



Reservoir Computing for Temporal Data Classification Using a Dynamic Solid Electrolyte ZnO Thin Film Transistor

Ankit Gaurav¹, Xiaoyao Song², Sanjeev Manhas¹, Aditya Gilra^{3,4}, Eleni Vasilaki³, Partha Roy⁵ and Maria Merlyne De Souza^{2*}

¹Department of Electronics and Communication, IIT Roorkee, Roorkee, India, ²Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, United Kingdom, ³Department of Computer Science, University of Sheffield, Sheffield, United Kingdom, ⁴Machine Learning Group, Centrum Wiskunde and Informatica, Amsterdam, Netherlands, ⁵Department of Computer Science, IIT Roorkee, Roorkee, India

OPEN ACCESS

Edited by:

Niko Münzenrieder,
Free University of Bozen-Bolzano, Italy

Reviewed by:

Matija Varga,
ABB, Switzerland
Wan Fazlida Hanim Abdullah,
MARA University of Technology,
Malaysia

*Correspondence:

Maria Merlyne De Souza
m.desouza@sheffield.ac.uk

Specialty section:

This article was submitted to
Flexible Electronics,
a section of the journal
Frontiers in Electronics

Received: 03 February 2022

Accepted: 18 March 2022

Published: 11 April 2022

Citation:

Gaurav A, Song X, Manhas S, Gilra A, Vasilaki E, Roy P and De Souza MM (2022) Reservoir Computing for Temporal Data Classification Using a Dynamic Solid Electrolyte ZnO Thin Film Transistor. *Front. Electron.* 3:869013. doi: 10.3389/felec.2022.869013

The processing of sequential and temporal data is essential to computer vision and speech recognition, two of the most common applications of artificial intelligence (AI). Reservoir computing (RC) is a branch of AI that offers a highly efficient framework for processing temporal inputs at a low training cost compared to conventional Recurrent Neural Networks (RNNs). However, despite extensive effort, two-terminal memristor-based reservoirs have, until now, been implemented to process sequential data by reading their conductance states only once, at the end of the entire sequence. This method reduces the dimensionality, related to the number of signals from the reservoir and thereby lowers the overall performance of reservoir systems. Higher dimensionality facilitates the separation of originally inseparable inputs by reading out from a larger set of spatiotemporal features of inputs. Moreover, memristor-based reservoirs either use multiple pulse rates, fast or slow read (immediately or with a delay introduced after the end of the sequence), or excitatory pulses to enhance the dimensionality of reservoir states. This adds to the complexity of the reservoir system and reduces power efficiency. In this paper, we demonstrate the first reservoir computing system based on a dynamic three terminal solid electrolyte ZnO/Ta₂O₅ Thin-film Transistor fabricated at less than 100°C. The inherent nonlinearity and dynamic memory of the device lead to a rich separation property of reservoir states that results in, to our knowledge, the highest accuracy of 94.44%, using electronic charge-based system, for the classification of hand-written digits. This improvement is attributed to an increase in the dimensionality of the reservoir by reading the reservoir states after each pulse rather than at the end of the sequence. The third terminal enables a read operation in the off state, that is when no pulse is applied at the gate terminal, via a small read pulse at the drain. This fundamentally allows multiple read operations without increasing energy consumption, which is not possible in the conventional two-terminal memristor counterpart. Further, we have also shown that devices do not saturate even after multiple write pulses which demonstrates the device's ability to process longer sequences.

Keywords: reservoir computing, solid electrolyte devices, temporal data, short-term memory, neural networks, thin-film transistor

INTRODUCTION

Artificial neural networks (ANNs) are computational models inspired by biological neural networks that have witnessed remarkable progress in real-world applications such as pattern recognition, classification, and forecasting time series events (Emmert-Streib et al., 2020). The network architectures of ANNs can be mainly grouped as feedforward neural networks (FNNs) (Schmidhuber, 2015) and recurrent neural networks (RNNs) (Güçlü and van Gerven, 2017; Hopfield, 2018). FNNs pass the data unidirectionally forward from the input to the output. The widely used convolutional neural network (CNN), a type of FNN, is mainly used for static (non-temporal) data processing. On the other hand, in RNNs, hidden neurons have cyclic connections, making the outputs dependent upon both the current inputs as well as the internal states of the neurons, thus making RNNs suitable for dynamic (temporal) data processing. The commonly known problem of exploding and vanishing gradients, arising in very deep FNNs and from cyclic connections in RNNs, results in network instability and less effective learning, making the training process complex and expensive. Vanishing gradients get smaller and approach zero as the backpropagation algorithm advances from the output layer towards the input, or past inputs in the case of RNN after the cyclic connections are unfolded in time, which eventually leaves the weights farthest from the output nearly unchanged. Whereas in exploding gradients, the gradients get larger, which ultimately causes huge weight updates. As a result, the gradient descent algorithm fails to converge to the optimum in both cases.

The concept of reservoir computing was proposed to address these problems. Initially, RC was introduced as RNN models of echo state networks (ESNs) (Jaeger, 2001) and liquid state machines (LSMs) (Maass et al., 2002). A conventional RC system consists of three sections: input, reservoir, and output consisting of a readout function. The role of the reservoir is to transform nonlinear sequential inputs into a high-dimensional feature space, consisting of spatiotemporal patterns, which are features of the input (or reservoir states). These patterns are subsequently used to train the readout section by a learning algorithm that maps every recorded reservoir state to an output. In an RC system, only the weights of the output layer need training, whereas weights connecting the input layer to the reservoir and the internal sparsely connected reservoir nodes are random and fixed, resulting in a significantly reduced cost, and avoiding the vanishing/exploding gradients, compared to a conventional RNN. Recently, the learning of the RC system was improved through sparse representations, where the dynamics of the reservoir was sampled across time to enhance its dimensionality see, for instance (Manneschi et al., 2021a; Manneschi et al., 2021b).

Other nonlinear dynamical systems can be used as reservoirs instead of the traditional model of RNNs. Their framework is called physical reservoir computing (PRC). One of the

motivations for PRC systems is a desire to develop fast information processing with minimal learning costs. Traditional hardware implementations of reservoirs often demand power-hungry neuromorphic hardware (Misra and Saha, 2010; Hasler and Marr, 2013). On the other hand, an implementation of a reservoir can also be accomplished by physical phenomena. Inherent nonlinearity of a dynamic physical memory maps the input signal onto a higher dimensional spatio-temporal feature space similar to a reservoir of artificial neurons. These features can be read out linearly to solve non-linearly separable problems. The essential requirements of the physical reservoir are a high dimensionality arising from the number of unique reservoir states (separation property) and short-term memory (fading memory) required to establish a temporal relationship of its sequential inputs. This determines the dependence of the reservoir state upon recent past inputs within a specified time interval (Nakajima and Fischer, 2021). This basic principle of PRC is versatile: One can solve many temporally dependent or temporally-independent tasks via converting it into a sequential task (i.e., providing spatial information over time) such as recognition of the spoken digit (Vandoorne et al., 2014), waveform generation and classification (Sillin et al., 2013; Torrejon et al., 2017), chaotic time-series forecasting (Moon et al., 2019), Hénon map (Zhong et al., 2021), classification of hand-written digits (Du et al., 2017) and short sentences (Sun et al., 2021).

The framework of PRC remains the same for different temporal input-output tasks. Only the representation of sample data before feeding into the reservoir varies from application to application. For example, in the classification of sine and square waveforms (Paquot et al., 2012) the goal is to classify each point of the input either as part of a sine or a square wave. Each period of the wave is discretized into eight points (before feeding into the reservoir) giving 16 different cases for classification. The temporal input is composed of 1,280 points (160 randomly arranged periods of sine or square). The first half of the input is used for training (to find the optimum weights of the readout function) and the second half for testing.

On the other hand, in the more complex and widely used “benchmark” case of voice recognition, the goal is to recognize digits from audio waveforms produced by different speakers. This task also requires the representation of sample data: for example, inputs for the memristor-based reservoir (Moon et al., 2019) are sound waveforms of isolated spoken digits (0–9 in English) from the NIST TI46 database (Texas Instruments, 1991). These are pre-processed into a set of 50 frequency channels with 40-time steps using Lyon’s passive ear model based on human cochlear channels (Lyon, 1982). The output of Lyon’s passive ear model, which is the firing probability, is then converted into digitized spike trains before applying to the reservoir. However, to increase the efficiency of a particular reservoir system, it is essential to properly represent the sample data and optimize the design of the RC system (Lukoševičius and Jaeger, 2009).

Temporal behaviour in image classification can be enforced by converting the pixel value (input) into spatio-temporal patterns by dividing the entire image into groups of pixels that are sequentially input into the reservoir. This approach was used to demonstrate the classification of hand-written digits from the MNIST dataset (LeCun et al., 1998) with an accuracy of 88% (Du et al., 2017) and 83% (Midya et al., 2019), respectively. (Du et al., 2017), used two different pulse rates to double the reservoir states for the same input sequence. Midya et al. (2019) demonstrated a diffusive memristor-based reservoir with a drift memristor-based readout function. In this case, a fast and slow read (done immediately around 5 and 300 μ s after the end of the input sequence), respectively, along with an excitatory pulse applied at the start of all sequences to pre-set the devices to a low resistance state were used to increase the dimensionality of reservoir states. In another demonstration (Sun et al., 2021) used memristors to demonstrate the classification of short Korean sentences with an accuracy of 91%. In these implementations, sequences consisting of 4 (Midya et al., 2019) or 5 bits (Du et al., 2017; Sun et al., 2021) were successfully demonstrated, ideally each resulting in a distinguishable state of the reservoir. Thus, a higher number of unique sequences representable in the reservoir adds to the dimensionality of reservoir states at the cost of complexity. Increasing the sequence length can result in complexity due to an inordinate number of input combinations that need to be recorded.

In all the above implementations, the dimensionality of the reservoir was restricted by reading the states of the reservoir only once at the end of each sequence, typically of sizes up to 4–5 bits. For example, an image with dimension 24×20 fed into a reservoir, column-wise will reduce the dimension of the image to 4×20 for a 6-bit sequence length. This reduces the dimensionality of the reservoir by extracting a smaller number of features, i.e., $80 = (4 \times 20)$ in this case, which is far less than the original 480 (24×20). The readout network with reduced features results in poor accuracy of recognition. In comparison, a conventional network that is similarly down-sampled for comparison with an RC system (for example in a ratio of 6 to 1 bits), results in a reduction of the dimensions of an image to the same extent as that of a reservoir. In such a case, the RC system performs better only for longer sequence lengths, which equates to a smaller size of the readout network. This outperformance of an RC system, for longer sequence lengths (and smaller readout networks) is due to the preservation of temporal information by the short-term memory of the device. The disadvantage of down sampling, however, is poorer accuracy. To improve accuracy, as sequence lengths reduce, there is less down-sampling of the image, which results in an RC system having no benefit compared to that of a conventional network (Du et al., 2017). The schematic of this approach of down-sampling and comparison of results of a down-sampled feature using RC system versus a conventional approach are shown in **Supplementary Figures S1 and S2**, respectively. This example highlights that the objective should not be the size of the readout network for comparison, but the richness of the reservoir.

Most memristors are two-terminal, never-the-less, three-terminal memory devices that can facilitate an immediate

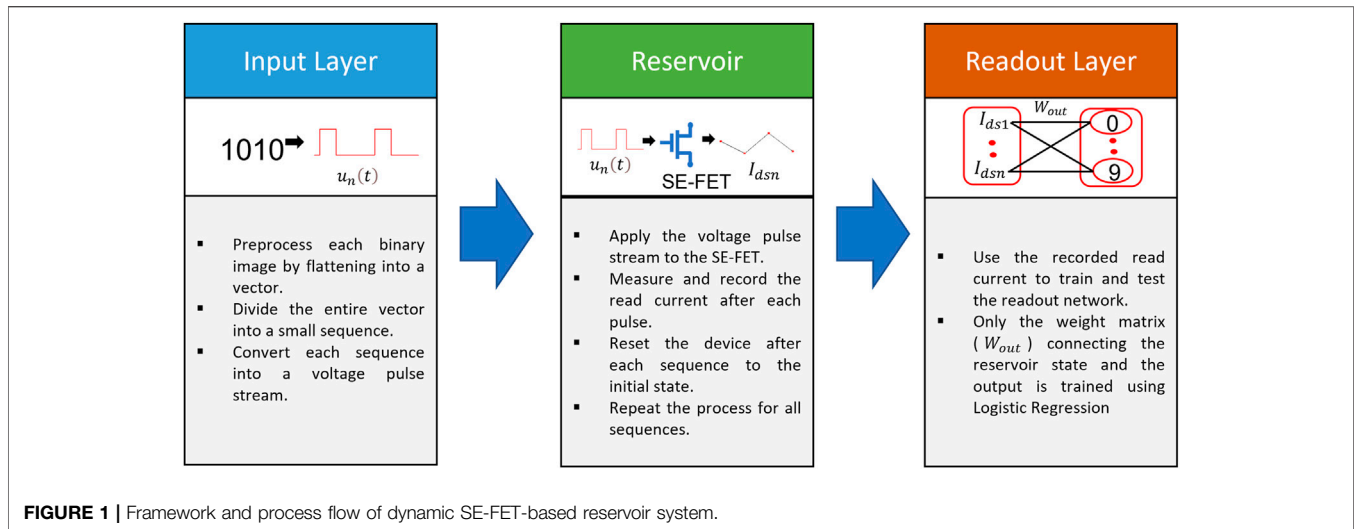
enrichment of reservoir states, via the gate electrode. Such thin-film transistors have wide applicability in upcoming applications of smart and seamlessly integrated sensors, capable of measurement and processing not on the cloud but on the edge (edge intelligence). Edge intelligence enables instantaneous decision-making and reduces the energy consumption of the system by minimizing data communication to and from the cloud. Typically, Ferroelectric field-effect transistors (FEFETs) have been widely touted as multi-state weight cells/analog synapses for deep neural network accelerators. The plasticity of ferroelectric polarization with sub-coercive voltages leads to multiple conductance states in FEFETs. Up to 100-fold modulation of conductance with 5 ns update pulses have been demonstrated (Ni et al., 2018). Typically, however, FEFETs are non-volatile, and do not have an inherent property of temporal behaviour which would imply that the ferroelectric mechanism is not naturally compatible with the concept of PRC.

METHODOLOGY

An alternative to a ferroelectric FET is a Solid-Electrolyte FET, first proposed for neuromorphic computing applications (Pillai and De Souza, 2017; Yang et al., 2018; Park et al., 2020; Qin et al., 2020). Our earlier study of the ZnO/Ta₂O₅ solid electrolyte FET (SE-FET) (Pillai and De Souza, 2017) shows versatility in terms of the number of conductance states, an inherently gradual and nonlinear conductance, which is an advantage in RC systems. Its short-term memory, due to an inherent mechanism based on diffusion of ions within the insulator of the device (Kumar et al., 2018), ensures the fading memory required to establish a temporal relationship needed for reservoir computing (Appeltant et al., 2011).

Our SE-FET ($W \times L = 100 \mu\text{m} \times 1.5 \mu\text{m}$) is a bottom-gated thin film transistor, (schematically illustrated in **Supplementary Figure S3**), fabricated via RF sputtering at room temperature and fully compatible with BEOL processing. A conducting Indium Tin Oxide (ITO, $20 \Omega/\text{square}$) is used as the gate electrode. 275 nm of tantalum oxide (Ta₂O₅) acts as a bottom gate insulator, over which 40 nm of ZnO is sputtered as a channel. Al contacts are evaporated over the ZnO as the top contact. Electrical characterization is undertaken using Keysight B2902A. In this paper, we experimentally demonstrate our new approach to generate high dimensionality by reading the response of the SE-FET after each pulse. The gate terminal is used for the write operation, whereas the drain terminal is used to perform read operations independently when the device is off, further reducing power consumption (Song et al., 2019). The use of the separate control terminal makes the RC system efficient and straightforward compared to any two-terminal-based artificial neural network. Moreover, its ease of fabrication, typically at room temperature, favours a wide range of wearable applications.

Our framework and process flow of the SE-FET based reservoir system is described in **Figure 1**. The connection between each temporal input pulse stream $u_n(t)$ to each SE-FET is fixed. SE-FETs inside our reservoir are not interconnected,



unlike in a conventional software-based RC system (Tanaka et al., 2019) because the inherent short-term memory of the device suffices to process temporal input consisting here of a sequence of voltage pulses $u_n(t)$ applied at the gate of the SE-FET. A small read voltage of -1 V is applied after each pulse to the drain terminal to read the conductance denoted by I_{dsn} (drain to source current), while the gate voltage is 0 V. Explicitly, an input pulse stream consisting of four pulses also results in 4 “read” current values. This process is repeated for all the input sequences, and after each input sequence, a reset pulse is applied to set the device back to the initial state. A negative read voltage is used for longer retention of the device state (conductance value). When using a positive read voltage (>0 V), when the gate of the device is off (0 V), this effectively puts the gate at a lower potential than the drain-source. This has a similar impact as applying a negative voltage on the gate, and for the high conductance state, the natural decay of the SE-FET reduces the retention time of the device memory. On the other hand, a small negative read voltage helps the device to hold its conductance state, but not too strong that it would write the device itself.

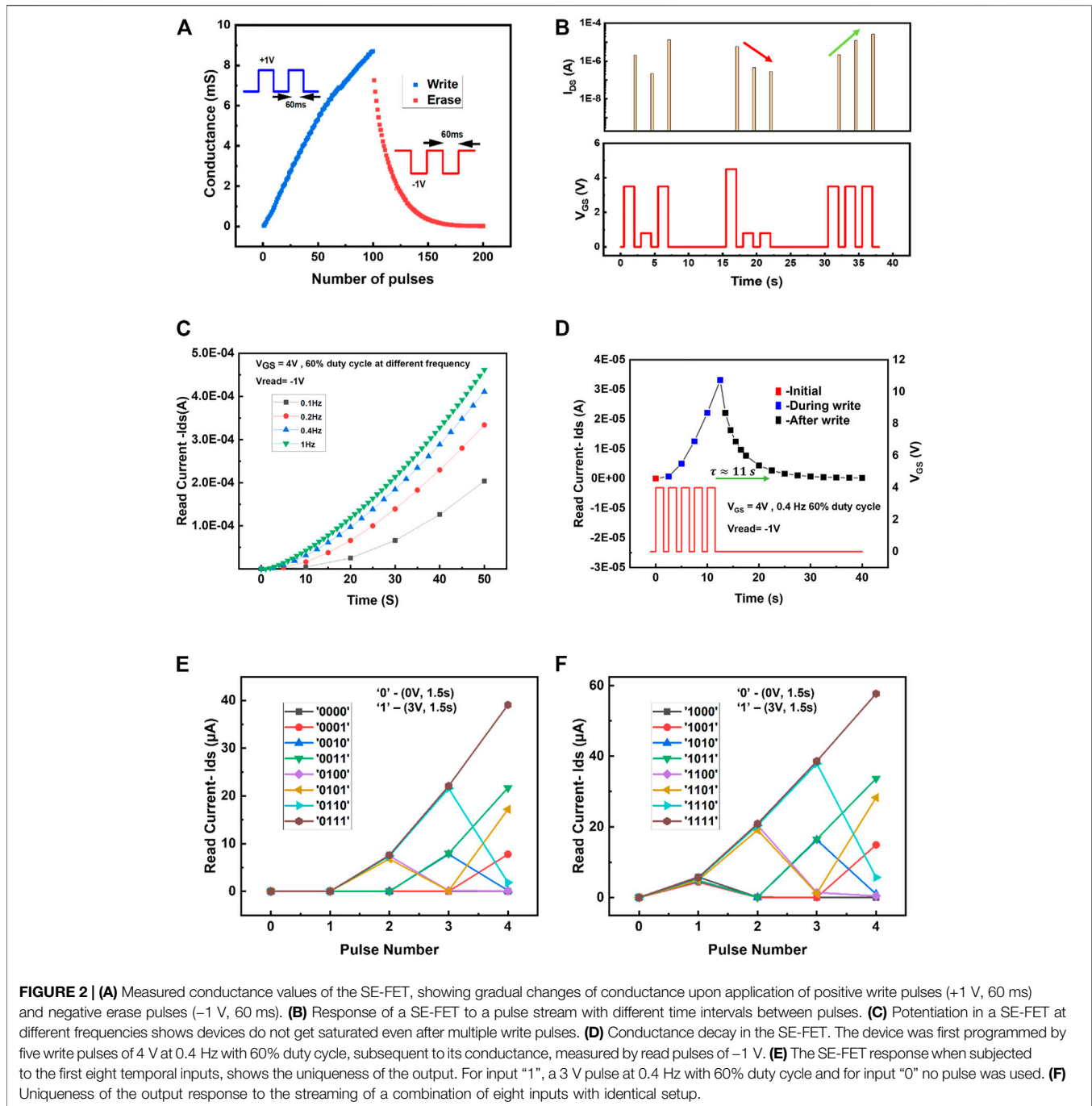
We used the measured and recorded read current values for training and testing the readout network offline. Only the weight matrix W_{out} connecting the reservoir states, and the output were trained using logistic regression. Logistic Regression (LR) is a supervised Machine Learning (ML) algorithm used in classification. The training algorithm uses the one-versus-rest scheme by splitting the multi-class dataset into multiple binary classification problems. For example, a multi-class classification problem with each class label as “0”, “1”, and “2” can be divided into three binary classification datasets as follows: Problem 1: “0” vs. [1, 2], Problem 2: “1” vs. [0, 2], and Problem 3: “2” vs. [0, 1]. The class index with the largest score is then used to predict a class. The probabilities describing the possible outcomes of a single trial are modelled using a logistic function, a common S-shaped curve defined as $f(x) = \frac{L}{1+e^{-k(x-x_0)}}$, where L is the curve’s maximum value, k is logistic growth rate or steepness of the curve, x_0 is the x value of the sigmoid midpoint, and x is a real number. The Logistic Regression with the liblinear Solver (Library for

Large Linear Classification) from Python’s scikit-learn library (Scikit-Learn, 2021) was used, which uses a gradient descent algorithm. Gradient descent minimizes a multivariate function (involving multiple dependent variables), resulting in a single outcome.

We trained our readout network using 60,000 image samples from the MNIST database, and to test the network, a separate 10,000 image sample set, not used in training was fed into the reservoir. Classification was performed from the readout network based on the reservoir state for each test case. To avoid the system from being overfitted to certain selections of the training and testing data, 7-fold cross-validation was used. In particular, training and testing were repeated seven times using 70,000 samples (60,000 samples for training and 10,000 samples for testing), with a different assignment for training and testing samples, every time. The classification accuracy was defined as the mean accuracy of all the test samples across the 7-fold cross-validation. Before feeding the MNIST image into the reservoir, the unused area of an original greyscale MNIST image was cropped and converted into binary. The actual 28×28 MNIST image and the pre-processed 24×24 binary images of digit 5, as an example, are shown in **Supplementary Figure S4**.

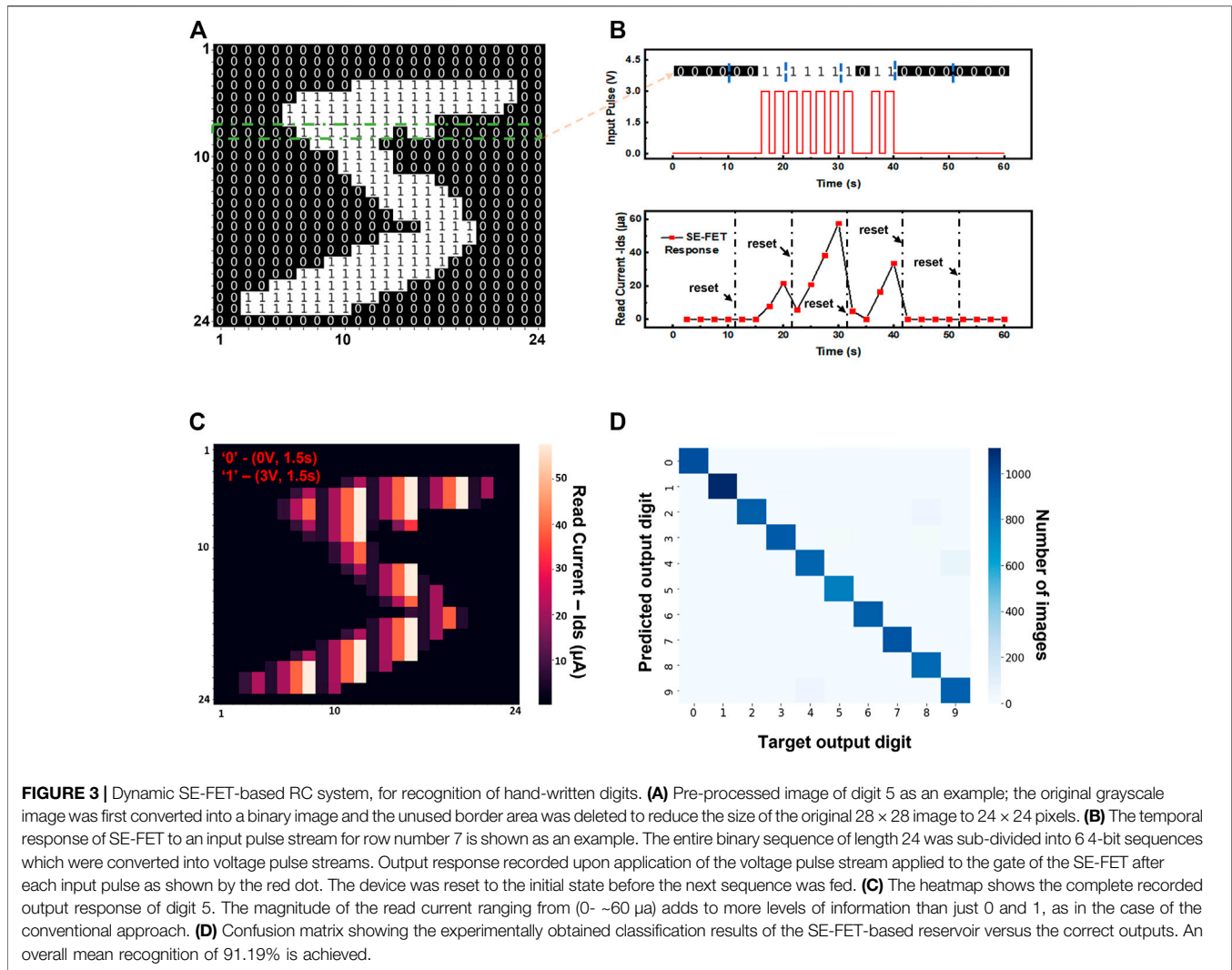
RESULTS AND DISCUSSION

The measured conductance values of the SE-FET with ± 1 V/60 ms pulses applied at the gate terminal are shown in **Figure 2A**. Positive write pulses gradually increase the conductance, whereas negative erase pulses gradually reduce it. This gradual change results in multiple conductance states, technically infinite, but for the present case, 100 write pulses of 1 V/60 ms were used to generate 100 conductance states. To demonstrate temporal dynamics, write pulses of different amplitude were applied to the device, and the read current is recorded as shown in **Figure 2B**. Two properties can be observed: 1) If multiple pulses are applied at short intervals, the conductance gradually increases (as indicated by the green arrow) 2) If no pulse or small



voltage pulse (i.e., 0.8 V), is applied to achieve bit 0, then the conductance decays towards the original resting state, as indicated by the red arrow. This temporal response of the SE-FET is attributed to the polarity-induced motion of oxygen vacancies in the gate insulator (Pillai and De Souza, 2017). The potentiation of conductance was further investigated at a frequency from 0.1 to 1 Hz with a V_{GS} (write pulse) of 4 V at 60% duty cycle, and V_{read} at -1 V. The devices do not saturate even after multiple write pulses as shown in **Figure 2C**. At 1 Hz, the SE-FET can process 50 write pulse without saturation as shown

by the green triangle in **Figure 2C**. The average change in conductance per pulse remains low at $\sim 25\%$ at a higher frequency, whereas it is $\sim 200\%$ at a lower frequency as shown in the **Supplementary Figure S5**. Therefore, the right balance is required in terms of speed of processing and conductance change per pulse to get an enhanced separation property of the reservoir. The conductance decay in a SE-FET is shown in **Figure 2D**. The device is first programmed by five write pulses of 4 V at 0.4 Hz with 60% duty cycle, subsequently, its conductance is monitored by read pulses of -1 V. In this case, a decay time of $\tau \approx 11$ s is



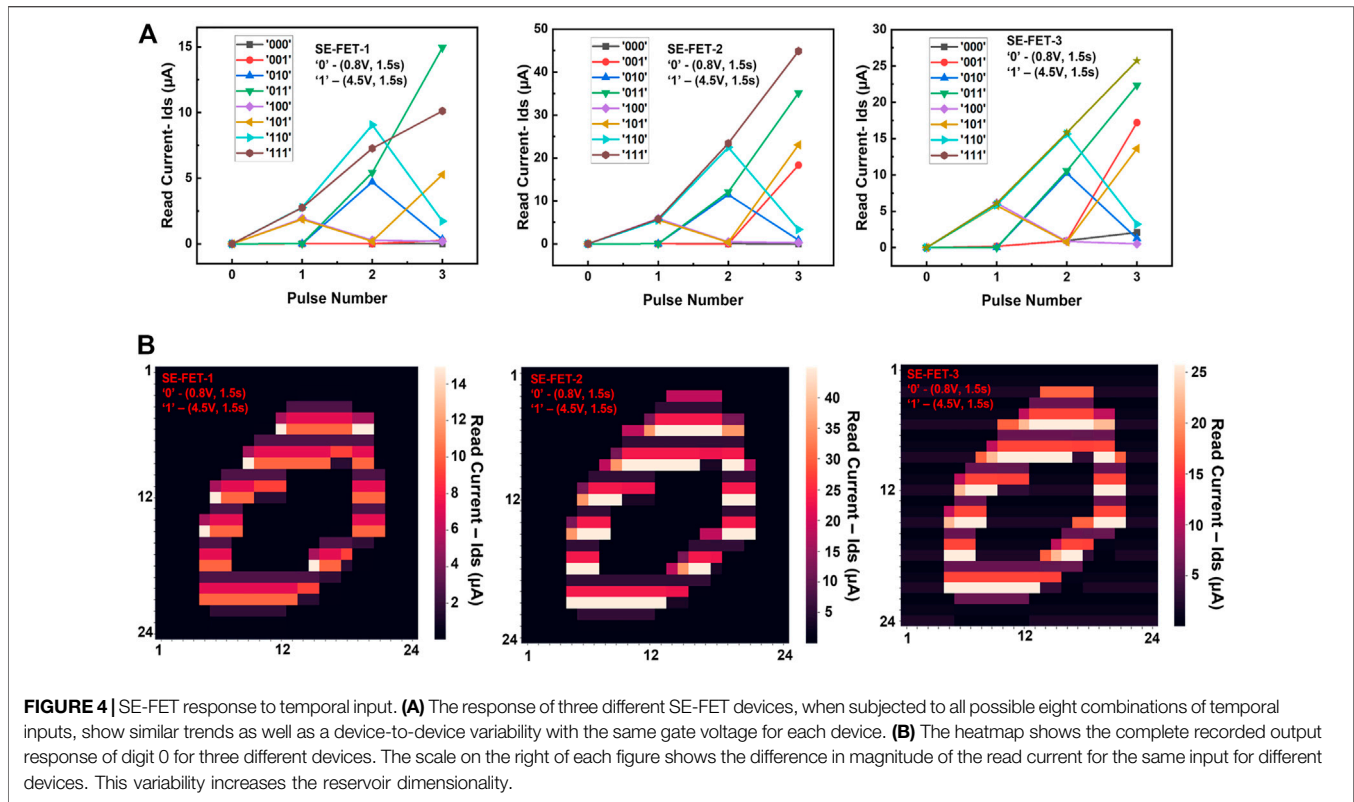
observed for devices to revert to the initial state as shown by the green arrow. Further, we have also monitored the device's short-term memory with variable read voltage as shown in the **Supplementary Figure S6**. These results show that negative read voltage can be used to adjust the time constant of memory fading. In case a higher reading current is required, a higher reading voltage can be applied for a shorter duration. Further, the response of a SE-FET device to all possible 16 temporal input combinations of 4-bit sequences is shown in **Figures 2E,F**. To carry out these tests, four input gate pulses, each of width (2.5 s), containing a write pulse of width (1.5 s), have been applied within a timeframe of 10 s. For input "1", a pulse of 3 V and for input "0" no pulse is applied (0 V). The SE-FET response when subject to all 16 combinations of temporal inputs reveals the uniqueness of the output, leading to distinct reservoir states.

Dynamic SE-FET Based RC System

The dynamic SE-FET-based RC system is tested for recognition of handwritten digits using the MNIST database initially with a

single device to demonstrate our new approach. Pre-processing has been performed as shown in **Figure 3A**. If the entire row is used as one input pulse stream, then, in theory, there can be 2^{24} different input patterns which may be too difficult for one SE-FET to distinguish. Therefore, to improve the ability of the reservoir, we have divided each row into six sub-sections, each containing four pixels, to allow better separation of the inputs. After each input sequence, a small reset pulse of -3 V with pulse width of (0.5 s) is applied to avoid impact from the previous sequence. Example of one such reset operation is shown in the **Supplementary Figure S7**. With these considerations, the image has been fed into the reservoir in 4-pixel sub-sections as input voltage pulse streams, and their output response is recorded after each input pulse, as shown by the red dot in **Figure 3B**.

Similarly, to extract features, each row of the image is fed into the reservoir by converting it into a pulse stream (as exemplified in **Figure 3B** for row number 8). As an example, all reservoir states corresponding to digit 5 are shown in **Figure 3C**. The different magnitudes of the read current ranging from (0 - $60 \mu\text{A}$) add to more levels of information than just 0 and 1,



as in the conventional approach. This shows the ability of our reservoir to solve linearly non-separable problems via mapping of the input signal into a higher dimensional space in a nonlinear relationship, which can be further decoded by a simple linear logistic regression machine learning algorithm (as discussed in the methodology section). For this implementation, an overall mean accuracy of 91.19% is achieved across 7-fold cross-validation of the test set. The performance of the readout network is summarized using a confusion matrix showing the experimentally obtained classification results of the 10,000-test image from the SE-FET-based RC system vs the target outputs digit as shown in **Figure 3D**. The occurrence of the predicted output for each test case is represented.

Further improvement in accuracy is obtained by using multiple devices (3 in our case) in two possible approaches for a reservoir with 3-bit sequence lengths. The first is to use each device with the same gate voltage pulse for input “1” and “0”. The second is to study the influence of gate voltage to achieve bit 1 with different devices. The response of each SE-FET to all possible combinations of eight temporal inputs for a 3-bit sequence is shown in **Figure 4A** (for the same gate voltage pulse). The complete recorded output response of digit 0 from all three devices is shown in **Figure 4B**. Similarly, it is done for all the train and test images. The complete recorded output response from three different devices for an image was concatenated in vector form of length 3×576 (24×24) and is then used to train or test the readout function. An overall mean recognition of 92.97% is achieved for the implementation with the same gate voltage, whereas, an

overall mean recognition of 92.44% is achieved for different gate voltages for bit “1”, with different devices.

The complete recorded output response of digit 0, as well as temporal response from all three devices, is shown in **Supplementary Figures S8 and S9A–C**. These results show that multiple devices increase dimensionality due to device-to-device variability. Moreover, the natural device-to-device variability can be kept intact using the same gate voltage. In contrast, different voltage levels either boost the conductance at a higher voltage compared to another device or suppress it at a lower voltage. This shows that the high dimensionality is mainly due to the natural device-to-device variability rather than the use of multiple gate voltages. Therefore, it is essential to properly represent the sample data and optimize the system design to allow a better separation of the inputs.

Further optimization has been undertaken by combining the horizontal (row-wise) and the vertical scan (column-wise). With this representation of sample data, a maximum mean accuracy of 94.44% using all three SE-FET devices with the same gate voltage (as shown in **Supplementary Figure S9D**) is achieved.

The performance of our dynamic SE-FET-based RC system is benchmarked to that of a fully connected conventional network, without any hidden layer, trained via logistic regression, whose schematic is shown in **Supplementary Figure S10**. An identically sized input is fed to the conventional network as the dynamic SE-FET-based RC system. The overall mean recognition accuracy of 90.82% is achieved across a 7-fold cross-validation for the test set using a fully connected conventional network. A comparison of the different approaches of the RC system to that of the fully

connected conventional network is shown in **Supplementary Table S1**. Our dynamic SE-FET-based RC system outperforms the conventional network by 3.62% for the same task of recognition of handwritten digits and does not depend on down-sampling.

In previous reports of classification of hand-written digits (Du et al., 2017; Midya et al., 2019), only one read pulse was applied at the end of the sequence to read the reservoir states, which restricted the dimensionality of these reservoirs. In contrast, we demonstrated that the dimensionality of the reservoir was increased significantly by reading the reservoir states after each pulse by operating the device in the off-state without increasing the energy consumption, which is not possible in the conventional two-terminal memristor counterpart for which we achieved a higher accuracy of 94.44%, compared to 88.1% (Du et al., 2017) and 83% (Midya et al., 2019) in two terminal memristor-based reservoir systems. Another work, which exploits magnetic domain wall dynamics for reservoir computing by modulating the amplitude of the applied magnetic field to inject time-multiplexed input signals into the reservoir, has also demonstrated an accuracy of ~87% (Ababei et al., 2021). Whereas large-scale spatiotemporal photonic RC systems have demonstrated an overall accuracy of ~97% attributed to a larger reservoir of 1,024 nodes (Nguimdo et al., 2020), and ~92% for a photonic extreme learning machine implemented by using an optical encoder and coherent wave propagation in free space (Pierangeli et al., 2021).

CONCLUSION

We experimentally showed that by reading out spatiotemporal dependencies of input pulses, after each pulse, significantly improves the dimensionality of a SE-FET-based reservoir. The spatiotemporal dependencies of inputs were captured by utilizing the device's inherent short-term memory based on diffusion kinetics of ions in the oxide. At the same time, the device nonlinearity allowed the reservoir to solve linearly non-separable problems via mapping the input signal into higher dimensional feature space in a nonlinear relationship, which was further decoded by a simple linear logistic regression

REFERENCES

- Ababei, R. V., Ellis, M. O. A., Vidamour, I. T., Devadasan, D. S., Allwood, D. A., Vasilaki, E., et al. (2021). Neuromorphic Computation with a Single Magnetic Domain Wall. *Sci. Rep.* 11, 1–13. doi:10.1038/s41598-021-94975-y
- Appeltant, L., Soriano, M. C., Van Der Sande, G., Danckaert, J., Massar, S., Dambre, J., et al. (2011). Information Processing Using a Single Dynamical Node as Complex System. *Nat. Commun.* 2, 466–468. doi:10.1038/ncomms1476
- Balakrishna Pillai, P., and De Souza, M. M. (2017). Nanoionics-Based Three-Terminal Synaptic Device Using Zinc Oxide. *ACS Appl. Mater. Inter.* 9, 1609–1618. doi:10.1021/acsami.6b13746
- Du, C., Cai, F., Zidan, M. A., Ma, W., Lee, S. H., and Lu, W. D. (2017). Reservoir Computing Using Dynamic Memristors for Temporal Information Processing. *Nat. Commun.* 8, 1–10. doi:10.1038/s41467-017-02337-y

machine learning algorithm. Our classification speed is currently ~24 bit/min because we use only one device in this work. Speed can be improved by an input sequence supplied to multiple devices. This would facilitate processing all the rows or the sequence simultaneously. The current device can also be further optimized significantly by scaling. In a practical implementation of the readout function we would expect to use an ADC with additional circuits to implement logistic functionality. Device-to-device variability and multiple voltages on the RC system give added dimensionality to the reservoir. We find that multiple devices with the same pulse significantly improves performance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MS conceived the project. MS and SM have directed the project. AGa performed the simulations and drafted the manuscript. XS performed the experimental measurements. EV and AGi suggested increased sampling to improve the dimensionality of the reservoir. All authors technically discussed the results and commented on the manuscript.

FUNDING

This project was funded by MHRD-SPARC under grant code No. P436 between University of Sheffield, UK and IIT - Roorkee India.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/felec.2022.869013/full#supplementary-material>

- Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., and Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models with Big Data. *Front. Artif. Intell.* 3, 1–23. doi:10.3389/felec.2020.00004
- Güçlü, U., and van Gerven, M. A. J. (2017). Modeling the Dynamics of Human Brain Activity with Recurrent Neural Networks. *Front. Comput. Neurosci.* 11, 1–14. doi:10.3389/fncom.2017.00007
- Hasler, J., and Marr, B. (2013). Finding a Roadmap to Achieve Large Neuromorphic Hardware Systems. *Front. Neurosci.* 7, 1–29. doi:10.3389/fnins.2013.00118
- Hopfield, J. J. (2018). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Natl. Acad. Sci. U S A.* 79, 2554–2558. doi:10.1201/978042950045910.1073/pnas.79.8.2554
- Jaeger, H. (2001). The “Echo State” Approach to Analysing and Training Recurrent Neural Networks – with an Erratum Note. *Bonn, Ger. Ger. Natl. Res. Cent. Inf. Technol. GMD Tech. Rep.* 148, 13. doi:10.1054/nepr.2001.0035

- Kumar, A., Balakrishna Pillai, P., Song, X., and De Souza, M. M. (2018). Negative Capacitance Beyond Ferroelectric Switches. *ACS Appl. Mater. Inter.* 10, 19812–19819. doi:10.1021/acsami.8b05093
- LeCun, Y., Cortes, C., and Burges, C. J. C. (1998). The MNIST Database of Handwritten Digits. Available at: <http://yann.lecun.com/exdb/mnist/> (Accessed December 1, 2021).
- Lukoševičius, M., and Jaeger, H. (2009). Reservoir Computing Approaches to Recurrent Neural Network Training. *Comput. Sci. Rev.* 3, 127–149. doi:10.1016/j.cosrev.2009.03.005
- Lyon, R. (1982). “A Computational Model of Filtering, Detection, and Compression in the Cochlea,” in ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., Paris, France, 3–5 May 1982, 1282–1285. doi:10.1109/ICASSP.1982.1171644
- Maass, W., Natschläger, T., and Markram, H. (2002). Real-Time Computing without Stable States: a New Framework for Neural Computation Based on Perturbations. *Neural Comput.* 14 (11), 2531–2560. doi:10.1162/089976602760407955
- Manneschi, L., Ellis, M. O. A., Gigante, G., Lin, A. C., Del Giudice, P., and Vasilaki, E. (2021a). Exploiting Multiple Timescales in Hierarchical Echo State Networks. *Front. Appl. Math. Stat.* 6, 1–15. doi:10.3389/fams.2020.616658
- Manneschi, L., Lin, A. C., and Vasilaki, E. (2021b). SparCe: Improved Learning of Reservoir Computing Systems through Sparse Representations. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–15. doi:10.1109/TNNLS.2021.3102378
- Midya, R., Wang, Z., Asapu, S., Zhang, X., Rao, M., Song, W., et al. (2019). Reservoir Computing Using Diffusive Memristors. *Adv. Intell. Syst.* 1, 1900084. doi:10.1002/aisy.201900084
- Misra, J., and Saha, I. (2010). Artificial Neural Networks in Hardware: A Survey of Two Decades of Progress. *Neurocomputing* 74, 239–255. doi:10.1016/j.neucom.2010.03.021
- Moon, J., Ma, W., Shin, J. H., Cai, F., Du, C., Lee, S. H., et al. (2019). Temporal Data Classification and Forecasting Using a Memristor-Based Reservoir Computing System. *Nat. Electron.* 2, 480–487. doi:10.1038/s41928-019-0313-3
- Nakajima, K., and Fischer, I. (2021). *Reservoir Computing*. Editors K. Nakajima and I. Fischer (Singapore: Springer Singapore). doi:10.1007/978-981-13-1687-6
- Nguimdo, R. M., Antonik, P., Marsal, N., and Rontani, D. (2020). Impact of Optical Coherence on the Performance of Large-Scale Spatiotemporal Photonic Reservoir Computing Systems. *Opt. Express* 28, 27989. doi:10.1364/oe.400546
- Ni, K., Smith, J. A., Grisafe, B., Rakshit, T., Obradovic, B., Kittl, J. A., et al. (2018). “SoC Logic Compatible Multi-Bit FeMFET Weight Cell for Neuromorphic Applications,” in 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 1–5 Dec. 2018. doi:10.1109/IEDM.2018.8614496
- Paquot, Y., Dupont, F., Smerieri, A., Dambre, J., Schrauwen, B., Haelterman, M., et al. (2012). Optoelectronic Reservoir Computing. *Sci. Rep.* 2, 1–6. doi:10.1038/srep00287
- Park, M.-J., Park, Y., and Lee, J.-S. (2020). Solution-Processed Multiterminal Artificial Synapses Based on Ion-Doped Solid Electrolytes. *ACS Appl. Electron. Mater.* 2, 339–345. doi:10.1021/acsaem.9b00788
- Pierangeli, D., Marcucci, G., and Conti, C. (2021). Photonic Extreme Learning Machine by Free-Space Optical Propagation. *Photon. Res.* 9, 1446. doi:10.1364/prj.423531
- Qin, J.-K., Zhou, F., Wang, J., Chen, J., Wang, C., Guo, X., et al. (2020). Anisotropic Signal Processing with Trigonal Selenium Nanosheet Synaptic Transistors. *ACS Nano* 14, 10018–10026. doi:10.1021/acsnano.0c03124
- Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks* 61, 85–117. doi:10.1016/j.neunet.2014.09.003
- Scikit-Learn (2021). Scikit-Learn sklearn.linear_model.LogisticRegression. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (Accessed December 1, 2021).
- Sillin, H. O., Aguilera, R., Shieh, H.-H., Avizienis, A. V., Aono, M., Stieg, A. Z., et al. (2013). A Theoretical and Experimental Study of Neuromorphic Atomic Switch Networks for Reservoir Computing. *Nanotechnology* 24, 384004. doi:10.1088/0957-4484/24/38/384004
- Song, X., Kumar, A., and De Souza, M. M. (2019). Off-State Operation of a Three Terminal Ionic FET for Logic-In-Memory. *IEEE J. Electron. Devices Soc.* 7, 1232–1238. doi:10.1109/JEDS.2019.2941076
- Sun, L., Wang, Z., Jiang, J., Kim, Y., Joo, B., Zheng, S., et al. (2021). In-Sensor Reservoir Computing for Language Learning via Two-Dimensional Memristors. *Sci. Adv.* 7 (20), eabg1455. doi:10.1126/sciadv.abg1455
- Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., et al. (2019). Recent Advances in Physical Reservoir Computing: A Review. *Neural Networks* 115, 100–123. doi:10.1016/j.neunet.2019.03.005
- Texas Instruments (1991). Texas Instruments-Developed 46-Word Speaker-Dependent Isolated Word Corpus (TI46) NIST Speech Disc 7-1.1
- Torrejon, J., Riou, M., Araujo, F. A., Tsunegi, S., Khalsa, G., Querlioz, D., et al. (2017). Neuromorphic Computing with Nanoscale Spintronic Oscillators. *Nature* 547, 428–431. doi:10.1038/nature23011
- Vandoorne, K., Mechet, P., Van Vaerenbergh, T., Fiers, M., Morthier, G., Verstraeten, D., et al. (2014). Experimental Demonstration of Reservoir Computing on a Silicon Photonics Chip. *Nat. Commun.* 5, 1–6. doi:10.1038/ncomms4541
- Yang, J.-T., Ge, C., Du, J.-Y., Huang, H.-Y., He, M., Wang, C., et al. (2018). Artificial Synapses Emulated by an Electrolyte-Gated Tungsten-Oxide Transistor. *Adv. Mater.* 30, 1801548. doi:10.1002/adma.201801548
- Zhong, Y., Tang, J., Li, X., Gao, B., Qian, H., and Wu, H. (2021). Dynamic Memristor-Based Reservoir Computing for High-Efficiency Temporal Signal Processing. *Nat. Commun.* 12, 408. doi:10.1038/s41467-020-20692-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gaurav, Song, Manhas, Gilra, Vasilaki, Roy and De Souza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.