSO. Basterrech et al. [65] tested the L2-Boost procedure [66] for initializing the matrix spectrum when the network is large. Maat et al. [67] and Cerina et al. [68] optimized the ESN hyper-parameters by using Bayesian optimization by further reduce the optimization cost.

There are also some new methods for searching ESN hyper-parameters. Ishu et al. [69] used a double evolutionary computation. The method first broad search the right parameters, then fine-tunes the the connectivity matrices of ESN. But it separated the topology and reservoir weights to reduce the search space. Thus, Ferreira et al. [70] designed an evolutionary method for simultaneous optimization of parameters, topology and reservoir weights. Liu et al.[71] proposed a covariance matrix adaption evolutionary strategy ESN (CMA-ES-ESN) for searching w^{in}, a and $\rho(W_{res})$. In [72], authors designed an improved grasshopper optimization algorithm (GOA) for the selection of ESN parameters. In [73], an improved GA was proposed with the immigration strategy. In [74], authors proposed a multi-step learning method for optimizing the hyper-parameters of multiple reservoirs ESNs. In [75], Thiede et al. presented a gradient based optimization algorithm to minimize the error on specific tasks and specific structures of ESNs.

Exiting other different novel optimization methods, like improved fruit fly optimization algorithm (IFOA) [76], fisher maximization based stochastic gradient descent (FM-SGD) [77], binary grey wolf algorithm (BGWO) [78] and so on.

2.2.4 Designs of regularization and training phase

Many efficient training method have propoed. For example, Shutin et al. [79] proposed variational Bayesian framework to train ESNs with automatic regularization and delay&sum readout adaptation. In [80], authors aimed to calculate the exact optimal output weights as a function of the structure of the system and the properties of the input and the target function. Scardapane et al. [81] decentralized algorithm in the case where data is distributed throughout a network of interconnected nodes. Chew et al. [82] employed Tikhonov regularization to train readouts and applied second-order proportional-integral-derivative feedback minimizes prediction bias. It reduced the number optimization parameters, while maintaining the estimation performance and following the excitation property of the estimator. Further, Couillet et al. [83] proposed a first theoretical performance analysis of the training phase of large dimensional linear ESNs based on random matrix theory.

The models discussed almost used supervised learning for training. Devert et al. [84] adapted the ESN with unsupervised learning at the first time. They tickled the problem of unclear ground truth by adaptive (1+1)-evolution strategy with an optimisation task: optimising an ESN amounts to optimising a real-valued vector representing the plastic weights of the network. Similarly, Jiang et al. [85] applied the methods in evolutionary continuous parameter optimization and the double pole balancing control problem to the evolutionary learning of ESN.

Many more efficient training schemes have been proposed to solve the ill-posed and over-fitting problem.

If the number of real-world training samples less than the size of the hidden layer, there may be an ill-posed problem. To overcome the ill-posed problem during the training process, Han et al. [86] employed the Laplacian eigenmaps to estimate the reservoir manifold and calculated output weights by the low-dimensional manifold. In [35], authors replaced the current component by the newly generated network with more compact structure. Qiao et al. [87] developed an adaptive Levenberg–Marquardt algorithm to achieve convergence and stability. The calculation of the W_{out} based on LM is equivalent to minimizing the objective function $E(W_{out}) = \frac{1}{2} \sum_{q=1}^{Q} (\hat{y}_q - y_q)^2$. Xu et al. [88] proposed a hybrid regularized ESN when there are a large number of unknown output weights. The method employed a sparse regression with the $L_{\frac{1}{2}}$ regularization, having properties of unbiasedness and sparsity, and the L_2 regularization, having ability on shrinking the amplitude of the output weights. Meanwhile, the $L_{\frac{1}{2}}$ norm regularization term could also used to overcome the iterative numerical oscillation problem, described in [89, 90].

For the widely existed over-fitting issue, Yang et al. [91] proposed an incremental ESN (IESN) with an leave-one-out cross-validation (LOO-CV) error, which was used to automatically identify the network architecture and avoid over-fitting. For the same goal, in [92], Bayesian Ridge ESN (BRESN), having a regularization effect is proposed; In [93], backtracking search optimization algorithm (BSA) determined the most appropriate output weights of ESN and avoid the over-fitted caused by the linear regression algorithm; In [94], the online sequential ESN with sparse recursive least squares (OSESN-SRLS) algorithm was proposed and two norm sparsity penalty constraints of output weights were separately employed to control the network size.

There are also some related studies about training phase. For example, good validation is key for good hyper-parameter tuning. Rather than using normal single validation split, Lukosevicius et al. [95] suggested k-fold cross-validation scheme, where the component that dominates the time complexity remained constant, not scaling up with k. Meanwhile, to provide a visualization of the modeled system dynamics, Lokse et al. [96] projected the output of the internal layer of ESNs on a lower dimensional space and enforced a regularization constraint that lead to better generalization capabilities. Further, Some works focused on how to make the results more accurate by changing the input more suitable for ESN. In [97], a series of linear parameter-varying models was used to extract feature as the input of ESNs. Li et al. [98] combined ESNs with a preprocessing based on the Hodrick-Prescott (HP) filter. HP filter can extract different components from a single time-series data.

3 Deep echo state networks

With the development of deep learning [99], Stacking architecture [100] has been introduced into reservoir computing. Deep echo state network (DeepESN) is the extension of ESNs to stacking multiple ESNs with deep learning framework. It was first introduced by Gallicchio et al. [11, 12] in 2017. DeepESNs combines the advantages of both ESNs and deep learning. The former pursues conciseness and effectiveness, but the latter focuses on the capacity to learn abstract, complex features in the service of the task.

3.1 Preliminaries

Definition 3 (Deep Echo State Networks (DeepESNs)) *DeepESN is a deep learning version of basic ESN. A Deep-ESN consists of an input layer, a dynamical stacking reservoirs component, and an output layer. The reservoir component of DeepESN is organized into a hierarchy of stacked recurrent layers, where the output of each layer acts as input for the next one.*

In a DeepESN, $u(t) \in R^D$ denotes the input value at time t of the time series, $x^{(l)}(t) \in R^N$ denotes the state of the reservoir layer i at time step t. Assuming L is the number of hidden layers, $x(t) = \{x^{(l)}(t)|i = l, 2, ...L\} \in R^{D+LN}$. The state transition equation is in 18. Where $W^{(1)} \in R^{N \times D}$ is the input weight matrix. $W^{(l)} \in R^{N \times N}$ for $l > 1$ is the
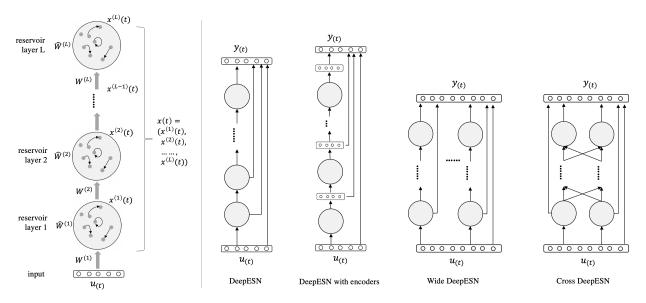
Figure 3: DeepESN and its variants

weight matrix for inter-layer connections from $(l-1)$-th layer to l-th layer. $\hat{W}^{(l)} \in R^{N \times N}$ is the recurrent weight matrix for layer l. $a^l \in [0,1]$ is the leaking rate for layer l. f is a nonlinear function such as tan and sigmoid.

$$
\begin{aligned}
x^{(l)}(t) &= F(x^{(l-1)}(t), x^{(l)}(t-1)) \\
&= (1 - a^{(l)})x^{(l)}(t-1) + a^{(l)}f(W^{(l)}i^{(l)}(t) + \hat{W}^{(l)}x^{(l)}(t-1)) \\
i^l(t) &= \begin{cases} u(t) & l = 1 \\ x^{l-1}(t) & l > 1 \end{cases}
\end{aligned}
\tag{18}
$$

In the basic DeepESN, the output turns to Equation 19. The matrix $W_out \in R^{M(D+LN)}$ contains the feedforward weights between reservoir neurons and the M output neurons.

$$
y(t) = W_{out}x(t) = W_{out}(x^{(1)}(t), ..., x^{(L)}(t))
\tag{19}
$$

Figure 3 shows the structure of a DeepESN and its variants. As for the standard shallow ESN model, DeepESN can embed the input into a rich state representation by the stacking reservoirs. Thus, DeepESN can show the intrinsic properties of state dynamics in layered RNN architectures [101], enable a multiple time-scales representation of the temporal information, naturally ordered along the network's hierarchy. DeepESN is also a efficiently trained deep neural network [9].

In [11], authors used the experiment results of state entropy and memory to show that DeepESNs can enhance the effect of known ESN factors of network design. In [102], authors pointed out even in the simplified linear setting, progressively higher layers focus on progressively lower frequencies by applying means of frequency analysis to investigate the hierarchically structured state representation in DeepESNs. Above two studty showed that higher layers in the hierarchy can effectively develop progressively slower dynamics.

Meanwhile, echo state property (ESP) have been generalized to DeepESNs in [103]. The authors gave the definition of a ESP based DeepESN in Equation 4 and proofed that a sufficient condition and a necessary condition to hold in case of deep RNN architectures, shown in Equation 21.

Definition 4 (Echo State Property of DeepESNs (ESP of DeepESNs)) *Assume a deepESN whose global dynamics are ruled by a function 18, Then the network has the echo state property if for each input sequence $U = [u(1), u(2), ...u(N)]$ and all couples of initial states x, x', it could hold the condition in Equation 20. $\hat{F}(U, x)$ is the global state of the network which has started in the initial state x and has been driven by the sequence U,*

$$||\hat{F}(U, x) - \hat{F}(U, x')|| \to 0 \quad as \quad N \to \infty$$

$$\hat{F}(U, x) = \begin{cases} x & if \quad U = \emptyset \\ F((u(N), \hat{F}([u(1), ..., u(N-1), x])) & if \quad U = [u(1), u(2), ...u(N)] \end{cases} \tag{20}$$

$$echo\ state\ property \Rightarrow \rho_g = \max_{k=1,...,L}(\rho((1-a^{(k)})I + a^{(k)}\hat{W}^{(k)})) < 1$$

$$C = \max_{k=1,...,L}((1-a^{(k)}) + a^{(i)}(C^{(i-1)}\overline{\sigma}(W^i) + \overline{\sigma}(\hat{W}^{(i)}))) < 1 \Rightarrow echo\ state\ property \tag{21}$$

3.2 Designs

Based on the basic DeepESN structure, many deformations have been propose.

Some existing works focused on the links between stacked reservoirs. In [104, 105], the deep structure was achieved along with encoders, shown in Figure 3, DeepESN with encoders. An encoder projected the representations in a reservoir into a lower-dimensional space, and then the new representations were processed by the next reservoir. Thus, the structure was built by stacking projection layers and encoding layers alternately. In addition to the advantages of DeepESNs, this design could also attenuate the effects of the collinearity problem in ESNs.

Like the idea of adding encoders, there are some other structures added different hidden layers between reservoirs. For example, Zhang et al. [106] used fuzzy clustering between reservoir to reinforce embedding features for classification enhancement. Arena et al. [107] added reduction layer between reservoirs to avoid the influence of noise data. Bo et al. [108] inserted some delayed modules between every two adjacent reservoirs. The reservoirs preserved the input characteristics by a relay mode and dealt with them asynchronously. This design achieved larger short-term memory capacity and richer dynamics.

Like the wide structure of multiple reservoirs in Figure 2, DeepESN can be organized by multiple modules of stacked reservoirs, shown in Figure 3, Wide DeepESN. Further, the reservoirs in different modules could have the cross connections, shown in Figure 3, Cross DeepESN. Based on the above connection structures, Carmichael et al. [109, 110] designed a DeepESN with modular architecture (Mod-DeepESN), which allowed for varying topologies of deep ESNs.

Like the multiple input structure in Figure 2, DeepESN can also be designed to receive different components of data. In [111], the additive decomposition (AD) is applied to the time series as a preprocessor. Bianchi et al. [112] implemented a bidirectional reservoir, whose last state captures also past dependencies in the input.

For the fundamental open issues of how to choose the number of layers in a deep RNN architecture, the work in [113] proposed an automatic method for the design of DeepESNs by the analysis of differentiation of the filtering effects of successive. The proposed approach allows to tailor the DeepESN architecture to the characteristics of the input signals.

4 Combinations of ESN and other models

4.1 Combinations of ESN and basic machine learning models

Shi et al. [114] combined ESN and support vector machine (SVM) as support vector echo-state machines (SVESMs). SVESMs performed linear support vector regression (SVR) in the high-dimension state space of reservoir. The new reservoir state variable of test data is generally written as Equation 22. Where $\alpha_s(t)$ is the Lagrange multiplier corresponding to the reservoir state vector $x(t)$. s is the number of support vectors. Scardapane et al. [115] combined the standard ESN with a semi-supervised support vector machine (S^{3}VM) for training ESN's adaptable connections in semi-supervised learning setting.

$$\hat{x}^{test}(t) = \sum_{i=1}^{s} \alpha_s(t)x(t)^T x^{test}(t) \tag{22}$$

Peng et al. [116] combined ESNs and multiplicative seasonal autoregressive integrated moving average (ARIMA) model [117] for mobile communication traffic series forecasting.

It is possible for ESN to handle tree-structure data, taking advantages from the compositionality of the structured representations both in terms of efficiency and in terms of effectiveness. The structure is in Figure 4 left. Gallicchio
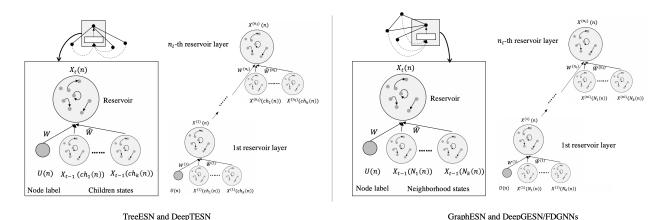
Figure 4: Tree echo state network and graph echo state network

et al. [118] presented the tree ESN (TreeESN) model to extend the applicability of the ESN approach from sequence data to tree structured data. In TreeESN, the state for a vertex n is computed from the input information attached to n itself, and the states of n's children nodes. Compared with the traditional ESN, the input of the children states takes the role of the state of n in previous time-step. Then, in 2018, the authors showed the advanced version - deep TreeESN (DeepTESN) [119]. Meanwhile, authors gave the theoretical properties, constraints, and empirical behavior of TreeESN and DeepTESN. Combining the deep learning, tree structure and reservoir computing take advantages from the layered architectural organization and from the compositionality of the structured representations both in terms of efficiency and in terms of effectiveness. The experimental results showed they outperformed the previous state-of-the-art in domains of document processing and computational biology.

It is also possible for ESN to handle graph-structure data. The structure is in Figure 4 right. The concept of reservoirs operating on graph-structure data (GraphESN) has been first introduced in [120]. GraphESN computed a state embedding for each vertex in an input graph. Similar to TreeESN, the state for a vertex n is computed from the input information attached to n itself, and the states of n's neighbors. Further, in [121], the authors of TreeESN demonstrated that the DeepESN approach had good performance in the case of learning with graph. They designed Deep GraphESN, enabling the development of fast and deep graph neural networks (FDGNNs).

4.2 Combinations of ESN and deep learning

- **Combinations of ESN and MLP**.

Babinec et al. [122] combined ESN with feedforward neural networks. As shown in Figure 5 left, the output layer of echo state network is substituted with feedforward neural network and the one-step training algorithm, pseudoinverse matrix in Equation 3, is replaced with backpropagation of error learning algorithm. The approach was tested in temperature forecasting and had better performance than basic ESN and multi-layered perceptron (MLP).

In [123], this idea was extended to multi reservoirs structure, shown as Figure 5 right. Similarly, in [124], different reservoirs were assigned to different wavelet extracted from time series.

- **Combinations of ESN and AE**.

Auto-encoder (AE) [13] structures are widely used in representation learning. It can extract the feature embedding in hidden layers with unsupervised learning technology. In [125, 126, 127], ESNs were used in AE structure, shown as Figure 5.

- **Combinations of ESN and GAN**.

Generative adversarial networks (GAN) [14], consisting of two networks, a discriminator to distinguish between teacher data and generated samples and a generator to deceive the discriminator, have good performance to both generate new data and classifying data. In [128], authors combined ESN and GAN for generating samples that reflect the dynamics in the teacher data. The discriminator in their model was a feedforward neural network so that backpropagation could not be used through training. The structure is in Figure 5.
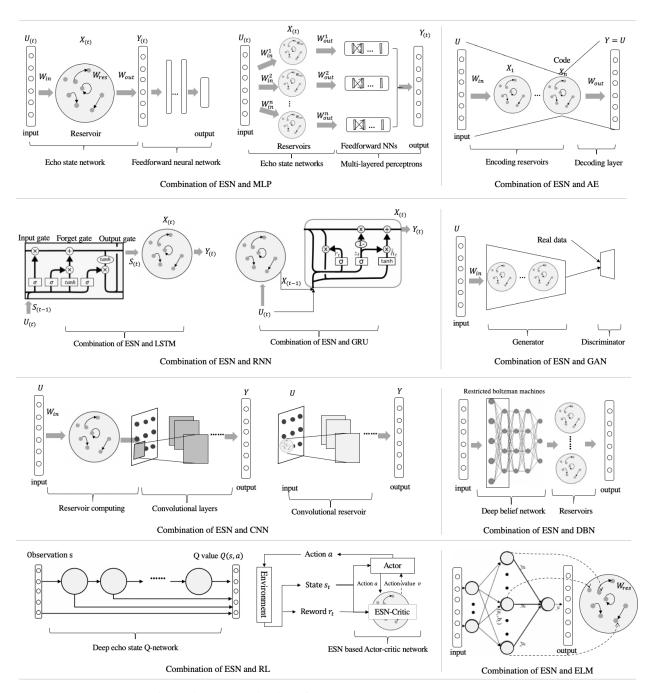
Figure 5: The combination of echo state network and other models

- **Combinations of ESN and RNN**.

Recurrent neural network (RNN) is a general term for a class of circularly linked neural networks. This type of neural network have shown excellent performance in processing sequence data. ESN is a special type of RNN. But at present, RNN refers to the deep learning network updated by error back-propagation. We use this expression in this section. In [22], authors compared RNN and ESN. Traditional gradient-descent-based RNN training methods adapt all connection weights, including input-to-RNN, RNN-internal, and RNN-to-output weights. In ESN, only the RNN-to-output weights are adapted.

RNNs were considered difficult to train, as their memory capabilities are often thwarted in practice by the exploding/vanishing gradients problem. As an attempt to solve these problems, long short-term memory cells (LSTM) [129] were designed to selectively forget information about old states and pay attention to new inputs. In [130], BiESN and BiLSTM were compared in the task of word sense disambiguation. The experiments confirmed the ability of the BiESN to capture more context information. And the main advantage of ESN over LSTM is the smaller number of trainable parameters and a simpler training algorithm. However, in ESN, the simplicity of not training its hidden layer may restrict reaching a better performance. In [131], the authors combined LSTM and ESN, replacing the hidden units of ESN by LSTM blocks. Initially all weights were randomly initialized and only the weights of output layer are learned. Then, whole network was trained by a gradient method but fixing the output weights. Yang et al. [132] also combined ESN and LSTM by applying the gate idea of LSTM to the reservoirs of ESN, aiming at the problem that information cannot be effectively retained, shown in Figure 5.

Gated recurrent unit (GRU) [133] is another enhanced version of RNN. In [134, 135], authors presented GRU-based ESNs to tackle complex real-world tasks while reducing the computational costs. The reservoir unit was replaced by the sparsely connected GRU neurons. Besides, the gating mechanisms improved the network's ability to deal with long-term dependencies within the data.

In addition to the direct combination of the two models, some works have incorporated the idea of RNN into ESN. Zheng et al. [136] found that in ESNs, no explicit mechanism captures the multi-scale characteristics of time series. They proposed long-short term echo state networks (LS-ESNs) consisting of three independent reservoirs. Each reservoir has recurrent connections of a specific time-scale to model the temporal dependencies of time series. Three state transition equations are in Equation 23.

$$
\begin{aligned}
x_{typical}(t) &= (1-a)x_{typical}(t) + af(W_{in}u(t) + W_{res}x_{typical}(t-1)) \\
x_{long}(t) &= (1-a)x_{long}(t) + af(W_{in}u(t) + W_{res}x_{long}(t-k)) \\
x_{short}(t) &= (1-a)x_{short}(t-1) + af(W_{in}u(t) + W_{res}x_{short}(t-1)) \quad until \quad x_{short}(t-m+1)
\end{aligned}
\tag{23}
$$

Li et al. [137] also found that ESN-based models have used only single-span features. Thus, they proposed multi-span DeepESN for features for considering the relations on different scales of a sequence. The state transition equation is in Equation 24. Where the input data is at the time t in the g-th group of the l-th reservoir layer. Group is a subsequence of raw sequence with a specific time span.

$$
x_{(g)}^{(l)}(t) = (1 - a_{(g)}^{(l)})x_{(g)}^{(l)}(t) + a_{(g)}^{(l)}f(W_{in(g)}^{(l)}u(t) + W_{res(g)}^{(l)}x_{(g)}^{(l)}(t-1))
\tag{24}
$$

• **Combinations of ESN and CNN**.

In recent years, convolutional neural networks (CNNs) have achieved great success in end-to-end pattern recognition. Many CNN-based models, such as VGG [15], ResNet [138] have been widely proposed. To combine the advantages of both ESN and CNN, in [139]. for series classification tasks, feature extraction was implemented by ESN, and the classification was obtained by the CNN. The connection of this structure is shown in Figure 5. Which had the advantage of the long-and-short term memory brought by the echo state and high-accuracy performance brought by the deep CNN.

In addition to connecting ESN and CNN directly, in [140], convolutional dynamics were built by extending randomly connected reservoirs to convolutional structures in the input-to-state transition, shown as Figure 5.The transition function of basic ESN changed to Equation 25. Where $(W_{res})^l, l = 1, ...L$ is a convolution at delay l and L is the length of time delays. This well optimized convolutional ESN can establish long-term time series predictions by ESN and short-term dependence by building the convolutional architecture.

$$
x(t) = f(W_{in}u(t) + \sum_{l=1}^{L}(W_{res})^l \otimes x(t-1) + W_{back}y(t))
\tag{25}
$$

• **Combinations of ESN and DBN**.

Restricted Boltzmann machine (RBM) [16, 141, 142] is the first multi-layer learning machine inspired by statistical mechanics. The adjustment of weights and the state evolution are determined by a certain probability distribution. In [143], authors proposed a fuzzy classification approach applying a combination of supervised ESN and unsupervised RBM, achieving two main strengths - hidden features extraction by RBM and dynamic patterns learning by ESN. The structure was the is to link two structures directly, the input of ESN is the output of RBM. In [144], an enhanced echo-state RBM (eERBM) was presented for network traffic prediction.

Deep belief network (DBN) is a typical unsupervised learning algorithm developed by Hinton et al. [145, 146]. DBN consisting of layers of RBM has strong ability to extract intrinsic nonlinear features of data through its hierarchical structure. Many works [147, 148] have explored ways to combine the advantages of DBN and ESN. Most of them is also the direct connection of them, show in Figure 5.

- **Combinations of ESN and RL**.

Reinforcement learning (RL) has been shown to be widely successful in many applications, ranging from playing games [149, 150] to robotics [151, 21]. RL algorithms can be divided into three categories - value-based, policy-based and actor-critic.

Deep Q-network (DQN) [150, 152, 153], one of the common value-based algorithms, has only one value function network. DeepESNs have abilities to replace the deep neural networks in DQN and compute Q-value [154, 155] with faster computing speed and global optimal solution.

Actor-critic network (AC) [156] is actually a combination of policy-based and value-based RL methods. Existing some works [157, 158, 159, 160] studied the application of ESN in AC. Most of them put ESN into Critic net. The main advantage is that ESN can learn a "memory task" through RL with back propagation [158]. Some tasks, like dynamic programming, need fast online training algorithms and robust modelisation of robot-environment interaction. Most of the method embeded RNN in RL networks for exhibiting continuous dynamics. To achieve fast online training and overcome the training difficulties of RNNs, ESN has been used in [157].

- **Combinations of ESN and ELM**.

Extreme learning machine (ELM) [161, 162] is a single hidden layer feedforward neural network which randomly chooses hidden nodes and analytically determines the output weights. Like ESN, the most significant characteristic of the ELM is that it provides good generalization performance at extremely fast learning speed than gradient-based learning algorithms [161]. Or In other words, ESN is a sort of ELM designed for time series. Some works [163, 164, 165, 166] have used ESN and ELM in real-time study application, like seismic multiple removal [165] and photovoltaic power prediction [164]. where the fast computation is required. In [167], An ESN-based ELM could model a non-sequential dataset or system with a high generalization capacity and no requirement of any optimization stage. The structure is shown in Figure 5.

5 Applications

For what regards the experimental analysis in applications, ESNs were shown to bring several advantages in both cases of synthetic and real-world tasks.

5.1 Benchmark datasets

- **Mackey-Glass System**.

Mackey-Glass (MG) system is a classical chaotic system that is often used evaluating time series prediction models. A standard discrete time approximation to the MG delay differential equation is in Equation 26

$$y(t+1) = y(t) + \delta(a\frac{y(t-\frac{\tau}{\delta})}{1+y(t-\frac{\tau}{\delta})^n} - by(t)) \tag{26}$$

Where the parameters δ, a, b, n are usually set to $0.1, 0.2, -0.1, 10$. The system becomes chaotic if $\tau > 16.8$ [27], most existing works [105] set $\tau = 17$ and initial $y(0) = 1.2$. The task can be predict the value in time series after step k. Which means using the data $u(1, ..., t)$ to forcast the value of $u(t + k)$.

- **Multiple Superimposed Oscillator series**.

Multiple Superimposed Oscillator Series (MSO) time series prediction problem is also a benchmark, the MSO time series data are generated by summing up several different frequency sine wave functions [168], which is derived as shown in Equation 27:

$$y(t) = \sum_{i=1}^{m} sin(\alpha_i t) \tag{27}$$

14

n general, the MSO can also be written as MSOm, where m represents the number of summed sine waves. The forecast difficulty is increased with the increase in the number of summed sine waves. To test ESN, most existing works set $s = 5$ and $\alpha_1 = 0.2, \alpha_2 = 0.311, \alpha_3 = 0.42, \alpha_4 = 0.51, \alpha_5 = 0.63$. [43, 134]

- **Lorenz system**.

The Lorenz system is one of the most typical benchmark tasks for time series prediction, described in Equation 28:

$$\begin{cases} \dfrac{dx}{dt} = -ax(t) + ay(t) \\ \dfrac{dy}{dt} = bx(t) - y(t) - x(t)z(t) \\ \dfrac{dz}{dt} = x(t)y(t) - cz(t) \end{cases} \tag{28}$$

Where $4a = 10$, $b = 28$, and $c = 8/3$. For getiting Lorenz time series, the Runge-Kutta method can generate sample of length L from the initial conditions (x_0, y_0, t_0) with a step size of 0.01. For prediction task, methods can predict the data in x-axis, y-axis or z-axis.

- **Rossler system**.

Rossler system a classical system which sufficiently chaotic phenomenon. It can be expressed in Equation 29

$$\begin{cases} \dfrac{dx}{dy} = -y - z \\ \dfrac{dy}{dz} = x + ay \\ \dfrac{dz}{dt} = b + z(x - c) \end{cases} \tag{29}$$

Where x, y, z are system variables, a, b, c are adjustable coefficients, and t denotes time dimension. When a = 0.2, b = 0.2, c = 5.7, Equation 29 produces chaos. For prediction task, methods can predict the data in x-axis, y-axis or z-axis.

- **NARMA system**.

Nonlinear auto-regressive moving average system (NARMA) is a highly nonlinear system incorporating memory. The tenth-order NARMA system depends on outputs and inputs 9 time steps back, which is difficult to identify. The tenth-order NARMA system identification problem is often used to test ESN models. The system is described in Equation

$$y(t+1) = 0.3y(t) + 0.05y(t)\sum_{i=0}^{9} y(t-i) + 1.5u(t-9)u(t) + 0.1 \tag{30}$$

where $u(t)$ is a random input of the system at time step t, drawn from a uniform distribution over [0, 0.5]. The output $y(t)$ is initialized by zeros for the first ten steps $(t = 1, ..., 10)$. One-step-ahead direct prediction on this time series is usually the task for ESN.

- **Sunspot datasets**.

The smoothed monthly mean sunspot numbers, which are one of the typical time series, are usually used as the benchmark prediction problems. The sunspot number is a dynamic manifestation of the strong magnetic field in the Sun's outer regions. Due to the complexity of potential solar activity and high non-linearity of the sunspot series, analyzing these series is a challenging task. There are two public datasets - SILSO and NOAA. The task can be forecasting the sunspot number in the future.

The World Data Center, Sunspot Index and Long-term Solar Observations (SILSO) [169], provides an open-source monthly sunspot series from 1749 to 2020.

Another dataset comes from the National Oceanic and Atmospheric Administration (NOAA) [170], which includes 3174 samples of sunspot activity numbers from 1750 to 2013.

15

- **Temperature datasets**.

A temperature series can show chaotic behavior, which makes them difficult to be forecasted accurately. There are two public benchmark temperature datasets - USHCN and DMTM. To test ESN model, the task can be climate forecasting after k days on these two datasets.

The U.S. Historical Climatology Network Monthly dataset (USHCN) [171] is publicly available and consists of daily meteorological data of stations in 48 states of the United States spanning from 1887 to 2014. It has everyday climate variables for each station, temperature, precipitation and snow precipitation.

Another temperature dataset is daily minimum temperatures in Melbourne (DMTM). There are 3650 minimum temperature points in Melbourne collected from January 1st, 1981 to December 31th, 1990.

- **Medical datasets**.

The Medical Information Mart for Intensive Care - three version dataset (MIMIC-III) [172] is a public dataset collected at Beth Israel Deaconess Medical Center from 2001 to 2012. It contains over 58,000 hospital admission records of 38,645 adults and 7,875 neonates. Potential tasks include mortality prediction, disease prediction, vital sign control, and patient subtyping.

- **Others**.

Henon map system [173]: a typical discrete-time dynamical sys-tem exhibiting a characteristic chaotic behavior; Labour dataset [174]: Monthly unemployment rate in US from 1948 to 1977; Gasoline dataset [175]: US finished motor gasoline product supplied; Santa Fe laser series [176]: the intensity of a NH3-FIR laser in a chaotic regime; Electricity dataset [177]: Half-hourly electricity demand in England; Photovoltaic power dataset [178]: Solar power data in Twentynine Palms and Whidbey Island; UCR database archive [179]: 358 publicly available synthetic and real time series databases.

5.2 Abnormal data orientated

In practice, the uncertainty in real-world dynamic system is quite prevalent such as data noise, data imbalance and multiple data formats. With the unclear data, the low accuracy of ESNs for real-world tasks showed up.

- **Data noises**.

Data noise is the most common problem. For example, the various noises are often introduced into an industrial system by sensors. The most intuitive method [168, 180] addressed the problem by data preprocessing in some specific fields such as nonlinear system modeling. For example, Xu et al. [181] gave a wavelet-denoising algorithm. Some works tickled the issue according to concerning the output uncertainty and internal states uncertainty causing by data noise. For example, in [182], noise addition was integrated into ESN to describe the additive noises of uncertainty of internal states the output, shown in Equation 31, where v and n are independent white Gaussian noise sequences. Some works used limitation in objective to alleviate the influence from noises. For example, Senn et al. [183] proposed abstract learning with the constrain $|W_{out}X_r| \leq Y_r$ of the objective function for the noise data, where Y_r is the maximum deviations could be tolerate and X_r is distance around the center value of X. Then Rigamonti et al. [184] gave the adaptive solution.

$$
\begin{aligned}
x(t) &= f(W_{in}u(t) + W_{res}x(t-1)) + v_{t-1} \\
y(t) &= W_{out}x(t) + n_t
\end{aligned}
\tag{31}
$$

For outliers data, in [185, 186], authors used Gaussian process to avoid the accumulative iteration error by establish the direct relation between the reservoir state and outputs. In [187], authors replaced the Gaussian distribution with Laplace distribution. In [188], Shen et al. used variational inference methods. In [189], Generalized correntropy induced metric (GCIM) was robust to outliers with a proper shape parameter. And Guo et al. [190] gave a correntropy induced loss function for ESN to improve its anti-noise capacity.

- **Imbalanced and incomplete data**.

Many actual datasets are mainly imbalanced. For example, in electronic health records (EHRs), the records of common diseases are much more than those of rare diseases. And for classification task, the lack of data in a minority class

may lead to uneven accuracy. Chen et al. [191] used well-trained ESN as the memory to replace the method of setting invariable thresholds between different classes.

Meanwhile, in the complex industrial environment, data missing often occurred in the process of data acquisition and transition, causing the incomplete dataset. The proposal of a deep bidirectional ESN (DBESN) [192] could model the incomplete dataset. It extracted the features along with forward and backward time scales and used deep auto-encoder to construct the incomplete output and input.

- **Dynamic data format**.

Under normal conditions, the inputs of existing ESNs are mostly limited to static and not dynamic temporal patterns.

For process dynamic data, Tanisaro et al. [193] used the idea of time warping invariant neural network (TWINN) [194]. Where the time warping in networks can be considered as a variation of the speed of the process. Zhang et al. [195] noticed the timeliness of data (time varying data) and proposed adaptive forgetting (AF) factor to ESN. AF could automatically tune the valid data window size according to prediction error magnitude. Tanaka et al. [196] addressed data that were generated by multiple dynamical systems switched with time by using time-varying reservoir. The reservoir adaptively changed depending on a statistical property of teacher data. For a class of discrete-time dynamic nonlinear systems (DDNS), Yao et al. [197] adjusted reservoir state adaptively according to the characteristics of input signals. Moreover, in dynamic data form, the reliability of former data also needs to be considered. In [198], the authors presented a recursive Bayesian linear regression ESN (RBLR-ESN) to change the confidence level of the prior data. The model was well-designed for long-range sequence.

The relation in time series data is mainly one-way time sequence, in natural language processing context, the relations reflect in bidirectional sequence. Schaetti [199] described Bidirectional ESN to model the order reverse order relations in sentences.

- **Expert knowledge**.

For a specific system, ESN may achieve better performance when incorporating the expert knowledge. For example, a physics-informed ESN [200, 201] could ensure that their predictions do not violate physical laws compared to conventional ESNs.

- **High dimensional data**.

Besides sequence data, ESNs can also applied in multi-dimensional data, such as those observed in the context of image recognition (2D imamge), renewable energy (3D wind modeling) and human centered computing (3-D inertial body sensors).

For image processing, in [202], ESNs were applied in classifying the digits of the MNIST, a basic database for computer vision; Wen et al. [203] designed an ensemble convolutional ESN for face recognition; And Duan et al. [204] used ESN for image restoration. For video data, in [205], ESNs were presented for the task of classifying high-level human activities. To cater for 3-D and 4-D processes, Quaternion-valued ESN (QESNs) [206] used quaternion nonlinear activation functions with local analytic properties on quaternion variable $q = q_r + \iota q_\iota + j q_j + \kappa q_\kappa$.

5.3 Real-world tasks orientated

As pertains to real-world problems, the ESN approaches recently proved effective in a variety of domains.

- **Industrial applications**.

ESNs are widely used in industry [207, 208, 17, 207], considering in particular the energy and manufacturing sectors. in energy fields, ESNs have applied in traditional energy forecasting [209], like prognostic of fuel cells [210, 97, 211], health of batteries [212], oil production platform control [213], gas prediction [132] and electric ship [214], and clean energy forecasting, like wind power generation[215, 216] and photovoltaic power prediction [164, 217, 217, 197]. ESN-based models have also been used in manufacturing, such as motor control [218], detection and diagnosis of anomaly and fault[219, 220, 221, 222, 106, 223, 89, 72], production system monitor [224, 225] and communications of sensors [226, 227, 228, 229, 230, 231], mobile [116, 232], satellite [233], wireless [234] and 5G Systems [235].

- **Medical applications**.

Medical field is also an important application of ESNs. At present, the main practices include feature learning of vital signs, such as electroencephalogram (EEG)-based feature extraction [125, 236, 237] and event detection [238, 239,

240, 241, 242], electrocardiogram (ECG)-based feature clustering [243], atrial fibrillation detection [244, 245], arterial blood pressure prediction [246] and magnetic resonance imaging (MRI)-based disease diagnosis [247]. Meanwhile, there are many methods for disease diagnosis. For example, ESNs methods have been applied in the diagnosis of parkinson'sdisease [248, 249], prediction of dialysis in critically ill patients [250], diagnosis of breast cancer [251], classification of ICU sepsis [18], detection of oral cancer [252], classification of bone marrow cell [253] and prediction of blood glucose concentration for type 1 diabetes [254]

- **Spatio-temporal applications**.

Many works [255, 256, 257] have designed ESN-based model for processing spatio-temporal data. Spatio-temporal data widely appear in traffic application. In this domain, ESNs can be used in traffic forecasting [258, 259, 47, 258], destination prediction [260], bike-sharing application [261, 111] and pedestrian counting system [262].

There are a lot of spatio-temporal data in astronomy and meteorology. sIn the field of astronomy, sunspot activity forecasting is a hot issue and the sunspot datasets is one of the benchmark dataset as we described in 5.1. Many works tested their models on this task [263]. Besides, there are also many ESN-based models designed for solar irradiance prediction [264, 265, 266, 267, 268], metereological forecasting[261, 111, 269], temperature prediction [198], water/stream flow forecasting [270, 271, 163] and wind speed Forecasting[272, 131, 273]

- **Financial applications**.

Financial forecasting is the focus of time series forecasting [261, 111], using ESNs model, many methods have proposed for stock data mining [274], stock price prediction [275], stock trading system control [276] and financial Data Forecasting [19].

- **Computer vision**.

In the field of computer vision (CV), ESN have been used in image processing, such as image segmentation [277, 278, 279, 280, 281], image restoration [204], facial expression recognition [203]. Many methods also could process radio audio data [282, 113, 261, 283], like video traffic [284], video annotation [285], audio Classification [115], speech recognition [286, 287] and emotion recognition [288, 289, 290, 291]. Meanwhile, there are also many applications based on 3D data, like 3D motion pattern indexing [292, 293, 294], activity /gesture recognition [295, 205, 202, 296] and human eye movements [297].

- **Nature language processing**.

In nature language processing (NLP) field, sentence processing [298, 299], existing methods have applied ESN idea in the tasks of grammatical structure learning [300, 301, 302], probabilistic language modeling [303], word embeddings [304] and word sense disambiguation [130].

- **Robotics**.

In the field of agent (robot) control, robot navigation/localization [305, 306], robotics trajectory control [307, 308, 20] and evolutionary Robotics [309] have been the applications of ESNs.

- **Ambient-assisted-living applications**.

Ambient assisted living (AAL) task aims to recognize the behavior of humans in their every-day environments. For example, by monitoring the regularity of people activities, enhancing the degree of personalization of smart home services. Wireless sensor networks is a mean to gather relevant data for the purpose of modeling the user's behavior and vast amount of temporal data is generated through the interaction of humans with the sensors. In this context, the interest in the adoption of ESN methodologies to discover relevant patterns from streams of sensorial information is constantly increasing. [310]

6 Discussions

6.1 Challenges and open questions

According to the analysis of the above research progress, there are many breakthroughs and good applications about ESN since it was proposed. However, some open questions still need to be solved.

Table 1: Category of methods related to ESN

Category	Subcategory	Researches	Achievements	Feature works
Basic ESN	Network analysis	[55] [56] [57] [58] [59] [60]	ESP,STM,MC,EOC...	More evaluation metrics; Relations between properties.
	Dynamic weights	[29] [30] [31] [32] [33] [35]	Self-prediction,critical, reinforcement,IP...	Auto-ML; Simplification complexity.
	Reservoir connections	[38] [39] [40] [41] [42] [43]	DLR,DLRB,SCR...	Dynamic study; Relation analysis.
	Multiple reservoirs	[45] [46] [47] [48] [49] [50] [51] [8] [52] [53] [54]	Wide,growing...	Auto-ML; Dynamic study; Relation analysis.
	Hyper-parameters optimization	[61] [62] [63] [61] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78]	GA,PSO,SA,DE...	Simplification complexity; Improve universality; More practical.
	Training phase designs	[85] [84] [79] [80] [81] [82] [81] [83] [96] [87] [91] [92] [88] [93] [94] [95] [89] [90] [97] [98]	Ill-posed,over-fitting	Simplification complexity; Improve universality; More practical.
Deep ESN	Network analysis	[11] [12] [101] [9] [11] [102] [103]	ESP,STM,MC,EOC...	More evaluation metrics; Relations between properties.
	Structure designs	[112] [113] [104] [105] [106] [107] [108] [109] [110] [111]	Stacked,wide,cross...	Auto-ML; Dynamic study; Relation analysis.
Combinations	Non-DL	[114] [115] [116] [118] [119] [120] [121]	SVM,ARIMA, Tree,Graph	More ML methods; Unstructured data
	DL	[122] [164] [123] [124] [125] [126] [127] [128] [130] [131] [132] [134] [135] [136] [137] [139] [140] [143] [144] [147] [148] [154] [155] [157] [158] [159] [160] [157] [163] [165] [166]	MLP,AE,GAN,RNN, CNN,DBN,RL,ELM...	More ML methods; Unstructured data; Simplification complexity; Improve accuracy; More practical.

1. **How the parameters of reservoir network structure effect the tasks? individually and collectively? generally and specifically?**

   The designs of network structures related to many considerations, such as hyper-parameters initialization and network connection. Some works have studied that the influence of three basic hyper-parameters and network topology to the final tasks. And they tried to suggest selections. But most of them tested the impact by changing one variable and fixing other variables. Meanwhile, in different application scenarios or downstream tasks, the impact seems to be different. Thus, three open questions are: 1) Do the parameters related to network design interact with each other? 2) What are their common and separate relations to the final tasks? 3) How the design affects common tasks and specific applications?

2. **What is the best setting of reservoir network structure? Is there an optimal reservoir network structure that we can construct from each task? How to connect different dynamics and different best reservoir structure?**

   Although some achievements have been made in the selection of hyper-parameters, but the best setting of weight matrix initialization and network topology is still an open question. Some theories have proved the setting method on some chaotic systems, like Mackey-Glass system, but for specific tasks, such as industrial and medical applications, the network settings are often determined by experience and experiments with not much theoretical guidance.

   Meanwhile, before we considered how to design a best network for a task, we can't prove whether the optimal design exists. Even if it does, it is unknown whether the idea of ESN can achieve it. To go further, we are not sure about the definition of the "best", both the results of the task and the performance of the network need to be considered.

Table 2: Applications of methods related to ESN

Category	Applications	Researches	Achievements	Feature works
Abnormal data	Noises	[168][180] [182] [181] [183] [184] [185] [186] [187] [188] [189] [190]	Noises Outliers	Wave oriented; Data enhancement.
	Imbalance	[191] [192] [86] [35]	Imbalanced Missing	Unequal intervals; Other data structure.
	Dynamic data format	[193] [195] [198] [196] [197] [311] [199]	Temporal Two direction	Real-time Changing
	Expert	[200] [201]	Physic law	Other knowledge
	High dimensional	[202] [205] [203] [204] [206]	2D,3D,4D	Unstructured data.
Real-world tasks	Industrial	[207] [208] [17] [207] [210] [97] [211] [209] [212] [213] [132] [214] [215] [216] [164] [217] [217] [197] [218] [219] [220] [221] [222][106] [223] [89] [72] [224] [225] [226] [227] [228] [229] [230] [231] [116] [232] [234] [235]	Energy Manufacturing	Others
	Medical	[125] [236] [237] [238] [239] [240] [241] [242] [243] [244] [245] [246] [247] [248] [249] [250] [251] [18] [252] [253] [254]	EEG,ECG,MRI Cancer Sepsis	Others
	Financial	[111] [274] [275] [276] [19]	Stock Prize	Others
	AAL	[310]	Representation	Others
	Spatio-temporal	[255] [256] [257] [258] [259] [47] [258] [260] [111] [262] [263] [264] [265] [266] [267] [268] [111] [269] [198] [270] [271] [163] [272] [131] [273]	Traffic Astronomy Meteorology	Others
	CV	[277] [278] [279] [280] [281] [204] [203] [282] [113] [283] [284] [285] [115] [286] [287] [288] [289] [290][291] [292] [293] [294] [295] [205] [202] [296] [297]	Image segmentation Emotion recognition gesture recognition	Others
	NLP	[298] [299] [300] [301] [302] [303] [304] [130]	Sentence processing Embedding	Others
	Robotics	[305] [306] [307] [308] [20] [309]	Navigation Localization	Others

For example, memory capacity and active information storage are both the measure to assess the computational properties of ESNs. Under different assessments, different optimal structures may be obtained.

3. **How to ensure the accuracy of ESN in fast calculation? On the contrary? Can ESNs be the plain substitute for BP RNN in the future?**

At present, there is little or no literature demonstrate that ESNs can completely replace other RNN methods. In most of complex tasks such as speech recognition, deep learning methods with gradient descent by BP algorithm still have

the best accuracy due to their highly nonlinear of 'black-box'. However, ESNs could be an alternative in some tasks that are designed to run on smaller, constrained devices where the size or performance of RNN is limited. But the complexity of ESNs is not always small. In order to get the best structure, some complex optimization algorithms are used, leading to heavy computational cost, which will weaken the competitiveness like training time of ESNs. However, reducing the complexity of the optimization algorithm often goes against improving the accuracy of the results. How to design a fast and accurate optimization algorithm is a problem.

6.2 Opportunities and future works

Based on the above three open questions, we propose some possible directions for future work.

1. **Study the relations between the parts of ESN.**

   The parts of ESN include inputs like data type and frequency, reservoir structure like network connections and number of layers, initialization parameters like matrix sparsity and scale, and outputs like results accuracy. Learning their relations is helpful to better design the network structure. For example, determination of the relations between input datasets and hyper-parameters helps adaptively adjust the structure and weight of reservoir; Studies on the frequency analysis of DeepESN dynamics allowed to develop an algorithm for the automatic setup of (the number of layers of) a DeepESN; The analysis of dynamic relationship and representation within ESN by using different technologies, like dimensionality reduction methods, time-varying graphs and other recurrence analysis tools, also provides insights for understanding the input system possibly.

2. **Study the optimization algorithm which is not only accurate but also fast.**

   Some optimization algorithms discard the advantage of fast computation of ESN when they pursue high accuracy. How to reduce the complexity of the optimization algorithm is a necessary future work. Meanwhile, the studies about speeding up the training process of ESNs, such as designing mini-batches, parallelizing calculation and taking advantage of GPU-based devices possibility, are required. Meanwhile, some basic issues still need to be addressed, like over-fitting problem.

3. **Study the method of automatically generating the optimal network structure of ESNs.**

   The method can not only improve the traditional parameter optimization method, but also refer to the deep learning method. For example, the techniques that can automate machine learning models both structurally and parametrically under the AutoML paradigm. Possible work can import the idea in AutoML area to reservoir-based realm and leverage the advances of modular topology. A fully modular ensemble architecture of ESN is also worth studying.

4. **Study the evaluation metrics of network property.**

   The network structure with the highest prediction accuracy does not necessarily have the best network performance. And different assessments lead to different optimal structure. The relations between network performance described by existing assessments needs to be studied. For example, a possible further research is the relationship between the maximal MC, the loss of the echo state property and the edge of chaos, which is rather complex nowadays. Another work worth studying is the relevance of network property and task outcome and the network dynamics in different tasks. These two kinds of studies can not only help to complete the task better, but also provide insights for network dynamic understanding and data relationship.

5. **From theory to practice.**

   At present, many existing practical aspects for successfully applying ESNs are not universal and should be filtered depending on a particular task. For example, most of the optimization algorithms are experimented on the simulation system, like MG system. But in the real-world industrial practice, the stability and regularity of the system cannot be guaranteed. Some rules and expert knowledge need to be considered. Although some methods are proposed for practical tasks, they are also not the only possible approaches, and can most likely be improved upon. Thus, the future work not only needs to put forward the solid theory, but also needs to put the method into practice.

6. **From specific applications to universal applications.**

   Some work is proposed for a particular point, but it is not suitable for most scenarios. For example, a noise-robust ESN can process univariable time series, but it can't deal with multivariable data. However, in the real world, multivariate time series are common. We do not deny the design of solutions to specific difficulties, we think that the scope of application of the method is a worthy research direction to improve the algorithm.

7. **From traditional applicable data to diversified data**

   Since ESNs was proposed, it has been used to model time series data. Although some methods propose to model noisy sequence data, there are many other irregularities in time series data such as unequal intervals between observations in two adjacent time points and different sampling rates between different time series. Meanwhile, the basic ESNs model short-term memory with no explicit mechanism to connect the multi-span memory or learn longer memory.

But time series may have different information in different spans, and have strong correlation with long-term data, such as periodic data. Therefore, future developments can be boosted by an enriched representation of the input dynamics, exploiting the time-scale differentiation in time granularity. At present, there are methods to apply ESN to tree-structured data and graph-structured data. In the future, ESN may also be used in unstructured data.

8. **Develop DeepESNs.**

DeepESNs combines the advantages of both ESNs and deep learning. The former pursues conciseness and effectiveness, but the latter focuses on the capacity of learning complex features. Thus, there is a gap between these two approaches, and a potentially future direction is to balance the efficiency and the feature learning capacity. But existing works focus more on the designs of stacked layers but ignore the basic theory. But better understanding the constitutive meaning of stacking layers of DeepESNs is the foundations for further model variants and extensive applications. Thus, more works about the basic theory are required.

7   Conclusion

In this paper, we categorize the ESN-based methods to basic ESNs, DeepESNs and combinations, then analyze them from the perspective of theoretical studies, network designs and specific applications. Finally, we discuss the challenges and opportunities this area, summarize three open questions and propose eight possible future works. This review aims to summarize the current ESN-based works and guide the future works with potential research directions.

References

[1] Williams R.J. Rumelhart D.E., Hinton G.E. Learning internal representations by error propagation. *Neurocomputing: Foundations of Research*, page pp. 673–695, 1988.

[2] Doya K. Bifurcations in the learning of recurrent neural networks. *Proceedings of IEEE International Symposium on Circuits and Systems*, 6:pp. 2777–2780, 1992.

[3] Sotirios P. Chatzis and Yiannis Demiris. The copula echo state network. *Pattern Recognit.*, 45(1):570–577, 2012.

[4] James Martens and Ilya Sutskever. Learning recurrent neural networks with hessian-free optimization. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 1033–1040, 2011.

[5] Mantas Lukosevicius. A practical guide to applying echo state networks. *Neural Networks: Tricks of the Trade - Second Edition*, pages 659–686, 2012.

[6] Herbert Jaeger. Adaptive nonlinear system identification with echo state networks. In *Advances in Neural Information Processing Systems 15 [Neural Information Processing Systems, NIPS 2002, December 9-14, 2002, Vancouver, British Columbia, Canada]*, pages 593–600, 2002.

[7] Peter Tiño, Barbara Hammer, and Mikael Bodén. Markovian bias of neural-based architectures with feedback connections. *Perspectives of Neural-Symbolic Integration*, pages 95–133, 2007.

[8] Claudio Gallicchio and Alessio Micheli. Architectural and markovian factors of echo state networks. *Neural Networks the Official Journal of the International Neural Network Society*, 24(5):440–456, 2011.

[9] Claudio Gallicchio and Alessio Micheli. Deep echo state network (deepesn): A brief survey. *CoRR*, abs/1712.04323, 2017.

[10] Izzet B. Yildiz, Herbert Jaeger, and Stefan J. Kiebel. Re-visiting the echo state property. *Neural Networks*, 35:1–9, 2012.

[11] Claudio Gallicchio and Alessio Micheli. Deep reservoir computing: A critical analysis. In *24th European Symposium on Artificial Neural Networks, ESANN 2016, Bruges, Belgium, April 27-29, 2016*, 2016.

[12] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 268(dec.11):87–99, 2017.

[13] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008*, 2008.

[14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014.

[15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[16] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cogn. Sci.*, 9(1):147–169, 1985.

[17] Muhammad Mansoor, Francesco Grimaccia, and Marco Mussetta. Echo state network performance in electrical and industrial applications. In *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, pages 1–7, 2020.

[18] Miquel Alfaras, Rui Varandas, and Hugo Gamboa. Ring-topology echo state networks for ICU sepsis classification. In *46th Computing in Cardiology, CinC 2019, Singapore, September 8-11, 2019*, pages 1–4, 2019.

[19] Junxiu Liu, Tiening Sun, Yuling Luo, Qiang Fu, Yi Cao, Jia Zhai, and Xuemei Ding. Financial data forecasting using optimized echo state network. In *Neural Information Processing - 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part V*, pages 138–149, 2018.

[20] Leandro Maciel, Fernando A. C. Gomide, David Santos, and Rosangela Ballini. Exchange rate forecasting using echo state networks for trading strategies. In *IEEE Conference on Computational Intelligence for Financial Engineering & Economics, CIFEr 2014, London, UK, March 27-28, 2014*, pages 40–47, 2014.

[21] Julian Whitman, Raunaq M. Bhirangi, Matthew J. Travers, and Howie Choset. Modular robot design synthesis with deep reinforcement learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 10418–10425, 2020.

[22] Mantas Lukosevicius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.*, 3(3):127–149, 2009.

[23] Michael R. Buehner and Peter Michael Young. A tighter bound for the echo state property. *IEEE Trans. Neural Networks*, 17(3):820–824, 2006.

[24] Manjunath Gandhi and Herbert Jaeger. Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural Comput.*, 25(3):671–696, 2013.

[25] Gilles Wainrib and Mathieu Galtier. A local echo state property through the largest lyapunov exponent. *Neural Networks*, 76:39–45, 2016.

[26] Cuili Yang, Junfei Qiao, Honggui Han, and Lei Wang. Design of polynomial echo state networks for time series prediction. *Neurocomputing*, 290:148–160, 2018.

[27] Jaeger and H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.

[28] Herbert Jaeger, Mantas Luko?Evi?Ius, Dan Popovici, and Udo Siewert. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Netw*, 20(3):335–352, 2007.

[29] Norbert Michael Mayer and Matthew Browne. Echo state networks and self-prediction. In *Biologically Inspired Approaches to Advanced Information Technology, First International Workshop, BioADIT 2004, Lausanne, Switzerland, January 29-30, 2004. Revised Selected Papers*, pages 40–48, 2004.

[30] Márton Albert Hajnal and András Lörincz. Critical echo state networks. In *Artificial Neural Networks - ICANN 2006, 16th International Conference, Athens, Greece, September 10-14, 2006. Proceedings, Part I*, pages 658–667, 2006.

[31] Hsiao-Tien Fan, Wei Wang, and Zhanpeng Jin. Performance optimization of echo state networks through principal neuron reinforcement. In *2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017*, pages 1717–1723, 2017.

[32] Stefan Babinec and Jiri Pospichal. Improving the prediction accuracy of echo state neural networks by anti-oja's learning. In *Artificial Neural Networks - ICANN 2007, 17th International Conference, Porto, Portugal, September 9-13, 2007, Proceedings, Part I*, pages 19–28, 2007.

[33] Joschka Boedecker, Oliver Obst, Norbert Michael Mayer, and Minoru Asada. Studies on reservoir initialization and dynamics shaping in echo state networks. In *ESANN 2009, 17th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 22-24, 2009, Proceedings*, 2009.

[34] J. Triesch. A gradient rule for the plasticity of a neuron's intrinsic excitability. In *ICANN Artificial Neural Networks: Biological Inspirations*, 2005.

[35] Junfei Qiao, Lei Wang, and Cuili Yang. Adaptive lasso echo state network based on modified bayesian information criterion for nonlinear system modeling. *Neural Comput. Appl.*, 31(10):6163–6177, 2019.

[36] Xinjie Wang, Yaochu Jin, and Kuangrong Hao. Evolving local plasticity rules for synergistic learning in echo state networks. *IEEE Trans. Neural Networks Learn. Syst.*, 31(4):1363–1374, 2020.

[37] Stefan Babinec and Jiri Pospichal. Gating echo state neural networks for time series forecasting. In *Advances in Neuro-Information Processing, 15th International Conference, ICONIP 2008, Auckland, New Zealand, November 25-28, 2008, Revised Selected Papers, Part I*, pages 200–207, 2008.

[38] Ali Rodan and Peter Tiño. Minimum complexity echo state network. *IEEE Trans. Neural Networks*, 22(1):131–144, 2011.

[39] Xiao-chuan Sun, Hongyan Cui, Ren Ping Liu, Jianya Chen, and Yunjie Liu. Modeling deterministic echo state network with loop reservoir. *J. Zhejiang Univ. Sci. C*, 13(9):689–701, 2012.

[40] Lijuan Sun, Xinyan Yang, Jian Zhou, Juan Wang, and Fu Xiao. Echo state network with multiple loops reservoir and its application in network traffic prediction. In *22nd IEEE International Conference on Computer Supported Cooperative Work in Design, CSCWD 2018, Nanjing, China, May 9-11, 2018*, pages 689–694. IEEE, 2018.

[41] Hongyan Cui, Chen Feng, Yuan Chai, Ren Ping Liu, and Yunjie Liu. Effect of hybrid circle reservoir injected with wavelet-neurons on performance of echo state network. *Neural Networks*, 57:141–151, 2014.

[42] Li Fan-jun and Li Ying. An approach to design growing echo state networks. In *Intelligent Data Engineering and Automated Learning - IDEAL 2016 - 17th International Conference, Yangzhou, China, October 12-14, 2016, Proceedings*, pages 220–230, 2016.

[43] Junfei Qiao, Fanjun Li, Honggui Han, and Wenjing Li. Growing echo-state network with multiple subreservoirs. *IEEE Trans. Neural Networks Learn. Syst.*, 28(2):391–404, 2017.

[44] Ying Li and Fanjun Li. Pso-based growing echo state network. *Appl. Soft Comput.*, 85, 2019.

[45] Wei-Biao Chen, Qian-Li Ma, and Hong Peng. Shared reservoir modular echo state networks for chaotic time series prediction. In *International Conference on Machine Learning and Cybernetics, ICMLC 2010, Qingdao, China, July 11-14, 2010, Proceedings*, pages 2439–2443, 2010.

[46] Seong Ik Han and Jang Myung Lee. Precise positioning of nonsmooth dynamic systems using fuzzy wavelet echo state networks and dynamic surface sliding mode control. *IEEE Trans. Ind. Electron.*, 60(11):5124–5136, 2013.

[47] Javier Del Ser, Ibai Laña, Miren Nekane Bilbao, and Eleni I. Vlahogianni. Road traffic forecasting using stacking ensembles of echo state networks. In *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, Auckland, New Zealand, October 27-30, 2019*, pages 2591–2597, 2019.

[48] Ying Han, Yuanwei Jing, Kun Li, and Georgi M. Dimirovski. Network traffic prediction using variational mode decomposition and multi- reservoirs echo state network. *IEEE Access*, 7:138364–138377, 2019.

[49] Yili Xia, Beth Jelfs, Marc M. Van Hulle, José C. Príncipe, and Danilo P. Mandic. An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals. *IEEE Trans. Neural Networks*, 22(1):74–83, 2011.

[50] Arnaud Rachez and Masafumi Hagiwara. Augmented echo state networks with a feature layer and a nonlinear readout. In *The 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, June 10-15, 2012*, pages 1–8. IEEE, 2012.

[51] Levy Boccato, Amauri Lopes, Romis Attux, and Fernando J. Von Zuben. An extended echo state network using volterra filtering and principal component analysis. *Neural Networks*, 32:292–302, 2012.

[52] Qian Zhang, Xiaojie Zhou, and Jian Tang. Orthogonal least squares based incremental echo state networks for nonlinear time series data analysis. *IEEE Access*, 7:185991–186003, 2019.

[53] Xianshuang Yao and Zhanshan Wang. Broad echo state network for multivariate time series prediction. *J. Frankl. Inst.*, 356(9):4888–4906, 2019.

[54] Yanbo Xue, Le Yang, and Simon Haykin. Decoupled echo state networks with lateral inhibition. *Neural Networks*, 20(3):365–376, 2007.

[55] H. Jaeger. Short term memory in echo state networks. *German National Research Center for Information Technology*, GMD Report 152, 2002.

[56] Qian-Li Ma, Wei-Biao Chen, Jia Wei, and Zhiwen Yu. Direct model of memory properties and the linear reservoir topologies in echo state networks. *Appl. Soft Comput.*, 22:622–628, 2014.

[57] Peter Barancok and Igor Farkas. Memory capacity of input-driven echo state networks at the edge of chaos. In *Artificial Neural Networks and Machine Learning - ICANN 2014 - 24th International Conference on Artificial Neural Networks, Hamburg, Germany, September 15-19, 2014. Proceedings*, pages 41–48, 2014.

[58] Igor Farkas, Radomír Bosák, and Peter Gergel. Computational analysis of memory capacity in echo state networks. *Neural Networks*, 83:109–120, 2016.

[59] Danil Koryakin, Johannes Lohmann, and Martin V. Butz. Balanced echo state networks. *Neural Networks*, 36:35–45, 2012.

[60] Aaron Stockdill and Kourosh Neshatian. Restricted echo state networks. In *AI 2016: Advances in Artificial Intelligence - 29th Australasian Joint Conference, Hobart, TAS, Australia, December 5-8, 2016, Proceedings*, pages 555–560, 2016.

[61] Ganesh K. Venayagamoorthy and Shishir Bashyal. Effects of spectral radius and settling time in the performance of echo state networks. *Neural Networks*, 22(7):861–863, 2009.

[62] Mao Cai, Xingming Fan, Chao Wang, Linlin Gao, and Xin Zhang. Research for parameters optimization of echo state network. In *11th International Symposium on Computational Intelligence and Design, ISCID 2018, Hangzhou, China, December 8-9, 2018, Volume 2*, pages 177–180. IEEE, 2018.

[63] Qiwu Luo, Jian Zhou, Yichuang Sun, and Oluyomi Simpson. Jointly optimized echo state network for short-term channel state information prediction of fading channel. In *16th International Wireless Communications and Mobile Computing Conference, IWCMC 2020, Limassol, Cyprus, June 15-19, 2020*, pages 1480–1484. IEEE, 2020.

[64] Charles E. Martin and James A. Reggia. Fusing swarm intelligence and self-assembly for optimizing echo state networks. *Comput. Intell. Neurosci.*, 2015:642429:1–642429:15, 2015.

[65] Sebastián Basterrech. An empirical study of l2-boost with echo state networks. In *13th International Conference on Intellient Systems Design and Applications, ISDA 2013, Salangor, Malaysia, December 8-10, 2013*, pages 295–300, 2013.

[66] P Bühlmann, B Yu, and P Bűhlmann. Boosting with the l2-loss: Regression and classification. *Journal of the American Statistical Association*, 98(462):324–339, 2003.

[67] Jacob Reinier Maat, Nikos Gianniotis, and Pavlos Protopapas. Efficient optimization of echo state networks for time series datasets. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–7. IEEE, 2018.

[68] Luca Cerina, Giuseppe Franco, and Marco Domenico Santambrogio. lightweight autonomous bayesian optimization of echo-state networks. In *27th European Symposium on Artificial Neural Networks, ESANN 2019, Bruges, Belgium, April 24-26, 2019*, 2019.

[69] K. Ishu, T. Van Der Zant, V. Becanovic, and P. Ploger. Identification of motion with echo state network. In *OCEANS '04. MTTS/IEEE TECHNO-OCEAN '04*, 2004.

[70] Aida A. Ferreira and Teresa Bernarda Ludermir. Evolutionary strategy for simultaneous optimization of parameters, topology and reservoir weights in echo state networks. In *International Joint Conference on Neural Networks, IJCNN 2010, Barcelona, Spain, 18-23 July, 2010*, pages 1–7, 2010.

[71] Kai Liu and Jie Zhang. Optimization of echo state networks by covariance matrix adaption evolutionary strategy. In *24th International Conference on Automation and Computing, ICAC 2018, Newcastle upon Tyne, United Kingdom, September 6-7, 2018*, pages 1–6. IEEE, 2018.

[72] Abubakar Bala, Idris Ismail, Rosdiazli Ibrahim, Sadiq M. Sait, and Diego Oliva. An improved grasshopper optimization algorithm based echo state network for predicting faults in airplane engines. *IEEE Access*, 8:159773–159789, 2020.

[73] Ruoyu Huang, Zetao Li, and Bin Cao. A soft sensor approach based on an echo state network optimized by improved genetic algorithm. *Sensors*, 20(17):5000, 2020.

[74] Takanori Akiyama and Gouhei Tanaka. Analysis on characteristics of multi-step learning echo state networks for nonlinear time series prediction. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14-19, 2019*, pages 1–8, 2019.

[75] Luca Anthony Thiede and Ulrich Parlitz. Gradient based hyperparameter optimization in echo state networks. *Neural Networks*, 115:23–29, 2019.

[76] Qingyong Zhang, Hao Qian, Yuepeng Chen, and Deming Lei. A short-term traffic forecasting model based on echo state network optimized by improved fruit fly optimization algorithm. *Neurocomputing*, 416:117–124, 2020.

[77] Muhammed Maruf Öztürk, Ibrahim Arda Çankaya, and Deniz Ipekçi. Optimizing echo state network through a novel fisher maximization based stochastic gradient descent. *Neurocomputing*, 415:215–224, 2020.

[78] Junxiu Liu, Tiening Sun, Yuling Luo, Su Yang, Yi Cao, and Jia Zhai. Echo state network optimization using binary grey wolf algorithm. *Neurocomputing*, 385:310–318, 2020.

[79] Dmitriy Shutin, Christoph Zechner, Sanjeev R. Kulkarni, and H. Vincent Poor. Regularized variational bayesian learning of echo state networks with delay&sum readout. *Neural Comput.*, 24(4):967–995, 2012.

[80] Alireza Goudarzi and Darko Stefanovic. Towards a calculus of echo state networks. In *5th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2014, Cambridge, MA, USA, November 7-9, 2014*, pages 176–181, 2014.

[81] Simone Scardapane, Dianhui Wang, and Massimo Panella. A decentralized training algorithm for echo state networks in distributed big data applications. *Neural Networks*, 78:65–74, 2016.

[82] Jouh Yeong Chew, Daisuke Kurabayashi, and Yudai Nakamura. Echo state networks with tikhonov regularization: optimization using integral gain. *Adv. Robotics*, 29(12):801–814, 2015.

[83] Romain Couillet, Gilles Wainrib, Harry Sevi, and Hafiz Tiomoko Ali. Training performance of echo state neural networks. In *IEEE Statistical Signal Processing Workshop, SSP 2016, Palma de Mallorca, Spain, June 26-29, 2016*, pages 1–4, 2016.

[84] Alexandre Devert, Nicolas Bredèche, and Marc Schoenauer. Unsupervised learning of echo state networks: A case study in artificial embryogeny. In *Artificial Evolution, 8th International Conference, Evolution Artificielle, EA 2007, Tours, France, October 29-31, 2007, Revised Selected Papers*, pages 278–290, 2007.

[85] Fei Jiang, Hugues Berry, and Marc Schoenauer. Unsupervised learning of echo state networks: balancing the double pole. In *Genetic and Evolutionary Computation Conference, GECCO 2008, Proceedings, Atlanta, GA, USA, July 12-16, 2008*, pages 869–870, 2008.

[86] Min Han and Meiling Xu. Laplacian echo state network for multivariate time series prediction. *IEEE Trans. Neural Networks Learn. Syst.*, 29(1):238–244, 2018.

[87] Junfei Qiao, Lei Wang, Cuili Yang, and Ke Gu. Adaptive levenberg-marquardt algorithm based echo state network for chaotic time series prediction. *IEEE Access*, 6:10720–10732, 2018.

[88] Meiling Xu, Min Han, Tie Qiu, and Hongfei Lin. Hybrid regularized echo state network for multivariate chaotic time series prediction. *IEEE Trans. Cybern.*, 49(6):2305–2315, 2019.

[89] Cheng Lu, Peng Xu, and Lin-Hu Cong. Fault diagnosis model based on granular computing and echo state network. *Eng. Appl. Artif. Intell.*, 94:103694, 2020.

[90] Qiwu Luo, Xiaoxin Fang, Yichuang Sun, Jiaqiu Ai, and Chunhua Yang. Self-learning hot data prediction: Where echo state network meets NAND flash memories. *IEEE Trans. Circuits Syst. I Regul. Pap.*, 67-I(3):939–950, 2020.

[91] Cuili Yang, Xinxin Zhu, Zohaib Ahmad, Lei Wang, and Junfei Qiao. Design of incremental echo state network using leave-one-out cross-validation. *IEEE Access*, 6:74874–74884, 2018.

[92] Hoang Minh Nguyen, Gaurav Kalra, Tae Joon Jun, and Daeyoung Kim. A novel echo state network model using bayesian ridge regression and independent component analysis. In *Artificial Neural Networks and Machine Learning - ICANN 2018 - 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part II*, volume 11140 of *Lecture Notes in Computer Science*, pages 24–34. Springer, 2018.

[93] Zhigang Wang, Yu-Rong Zeng, Sirui Wang, and Lin Wang. Optimizing echo state network with backtracking search optimization algorithm for time series forecasting. *Eng. Appl. Artif. Intell.*, 81:117–132, 2019.

[94] Cuili Yang, Junfei Qiao, Zohaib Ahmad, Kaizhe Nie, and Lei Wang. Online sequential echo state network with sparse RLS algorithm for time series prediction. *Neural Networks*, 118:32–42, 2019.

[95] Mantas Lukosevicius and Arnas Uselis. Efficient cross-validation of echo state networks. In *Artificial Neural Networks and Machine Learning - ICANN 2019 - 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17-19, 2019, Proceedings - Workshop and Special Sessions*, pages 121–133, 2019.

[96] Sigurd Løkse, Filippo Maria Bianchi, and Robert Jenssen. Training echo state networks with regularization through dimensionality reduction. *Cogn. Comput.*, 9(3):364–378, 2017.

[97] Zhongliang Li, Zhixue Zheng, and Rachid Outbib. Adaptive prognostic of fuel cells by implementing ensemble echo state networks in time-varying model space. *IEEE Trans. Ind. Electron.*, 67(1):379–389, 2020.

[98] Ziqiang Li and Gouhei Tanaka. HP-ESN: echo state networks combined with hodrick-prescott filter for nonlinear time-series prediction. In *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, pages 1–9. IEEE, 2020.

[99] Y. Bengio Y.LeCun and G Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[100] Li Deng, Dong Yu, and John C. Platt. Scalable stacking and learning for building deep architectures. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2012, Kyoto, Japan, March 25-30, 2012*, pages 2133–2136, 2012.

[101] Claudio Gallicchio and Alessio Micheli. Why layering in recurrent neural networks? A deepesn survey. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.

[102] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Hierarchical temporal representation in linear reservoir computing. *CoRR*, abs/1705.05782, 2017.

[103] Claudio Gallicchio and Alessio Micheli. Echo state property of deep reservoir computing networks. *Cogn. Comput.*, 9(3):337–350, 2017.

[104] Qianli Ma, Lifeng Shen, and Garrison W. Cottrell. Deep-esn: A multiple projection-encoding hierarchical reservoir computing framework. *CoRR*, abs/1711.05255, 2017.

[105] Qianli Ma, Lifeng Shen, and Garrison W. Cottrell. Deepr-esn: A deep projection-encoding echo-state network. *Inf. Sci.*, 511:152–171, 2020.

[106] Shaohui Zhang, Zhenzhong Sun, Man Wang, Jianyu Long, Yun Bai, and Chuan Li. Deep fuzzy echo state networks for machinery fault diagnosis. *IEEE Trans. Fuzzy Syst.*, 28(7):1205–1218, 2020.

[107] Paolo Arena, Luca Patanè, and Angelo Giuseppe Spinosa. Robust modelling of binary decisions in laplacian eigenmaps-based echo state networks. *Eng. Appl. Artif. Intell.*, 95:103828, 2020.

[108] Ying-Chun Bo, Ping Wang, and Xin Zhang. An asynchronously deep reservoir computing for predicting chaotic time series. *Appl. Soft Comput.*, 95:106530, 2020.

[109] Zachariah Carmichael, Humza Syed, and Dhireesha Kudithipudi. Analysis of wide and deep echo state networks for multiscale spatiotemporal time series forecasting. *CoRR*, abs/1908.08380, 2019.

[110] Zachariah Carmichael, Humza Syed, Stuart Burtner, and Dhireesha Kudithipudi. Mod-deepesn: Modular deep echo state network. *CoRR*, abs/1808.00523, 2018.

[111] Taehwan Kim and Brian R. King. Time series prediction using deep echo state networks. *Neural Comput. Appl.*, 32(23):17769–17787, 2020.

[112] Filippo Maria Bianchi, Simone Scardapane, Sigurd Løkse, and Robert Jenssen. Bidirectional deep-readout echo state networks. In *26th European Symposium on Artificial Neural Networks, ESANN 2018, Bruges, Belgium, April 25-27, 2018*, 2018.

[113] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Design of deep echo state networks. *Neural Networks*, 108:33–47, 2018.

[114] Zhiwei Shi and Min Han. Support vector echo-state machine for chaotic time-series prediction. *IEEE Trans. Neural Networks*, 18(2):359–372, 2007.

[115] Simone Scardapane and Aurelio Uncini. Semi-supervised echo state networks for audio classification. *Cogn. Comput.*, 9(1):125–135, 2017.

[116] Yu Peng, Miao Lei, Jun-Bao Li, and Xiyuan Peng. A novel hybridization of echo state networks and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting. *Neural Comput. Appl.*, 24(3-4):883–890, 2014.

[117] G. E. P. Box and David A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Publications of the American Statistical Association*, 65(332):1509–1526, 1970.

[118] Claudio Gallicchio and Alessio Micheli. Tree echo state networks. *Neurocomputing*, 101:319–337, 2013.

[119] Claudio Gallicchio and Alessio Micheli. Deep reservoir neural networks for trees. *Information Sciences*, 2018.

[120] Claudio Gallicchio and Alessio Micheli. Graph echo state networks. In *International Joint Conference on Neural Networks, IJCNN 2010, Barcelona, Spain, 18-23 July, 2010*, pages 1–8, 2010.

[121] Claudio Gallicchio and Alessio Micheli. Fast and deep graph neural networks. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 3898–3905, 2020.

27

[122] Stefan Babinec and Jiri Pospichal. Merging echo state and feedforward neural networks for time series forecasting. In *Artificial Neural Networks - ICANN 2006, 16th International Conference, Athens, Greece, September 10-14, 2006. Proceedings, Part I*, pages 367–375, 2006.

[123] Ali Rodan and Peter Tiño. Negatively correlated echo state networks. In *ESANN 2011, 19th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 27-29, 2011, Proceedings*, 2011.

[124] Jianmin Wang, Yu Peng, and Xiyuan Peng. Anti boundary effect wavelet decomposition echo state networks. In *Advances in Neural Networks - ISNN 2011 - 8th International Symposium on Neural Networks, ISNN 2011, Guilin, China, May 29-June 1, 2011, Proceedings, Part I*, pages 445–454, 2011.

[125] Leilei Sun, Bo Jin, Haoyu Yang, Jianing Tong, Chuanren Liu, and Hui Xiong. Unsupervised EEG feature extraction based on echo state network. *Inf. Sci.*, 475:1–17, 2019.

[126] Heshan Wang, Q. M. Jonathan Wu, Jianbin Xin, Jie Wang, and Heng Zhang. Optimizing deep belief echo state network with a sensitivity analysis input scaling auto-encoder algorithm. *Knowl. Based Syst.*, 191:105257, 2020.

[127] Naima Chouikhi, Boudour Ammar, Amir Hussain, and Adel M. Alimi. Bi-level multi-objective evolution of a multi-layered echo-state network autoencoder for data representations. *Neurocomputing*, 341:195–211, 2019.

[128] Takanori Akiyama and Gouhei Tanaka. Echo state network with adversarial training. In *Artificial Neural Networks and Machine Learning - ICANN 2019 - 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17-19, 2019, Proceedings - Workshop and Special Sessions*, pages 82–88, 2019.

[129] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997.

[130] Alexander Popov, Petia D. Koprinkova-Hristova, Kiril Simov, and Petya Osenova. Echo state vs. LSTM networks for word sense disambiguation. In *Artificial Neural Networks and Machine Learning - ICANN 2019 - 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17-19, 2019, Proceedings - Workshop and Special Sessions*, pages 94–109, 2019.

[131] Erick López, Carlos Valle, Héctor Allende, and Esteban Gil. Long short-term memory networks based in echo state networks for wind speed forecasting. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 22nd Iberoamerican Congress, CIARP 2017, Valparaíso, Chile, November 7-10, 2017, Proceedings*, pages 347–355, 2017.

[132] Yinghua Yang, Xin Zhao, and Xiaozhi Liu. A novel echo state network and its application in temperature prediction of exhaust gas from hot blast stove. *IEEE Trans. Instrum. Meas.*, 69(12):9465–9476, 2020.

[133] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1724–1734, 2014.

[134] Xinjie Wang, Yaochu Jin, and Kuangrong Hao. A gated recurrent unit based echo state network. In *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, pages 1–7, 2020.

[135] Daniele Di Sarli, Claudio Gallicchio, and Alessio Micheli. Gated echo state networks: a preliminary study. In *International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2020, Novi Sad, Serbia, August 24-26, 2020*, pages 1–5, 2020.

[136] Kaihong Zheng, Bin Qian, Sen Li, Yong Xiao, Wanqing Zhuang, and Qianli Ma. Long-short term echo state network for time series prediction. *IEEE Access*, 8:91961–91974, 2020.

[137] Ziqiang Li and Gouhei Tanaka. Deep echo state networks with multi-span features for nonlinear time series prediction. In *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, pages 1–9, 2020.

[138] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016.

[139] Anguo Zhang, Wei Zhu, and Juanyu Li. Spiking echo state convolutional neural network for robust time series classification. *IEEE Access*, 7:4927–4935, 2019.

[140] Gege Zhang, Chao, Zheqing Li, and Weidong Zhang. A new PSOGSA inspired convolutional echo state network for long-term health status prediction. In *IEEE International Conference on Robotics and Biomimetics, ROBIO 2018, Kuala Lumpur, Malaysia, December 12-15, 2018*, pages 1298–1303, 2018.

[141] Yee Whye Teh and Geoffrey E. Hinton. Rate-coded restricted boltzmann machines for face recognition. In *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS) 2000, Denver, CO, USA*, pages 908–914, 2000.

[142] Geoffrey E. Hinton. Boltzmann machines. In *Encyclopedia of Machine Learning and Data Mining*, pages 164–168. Springer, 2017.

[143] Olga Fink, Enrico Zio, and Ulrich Weidmann. Fuzzy classification with restricted boltzman machines and echo-state networks for predicting potential railway door system failures. *IEEE Trans. Reliab.*, 64(3):861–868, 2015.

[144] Xiaochuan Sun, Shuhao Ma, Yingqi Li, Duo Wang, Zhigang Li, Ning Wang, and Guan Gui. Enhanced echo-state restricted boltzmann machines for network traffic prediction. *IEEE Internet Things J.*, 7(2):1287–1297, 2020.

[145] Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, 2006.

[146] Geoffrey E. Hinton. Deep belief nets. In *Encyclopedia of Machine Learning*, pages 267–269. Springer, 2010.

[147] Xu Cheng and Chenyuan Zhao. Prediction of tourist flow based on deep belief network and echo state network. *Rev. d'Intelligence Artif.*, 33(4):275–281, 2019.

[148] Xiaochuan Sun, Tao Li, Qun Li, Yue Huang, and Yingqi Li. Deep belief echo-state network and its application to time series prediction. *Knowl. Based Syst.*, 130:17–29, 2017.

[149] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[150] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy P. Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[151] Chelsea Finn and Sergey Levine. Deep visual foresight for planning robot motion. In *2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017*, pages 2786–2793, 2017.

[152] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin A. Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature.*, 518(7540):529–533, 2015.

[153] Hado van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 2094–2100, 2016.

[154] Istvan Szita, Viktor Gyenes, and András Lörincz. Reinforcement learning with echo state networks. In *Artificial Neural Networks - ICANN 2006, 16th International Conference, Athens, Greece, September 10-14, 2006. Proceedings, Part I*, pages 830–839, 2006.

[155] Hao-Hsuan Chang, Lingjia Liu, and Yang Yi. Deep echo state q-network (DEQN) and its application in dynamic spectrum sharing for 5g and beyond. *CoRR*, abs/2010.05449, 2020.

[156] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 1928–1937, 2016.

[157] Mohamed Oubbati. Anticipating rewards in continuous time and space with echo state networks and actor-critic design. In *ESANN 2011, 19th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 27-29, 2011, Proceedings*, 2011.

[158] Nico M. Schmidt, Matthias Baumgartner, and Rolf Pfeifer. Actor-critic design using echo state networks in a simulated quadruped robot. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14-18, 2014*, pages 2224–2229, 2014.

[159] Toshitaka Matsuki and Katsunari Shibata. Reinforcement learning of a memory task using an echo state network with multi-layer readout. In *Robot Intelligence Technology and Applications 5 - Results from the 5th International Conference on Robot Intelligence Technology and Applications, RiTA 2017, Daejeon, Korea, December 13-15, 2017*, pages 17–26, 2017.

[160] Petia D. Koprinkova-Hristova, Mohamed Oubbati, and Günther Palm. Adaptive critic design with echo state network. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10-13 October 2010*, pages 1010–1015, 2010.

[161] Guang-Bin Huang, Qin-Yu Zhu, and Chee Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501, 2006.

[162] Shifei Ding, Han Zhao, Yanan Zhang, Xinzheng Xu, and Ru Nie. Extreme learning machine: algorithm, theory and applications. *Artif. Intell. Rev.*, 44(1):103–115, 2015.

[163] Hugo Valadares Siqueira, Levy Boccato, Romis Attux, and Christiano Lyra. Echo state networks and extreme learning machines: A comparative study on seasonal streamflow series prediction. In *Neural Information Processing - 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part II*, pages 491–500, 2012.

[164] Iroshani Jayawardene and Ganesh K. Venayagamoorthy. Comparison of echo state network and extreme learning machine for PV power prediction. In *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid, CIASG 2014, Orlando, FL, USA, December 9-12, 2014*, pages 14–21, 2014.

[165] Heitor S. Carvalho, Farzin Shams, Rafael Ferrari, and Levy Boccato. Application of extreme learning machines and echo state networks to seismic multiple removal. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.

[166] Victor Henrique Alves Ribeiro, Gilberto Reynoso-Meza, and Hugo Valadares Siqueira. Multi-objective ensembles of echo state networks and extreme learning machines for streamflow series forecasting. *Eng. Appl. Artif. Intell.*, 95:103910, 2020.

[167] Ömer Faruk Ertugrul. A novel randomized machine learning approach: Reservoir computing extreme learning machine. *Appl. Soft Comput.*, 94:106433, 2020.

[168] Yanbo Xue, Le Yang, and Simon Haykin. Decoupled echo state networks with lateral inhibition. *Neural Networks the Official Journal of the International Neural Network Society*, 20(3):365–376, 2007.

[169] SILSO World Data Center. The international sunspot number, int. sunspot number monthly bull. online catalogue (1749-2016). *[Online]*, 2016.

[170] National Oceanic and Atmospheric Administration. Sunspot numbers. *[Online]*, 2014.

[171] Williams C. Menne, M. and Vose R. Long-term daily and monthly climate records from stations across the contiguous united states. *[Online]*, 2010.

[172] A. et al. Johnson. Mimic-iii, a freely accessible critical care database. *SCI. data*, 2016.

[173] M. Hénon. A two-dimensional mapping with a strange attractor. *[Online]*, 1976.

[174] U.S. Bureau of Labor Statistics. Monthly unemployment rate in us from 1948 to 1977, 2020. `www.data.bls.gov/timeseries/lns14000000`.

[175] U.S. Energy Information Administration. Weelky u.s. product supplied of finished motor gasoline (thousand barrels per day), 2020. `www.eia.gov/dnav/pet/hist//LeafHandler.ashx?n=PET&s=wgfupus2&f=W`.

[176] N.A. Gershenfeld and A.S. Weigend. Santa fe time series competition data. *[Online]*, 1994.

[177] J. W. Taylor. Short-term electricity demand forecasting using double seasonal exponential smoothing. *J. Oper. Res. Soc.*, 54(8):799–805, 2003.

[178] National Renewable Energy Laboratory. Solar power data for integration studies. *[Online]*, 2006.

[179] Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, and Gustavo Batista. The ucr time series classification archive, July 2015. `www.cs.ucr.edu/~eamonn/time_series_data/`.

[180] Jochen J. Steil. Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning. *Neural Networks*, 20(3):353–364, 2007.

[181] Meiling Xu, Min Han, and Hongfei Lin. Wavelet-denoising multiple echo state networks for multivariate time series prediction. *Inf. Sci.*, 465:439–458, 2018.

[182] Chunyang Sheng, Jun Zhao, Ying Liu, and Wei Wang. Prediction for noisy nonlinear time series by echo state network based on dual estimation. *Neurocomputing*, 82:186–195, 2012.

[183] Christoph Walter Senn and Itsuo Kumazawa. Abstract echo state networks. In Frank-Peter Schilling and Thilo Stadelmann, editors, *Artificial Neural Networks in Pattern Recognition - 9th IAPR TC3 Workshop, ANNPR 2020, Winterthur, Switzerland, September 2-4, 2020, Proceedings*, volume 12294 of *Lecture Notes in Computer Science*, pages 77–88. Springer, 2020.

[184] Marco Rigamonti, Piero Baraldi, Enrico Zio, Indranil Roychoudhury, Kai Goebel, and Scott Poll. Ensemble of optimized echo state networks for remaining useful life prediction. *Neurocomputing*, 281:121–138, 2018.

[185] Y. Liu, J. Zhao, and W. Wang. A gaussian process echo state networks model for time series forecasting. In *Joint IFSA World Congress and NAFIPS Annual Meeting, IFSA/NAFIPS 2013, Edmonton, Alberta, Canada, June 24-28, 2013*, pages 643–648, 2013.

[186] Xiaofeng Yang and Feng Zhao. Echo state network and echo state gaussian process for non-line-of-sight target tracking. *IEEE Syst. J.*, 14(3):3885–3892, 2020.

[187] Decai Li, Min Han, and Jun Wang. Chaotic time series prediction based on a novel robust echo state network. *IEEE Trans. Neural Networks Learn. Syst.*, 23(5):787–799, 2012.

[188] Lihua Shen, Jihong Chen, Zhigang Zeng, Jianzhong Yang, and Jian Jin. A novel echo state network for multivariate and nonlinear time series prediction. *Appl. Soft Comput.*, 62:524–535, 2018.

[189] Changhao Zhang, Yu Guo, Fei Wang, and Badong Chen. Generalized maximum correntropy-based echo state network for robust nonlinear system identification. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–6. IEEE, 2018.

[190] Yu Guo, Fei Wang, Badong Chen, and Jingmin Xin. Robust echo state networks based on correntropy induced loss function. *Neurocomputing*, 267:295–303, 2017.

[191] Qing Chen, Anguo Zhang, Tingwen Huang, Qianping He, and Yongduan Song. Imbalanced dataset-based echo state networks for anomaly detection. *Neural Comput. Appl.*, 32(8):3685–3694, 2020.

[192] Qiang Wang, Linqing Wang, Ying Liu, Jun Zhao, and Wei Wang. Time series prediction with incomplete dataset based on deep bidirectional echo state network. *IEEE Access*, 7:152533–152544, 2019.

[193] Pattreeya Tanisaro and Gunther Heidemann. Time series classification using time warping invariant echo state networks. In *15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, Anaheim, CA, USA, December 18-20, 2016*, pages 831–836, 2016.

[194] Guo Zheng Sun, Hsing Hen Chen, and Yee Chun Lee. Time warping invariant neural networks. In *Advances in Neural Information Processing Systems*, 1992.

[195] Zhang Song-lin and Li Xue. Adaptive forgetting factor echo state networks for time series prediction. *Int. J. Intell. Syst. Technol. Appl.*, 16(1):80–93, 2017.

[196] Gouhei Tanaka, Ryosho Nakane, and Akira Hirose. Echo state networks composed of units with time-varying nonlinearity. *Aust. J. Intell. Inf. Process. Syst.*, 17(2):34–39, 2019.

[197] Xianshuang Yao, Zhanshan Wang, and Huaguang Zhang. Prediction and identification of discrete-time dynamic nonlinear systems based on adaptive echo state network. *Neural Networks*, 113:11–19, 2019.

[198] Biaobing Huang, Guihe Qin, Rui Zhao, Qiong Wu, and Alireza Shahriari. Recursive bayesian echo state network with an adaptive inflation factor for temperature prediction. *Neural Comput. Appl.*, 29(12):1535–1543, 2018.

[199] Nils Schaetti. Bidirectional echo state network-based reservoir computing for cross-domain authorship attribution: Notebook for PAN at CLEF 2018. In *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018*, volume 2125 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

[200] Nguyen Anh Khoa Doan, Wolfgang Polifke, and Luca Magri. Physics-informed echo state networks for chaotic systems forecasting. In *Computational Science - ICCS 2019 - 19th International Conference, Faro, Portugal, June 12-14, 2019, Proceedings, Part IV*, pages 192–198, 2019.

[201] Dong-keun Oh. Toward the fully physics-informed echo state network - an ODE approximator based on recurrent artificial neurons. *CoRR*, abs/2011.06769, 2020.

[202] Nils Schaetti, Michel Salomon, and Raphaël Couturier. Echo state networks-based reservoir computing for MNIST handwritten digits recognition. In *2016 IEEE Intl Conference on Computational Science and Engineering, CSE 2016, and IEEE Intl Conference on Embedded and Ubiquitous Computing, EUC 2016, and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering, DCABES 2016, Paris, France, August 24-26, 2016*, pages 484–491, 2016.

[203] Guihua Wen, Huihui Li, and Danyang Li. An ensemble convolutional echo state networks for facial expression recognition. In *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015, Xi'an, China, September 21-24, 2015*, pages 873–878, 2015.

[204] Haibin Duan and Xiaohua Wang. Echo state networks with orthogonal pigeon-inspired optimization for image restoration. *IEEE Trans. Neural Networks Learn. Syst.*, 27(11):2413–2425, 2016.

[205] Luiza Mici, Xavier Hinaut, and Stefan Wermter. Activity recognition with echo state networks using 3d body joints and objects category. In *24th European Symposium on Artificial Neural Networks, ESANN 2016, Bruges, Belgium, April 27-29, 2016*, 2016.

[206] Yili Xia, Cyrus Jahanchahi, and Danilo P. Mandic. Quaternion-valued echo state networks. *IEEE Trans. Neural Networks Learn. Syst.*, 26(4):663–673, 2015.

[207] Stefano Dettori, Ismael Matino, Valentina Colla, and Ramon Speets. Deep echo state networks in industrial applications. In *Artificial Intelligence Applications and Innovations - 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5-7, 2020, Proceedings, Part II*, pages 53–63, 2020.

[208] Valentina Colla, Ismael Matino, Stefano Dettori, Silvia Cateni, and Ruben Matino. Reservoir computing approaches applied to energy management in industry. In *Engineering Applications of Neural Networks - 20th International Conference, EANN 2019, Hersonissos, Crete, Greece, May 24-26, 2019, Proceedings*, pages 66–79, 2019.

[209] S.-X. Lv H. Hu, L. Wang. Forecasting energy consumption and wind power generation using deep echo state network. *Renewable Energy*, 154:598–613, 2020.

[210] Simon Morando, Samir Jemei, Daniel Hissel, Rafael Gouriveau, and Noureddine Zerhouni. ANOVA method applied to proton exchange membrane fuel cell ageing forecasting using an echo state network. *Math. Comput. Simul.*, 131:283–294, 2017.

[211] Simon Morando, Samir Jemei, Rafael Gouriveau, Noureddine Zerhouni, and Daniel Hissel. Fuel cells prognostics using echo state network. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, Vienna, Austria, November 10-13, 2013*, pages 1632–1637, 2013.

[212] Luciano Sánchez, David Anseán, José Otero, and Inés Couso. Assessing the health of lifepo$_4$ traction batteries through monotonic echo state networks. *Sensors*, 18(1):9, 2018.

[213] Jean P. Jordanou, Eric Aislan Antonelo, and Eduardo Camponogara. Online learning control with echo state networks of an oil production platform. *Eng. Appl. Artif. Intell.*, 85:214–228, 2019.

[214] Jing Dai, Ganesh K. Venayagamoorthy, and Ronald G. Harley. Harmonic identification using an echo state network for adaptive control of an active filter in an electric ship. In *International Joint Conference on Neural Networks, IJCNN 2009, Atlanta, Georgia, USA, 14-19 June 2009*, pages 634–640, 2009.

[215] Ronaldo R. B. de Aquino, Otoni Nóbrega Neto, Ramon B. Souza, Milde M. S. Lira, Manoel A. Carvalho, Teresa Bernarda Ludermir, and Aida A. Ferreira. Investigating the use of echo state networks for prediction of wind power generation. In *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions, CIES 2014, Orlando, FL, USA, December 9-12, 2014*, pages 148–154, 2014.

[216] Manuel Dorado-Moreno, Pedro Antonio Gutiérrez, Sancho Salcedo-Sanz, Luis Prieto, and César Hervás-Martínez. Wind power ramp events ordinal prediction using minimum complexity echo state networks. In *Intelligent Data Engineering and Automated Learning - IDEAL 2018 - 19th International Conference, Madrid, Spain, November 21-23, 2018, Proceedings, Part II*, pages 180–187, 2018.

[217] Jing Zhang, Yuxi Liu, Yan Chen, Bao Yuan, Jiakui Zhao, and Di Liu. Photovoltaic output prediction model based on echo state networks with weather type index. In *ICIAI 2019: The 3rd International Conference on Innovation in Artificial Intelligence, Suzhou, China, March 15-18, 2019*, pages 91–95, 2019.

[218] Matthias Salmen and Paul-Gerhard Plöger. Echo state networks used for motor control. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, April 18-22, 2005, Barcelona, Spain*, pages 1953–1958, 2005.

[219] Oliver Obst, X. Rosalind Wang, and Mikhail Prokopenko. Using echo state networks for anomaly detection in underground coal mines. In *Proceedings of the 7th International Conference on Information Processing in Sensor Networks, IPSN 2008, St. Louis, Missouri, USA, April 22-24, 2008*, pages 219–229, 2008.

[220] Marcus Chang, Andreas Terzis, and Philippe Bonnet. Mote-based online anomaly detection using echo state networks. In *Distributed Computing in Sensor Systems, 5th IEEE International Conference, DCOSS 2009, Marina del Rey, CA, USA, June 8-10, 2009. Proceedings*, pages 72–86, 2009.

[221] Adam J. Wootton, Charles R. Day, and Peter W. Haycock. Fault detection in steel-reinforced concrete using echo state networks. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8, 2018.

[222] Jianyu Long, Shaohui Zhang, and Chuan Li. Evolving deep echo state networks for intelligent fault diagnosis. *IEEE Trans. Ind. Informatics*, 16(7):4928–4937, 2020.

[223] Mingjing Xu, Piero Baraldi, Sameer Al-Dahidi, and Enrico Zio. Fault prognostics by an ensemble of echo state networks in presence of event based measurements. *Eng. Appl. Artif. Intell.*, 87, 2020.

[224] Ganesh K. Venayagamoorthy. Online design of an echo state network based wide area monitor for a multimachine power system. *Neural Networks*, 20(3):404–413, 2007.

[225] Eric A. Antonelo, Eduardo Camponogara, and Bjarne Foss. Echo state networks for data-driven downhole pressure estimation in gas-lift oil wells. *Neural Networks*, 85:106–117, 2017.

[226] Emi Mathews and Axel Poigné. An echo state network based pedestrian counting system using wireless sensor networks. In *International Workshop on Intelligent Solutions in Embedded Systems, WISES 2008, Regensburg, Germany, July 10-11, 2008*, pages 1–14, 2008.

[227] Ling Qin, Rongqiang Hu, and Qi Zhang. Saving energy in wireless sensor networks based on echo state networks. In *Advances in Neural Networks - ISNN 2009, 6th International Symposium on Neural Networks, ISNN 2009, Wuhan, China, May 26-29, 2009, Proceedings, Part III*, pages 923–928, 2009.

[228] André Frank Krause, Bettina Bläsing, Volker Dürr, and Thomas Schack. Direct control of an active tactile sensor using echo state networks. In *Human Centered Robot Systems, Cognition, Interaction, Technology*, pages 11–21, 2009.

[229] Qingsong Song, Zuren Feng, Liangjun Ke, and Min Li. Echo state network for abrupt change detections in non-stationary signals. In *Ninth International Conference on Intelligent Systems Design and Applications, ISDA 2009, Pisa, Italy , November 30-December 2, 2009*, pages 226–231, 2009.

[230] Nalin Harischandra and Volker Dürr. A forward model for an active tactile sensor using echo state networks. In *IEEE International Symposium on Robotic and Sensors Environments Proceedings, ROSE 2012, Magdeburg, Germany, November 16-18, 2012*, pages 103–108, 2012.

[231] Junichi Kuwabara, Kohei Nakajima, Rongjie Kang, David T. Branson, Emanuele Guglielmino, Darwin G. Caldwell, and Rolf Pfeifer. Timing-based control via echo state network for soft robotic arm. In *The 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, June 10-15, 2012*, pages 1–8, 2012.

[232] Filippo Maria Bianchi, Simone Scardapane, Aurelio Uncini, Antonello Rizzi, and Alireza Sadeghian. Prediction of telephone calls load using echo state network with exogenous variables. *Neural Networks*, 71:204–213, 2015.

[233] Marc Bauduin, Anteo Smerieri, Serge Massar, and François Horlin. Equalization of the non-linear satellite communication channel with an echo state network. In *IEEE 81st Vehicular Technology Conference, VTC Spring 2015, Glasgow, United Kingdom, 11-14 May, 2015*, pages 1–5, 2015.

[234] Kenneth Gideon, Clement N. Nyirenda, and Clement Temaneh-Nyah. Echo state network-based radio signal strength prediction for wireless communication in northern namibia. *IET Commun.*, 11(12):1920–1926, 2017.

[235] Kangjun Bai, Yang Yi, Zhou Zhou, Shashank Jere, and Lingjia Liu. Moving toward intelligence: Detecting symbols on 5g systems through deep echo state network. *IEEE J. Emerg. Sel. Topics Circuits Syst.*, 10(2):253–263, 2020.

[236] Hoon-Hee Kim and Jaeseung Jeong. Decoding electroencephalographic signals for direction in brain-computer interface using echo state network and gaussian readouts. *Comput. Biol. Medicine*, 110:254–264, 2019.

[237] Rahma Fourati, Boudour Ammar, Yaochu Jin, and Adel M. Alimi. EEG feature learning with intrinsic plasticity based deep echo state network. In *2020 International Joint Conference on Neural Networks, IJCNN 2020, Glasgow, United Kingdom, July 19-24, 2020*, pages 1–8, 2020.

[238] Sudhanshu S. D. P. Ayyagari, Richard D. Jones, and Stephen J. Weddell. Eeg-based event detection using optimized echo state networks with leaky integrator neurons. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014, Chicago, IL, USA, August 26-30, 2014*, pages 5856–5859, 2014.

[239] Sudhanshu S. D. P. Ayyagari, Richard D. Jones, and Stephen J. Weddell. Optimized echo state networks with leaky integrator neurons for eeg-based microsleep detection. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2015, Milan, Italy, August 25-29, 2015*, pages 3775–3778, 2015.

[240] Rahma Fourati, Boudour Ammar, Chaouki Aouiti, Javier J. Sánchez Medina, and Adel M. Alimi. Optimized echo state network with intrinsic plasticity for eeg-based emotion recognition. In *Neural Information Processing - 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II*, pages 718–727, 2017.

[241] Fuji Ren, Yindong Dong, and Wei Wang. Emotion recognition based on physiological signals using brain asymmetry index and echo state network. *Neural Comput. Appl.*, 31(9):4491–4501, 2019.

[242] Fuji Ren, Yindong Dong, and Wei Wang. Correction to: Emotion recognition based on physiological signals using brain asymmetry index and echo state network. *Neural Comput. Appl.*, 32(10):6415, 2020.

[243] Miguel A. Atencia Ruiz, Claudio Gallicchio, Gonzalo Joya, and Alessio Micheli. Time series clustering with deep reservoir computing. In *Artificial Neural Networks and Machine Learning - ICANN 2020 - 29th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 15-18, 2020, Proceedings, Part II*, pages 482–493, 2020.

[244] Andrius Petrenas, Vaidotas Marozas, and Arunas Lukosevicius. Ventricular activity cancellation in ECG using an adaptive echo state network. In *IEEE 6th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2011, Prague, Czech Republic, September 15-17, 2011, Volume 1*, pages 378–382, 2011.

[245] Andrius Petrenas, Vaidotas Marozas, Leif Sörnmo, and Arunas Lukosevicius. An echo state neural network for QRST cancellation during atrial fibrillation. *IEEE Trans. Biomed. Eng.*, 59(10):2950–2957, 2012.

[246] Allan Fong, Ranjeev Mittu, Raj M. Ratwani, and James A. Reggia. Predicting electrocardiogram and arterial blood pressure waveforms with different echo state network architectures. In *AMIA 2014, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 15-19, 2014*, 2014.

[247] Axel Wismüller, Adora M. DSouza, Anas Z. Abidin, Xixi Wang, Susan K. Hobbs, and Mahesh B. Nagarajan. Functional connectivity analysis in resting state fmri with echo-state networks and non-metric clustering for network structure recovery. In *Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging, Orlando, Florida, United States, 21-26 February 2015*, page 94171M, 2015.

[248] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Deep echo state networks for diagnosis of parkinson's disease. In *26th European Symposium on Artificial Neural Networks, ESANN 2018, Bruges, Belgium, April 25-27, 2018*, 2018.

[249] Stuart E. Lacy, Stephen L. Smith, and Michael A. Lones. Using echo state networks for classification: A case study in parkinson's disease diagnosis. *Artif. Intell. Medicine*, 86:53–59, 2018.

[250] Thierry Verplancke, Stijn Van Looy, Kristof Steurbaut, Dominique Benoit, Filip De Turck, Georges De Moor, and Johan Decruyenaere. A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks. *BMC Medical Informatics Decis. Mak.*, 10:4, 2010.

[251] Summrina Kanwal Wajid, Amir Hussain, and Bin Luo. An efficient computer aided decision support system for breast cancer diagnosis using echo state network classifier. In *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health, CICARE 2014, Orlando, FL, USA, December 9-12, 2014*, pages 17–24, 2014.

[252] Mohammed Al-Maitah and Ahmad Ali AlZubi. Enhanced computational model for gravitational search optimized echo state neural networks based oral cancer detection. *J. Medical Syst.*, 42(11):205:1–205:7, 2018.

[253] Philipp Kainz, Harald Burgsteiner, Martin Asslaber, and Helmut Ahammer. Training echo state networks for rotation-invariant bone marrow cell classification. *Neural Comput. Appl.*, 28(6):1277–1292, 2017.

[254] Ning Li, Jianyong Tuo, Youqing Wang, and Menghui Wang. Prediction of blood glucose concentration for type 1 diabetes based on echo state networks embedded with incremental learning. *Neurocomputing*, 378:248–259, 2020.

[255] Harold Soh and Yiannis Demiris. Spatio-temporal learning with the online finite and infinite echo-state gaussian processes. *IEEE Trans. Neural Networks Learn. Syst.*, 26(3):522–536, 2015.

[256] Erik Schaffernicht, Victor Manuel Hernandez Bennetts, and Achim J. Lilienthal. Mobile robots for learning spatio-temporal interpolation models in sensor networks - the echo state map approach. In *2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017*, pages 2659–2665, 2017.

[257] Patrick L. McDermott and Christopher K. Wikle. Hierarchical (deep) echo state networks with uncertainty quantification for spatio-temporal forecasting. *CoRR*, abs/1806.10728, 2018.

[258] Javier Del Ser, Ibai Lana, Eric L. Manibardo, Izaskun Oregi, Eneko Osaba, Jesus L. Lobo, Miren Nekane Bilbao, and Eleni I. Vlahogianni. Deep echo state networks for short-term traffic forecasting: Performance comparison and statistical assessment. *CoRR*, abs/2004.08170, 2020.

[259] Yisheng An, Qingsong Song, and Xiangmo Zhao. Short-term traffic flow forecasting via echo state neural networks. In *Seventh International Conference on Natural Computation, ICNC 2011, Shanghai, China, 26-28 July, 2011*, pages 844–847, 2011.

[260] Zuohua Song, Keyu Wu, and Jie Shao. Destination prediction using deep echo state network. *Neurocomputing*, 406:343–353, 2020.

[261] Meysam Alizamir, Sungwon Kim, Ozgur Kisi, and Mohammad Zounemat-Kermani. Deep echo state network: a novel machine learning approach to model dew point temperature using meteorological variables. *Hydrological ences Journal*, 2020.

[262] Emi Mathews and Axel Poigné. Evaluation of a "smart" pedestrian counting system based on echo state networks. *EURASIP J. Embed. Syst.*, 2009, 2009.

[263] Friedhelm Schwenker and Amr Labib. Echo state networks and neural network ensembles to predict sunspots activity. In *ESANN 2009, 17th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 22-24, 2009, Proceedings*, 2009.

[264] Sebastián Basterrech, Ladislav Zjavka, Lukás Prokop, and Stanislav Misák. Irradiance prediction using echo state queueing networks and differential polynomial neural networks. In *13th International Conference on Intellient Systems Design and Applications, ISDA 2013, Salangor, Malaysia, December 8-10, 2013*, pages 271–276, 2013.

[265] Qian Li, Zhou Wu, Rui Ling, Liang Feng, and Kai Liu. Multi-reservoir echo state computing for solar irradiance prediction: A fast yet efficient deep learning approach. *Appl. Soft Comput.*, 95:106481, 2020.

[266] Sebastián Basterrech and Tomás Buriánek. Solar irradiance estimation using the echo state network and the flexible neural tree. In *Intelligent Data analysis and its Applications, Volume I - Proceeding of the First Euro-China Conference on Intelligent Data Analysis and Applications, ECC 2014, June 13-15, 2014, Shenzhen, China*, pages 475–484, 2014.

[267] Tibor Kmet and Maria Kmetova. A 24h forecast of solar irradiance using echo state neural networks. In *Workshop Proceedings of the 16th International Conference on Engineering Applications of Neural Networks, EANN 2015, Rhodes Island, Greece, September 25-28, 2015*, pages 6:1–6:5, 2015.

[268] Zhou Wu, Qian Li, and Xiaohua Xia. Multi-timescale forecast of solar irradiance based on multi-task learning and echo state network approaches. *IEEE Trans. Ind. Informatics*, 17(1):300–310, 2021.

[269] Meiling Xu, Yuanzhe Yang, Min Han, Tie Qiu, and Hongfei Lin. Spatio-temporal interpolated echo state network for meteorological series prediction. *IEEE Trans. Neural Networks Learn. Syst.*, 30(6):1621–1634, 2019.

[270] Rodrigo Sacchi, Mustafa C. Ozturk, José C. Príncipe, Adriano A. F. Carneiro, and Ivan Nunes da Silva. Water inflow forecasting using the echo state network: a brazilian case study. In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2007, Celebrating 20 years of neural networks, Orlando, Florida, USA, August 12-17, 2007*, pages 2403–2408, 2007.

[271] Hugo Valadares Siqueira, Levy Boccato, Romis Attux, and Christiano Lyra Filho. Echo state networks for seasonal streamflow series forecasting. In *Intelligent Data Engineering and Automated Learning - IDEAL 2012 - 13th International Conference, Natal, Brazil, August 29-31, 2012. Proceedings*, pages 226–236, 2012.

[272] Wei Qiao. Echo-state-network-based real-time wind speed estimation for wind power generation. In *International Joint Conference on Neural Networks, IJCNN 2009, Atlanta, Georgia, USA, 14-19 June 2009*, pages 2572–2579, 2009.

[273] Mohammad Amin Chitsazan, M. Sami Fadali, Amanda K. Nelson, and Andrzej M. Trzynadlowski. Wind speed forecasting using an echo state network with nonlinear output functions. In *2017 American Control Conference, ACC 2017, Seattle, WA, USA, May 24-26, 2017*, pages 5306–5311, 2017.

[274] Xiaowei Lin, Zehong Yang, and Yixu Song. The application of echo state network in stock data mining. In *Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference, PAKDD 2008, Osaka, Japan, May 20-23, 2008 Proceedings*, pages 932–937, 2008.

[275] Xiaowei Lin, Zehong Yang, and Yixu Song. Short-term stock price prediction based on echo state networks. *Expert Syst. Appl.*, 36(3):7313–7317, 2009.

[276] Xiaowei Lin, Zehong Yang, and Yixu Song. Intelligent stock trading system based on improved technical analysis and echo state network. *Expert Syst. Appl.*, 38(9):11347–11354, 2011.

[277] Abdelkerim Souahlia, Ammar Belatreche, Abdelkader Benyettou, and Kevin Curran. An experimental evaluation of echo state network for colour image segmentation. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 1143–1150, 2016.

[278] Boudjelal Meftah, Olivier Lézoray, and Abdelkader Benyettou. Novel approach using echo state networks for microscopic cellular image segmentation. *Cogn. Comput.*, 8(2):237–245, 2016.

[279] Abdelkerim Souahlia, Ammar Belatreche, Abdelkader Benyettou, and Kevin Curran. Blood vessel segmentation in retinal images using echo state networks. In *Ninth International Conference on Advanced Computational Intelligence, ICACI 2017, Doha, Qatar, February 4-6, 2017*, pages 91–98, 2017.

[280] Charles Donkor, Emmanuel Sam, and Sebastián Basterrech. Analysis of tensor-based image segmentation using echo state networks. In *Modelling and Simulation for Autonomous Systems - 5th International Conference, MESAS 2018, Prague, Czech Republic, October 17-19, 2018, Revised Selected papers*, pages 490–499, 2018.

[281] Abdelkerim Souahlia, Ammar Belatreche, Abdelkader Benyettou, Zoubir Ahmed-Foitih, Elhadj Benkhelifa, and Kevin Curran. Echo state network-based feature extraction for efficient color image segmentation. *Concurr. Comput. Pract. Exp.*, 32(21), 2020.

[282] Stefano Squartini, Stefania Cecchi, Michele Rossini, and Francesco Piazza. Echo state networks for real-time audio applications. In *Advances in Neural Networks - ISNN 2007, 4th International Symposium on Neural Networks, ISNN 2007, Nanjing, China, June 3-7, 2007, Proceedings, Part III*, pages 731–740, 2007.

[283] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Comparison between deepesns and gated rnns on multivariate time-series prediction. In *27th European Symposium on Artificial Neural Networks, ESANN 2019, Bruges, Belgium, April 24-26, 2019*, 2019.

[284] Xiao-chuan Sun, Hongyan Cui, Ren Ping Liu, Jianya Chen, and Yunjie Liu. Multistep ahead prediction for real-time VBR video traffic using deterministic echo state network. In *2nd IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS 2012, Hangzhou, China, October 30 - November 1, 2012*, pages 928–931, 2012.

[285] Sohini Roychowdhury and L. Srikar Muppirisetty. Fast proposals for image and video annotation using modified echo state networks. In *17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Orlando, FL, USA, December 17-20, 2018*, pages 1225–1230, 2018.

[286] Mark D. Skowronski and John G. Harris. Noise-robust automatic speech recognition using a discriminative echo state network. In *International Symposium on Circuits and Systems (ISCAS 2007), 27-20 May 2007, New Orleans, Louisiana, USA*, pages 1771–1774, 2007.

[287] Qingchun Zhao, Hongxi Yin, Xiaolei Chen, and Wenbo Shi. Performance optimization of the echo state network for time series prediction and spoken digit recognition. In *11th International Conference on Natural Computation, ICNC 2015, Zhangjiajie, China, August 15-17, 2015*, pages 502–506, 2015.

[288] Stefan Scherer, Mohamed Oubbati, Friedhelm Schwenker, and Günther Palm. Real-time emotion recognition using echo state networks. In *Perception in Multimodal Dialogue Systems, 4th IEEE Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems, PIT 2008, Kloster Irsee, Germany, June 16-18, 2008, Proceedings*, pages 200–204, 2008.

[289] Edmondo Trentin, Stefan Scherer, and Friedhelm Schwenker. Maximum echo-state-likelihood networks for emotion recognition. In *Artificial Neural Networks in Pattern Recognition, 4th IAPR TC3 Workshop, ANNPR 2010, Cairo, Egypt, April 11-13, 2010. Proceedings*, pages 60–71, 2010.

[290] Qutaiba Saleh, Cory E. Merkel, Dhireesha Kudithipudi, and Bryant T. Wysocki. Memristive computational architecture of an echo state network for real-time speech-emotion recognition. In *2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2015, Verona, NY, USA, May 26-28, 2015*, pages 1–5, 2015.

[291] Lachezar Bozhkov, Petia D. Koprinkova-Hristova, and Petia Georgieva. Learning to decode human emotions with echo state networks. *Neural Networks*, 78:112–119, 2016.

[292] Boris Bacic. Echo state network for 3d motion pattern indexing: A case study on tennis forehands. In *Image and Video Technology - 7th Pacific-Rim Symposium, PSIVT 2015, Auckland, New Zealand, November 25-27, 2015, Revised Selected Papers*, pages 295–306, 2015.

[293] Boris Bacic. Echo state network ensemble for human motion data temporal phasing: A case study on tennis forehands. In *Neural Information Processing - 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16-21, 2016, Proceedings, Part IV*, pages 11–18, 2016.

[294] Pattreeya Tanisaro, Constantin Lehman, Leon René Sütfeld, Gordon Pipa, and Gunther Heidemann. Classifying bio-inspired model of point-light human motion using echo state networks. In *Artificial Neural Networks and Machine Learning - ICANN 2017 - 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part I*, pages 84–91, 2017.

[295] Doreen Jirak, Pablo V. A. Barros, and Stefan Wermter. Dynamic gesture recognition using echo state networks. In *23rd European Symposium on Artificial Neural Networks, ESANN 2015, Bruges, Belgium, April 22-24, 2015*, 2015.

[296] Weipeng Wang, Xiangpeng Liang, Maher Assaad, and Hadi Heidari. Wearable wristworn gesture recognition using echo state network. In *26th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2019, Genoa, Italy, November 27-29, 2019*, pages 875–878, 2019.

[297] Petia D. Koprinkova-Hristova, Miroslava Stefanova, Bilyana Genova, and Nadejda Bocheva. Echo state network for classification of human eye movements during decision making. In *Artificial Intelligence Applications and Innovations - 14th IFIP WG 12.5 International Conference, AIAI 2018, Rhodes, Greece, May 25-27, 2018, Proceedings*, pages 337–348, 2018.

[298] Stefan L. Frank. Strong systematicity in sentence processing by an echo state network. In *Artificial Neural Networks - ICANN 2006, 16th International Conference, Athens, Greece, September 10-14, 2006. Proceedings, Part I*, pages 505–514, 2006.

[299] Stefan L. Frank and Henrik Jacobsson. Sentence-processing in echo state networks: a qualitative analysis by finite state machine extraction. *Connect. Sci.*, 22(2):135–155, 2010.

[300] Matthew H. Tong, Adam D. Bickett, Eric M. Christiansen, and Garrison W. Cottrell. Learning grammatical structure with echo state networks. *Neural Networks*, 20(3):424–432, 2007.

[301] Xavier Hinaut, Maxime Petit, Grégoire Pointeau, and Peter F. Dominey. Exploring the acquisition and production of grammatical constructions through human-robot interaction with echo state networks. *Frontiers Neurorobotics*, 8:16, 2014.

[302] Zongying Liu, Chu Kiong Loo, and Min Jiang. Grammatical structure detection with intrinsic plasticity based echo state networks for cognitive robot. In *IEEE Symposium Series on Computational Intelligence, SSCI 2019, Xiamen, China, December 6-9, 2019*, pages 2207–2214, 2019.

[303] Yukinori Homma and Masafumi Hagiwara. An echo state network with working memories for probabilistic language modeling. In *Artificial Neural Networks and Machine Learning - ICANN 2013 - 23rd International Conference on Artificial Neural Networks, Sofia, Bulgaria, September 10-13, 2013. Proceedings*, pages 595–602, 2013.

[304] Kiril Ivanov Simov, Petia D. Koprinkova-Hristova, Alexander Popov, and Petya Osenova. Word embeddings improvement via echo state networks. In *IEEE International Symposium on INnovations in Intelligent SysTems and Applications, INISTA 2019, Sofia, Bulgaria, July 3-5, 2019*, pages 1–6, 2019.

[305] Cédric Hartland and Nicolas Bredèche. Using echo state networks for robot navigation behavior acquisition. In *IEEE International Conference on Robotics and Biomimetics, ROBIO 2007, Sanya, China, 15-28 December 2007*, pages 201–206, 2007.

[306] Stefano Chessa, Claudio Gallicchio, Roberto Guzmán, and Alessio Micheli. Robot localization by echo state networks using RSS. In *Recent Advances of Neural Network Models and Applications - Proceedings of the 23rd Workshop of the Italian Neural Networks Society, WIRN 2013, Vietri sul Mare, Salerno, Italy, May 23-25, 2013*, pages 147–154, 2013.

[307] Qingsong Song and Zuren Feng. Stable trajectory generator-echo state network trained by particle swarm optimization. In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA 2009, 15-18 December 2009, Daejeon, Korea*, pages 21–26, 2009.

[308] Cesar H. Valencia, Marley Maria Bernardes Rebuzzi Vellasco, and Karla Figueiredo. Trajectory tracking control using echo state networks for the corobot's arm. In *Robot Intelligence Technology and Applications 2 - Results from the 2nd International Conference on Robot Intelligence Technology and Applications, RiTA 2013, Denver, Colorado, USA, December 18-20, 2013*, pages 433–447, 2013.

[309] Cédric Hartland, Nicolas Bredèche, and Michèle Sebag. Memory-enhanced evolutionary robotics: The echo state network approach. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2009, Trondheim, Norway, 18-21 May, 2009*, pages 2788–2795, 2009.

[310] Claudio Gallicchio and Alessio Micheli. Experimental analysis of deep echo state networks for ambient assisted living. In *Proceedings of the Third Italian Workshop on Artificial Intelligence for Ambient Assisted Living 2017 co-located with 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017), Bari, Italy, November 16th and 17th, 2017*, pages 44–57, 2017.

[311] Kui Xiang, Bing Nan Li, Liyan Zhang, Muye Pang, Meng Wang, and Xuelong Li. Regularized taylor echo state networks for predictive control of partially observed systems. *IEEE Access*, 4:3300–3309, 2016.