

## SUPPLEMENTARY MATERIAL

This supplement includes details about the ESN and HESN implementations along with additional tables and figures referenced in the main text.

### ESN and HESN Implementation Details

Figure S1 exhibits the ESN and HESN architectures employed for prediction of action potential time series, which are a special case of the general topology illustrated in Fig. 1, where the input to the network is a multivariate time series including the cardiac action potential and the pacing stimulus driving the network.

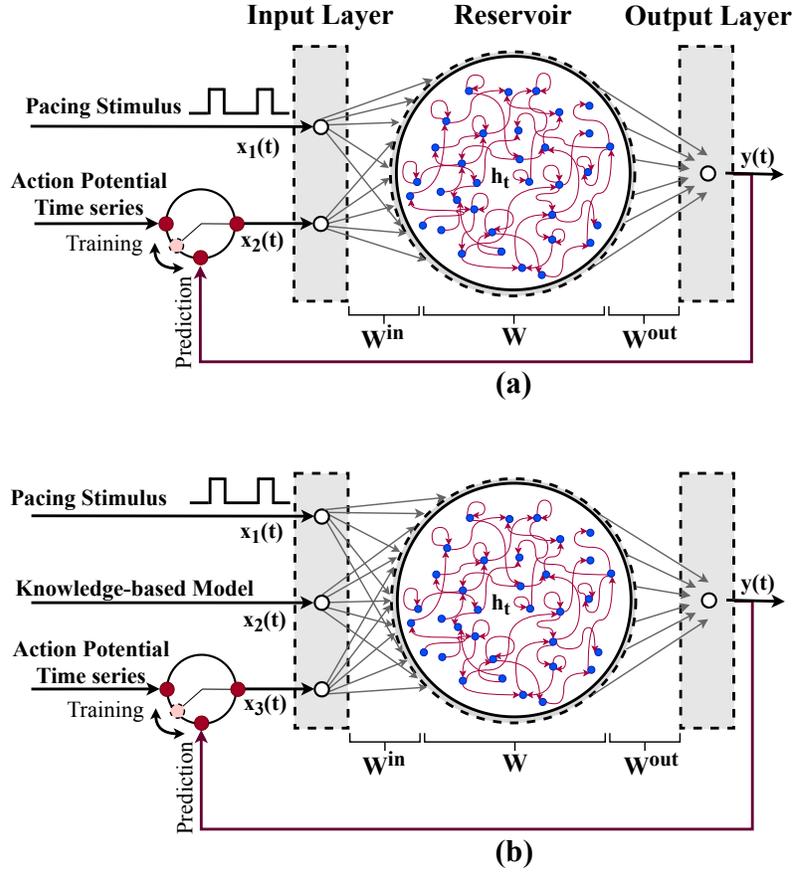


FIG. S1: The main components of (a) the ESN and (b) HESN for prediction of cardiac action potential time series

Moreover, in the HESN, an additional input is required for the knowledge-based model time series. The knowledge-based model and the ESN operate simultaneously during the training and prediction. Therefore, the input layer in HESN accommodates three input variables: (i) the pacing stimulus exciting the network at prescribed interval which is denoted by  $x_1(t)$ , (ii) the knowledge-based model denoted by  $x_2(t) = V_{KB}(t)$ , which can be any imperfect model providing an approximation of the cardiac cell voltage dynamics, and (iii) the time series representing voltage measurements denoted by  $x_3(t) = V(t)$ , where in this work, it is the synthetic measurements generated by the FK model in the first example, or the experimental data in the second example. Therefore, the input signal vector can be written as,

$$x_t = [(x_1(t); x_2(t); x_3(t))]. \quad (1)$$

Accordingly, the update equation of the reservoir state  $h_t$  is identical to Equation 1. In this work, we used the Corrado-Niederer<sup>1</sup> update of the two-variable Mitchell-Schaeffer<sup>2</sup> model with  $\tau_{in} = 0.3$  ms,  $\tau_{out} = 6$  ms,  $\tau_{open} = 40$  ms,  $\tau_{close} = 20$  ms, and  $v_{gate} = 0.13$ .

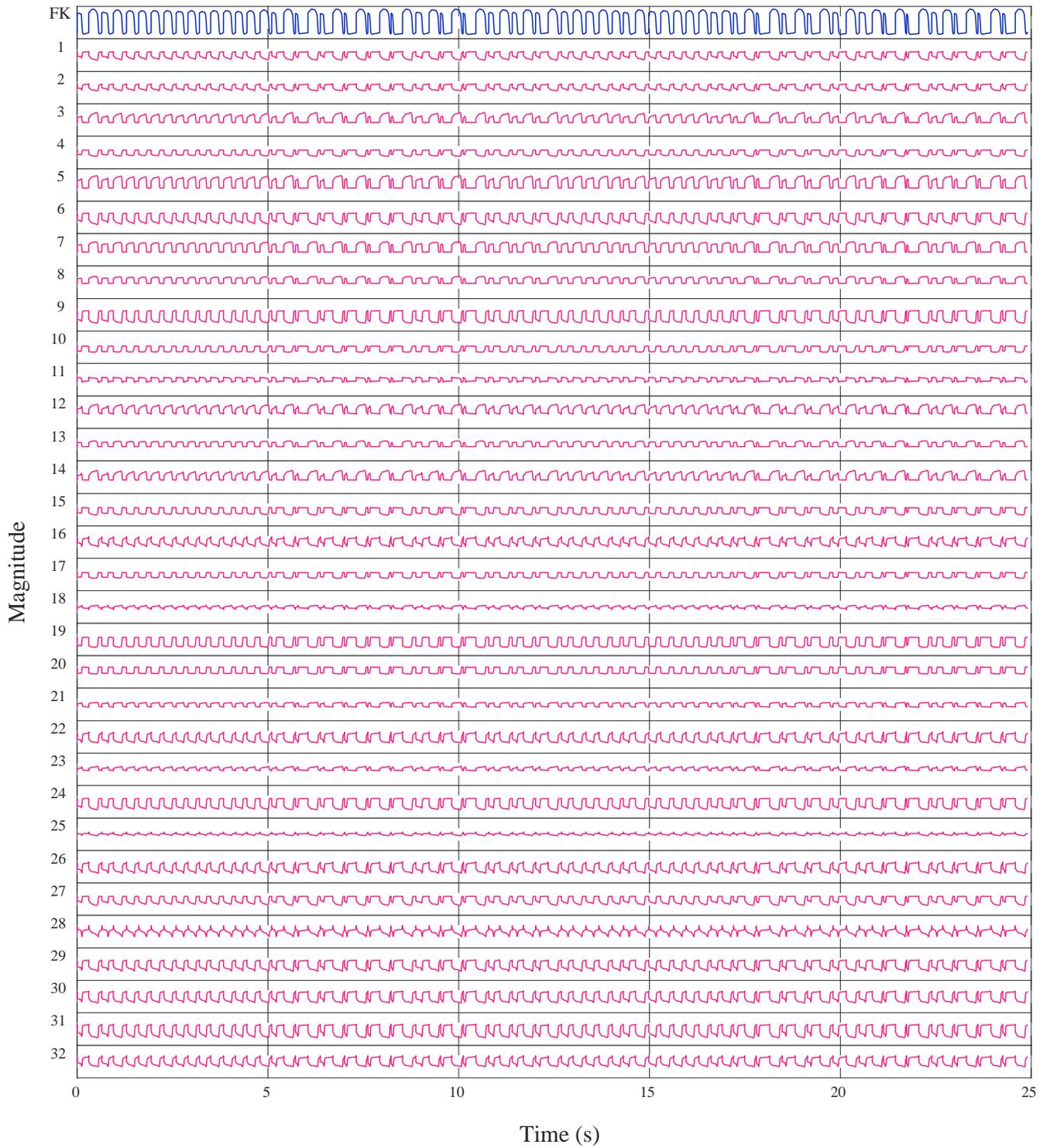


FIG. S2: Visualizing the output of the trained encoder for the FK time series. The range of the y-axis for the input time series is  $[-0.2, 1.2]$ , and for the rest of the subplots representing the trained encoder outputs the y-axis range is set to  $[-1, 1]$ .

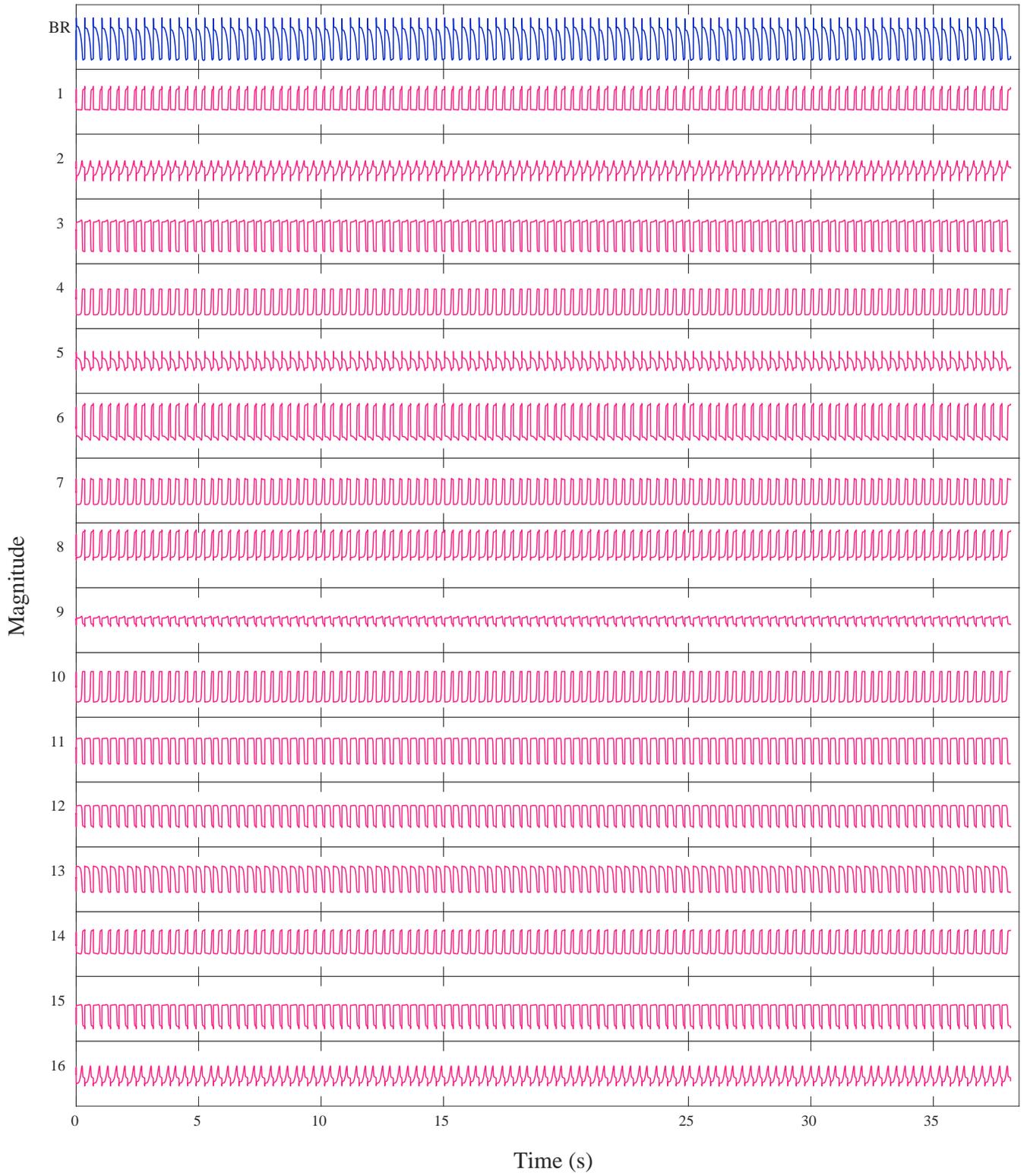


FIG. S3: Visualizing the output of the trained encoder for the BR time series. The range of the y-axis for the input time series is  $[-0.2, 1.2]$ , and for the rest of the subplots representing the trained encoder outputs the y-axis range is set to  $[-1, 1]$ .



FIG. S4: Visualizing the output of the trained encoder for the experimental time series. The range of the y-axis for the input time series is  $[-0.2, 1.2]$ , and for the rest of the subplots representing the trained encoder outputs the y-axis range is set to  $[-1, 1]$ .

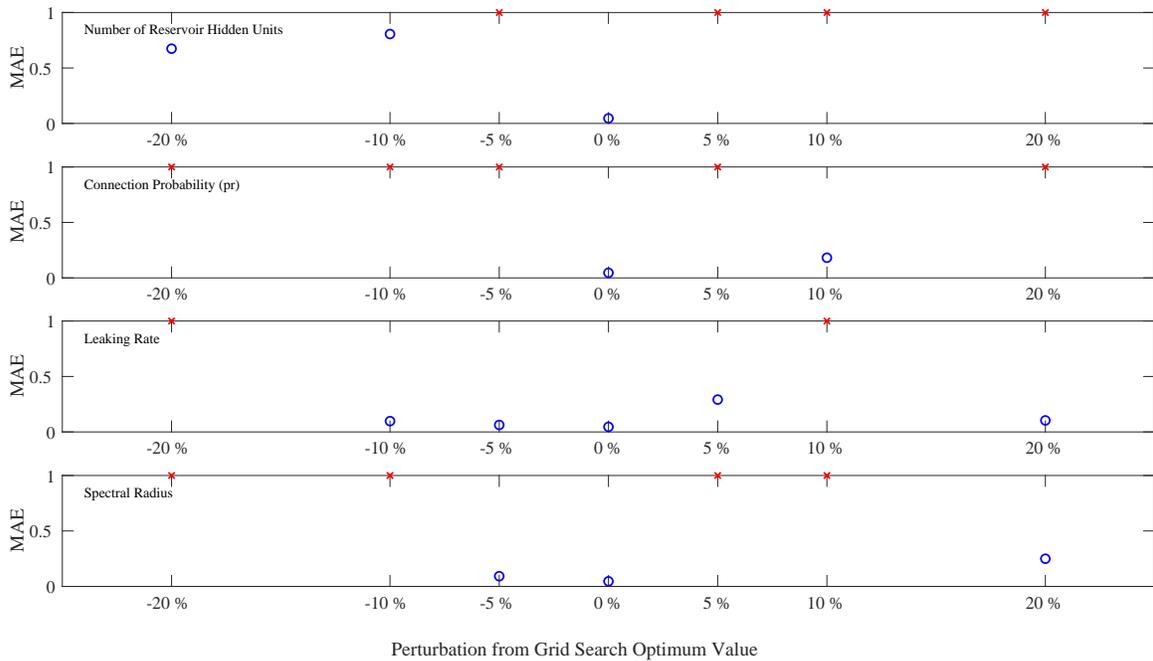


FIG. S5: Sensitivity of the ESN hyperparameters for predicting the FK time series. To facilitate visualization of the low-error cases, red x's represent cases where the error values are above 1 and the network provides poor predictions.

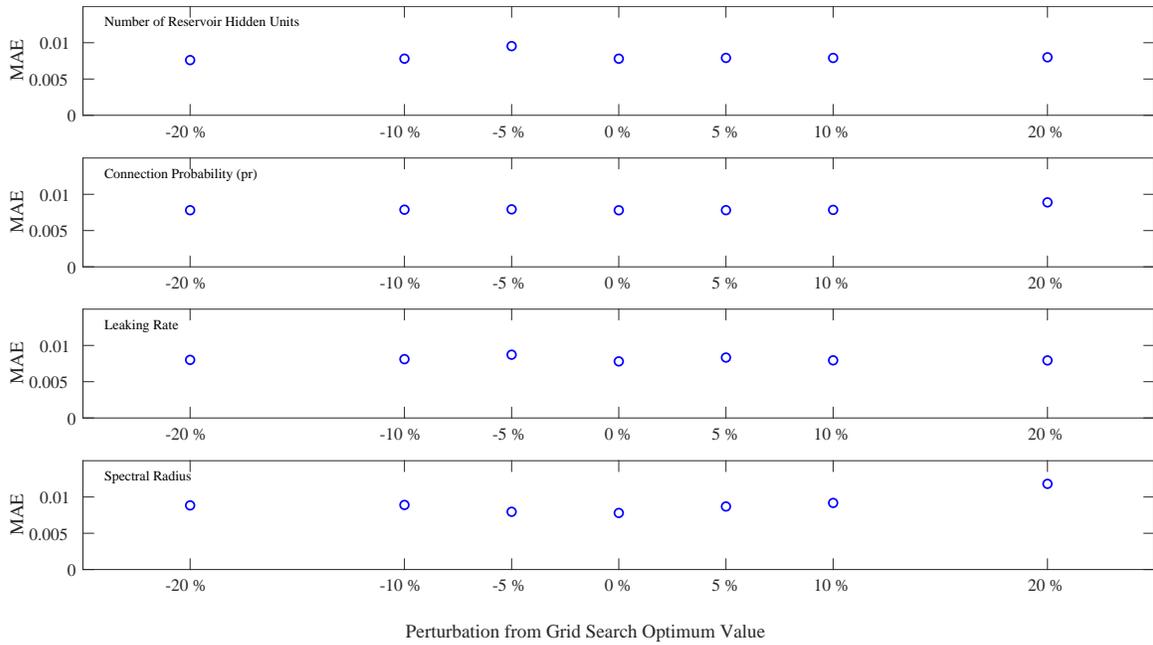


FIG. S6: Sensitivity of the AE-ESN hyperparameters for predicting the FK time series.

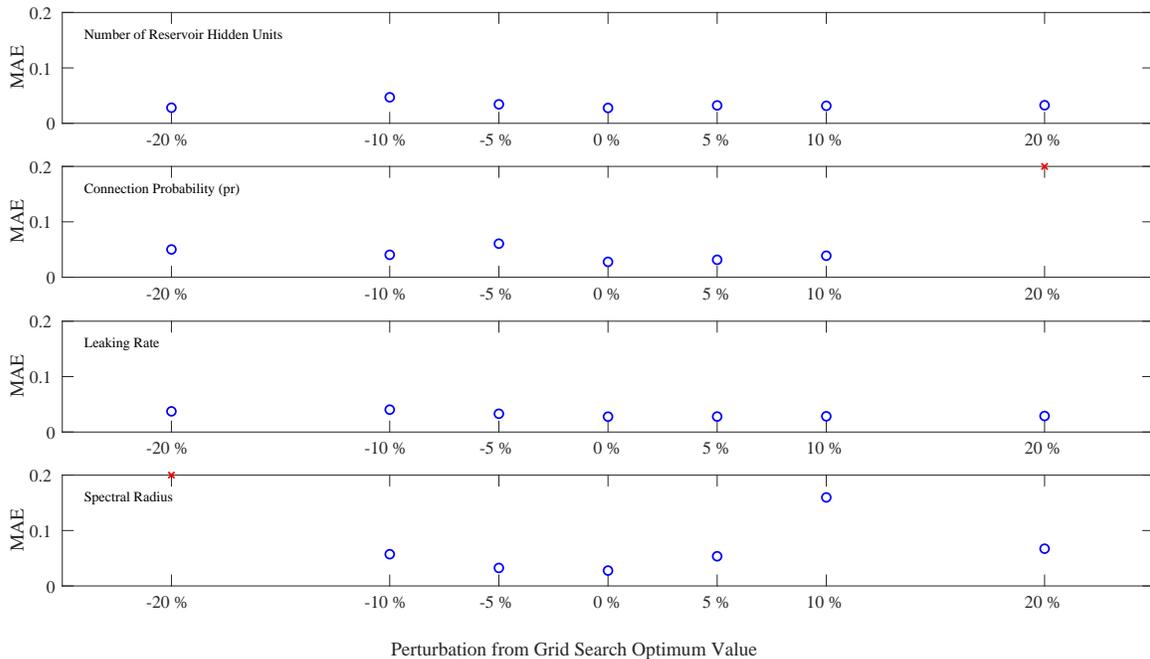


FIG. S7: Sensitivity of the ESN hyperparameter for predicting the BR time series. To facilitate visualization of the low-error cases, red x's represent cases where the error values are above 0.2 and the network provides poor predictions.

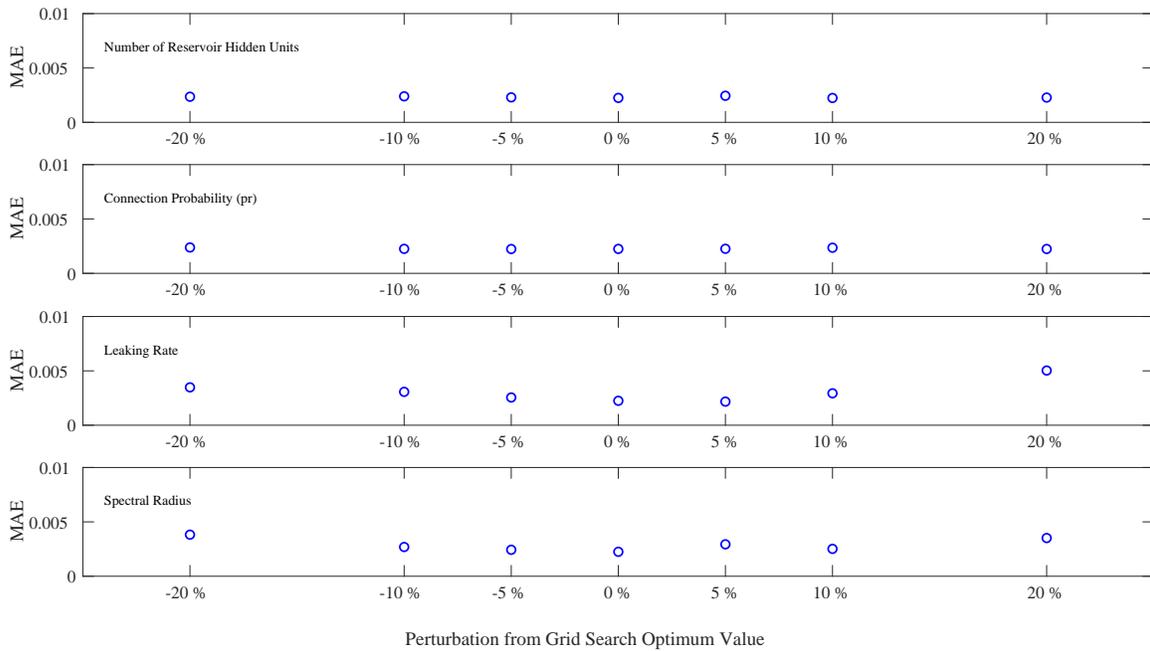


FIG. S8: Sensitivity of the AE-ESN hyperparameter for predicting the BR time series.

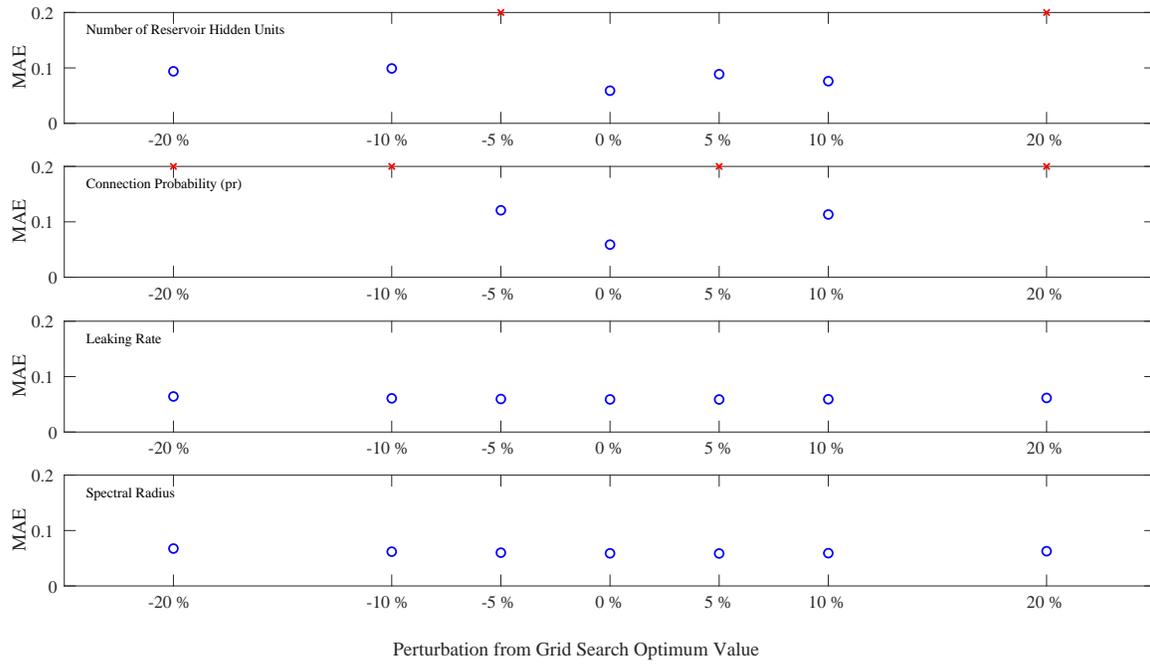


FIG. S9: Sensitivity of the ESN hyperparameter for predicting the experimental time series. To facilitate visualization of the low-error cases, red x's represent cases where the error values are above 0.2 and the network provides poor predictions.

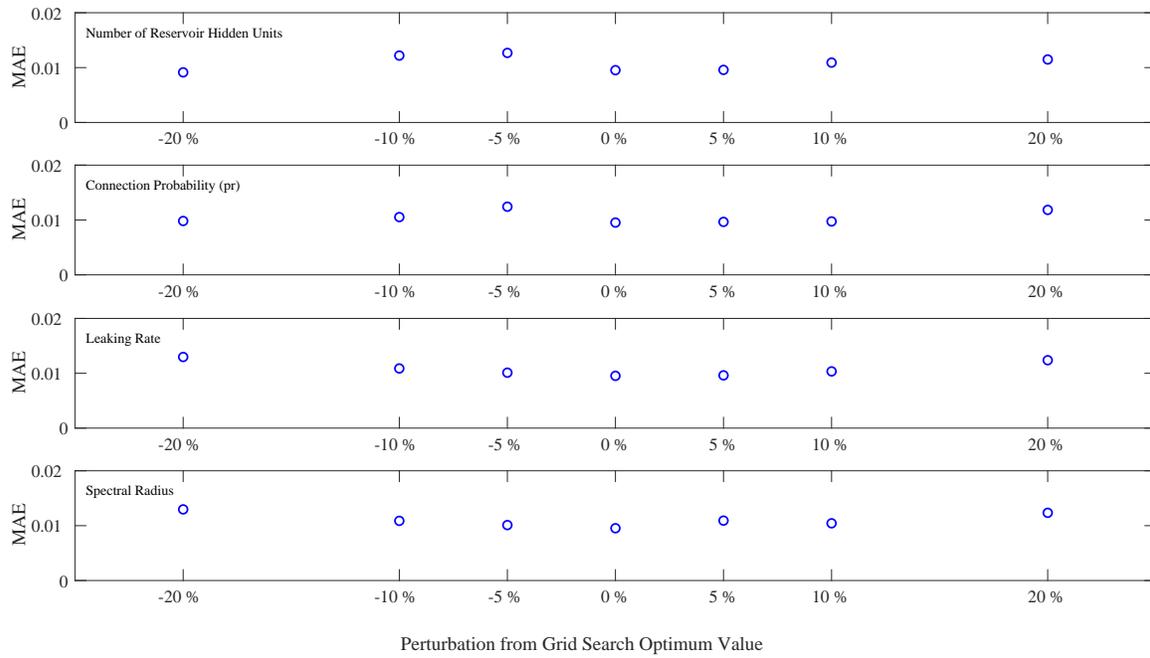


FIG. S10: Sensitivity of the AE-ESN hyperparameter for predicting the experimental time series.

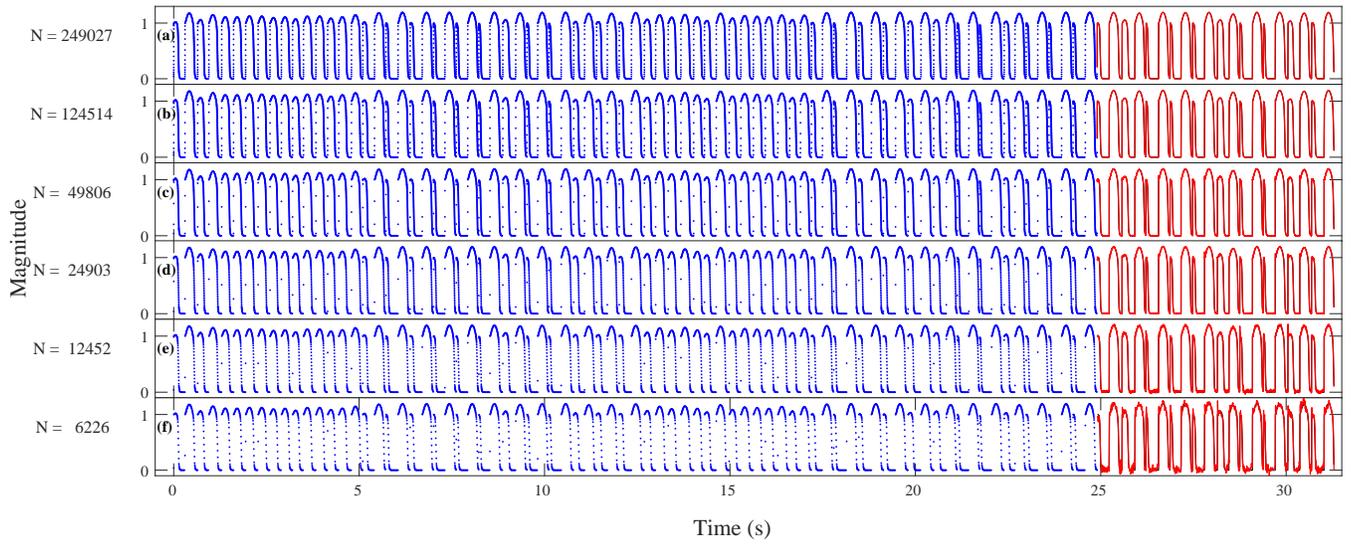


FIG. S11: Application of the AE-ESN to the FK dataset with different temporal resolutions. The training datasets are shown in blue and the prediction results in red. The number of data points in each time series is denoted by  $N$ . The resolutions shown represent (a) an increase by a factor of 2, (b) the baseline resolution, and (c)-(f) decreases in resolution by factors of 2.5, 5, 10, and 20.

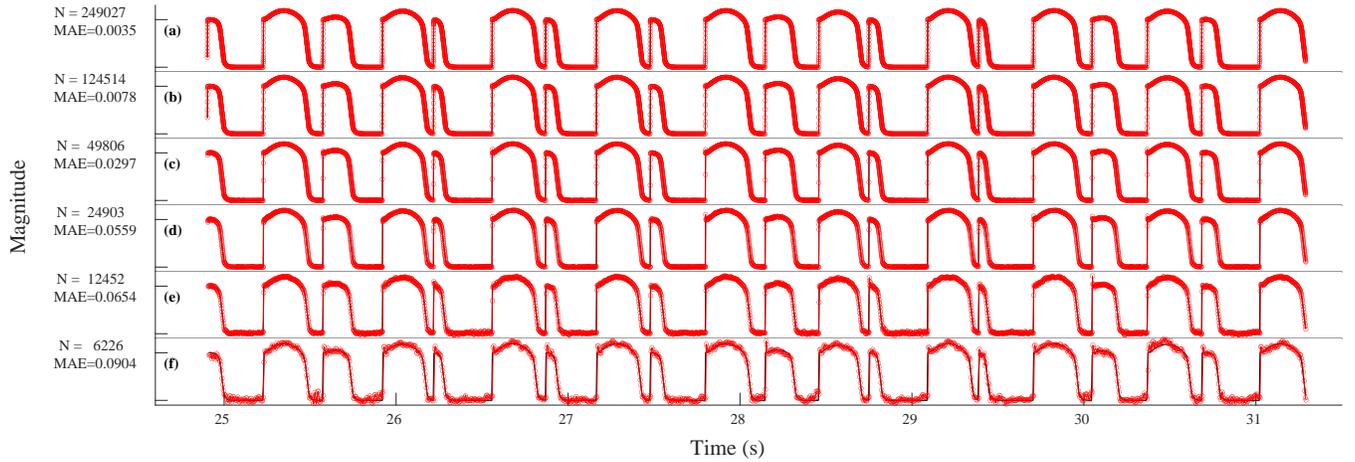


FIG. S12: Results of applying the AE-ESN for forecasting the FK dataset with different temporal resolutions. The size of each training dataset,  $N$ , and the mean absolute error of the prediction results, MAE, are also given. The resolutions shown represent (a) an increase by a factor of 2, (b) the baseline resolution, and (c)-(f) decreases in resolution by factors of 2.5, 5, 10, and 20.

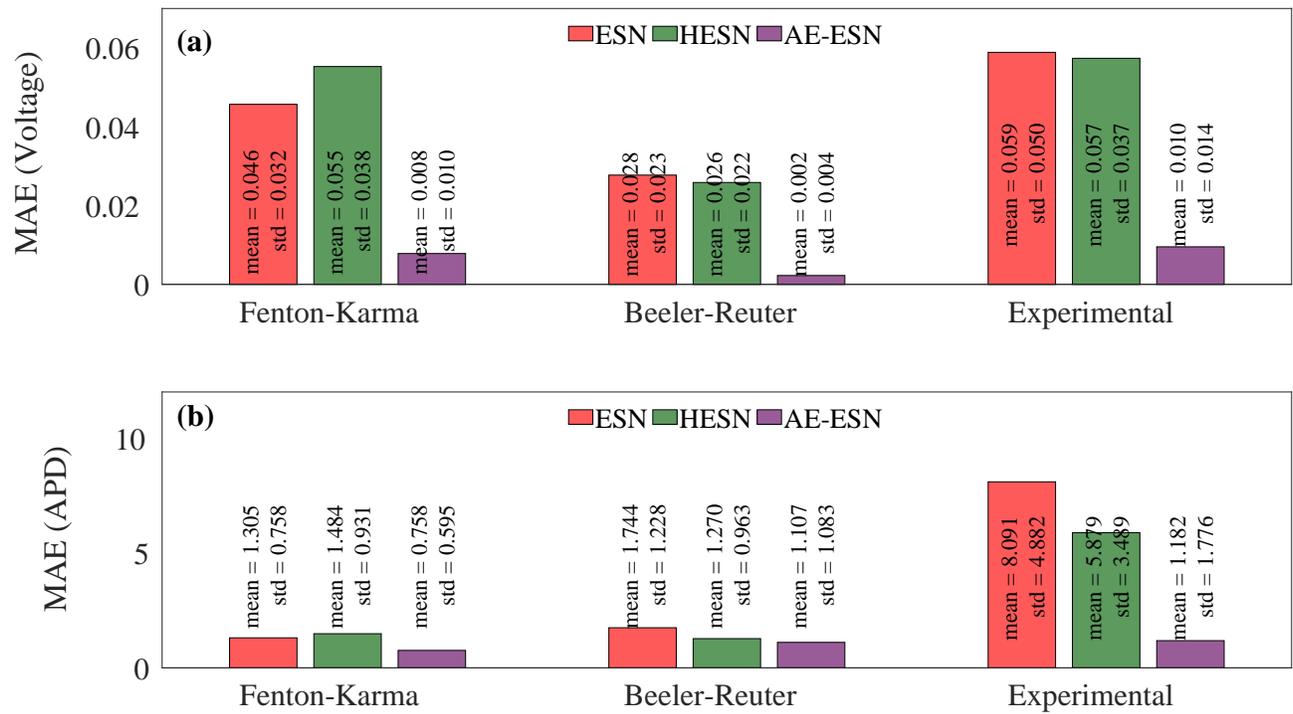


FIG. S13: Summary of the prediction results across the three datasets and three methods used. (a) Prediction error of the voltage values. (b) Prediction error of the APD values.

TABLE S1: Hyperparameter values used for the grid search optimization for each prediction method to predict the FK time series. The selected optimum values are indicated in bold.

| Methods                                      | Parameters  | Values  |
|--|---|---|
| ESN  | Number of reservoir hidden units ( $n_r$ )  | {60, 100, <b>200</b> , 300, 400, 500}   |
|  | Resampling voltage threshold <sup>a</sup> ( $\delta$ )                                      | {0.00, 0.01, 0.02, <b>0.03</b> , 0.04}  |
|  | Resampling time threshold <sup>b</sup> ( $\tau$ )   | {20, 30, <b>40</b> , 50}  |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Input weight scale (pacing stimulus, $\sigma_{in}^2$ )                                      | {0.02, <b>0.05</b> , 0.10, 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, 0.90, 0.99, <b>1.05</b> , 1.15, 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, <b>0.70</b> , 0.80, 0.90, 1.00}                              |
|  | Regularization <sup>d</sup> ( $\lambda$ )   | { $10^{-7}$ , <b><math>10^{-6}</math></b> , $10^{-5}$ , $10^{-4}$ , $10^{-3}$ , $10^{-2}$ } |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, 0.05, 0.10, <b>0.15</b> , 0.20}  |   |
| HESN   | Number of reservoir hidden units ( $n_r$ )  | {60, 100, <b>200</b> , 300, 400, 500}   |
|  | Resampling voltage threshold <sup>a</sup> ( $\delta$ )                                      | {0.00, 0.01, 0.02, <b>0.03</b> , 0.04}  |
|  | Resampling time threshold <sup>b</sup> ( $\tau$ )   | {20, 30, <b>40</b> , 50}  |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Input weight scale (pacing stimulus, $\sigma_{in}^2$ )                                      | {0.02, <b>0.05</b> , 0.10, 0.20, 0.50, 0.80}  |
|  | Input weight scale (knowledge based model, $\sigma_{in}^3$ )                                | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, 0.90, 0.99, <b>1.05</b> , 1.15, 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, <b>0.30</b> , 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00}                              |
| Regularization <sup>d</sup> ( $\lambda$ )    | { $10^{-7}$ , $10^{-6}$ , $10^{-5}$ , <b><math>10^{-4}</math></b> , $10^{-3}$ , $10^{-2}$ } |   |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, 0.05, <b>0.10</b> , 0.15, 0.20}  |   |
| AE-ESN                                       | Layers and cells <sup>f</sup>   | {[64, 32], [64, 16], [64, 4], [ <b>128, 32</b> ], [128, 16], [128, 4]}                      |
|  | Learning rate ( $\eta$ )  | {0.001, 0.002, <b>0.005</b> , 0.010, 0.150}   |
|  | Number of reservoir hidden units ( $n_r$ )  | {60, 100, 200, 300, 400, <b>500</b> }   |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, 0.90, 0.99, 1.05, <b>1.15</b> , 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, <b>0.90</b> , 1.00}                              |
|  | Regularization <sup>d</sup> ( $\lambda$ )   | { $10^{-7}$ , $10^{-6}$ , $10^{-5}$ , $10^{-4}$ , <b><math>10^{-3}</math></b> , $10^{-2}$ } |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, <b>0.05</b> , 0.10, 0.15, 0.20}  |   |

- <sup>a</sup> The minimum difference between the voltage values of each two consecutive data points used as the first criterion for resampling the action potential time series in ESN and HESN methods.
- <sup>b</sup> The maximum time gap (in ms) between each two consecutive data points used for resampling in ESN and HESN.
- <sup>c</sup> The reservoir weight matrix is scaled such that its spectral radius, defined as the largest among the absolute values of the eigenvalues, is equal to the selected value.
- <sup>d</sup> The ridge regression regularization factor used for calculation of the readout weights.
- <sup>e</sup> The probability of having an edge between each two neurons in the reservoir controlling the sparsity of the reservoir graph.
- <sup>f</sup> The number of LSTM cells in the layers of encoder part. The decoder includes the same values but in reverse order to form a mirrored topology.

TABLE S2: Hyperparameter values used for the grid search optimization for each prediction method to predict the BR time series. The selected optimum values are indicated in bold.

| Methods                                      | Parameters  | Values  |
|--|---|---|
| ESN  | Number of reservoir hidden units ( $n_r$ )  | {60, 100, 200, <b>300</b> , 400, 500}   |
|  | Resampling voltage threshold <sup>a</sup> ( $\delta$ )                                      | {0.00, 0.01, <b>0.02</b> , 0.03, 0.04}  |
|  | Resampling time threshold <sup>b</sup> ( $\tau$ )   | {20, <b>30</b> , 40, 50}  |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Input weight scale (pacing stimulus, $\sigma_{in}^2$ )                                      | {0.02, <b>0.05</b> , 0.10, 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, <b>0.90</b> , 0.99, 1.05, 1.15, 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, <b>0.80</b> , 0.90, 1.00}                              |
|  | Regularization <sup>d</sup> ( $\lambda$ )   | { <b><math>10^{-7}</math></b> , $10^{-6}$ , $10^{-5}$ , $10^{-4}$ , $10^{-3}$ , $10^{-2}$ } |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, 0.05, 0.10, <b>0.15</b> , 0.20}  |   |
| HESN   | Number of reservoir hidden units ( $n_r$ )  | {60, 100, 200, 300, 400, <b>500</b> }   |
|  | Resampling voltage threshold <sup>a</sup> ( $\delta$ )                                      | {0.00, 0.01, <b>0.02</b> , 0.03, 0.04}  |
|  | Resampling time threshold <sup>b</sup> ( $\tau$ )   | {20, <b>30</b> , 40, 50}  |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Input weight scale (pacing stimulus, $\sigma_{in}^2$ )                                      | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Input weight scale (knowledge based model, $\sigma_{in}^3$ )                                | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, 0.90, 0.99, <b>1.05</b> , 1.15, 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, <b>0.80</b> , 0.90, 1.00}                              |
| Regularization <sup>d</sup> ( $\lambda$ )    | { $10^{-7}$ , $10^{-6}$ , $10^{-5}$ , $10^{-4}$ , <b><math>10^{-3}</math></b> , $10^{-2}$ } |   |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, 0.05, <b>0.10</b> , 0.15, 0.20}  |   |
| AE-ESN                                       | Layers and cells <sup>f</sup>   | {[64, 32], [ <b>64, 16</b> ], [64, 4], [128, 32], [128, 16], [128, 4]}                      |
|  | Learning rate ( $\eta$ )  | {0.001, <b>0.002</b> , 0.005, 0.010, 0.150}   |
|  | Number of reservoir hidden units ( $n_r$ )  | {60, 100, 200, 300, 400, <b>500</b> }   |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, 0.90, <b>0.99</b> , 1.05, 1.15, 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, <b>0.90</b> , 1.00}                              |
|  | Regularization <sup>d</sup> ( $\lambda$ )   | { $10^{-7}$ , $10^{-6}$ , $10^{-5}$ , $10^{-4}$ , $10^{-3}$ , <b><math>10^{-2}</math></b> } |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, <b>0.05</b> , 0.10, 0.15, 0.20}  |   |

<sup>a</sup> The minimum difference between the voltage values of each two consecutive data points used as the first criterion for resampling the action potential time series in ESN and HESN methods.

<sup>b</sup> The maximum time gap (in ms) between each two consecutive data points used for resampling in ESN and HESN.

<sup>c</sup> The reservoir weight matrix is scaled such that its spectral radius, defined as the largest among the absolute values of the eigenvalues, is equal to the selected value.

<sup>d</sup> The ridge regression regularization factor used for calculation of the readout weights.

<sup>e</sup> The probability of having an edge between each two neurons in the reservoir controlling the sparsity of the reservoir graph.

<sup>f</sup> The number of LSTM cells in the layers of encoder part. The decoder includes the same values but in reverse order to form a mirrored topology.

TABLE S3: Hyperparameter values used for the grid search optimization for each prediction method to predict the experimental time series. The selected optimum values are indicated in bold.

| Methods                                      | Parameters  | Values  |
|--|---|---|
| ESN  | Number of reservoir hidden units ( $n_r$ )  | {60, <b>100</b> , 200, 300, 400, 500}   |
|  | Resampling voltage threshold <sup>a</sup> ( $\delta$ )                                      | {0.00, 0.01, 0.02, <b>0.03</b> , 0.04}  |
|  | Resampling time threshold <sup>b</sup> ( $\tau$ )   | {20, 30, <b>40</b> , 50}  |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, 0.10, 0.20, <b>0.50</b> , 0.80}  |
|  | Input weight scale (pacing stimulus, $\sigma_{in}^2$ )                                      | {0.02, 0.05, 0.10, 0.20, <b>0.50</b> , 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, <b>0.85</b> , 0.90, 0.99, 1.05, 1.25, 1.50}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, <b>0.50</b> , 0.60, 0.70, 0.80, 0.90, 1.00}                              |
|  | Regularization <sup>d</sup> ( $\lambda$ )   | { $10^{-7}$ , <b><math>10^{-6}</math></b> , $10^{-5}$ , $10^{-4}$ , $10^{-3}$ , $10^{-2}$ } |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, 0.05, <b>0.10</b> , 0.15, 0.20}  |   |
| HESN   | Number of reservoir hidden units ( $n_r$ )  | {60, 100, <b>200</b> , 300, 400, 500}   |
|  | Resampling voltage threshold <sup>a</sup> ( $\delta$ )                                      | {0.00, 0.01, 0.02, <b>0.03</b> , 0.04}  |
|  | Resampling time threshold <sup>b</sup> ( $\tau$ )   | {20, 30, <b>40</b> , 50}  |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, 0.10, 0.20, 0.50, <b>0.80</b> }  |
|  | Input weight scale (pacing stimulus, $\sigma_{in}^2$ )                                      | {0.02, 0.05, 0.10, 0.20, <b>0.50</b> , 0.80}  |
|  | Input weight scale (knowledge based model, $\sigma_{in}^3$ )                                | {0.02, 0.05, 0.10, 0.20, <b>0.50</b> , 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | { <b>0.80</b> , 0.85, 0.90, 0.99, 1.05, 1.25, 1.50}   |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00}                                      |
| Regularization <sup>d</sup> ( $\lambda$ )    | { $10^{-7}$ , $10^{-6}$ , <b><math>10^{-5}</math></b> , $10^{-4}$ , $10^{-3}$ , $10^{-2}$ } |   |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, 0.05, <b>0.10</b> , 0.15, 0.20}  |   |
| AE-ESN                                       | Layers and cells <sup>a</sup>   | {[64, 32], [64, 16], [64, 4], [ <b>128, 32</b> ], [128, 16], [128, 4]}                      |
|  | Learning rate ( $\eta$ )  | {0.001, 0.002, <b>0.005</b> , 0.010, 0.150}   |
|  | Number of reservoir hidden units ( $n_r$ )  | {60, 100, 200, 300, 400, <b>500</b> }   |
|  | Input weight scale (action potential, $\sigma_{in}^1$ )                                     | {0.02, 0.05, <b>0.10</b> , 0.20, 0.50, 0.80}  |
|  | Spectral radius <sup>c</sup> ( $\rho$ )   | {0.80, 0.85, 0.90, <b>0.99</b> , 1.05, 1.15, 1.25}  |
|  | Leaking rate ( $\alpha$ )   | {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, <b>0.90</b> , 1.00}                              |
|  | Regularization <sup>d</sup> ( $\lambda$ )   | { $10^{-7}$ , $10^{-6}$ , $10^{-5}$ , $10^{-4}$ , $10^{-3}$ , <b><math>10^{-2}</math></b> } |
| Connection probability <sup>e</sup> ( $pr$ ) | {0.01, 0.02, <b>0.05</b> , 0.10, 0.15, 0.20}  |   |

<sup>a</sup> The minimum difference between the voltage values of each two consecutive data points used as the first criterion for resampling the action potential time series in ESN and HESN methods

<sup>b</sup> The maximum time gap (in ms) between each two consecutive data points used for resampling in ESN and HESN

<sup>c</sup> The reservoir weight matrix is scaled such that its spectral radius, defined as the largest among the absolute values of the eigenvalues, is equal to the selected value.

<sup>d</sup> The ridge regression regularization factor used for calculation of the readout weights.

<sup>e</sup> The probability of having an edge between each two neurons in the reservoir controlling the sparsity of the reservoir graph.

<sup>f</sup> The number of LSTM cells in the layers of encoder part. The decoder includes the same values but in reverse order to form a mirrored topology.

**REFERENCES**

- <sup>1</sup>C. Corrado and S. A. Niederer, "A two-variable model robust to pacemaker behaviour for the dynamics of the cardiac action potential," *Mathematical Biosciences* **281**, 46–54 (2016).
- <sup>2</sup>C. C. Mitchell and D. G. Schaeffer, "A two-current model for the dynamics of cardiac membrane," *Bulletin of Mathematical Biology* **65**, 767–793 (2003).