IEEE Open Journal of the
Industrial Electronics Society

Received 10 April 2024; accepted 30 April 2024. Date of publication 9 May 2024;
date of current version 20 May 2024. The review of this article was arranged by Associate Editor Guanghui Wen.

Digital Object Identifier 10.1109/OJIES.2024.3397789

Explainable Artificial Intelligence for Crowd Forecasting Using Global Ensemble Echo State Networks

CHAMOD SAMARAJEEWA [1], DASWIN DE SILVA [1] (Senior Member, IEEE), MILOS MANIC [2] (Fellow, IEEE), NISHAN MILLS [1] (Member, IEEE), PRABOD RATHNAYAKA [1], AND ANDREW JENNINGS [1]

[1]Centre for Data Analytics and Cognition, La Trobe University, Bundoora, VIC 3086, Australia
[2]Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284 USA

CORRESPONDING AUTHOR: DASWIN DE SILVA (e-mail: d.desilva@latrobe.edu.au).

ABSTRACT Crowd monitoring is a primary function in diverse industrial domains, such as smart cities, public transport, and public safety. Recent advancements in low-energy devices and rapid connectivity have enabled the generation of real-time data streams suitable for crowd-monitoring applications. Crowd forecasting is typically achieved using deep learning models that learn the evolving nature of data streams. The computational complexity, execution time, and opaqueness are inherent challenges of deep learning models that also overlook the latent relationships between multiple real-time data streams for improved accuracy. To address these challenges, we propose the global ensemble echo state network approach for explainable crowd forecasting using multiple WiFi data streams. This approach replaces the random input mapping layer with a clustering layer, allowing the network to learn input projections on cluster centroids. It incorporates an ensemble readout comprising a stack of reservoir layers that provide model explainability. It also learns multiple related time series in parallel to construct a global model that leverage latent relationships across the data streams. This approach was empirically evaluated in a multicampus, mixed-use tertiary education setting. The results of which confirm the effectiveness and interpretability of the proposed approach for industrial applications of crowd forecasting.

INDEX TERMS Artificial intelligence (AI), crowd forecasting, echo state networks, global forecasting models, model explainability.

I. INTRODUCTION

Crowd monitoring is an increasingly important activity for the operational efficiency, security, safety, and optimization of industrial, social, and economical systems. Crowd monitoring or the modeling of crowd dynamics spans across a number of subfields, such as people counting, density mapping, crowd recognition, crowd characterization, crowd tracking, and crowd forecasting [1]. Out of these subfields, crowd forecasting is crucial for the planning and preparation of smart cities as well as safety and security protocols and policies of related services. For instance, the injuries and loss of life due to poor crowd control of mass gatherings in smart city settings [2], [3], [4], could be mitigated with accurate and timely predictions. Besides crowd safety, event scheduling, resource allocation, and service optimization can also be enabled and supported with such crowd forecasts [5].

The source data streams used for crowd monitoring tasks are diverse. This includes static or manual-operated counters, Bluetooth sensors, closed-circuit television (CCTV) cameras, WiFi sensors, infrared sensors, and Internet-of-Things (IoT) devices. The usage of Bluetooth sensors is prevalent in indoor settings and large-scale events. These sensors are installed in key locations, and a monitoring system generates a real-time population estimation [6], [7], [8], [9]. With infrared sensors [10], crowd count is determined indirectly by temperature differences surrounding the sensors. The channel state information (CSI) of WiFi-enabled IoT devices [11], [12] estimates crowd count by analyzing the strength of WiFi signals

that devices in a certain region are sending out. Video feeds from CCTV cameras have been widely used for monitoring and surveillance [13], [14], [15]. However, the effectiveness of such methods depends on face detection capability, and require large training datasets to be accurate. In contrast, WiFi data streams overcome this issue of obscured image frames, as well as being a low-cost, widely deployed base communication technology [16], [17].

Despite the diversity of data streams, the process for crowd monitoring use cases follows a consistent path. After the raw data streams have been identified, acquired, preprocessed, and staged, then the computational tasks of crowd monitoring are initiated. Following our focus in this article on crowd forecasting, we present an overview of such techniques that represent the crowd counts as time-series forecasting. Although early work consisted of statistical models, such as auto regression integrated moving average (ARIMA) [18], Gaussian models [19], and Markov models [20], more recently, artificial intelligence (AI) models have outperformed the conventional approaches as they are able to better capture complex temporal dynamics of crowd counts across geographical regions. For instance, recurrent neural networks (RNNs) [21] and long short-term memory (LSTM) [22] have been highly effective. However, they also report on the complexity of the learning process and the inability to learn efficiently under computing constraints as limitations. Reservoir computing approaches, such as echo-state machines (ESNs) [23], [24] and liquid state machines [25], have been proposed as a viable alternative that does not compromise the accuracy of capturing temporal dynamics while also maintaining a lower computational load. Noting the multiple, competing challenges of crowd forecasting, such as weather, traffic, event format, and unpredictability of human behaviors, ESNs are technically capable of addressing these challenges while also providing accurate forecasts as they contain few parameters, which makes them less prone to overfitting and more robust to changes in the underlying data streams.

WiFi data, readily available in commercial settings, presents an underutilized resource. Acknowledging the significance of crowd forecasting and monitoring in averting congestion, managing security risks, and optimizing resource allocation, this data hold promise in approximating crowd behavior [5], [26]. Leveraging this information for crowd forecasting seems pertinent, considering its intrinsic availability. However, treating this as a time-series problem and employing conventional time-series models encounters challenges in capturing the intricate temporal dynamics of human behavior. Hence, the necessity for a model resilient to such complex and erratic behavior fuels the use of echo state networks [27]. Yet, existing models, including echo state networks, lack interpretability for evaluating the quality of crowd-forecasting tasks. The attribution of the reasoning behind forecasts motivates the development of an explainable AI model for improving the standard echo state network models. This development aims to provide insight and understanding behind the forecasts, adding a layer of interpretability to the predictions.

Effectively leveraging WiFi data for crowd behavior estimation while safeguarding user privacy remains a pivotal challenge in this domain. In addition, capturing the dynamic and often chaotic nature of human behavior within crowded settings necessitates innovative methodologies capable of comprehending these intricate patterns. Finding the balance between model accuracy and its interpretability is a key challenge for explainability without compromising predictive capabilities. Moreover, ensuring the scalability and applicability of these models [28] across various urban settings and crowd behaviors presents an unresolved challenge in this evolving field. Addressing these multifaceted challenges stands as a crucial step toward enhancing the efficacy of crowd behavior predictions.

In this article, we present a novel explainable AI approach based on ESN, the global ensemble echo state network for explainable crowd forecasting using WiFi data streams. The key contributions are as follows.

1) A novel echo state network learning approach with a clustering layer to replace the input mappings, multiple reservoirs allocated to each cluster, and an ensemble readout layer that completes the time-series forecasts.
2) Model explainability of this novel ESN learning approach that interprets the forecast and an explainability metric that evaluates model performance quality.
3) Global model of this novel ESN learning approach that is trained on multiple time series to enhance the prediction accuracy by taking advantage of latent relationships across the data streams.

The rest of this article is organized as follows. Section II presents related research in crowd forecasting and the use of echo state networks for similar use cases. The proposed global ensemble echo state network is presented in Section III followed by the experiments carried out on crowd forecasting and results in Section IV. Section V concludes this article.

II. RELATED WORK

The global rise of smart cities showcases the integration of AI, specifically machine learning (ML) and deep learning, alongside the incorporation of IoT devices and their real-time data streams. This integration significantly enhances various aspects, particularly in ensuring safety, where dedicated efforts combat cyber attacks. For instance, Eddin [29] proposed an efficient multitask deep learning-based detector aimed at thwarting electricity theft attacks on smart meters. Similarly, Keliris et al. [30] introduced a novel ML-based intrusion detection system tailored to combat cyber attacks on smart grids. Magaia [31] examined the strengths and weaknesses of security-related approaches utilizing deep learning methods, shedding light on open issues and future directions in leveraging these techniques to fortify IoT security within smart city applications. Liu et al. [32] presented a hybrid forecasting method that combines TimeGAN for data expansion with a CNN-enhanced LSTM network, addressing challenges associated with sparse power load data. Throughout literature, discussions on harnessing deep learning and ML approaches

to bolster IoT, wireless devices, and smart city services have been pervasive (Mohammadi et al. [33]; Ameer et al. [34]; Bellavista et al. [35]).

Despite the inherent limitations of extracting nonlinear temporal dynamics, statistical models like ARIMA [18], Gaussian models [19], and Markov models have been shown to be effective approximations. Tan et al. [36] described a method for estimating the population of a shopping centre using WiFi sensors and forecasting future volume using an ARIMA model. For tasks involving time-series prediction, the seasonal autoregressive integrated moving average model (SARIMA), an enhanced variant of the ARIMA model, has also been employed. A differential process is present in this model and is utilized to continuously extract trend and period information. Wu et al. [37] described how to use WiFi probe requests to estimate the number of customers inside a business and forecast using an SARIMA model. Modeled as a time-series, crowd forecasting has been addressed using deep learning and artificial neural networks methods. The use of RNNs [21], LSTMs [22], and GRUs [38] are among the most widely used. In [17], an LSTM model is used to predict crowd size for a large public gathering. A variety of LSTMs, including VLSTM, BiLSTM, EDLSTM, CNNLSTM, and ConvLSTM, have been used to make predictions in this case. The crowd size is approximated via WiFi probe requests. However, only the crowd count is used as an input; other factors, such as the weather and traffic updates, are not taken into account. Echo state networks for crowd forecasting have been overlooked despite its potential being demonstrated in several other application domains of time-series forecasting. A clustered complex echo state network has been proposed for predicting mobile communication traffic with prior knowledge [39]. By using the Fourier spectrum as the prior knowledge to create several clusters within the reservoir, the prediction accuracy of a traditional ESN is outperformed. In [40], authors presented another comparable example of using an ESN model for forecasting network traffic. For both short- and long-term forecasting, the standard ESN is evaluated against LSTM, CNN, and SARIMA models in this study, which also draws on the primary elements of the AI lifecycle [41]. Road traffic forecasting is another application scenario modeled by echo state networks. Del Ser et al. [42] presented a stacking ensemble learning approach to an ESN model to predict road traffic. Zhang et al. [43] suggested using an ESN in a similar method where a fruit fly optimization algorithm is used to predict the volume of traffic every five minutes with ESN. Global models for time-series forecasting are a recent development that simultaneously learns from several time series with relatedness properties. This single model with a set of global parameters across many series is in contrast to local models, as it trains on inter-series fluctuations using a smaller number of parameters [44]. It has also been demonstrated that the complexity of local, individual models increase as a new model is trained for each series, unlike global models where the complexity is invariable to the number of time series trained [45]. Global models have been used for energy consumption forecasting [46]. ESNs maintain common properties for time-series forecasting type problems, with application-level difference due to the nature of the entities and their movements. For instance, data and network traffic vary temporally and not spatially, road traffic has limited spatial variability, and crowd movement varies both spatially and temporally.

With the proliferation of real-world AI applications, there is a growing imperative to move beyond black box (opaque) models that cannot justify decision-making processes. As a response, significant research has delved into explainable AI. Broadly, two major categories have emerged: 1) intrinsically interpretable models and 2) model-agnostic approaches [47]. Models like linear regression and logistic regression fall under the category of interpretable models due to their simple equation-based nature, providing coefficients for each feature. However, their limitation lies in their capacity to model only linear relationships. On the other end, decision trees offer interpretability by explicitly showcasing feature importance based on the sequence of feature splits and a tree structure. Features higher in the tree typically hold more influence in predictions by contributing more to information gain or reducing impurity.

While these models may not be suitable for complex problems, model-agnostic methods have gained traction, separating explanation from the ML model, thus ensuring compatibility with various models. Partial dependence plot (PDP) [48] is one such method that offers a function dependent solely on the plotted feature(s) by marginalizing over other features, considering their interactions. PDP offers a clear and causal interpretation by illustrating how changes in specific features affect predictions. Another method, individual conditional expectation [49], differs from PDP by plotting one line per instance, illustrating how a feature influences prediction changes. While these approaches provide global explanations regarding how specific features impact model predictions, literature also features several local explainable methods that offer explanations for individual predictions. Local interpretable model-agnostic explanations [50] is one such method, training an inherently interpretable model on a new dataset formed by permuting samples and their corresponding predictions from the opaque model. Shapely [51], another local explanation approach rooted in the cooperative game theory, fairly distributes feature importance among participating features. Despite the existence of these approaches in literature, as far as our knowledge extends, no proper explainable echo state network model has been proposed, and no model-agnostic method has been applied. Therefore, our current focus revolves around developing an interpretable echo state network model.

III. PROPOSED METHOD: GLOBAL ENSEMBLE ECHO STATE NETWORK

Echo state networks typically use reservoir weights and input mapping weight matrices that are randomly initialized, that carry the intrinsic advantage of random connections that learn nonlinear temporal dynamics [52]. However, this also

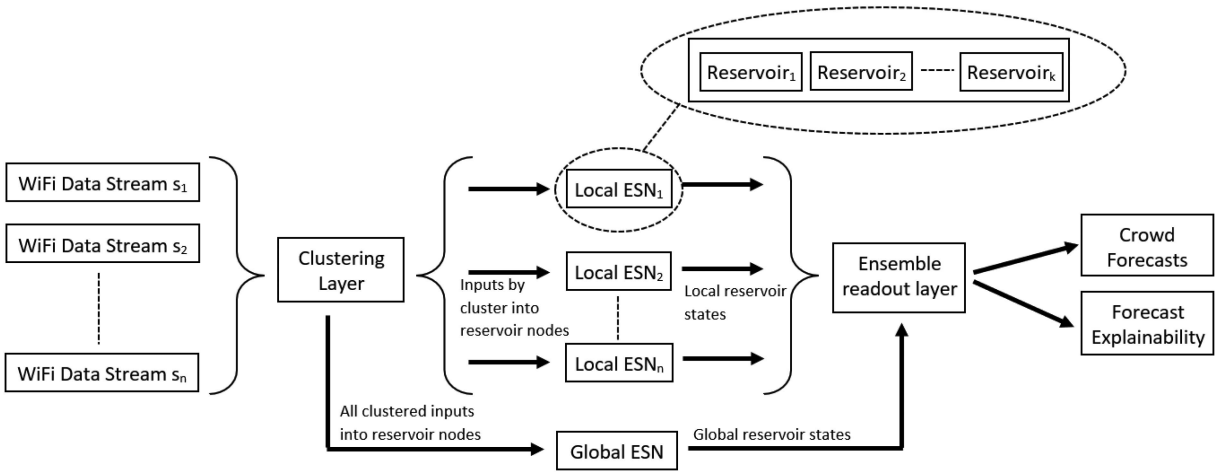

FIGURE 1. **High-level architecture of the proposed global ensemble echo state network.**

means that input mapping weights cause fluctuations (diminish or amplify) of the input itself during mapping to the reservoir nodes. We attempt to address this by introducing a premapping layer for clustering the inputs into segments based on common properties. The centroids of these clusters then become the weights for the mapping between inputs and reservoir nodes. In contrast to a similar method proposed by Steiner et al. [53] where there are restrictions placed on assigning inputs to particular nodes, in our method we generate a unique reservoir network for each recognized cluster. Each input is projected on to centroids, and the projections of the inputs are then injected into reservoirs. By doing this, each reservoir only receives information about input data projections that are pertinent to the cluster. A larger learning space for the input patterns is offered by allocating the full reservoir to each cluster, while the separation of reservoirs prevents inputs from across different centroids. A readout layer follows each reservoir, which generates its output solely based on learned patterns of relevant cluster projection. In order to forecast the outcome, a second readout layer is trained based on the predictions made by each reservoir. Fig. 1 illustrates the high-level view of the proposed global ensemble echo state network approach, which are further delineated in the following subsections.

A. CLUSTERING LAYER

In contrast to typical echo-state networks that use random weights to map inputs to a number of reservoir nodes, the clustering layer groups the inputs into segments based on common properties. Given that clustering aims to reduce the workload on the reservoirs, the main selection criteria for a suitable algorithm in time-efficiency. Therefore, we use the k-means algorithm, which maintains a time complexity of $O(\text{NTK})$, where N is the total number of data sets, K is the total number of partitions, and T is the number of iterations. The elbow approach selects the optimal number of clusters from the range of [1, 20] while evaluating the quality of clustering using the sum of the squared distances between points in a cluster and the centroid.

The vector dot product between the input vectors and cluster centroids is used to project the input to the reservoir layers once the number of clusters k has been determined. Let $X = [x_1, x_2, x_3, \ldots, x_n]$ be an input dataset with n data points of m dimension. Let the set of cluster centroids be $C = [c_1, c_2, c_3, \ldots, c_k]$ then, the portion of x_i input received to each of the reservoir nodes in the jth layer would be $c_j \cdot x_i$, as follows:

$$c_j \cdot x_i = |c_j||x_i| \cos\theta = |c_j| \left(\text{Projection of } x_i \text{ on } c_j\right). \quad (1)$$

The input is transferred into the reservoir nodes (1), which projects the input on to the cluster centroid. Although this is compounded by an additional term called $|cj|$, it is constant for all the inputs in the reservoir and offsets searching for input patterns.

B. RESERVOIR NETWORK

Each cluster detected in the clustering layer is represented by a reservoir in the reservoir network. An r number of randomly connected recurrent nodes make up each reservoir. The reservoir acts as a higher dimensional space for inputs. The random recurrent connections sustain past inputs and aid in exploiting linear dependencies between outputs. In each training iteration, the states of nodes are updated. The proportion of input and proportion of already learned states of other nodes are transferred to the node state based on the leaking rate(α). The jth reservoir layer's current state($h_{t,j}$) is updated as follows:

$$h_{t,j} = (1-\alpha)h_{t-1,j} + \alpha \tanh\left(W_{\text{res}} \cdot h_{t-1,j} + c_j \cdot x_t\right). \quad (2)$$

The random weights that make up the W_{res}, $r \times r$-dimensional matrix in this case would determine the random connections between reservoir nodes. To generate a final state matrix H of dimension $k \times r \times t_{\text{train}}$, the reservoir layer is fed

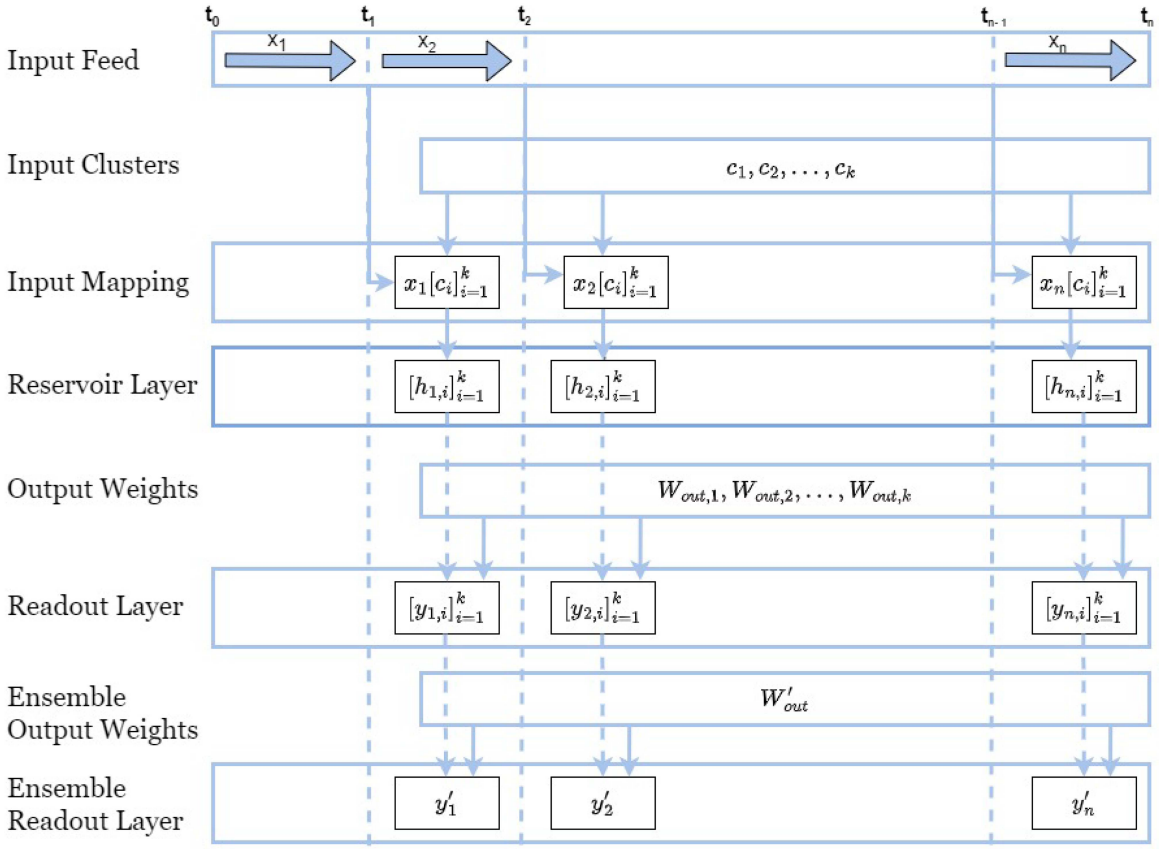

FIGURE 2. Functional view of the proposed global ensemble echo-state network.

with all of the training inputs from ($i = 1$ to t_{train})

$$H = \begin{bmatrix} h_{1,1} & h_{2,1} & \dots & h_{t_{\text{train}},1} \\ h_{1,2} & h_{2,2} & \dots & h_{t_{\text{train}},2} \\ \dots & & & \\ \dots & & & \\ h_{1,k} & h_{2,k} & \dots & h_{t_{\text{train}},k} \end{bmatrix}. \tag{3}$$

The state matrix in (3) represents the learned input patterns of input data in high-dimensional space. Furthermore, as is the case with traditional ESN models, this state matrix can be linearly related to the output values to complete the learning process.

C. ENSEMBLE READOUT LAYER

The output of each reservoir layer is generated as follows:

$$y_{t,j} = W_{\text{out},j} \cdot h_{t,j} \tag{4}$$

where $y_{t,j}$ is network output and $W_{\text{out},j}$ is the output weight matrix. The values of $W_{\text{out},j}$ for each reservoir layer are determined by applying ridge regression as follows:

$$W_{\text{out},j} = Y_{\text{target}} \cdot H(j)^T \left(H(j) \cdot H(j)^T + \beta I\right)^{-1}. \tag{5}$$

Here, $H(j)$ represents the states responsible for the jth layer and β is the regularization coefficient. The output generated by each reservoir layer is stacked to generate the input Y' to the ensemble output layer

$$Y' = \begin{bmatrix} y_{1,1} & y_{2,1} & \dots & y_{t_{\text{train}},1} \\ y_{1,2} & y_{2,2} & \dots & y_{t_{\text{train}},2} \\ \dots & & & \\ \dots & & & \\ y_{1,k} & y_{2,k} & \dots & y_{t_{\text{train}},k} \end{bmatrix}. \tag{6}$$

The combined final output is generated as follows:

$$y'_t = W'_{\text{out}} \cdot y_t \tag{7}$$

where $y'_{t,j}$ is the final network output and W'_{out} is the final output weight matrix. The values of W'_{out} are determined by applying ridge regression as follows:

$$W'_{\text{out}} = Y_{\text{target}} \cdot Y'^T \left(Y' \cdot Y'^T + \beta I\right)^{-1}. \tag{8}$$

At the end of the training, two weight matrices W_{out} and W'_{out} responsible for mapping the reservoir states into the final output are learned and locked for the prediction phase.

A functional view of the proposed global ensemble echo state network is presented in Fig. 2, which shows how the inputs propagate through the input mapping layer consisting of cluster centroids, stacked reservoir layer, and the two readout layers of that ensemble output from each reservoir to produce the final output.

D. EXPLAINABILITY LAYER

Traditional echo state network models often lack interpretability due to their complex and random connections between input nodes and reservoir neurons, limiting the ability to discern feature importance regarding predictions. The proposed ensemble ESN model addresses this limitation by incorporating a clustered input mapping layer and an ensemble output. The clusters generated represent the key feature groupings of the data. Input data are then projected onto these clusters and learned within allocated reservoir spaces, effectively isolating their impact on the final prediction. The ultimate prediction results from combining outputs from each reservoir, utilizing the trained weights during the final ensemble readout layer, as depicted in (8). Consequently, relevant feature importance metrics can be derived from the weight matrix. This allows for the calculation of the explainability value $E_{i,t}$ of the reservoir at timestamp t, as shown in (7). This equation helps to elucidate the interpretability of the model by associating specific feature importance with reservoir behavior at distinct time points

$$E_{i,t} = W'_{\text{out}}[:, i] \cdot y_t. \tag{9}$$

Here, $W'_{\text{out}}[:, i]$ is the ith column of the W'_{out} matrix which corresponds to contribution coefficient of the ith cluster.

IV. EXPERIMENTS AND RESULTS

The proposed global ensemble echo state network approach was empirically evaluated on the real-world, multicampus, mixed-use tertiary education setting of La Trobe University in Victoria, Australia. The La Trobe Energy AI Platform (LEAP), the core AI and data analytics technology stack in La Trobe University's "Net Zero Carbon Emissions Program", aims to reduce the University's carbon footprint to net zero emissions by 2029, alongside improved energy efficiency and increased resource utilization [54], [55]. Crowd forecasting is a critical factor in reducing energy costs and achieving this net zero emissions goal. The experiments were designed to evaluate the three main contributions, the novel ESN learning approach with a clustering layer, multiple reservoirs, and an ensemble readout layer for time-series forecasts, model explainability metrics, and the global ESN model that enhances the prediction accuracy through latent relationships across the raw data streams.

A. MODEL TRAINING AND HYPER PARAMETERS

The training dataset consisted of timestamp, WiFi sensor data streams, 1-h interval weather data, and 15-min interval energy consumption data collected, preprocessed, and stored in the central data lake of the LEAP. Weather data significantly influence people's presence at various times and days, making its integration crucial. We accessed weather data from the Australian Bureau of Meteorology (BOM), utilizing the closest weather station to our university network's multiple locations. The BOM records key parameters in 1-min granularity, from which we extracted temperature, humidity, wind speed, and wind direction information for our dataset. Date and time-related details, including holidays and academic calendar specifics such as working, nonworking, semester days, and exam periods, were also extracted. Each campus building is equipped with a separate smart meter, from which we extracted the energy consumption data. More comprehensive information about our data collection and storage mechanisms can be found in [55]. In addition to this, we utilized WiFi requests from routers situated across university locations to approximate crowd presence. Each WiFi request contained a user ID, timestamp, and router location, allowing us to estimate the number of WiFi connections within specific buildings at a specific time as a representative measure of the crowd or the number of individuals present. Before inputting the data into the model, we employed a min-max scalar for data normalization. Our model operates with a delayed input of 24 time steps, forecasting 24 h. Moreover, an overview of the dataset's metadata is detailed in Table 1, and the generated dataset from this study is now publicly available on GitHub [56].

TABLE 1 Overview of Meta Data

Data Field	Description	Unit
DateKey	Date in YYYYMMDD format	-
HourKey	Hour in HHMM format	-
Weekday	Day of the week	-
Month	Month of the year	-
Month	Month of the year	-
Holiday	Whether a holiday or not	-
ApparentTemperature	Apparent temperature	°C
AirTemperature	Air Temperature	°C
DewPointTemperature	Dew Point Temperature	°C
RelativeHumidity	Relative humidity of air	%
ElectricityConsumption	Electricity consumption	kWh
PeopleCount	Number of people	-

The evaluated ensemble echo state network comprised three reservoir network layers, based on the clustering layer's outcome with $k = 3$ clusters derived through the elbow method. Consequently, the model architecture featured three stacked reservoir layers, each containing 300 randomly connected neurons. This was followed by a readout layer comprising three output neurons, each randomly connected to the 300 neurons within its relevant reservoir. Finally, the ensemble readout layer comprised a final output neuron connecting to all three neurons from the readout layer. To determine the optimal hyperparameters for the proposed ensemble echo state network, a grid search methodology was employed. The model parameters utilized during training were as follows: A leaking rate of 0.56, spectral radius of 0.61, input scaling of 0.6, and a regularization coefficient of $1 \times 1e^{-5}$ for ridge regression. The leaking rate plays a critical role in governing the extent of information decay or retention within reservoir neurons across time steps. It dictates how much of the prior activation state persists or carries forward when neurons receive new input or process information. On the other hand, the spectral radius serves as an amplification factor for signals propagating through the reservoir. A spectral radius greater than 1 tends to amplify signals, potentially inducing chaotic

TABLE 2 Results of Experiment 1: RMSE and MAE for Local Models and Global Models

Building	RMSE-Local Model	RMSE(Peak)-Local Model	RMSE-Global Model	RMSE(Peak)-Global Model	MAE-Local Model	MAE(Peak)-Local Model	MAE-Global Model	MAE(Peak)-Global Model
Library	6.46	9.79	**5.80**	8.65	3.98	6.91	**3.96**	7.90
Lecture Halls	1.78	2.56	**1.66**	2.54	1.03	1.90	**1.01**	1.85
Sport Facilities	**1.29**	1.57	1.31	1.67	**0.78**	1.12	0.92	1.26
Administration	1.60	1.18	**0.67**	1.11	0.45	1.16	**0.32**	0.86
Accommodation	0.41	0.94	0.41	0.98	**0.21**	0.66	0.26	0.81
Mixed-Use	3.33	4.87	**3.14**	3.93	1.63	2.91	**1.19**	2.09

The bold values are the lowest by metric, by category.

behavior. Conversely, when the spectral radius is less than 1, signals tend to decay within the network. The input scaling parameter, involves weighting the input connections entering the reservoir layer of the ESN. This parameter holds significance in maintaining a balance between the importance of incoming input signals and preserving the intrinsic dynamics of the network, ensuring that the inputs do not overpower the network's inherent behavior.

The same hyperparameters were used in evaluation where the metrics were root mean square error (RMSE) and mean absolute error (MAE). For benchmark comparisons, we employed an LSTM model, a BiLSTM model, and a GRU model, each configured with 50 hidden units. In addition, an ESN model with identical parameters to the global ensemble ESN model was included for comparative evaluation.

B. EXPERIMENT 1 - CROWD FORECASTING

The proposed method was used to generate crowd forecasts for six building types: 1) the library, 2) lecture halls, 3) sport facilities, 4) administration-only building, 4) student accommodation, and 5) a mixed-use building. We used a "3:1" split of training and testing data, to build local models per building and a global model trained all buildings simultaneously. The results are shown in Table 2.

For most of the buildings, the global model outperformed the local models. The prediction of the global model for the library which in general has the highest crowd density is notably improved with a 30% increase in RMSE value and a 25% improvement in MAE value compared to the local model. The global model's forecast on the lecture halls, mixed-use, and administration buildings were other noticeable improvements from the local models. However, the local models outperformed the global model when trained on data from the sports facilities and the accommodation building. This is due to inverse use of these buildings during off-peak and weekends, compared to the others which are associated with work-hours and workdays.

The crowd counts of three separate buildings for the chosen time period were projected by the global model and the local models, as shown in Fig. 3. The busiest days for all the buildings represented in the model are Monday, Tuesday, and Wednesday. Furthermore, Thursday and Friday are less congested due to less classes. A peak can be seen in the library and teaching building between 9–10 A.M., when work and study begins. As expected, the library and lecture halls have limited to no activity over the weekend, whereas the sports facility sustains the same amount of activity on weekends as weekdays.

To evaluate the model's performance on peak values, separate forecasts specifically focused on these instances were generated. Peak values were identified by employing a moving window of five values centered around each data point and determining whether the data point was the maximum within that window. Forecasts were then generated using both local and global models specifically for these peak instances. The resulting error metrics are summarized in Table 2. Both the global and local models exhibited relatively minor increases in error when predicting peak values for the library. Specifically, the RMSE for peak value predictions in the library was 8.65 and 9.79, representing only a marginal increase of 2.85 and 3.33 compared to off-peak predictions, respectively. Across other buildings, the increase in error for peak value predictions compared to off-peak predictions remained within the range of 0.28 to 1.54. This indicates a notably small deviation in error metrics for peak values compared to off-peak values. However, it is important to note that the relatively higher error observed in the library building is likely attributed to the presence of a high number of people.

Another experiment was conducted to assess the effectiveness of clustered input mappings in the proposed ESN model, comparing its performance with and without these mappings. The results, outlined in Table 3, showcase the RMSE and MAE values for each building. Interestingly, for most buildings, employing clustered input mappings yielded improved results. Notably, the model without cluster mappings performed relatively better in the sports centre. This anomaly is attributed to the irregular pattern of crowd movement and behaviors in the sports centre, leading to less effective clusters.

Subsequently, another experiment compared the performance of our proposed model against benchmark time-series forecasting models: LSTM, BiLSTM, GRU, and traditional ESN. Each model underwent training with the same dataset encompassing six buildings utilized in prior experiments. The resulting RMSE and MAE values for the test dataset are presented in Table 4. In most cases, our model outperformed the benchmark models significantly. One key advantage of our approach lies in training with a smaller amount of data due

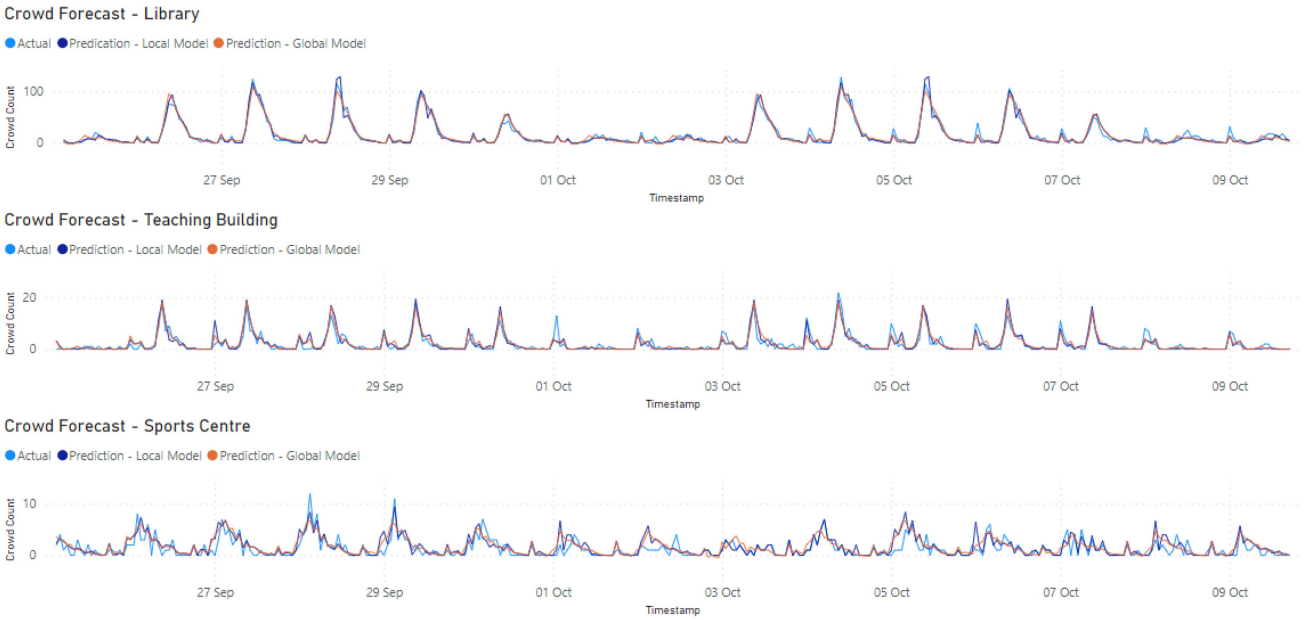

FIGURE 3. Crowd forecast of the global ensemble echo state network and local models for the library, teaching building, and sports centre.

TABLE 3 Results of Experiment 1: RMSE and MAE for ESN With and Without Cluster Mappings

Building	RMSE-with cluster mappings	RMSE-without cluster mappings	MAE-with cluster mappings	MAE-without cluster mappings
Library	**9.72**	9.88	6.53	**6.16**
Lecture Halls	**2.88**	3.16	**1.89**	2.09
Sport Facilities	2.59	**2.51**	1.61	**1.53**
Administration	**1.44**	1.43	**0.80**	0.83
Accommodation	**0.52**	0.54	**0.37**	0.35
Mixed-Use	**4.67**	5.36	**3.29**	3.38

The bold values are the lowest by metric, by category.

TABLE 4 Results of Experiment 1: RMSE and MAE for Ensemble ESN With Benchmark Models

Building	LSTM		Bi-LSTM		GRU		ESN		Ensemble ESN(Ours)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Library	13.82	8.69	13.80	8.77	14.73	8.52	9.88	**6.16**	**9.75**	6.62
Lecture Halls	2.66	1.55	2.64	1.60	2.68	1.49	3.16	2.09	**2.29**	**1.43**
Sport Facilities	2.67	1.69	2.66	1.68	2.68	1.78	2.51	1.53	**2.44**	**1.52**
Administration	1.49	0.79	1.46	0.80	1.50	**0.78**	1.43	0.83	**1.33**	**0.78**
Accommodation	0.57	0.39	0.57	0.35	0.58	0.35	**0.54**	0.35	**0.54**	**0.34**
Mixed-Use	4.62	3.03	4.50	2.89	4.71	3.05	5.36	3.38	**3.82**	**2.43**

The bold values are the lowest by metric, by category.

to fewer parameters, leveraging fixed weights in input mapping and reservoir. Conversely, RNN models, such as LSTM, BiLSTM, and GRU tend to struggle due to their higher parameter counts, resulting in comparatively lesser performance. Even when compared against the traditional ESN model, our ensemble model demonstrated superior performance, underscoring the efficacy of the clustering layer and the ensemble approach.

C. EXPERIMENT 2 - FORECAST ROBUSTNESS

The purpose of this experiment was to evaluate the robustness of the forecasting models generated by the proposed global ensemble echo state network. For both local and global models, we investigate the number of training data points needed for acceptable performance as measured by RMSE and MAE. As depicted by the results in Table 5, the prediction accuracy increases with the number of the training points. Similar to the

TABLE 5 Results of Global Model vs Local Model for Experiment 2

	Training Datapoints	RMSE-Local Model	RMSE-Global Model	MAE-Local Model	MAE-Global Model
Library	100	27.87	**22.10**	22.56	**11.87**
	300	18.26	**16.49**	13.56	**9.49**
	500	20.30	**12.46**	13.72	**8.03**
	700	15.60	**11.79**	10.01	**7.30**
Lecture halls	100	5.25	**2.71**	4.31	**1.64**
	300	3.88	**2.75**	3.05	**1.64**
	500	3.85	**2.76**	2.66	**1.67**
	700	2.91	**2.76**	2.13	**1.67**
Sports Facilities	100	3.40	**2.74**	2.03	**1.62**
	300	**2.61**	2.77	1.91	**1.64**
	500	**2.41**	2.80	**1.56**	1.65
	700	**2.38**	2.79	**1.42**	1.64
Mixed-Use	100	8.42	**3.98**	7.50	**2.69**
	300	7.67	**3.97**	6.70	**2.60**
	500	5.31	**4.03**	3.76	**2.51**
	700	5.18	**3.99**	3.55	**2.55**
Administration	100	1.64	**1.59**	**0.79**	0.93
	300	1.61	**1.57**	0.95	**0.84**
	500	1.59	**1.44**	0.93	**0.87**
	700	1.63	**1.45**	0.97	**0.88**
Accommodation	100	0.76	**0.69**	**0.46**	0.58
	300	**0.66**	0.76	0.57	**0.46**
	500	**0.52**	0.84	**0.33**	0.52
	700	**0.55**	0.84	**0.40**	0.52

The bold values are the lowest by metric, by category.

previous experiment, the global model's ability to adapt to the data from the sports facilities and accommodation was inferior to that of the local model. However, the global model accurately predicts the teaching and mixed-use buildings with just 100 data points, in contrast to the local models that required training on 700 data points. This further validates that the global model is robust because it learns the latent relationships between similar time-series data streams.

D. EXPERIMENT 3 - FORECAST EXPLAINABILILTY

For the forecast model explainability experiment, clusters from the initial layer were examined for the main input variables and values that each cluster represented. This depiction of input variables and values represented by the discovered clusters is shown in Fig. 4.

Fig. 4(a), (b), and (d) depict that no clusters are defined based on the values of air temperature, relative humidity, and crowd count. However, cluster-2(orange) depicts the morning hours while cluster-1(dark blue) represents the afternoon hours. Plot (c) further indicates that cluster-0(light blue) is created for weekends.

After labeling the clusters, we extracted the partial contributions from each reservoir network in order to better understand how each cluster's respective reservoir network influenced the result. For the Library building, the explainability values for each prediction were extracted using (9) and plotted alongside the values of the input attributes that dominantly represent the clusters (see Fig. 5).

It is noted that the afternoon cluster (cluster-1) has a significant contribution for predictions made in the afternoon, whereas the morning cluster has the opposing pattern (cluster-2). In addition, the weekend cluster (cluster-0) has a higher impact on weekend predictions. We also proposed an evaluation metric called contribution quality (CQ_i) to determine the impact of each cluster, as follows:

$$CQ_0 = \frac{n_2\left(\sum^{n}_{\substack{i=0\\5\le day\le 6}} CF\right)}{n_1\left(\sum^{n}_{\substack{i=0\\0\le day\le 4}} CF\right)} \tag{10}$$

$$CQ_1 = \frac{n_2\left(\sum^{n}_{\substack{i=0\\0\le hour\le 10}} CF\right)}{n_1\left(\sum^{n}_{\substack{i=0\\11\le hour\le 23}} CF\right)} \tag{11}$$

$$CQ_2 = \frac{n_2\left(\sum^{n}_{\substack{i=0\\11\le hour\le 23}} CF\right)}{n_1\left(\sum^{n}_{\substack{i=0\\0\le hour\le 10}} CF\right)}. \tag{12}$$

Here, the n_1 represents the number of data points captured under the attribute of interest used to define the cluster and n_2 is the number of remaining data points. As an example, for CQ_0, n_1 is the number of data points in which the day attribute was 5 or 6 and n_2 is the number of data points where day values were equal or between 0 and 4. CQ_1 is defined for the morning cluster, hence the attribute of interest, hour value is kept between 0 and 10. The opposite is true for the afternoon cluster, hence the hour is kept between 11 and 23 to calculate CQ_2. The metric "Contribution Quality" refers to an assessment of how well a reservoir network assigned to

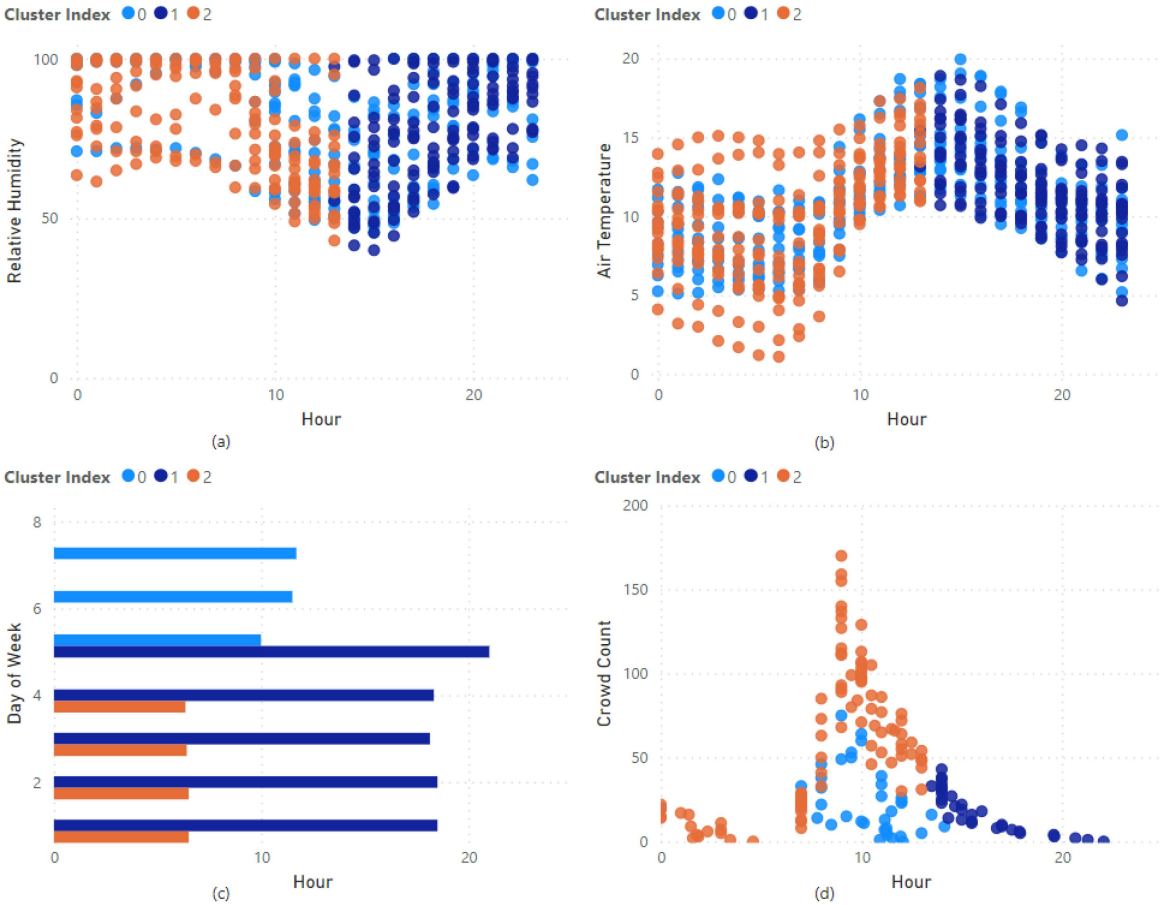

FIGURE 4. **Input variable by hour, with colour depicting cluster ID. (a) Relative humidity by hour. (b) Air temperature by hour. (c) Day of the week by hour. (d) Crowd count by hour.**

TABLE 6 **Results of the Contribution Quality (CQ) Metric**

Building	CQ_0	CQ_1	CQ_2
Library	1.08	1.13	1.01
Lecture Halls	0.96	1.21	1.67
Sports Facilities	1.11	0.88	1.10
Mixed-Use	1.77	1.09	0.87
Administration	1.01	1.18	1.03
Accommodation	1.07	1.12	0.79

a cluster has contributed to the predictions when the input attributes are within the cluster's defined parameter values versus when they are not. Since the metric is defined as a ratio of contributions, a value more than 1 denotes that the cluster's reservoir network has performed better within the boundaries of the defined input attribute values. Accordingly, it can be inferred that the model's explainability is comparatively high and that an accurate assessment of the model's decision-making process can be made using the explainability values that were taken from the model.

Table 6 shows the calculated contribution quality values for the data from each building. The results of Table 6 show that the contribution quality values for all the buildings are greater than 1 in most cases. Higher the value, the higher the confidence in explainability. As an example, CQ_0 of mixed-use building value, 1.77 is the highest among other buildings. This suggests that the predictions on weekends (related to cluster-0) are more explainable for the mixed-use building than the other buildings. Collectively, these results demonstrate the explainability capability of the proposed method as well as that different reservoir networks have based their learning on the initial clustering method when searching input patterns for attribute values.

For a quantitative analysis of the explainability of the proposed ensemble ESN model, two additional experiments were conducted. The first experiment assessed the fidelity [57] of the model, which evaluates how well the explanation approximates the prediction of the opaque model. To quantify fidelity, we compared our explainable model against a traditional ESN model in a forecasting experiment, recording the RMSE/MAE values. The results, detailed in Table 7, showcased relatively higher forecasting MAE/RMSE values for each building in the explainable ensemble ESN model compared to the traditional ESN model without the explainable capability. Notably, this suggests that even with explainable capabilities, our model

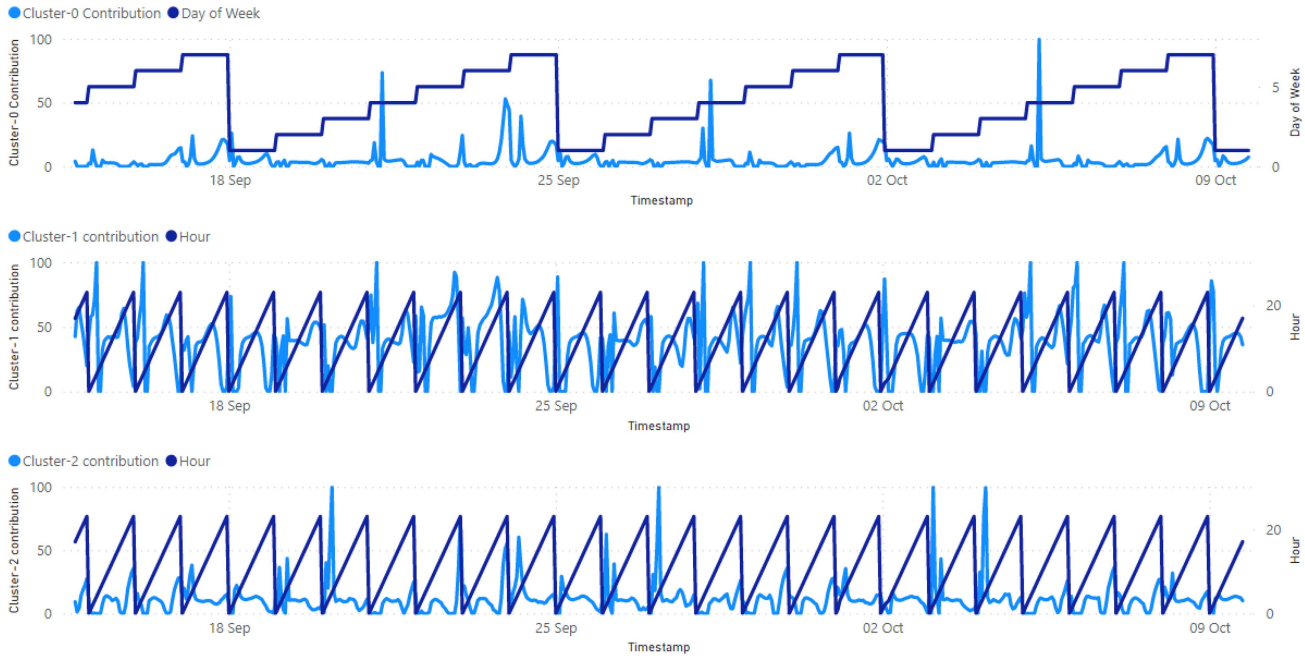

FIGURE 5. For the library building, explainability values related to each cluster along with the primary contributing input.

TABLE 7 Results of Experiment 3: Fidelity Measures for Explainable Ensemble ESN Model Against Traditional ESN

Building	RMSE-Ensemble ESN	RMSE-Traditional ESN	MAE-Ensemble ESN	MAE-Traditional ESN
Library	**9.75**	9.88	6.62	**6.16**
Lecture Halls	**2.29**	3.16	**1.43**	2.09
Sport Facilities	**2.44**	2.51	**1.52**	1.53
Administration	**1.33**	1.43	**0.78**	0.83
Accommodation	0.54	0.54	**0.34**	0.35
Mixed-Use	**3.82**	5.36	**2.43**	3.38

The bold values are the lowest by metric, by category.

TABLE 8 Results of Experiment 3: Stability Measures for Explainable Ensemble ESN Model

Building	Percentage Agreement (%)
Library	97.49
Lecture Halls	90.01
Sport Facilities	98.86
Administration	77.20
Accommodation	99.01
Mixed-Use	93.31

outperforms the opaque model (traditional ESN), indicating a high fidelity level.

The second experiment aimed at evaluating the stability [57] of the explainable approach. Stability refers to a model's ability to consistently produce same explanations despite slight alterations or perturbations in the input. To quantify stability, small noise values in the range of 0–5 were introduced to the crowd count variable of each data point in the test dataset used for the forecasting experiment. These experiments were conducted five times for each building, and the percentage agreement across each experiment in providing the most explainable feature for the output of each data point was recorded. The percentage agreement values, presented in Table 8, indicate high stability, suggesting that the model is robust against noise values and consistently provides the same explanation for the prediction. Except for the administration building, all other buildings demonstrated a percentage agreement over 90%, indicating the high stability of our explainable model. The inconsistent pattern of crowd movement in the administration building resulted in less meaningful clusters, lowering the explainable capabilities and justifying the lower stability observed in this specific case.

V. CONCLUSION

In this article, we present the design, development, and evaluation of the global ensemble echo state network for explainable crowd forecasting using WiFi data streams. This is a novel ESN learning approach with a clustering layer that replaces the input mappings, multiple reservoirs allocated to each cluster and an ensemble readout layer that completes the time-series forecasts. It also consists of a model explainability

feature for interpreting the forecast in terms of clusters and corresponding input variables and an explainability metric that evaluates model performance quality. The improved accuracy and robustness of the global model that is trained on multiple time series is a further contribution of the proposed approach.

The proposed method was empirically evaluated on the multicampus, mixed-use tertiary education setting of La Trobe University, in the context of the LEAP, for reducing the University's carbon footprint to net zero emissions by 2029. Crowd forecasting is a primary function in energy conservation and optimized energy usage opportunities. The three experiments reported in this article evaluated the performance accuracy of the forecasts of both local and global models, the robustness of the proposed method in situations where the available data are limited and the forecast explainability using the layer approach of clusters and input variables contributing to the forecast generation.

When comparing the performance over crowd forecasting between the global ensemble model and the local model across buildings, the local model demonstrated an RMSE of 2.47, while the global model exhibited an RMSE of 2.16, indicating a 12% reduction in error. At peak values, these figures were 3.49 and 3.15 for local and global models, respectively, still showing a 10% improvement in the global model. Conducting experiments with and without cluster mappings in the global model revealed a 4.5% decrease in error with cluster mappings, underscoring the importance of the cluster layer. Comparing the average performance across buildings with benchmark models, such as LSTM, Bidirectional-LSTM, GRU, and a traditional ESN, our model outperformed them by 22%, 21%, 25%, and 12%, respectively, based on RMSE. Further experiments were carried out to evaluate the robustness of the global ensemble echo state model concerning adaptability with varying numbers of data points. The comparison between the global and local models as the number of data points increased from 100 to 700 demonstrated that the global model could predict with fewer data points, showing an average improvement of 28.5% in error (RMSE of 7.89 in the local model; RMSE of 5.64 in the global model) across buildings with just 100 data points compared to a 20.1% improvement in error (RMSE of 4.71 in the local model; RMSE of 3.72 in the global model) when utilizing 700 data points. Further experiments were conducted to assess the explainability of the proposed global ensemble ESN model. The fidelity metrics based on RMSE across buildings for the proposed model averaged 3.36, compared to an RMSE of 3.81 in the traditional ESN model, indicating a 12% improvement. Stability measures based on percentage agreement demonstrated a 92.65% agreement across buildings.

The results reveal that the proposed method achieves high performance for the crowd forecasting task and that the global model improves forecast accuracy compared to models that are separately trained as it takes advantage of the latent relationships between the data streams. The global model is also robust in comparison to local models. For explainability, the results reveal how the proposed method is able to distinguish contributing clusters and attributes, as well as a new metric that assesses the model explainability with a qualitative explanation of the values.

The main technical limitation of the proposed method is in the lack of network optimization strategies due to the layered approach. The effectiveness of the clustering algorithm influences model's explainability. The inherent limitations of echo state networks persist due to the random initialization of reservoir weights, which are crucial for capturing complex temporal dynamics. Consequently, careful selection of hyperparameters and autotuning is necessary. In future work, we aim to address these inherent limitations of echo state networks and enhance the optimization and autotuning procedures to improve model training. Effective reservoir design poses a significant challenge in echo state networks. While our current approach involves stacked layers of reservoirs, there is potential to explore alternative configurations with a multitude of reservoirs connected in various ways. Thus, future work will involve experimenting with reservoir design to discover more optimal solutions for explainable echo state networks. We also plan to extend beyond the crowd forecasting use case to more complex scenarios, such as crowd tracking, hotspot identification, and movement predictions. These applications predominantly involve analyzing crowd movements over time, leveraging the temporal dependencies learned in the reservoir nodes of echo state networks.

REFERENCES

[1] U. Singh, J. Determe, F. Horlin, and P. D. Doncker, "Crowd monitoring: State-of-the-art and future directions," *IETE Tech. Rev.*, vol. 38, pp. 578–594, 2021.

[2] A. Sharma et al., "Global mass gathering events and deaths due to crowd surge, stampedes, crush and physical injuries-lessons from the Seoul halloween and other disasters," *Travel Med. Infect. Dis.*, vol. 52, 2023, Art. no. 102524.

[3] M. Yamin, "Managing crowds with technology: Cases of Hajj and Kumbh Mela," *Int. J. Inf. Technol.*, vol. 11, pp. 229–237, 2019.

[4] K. Ngai, F. Burkle, A. Hsu, and E. Hsu, "Human stampedes: A systematic review of historical and peer-reviewed sources," *Disaster Med. Public Health Preparedness*, vol. 3, pp. 191–195, 2009.

[5] S. Ruan et al., "Dynamic public resource allocation based on human mobility prediction," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–22, 2020.

[6] M. Versichele, T. Neutens, M. Delafontaine, and N. Weghe, "The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent festivities," *Appl. Geography*, vol. 32, pp. 208–220, 2012.

[7] A. Kurkcu and K. Ozbay, "Estimating pedestrian densities, wait times, and flows with Wi-Fi and bluetooth sensors," *Transp. Res. Rec.*, vol. 2644, pp. 72–82, 2017.

[8] M. Bessho and K. Sakamura, "Sensing street-level crowd density by observing public bluetooth low energy advertisements from contact tracing applications," in *Proc. IEEE Int. Smart Cities Conf.*, 2021, pp. 1–7.

[9] H. Choi, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-CaL: WiFi sensing and machine learning based device-free crowd counting and localization," *IEEE Access*, vol. 10, pp. 24395–24410, 2022.

[10] Y. Ding, W. Chen, S. Wei, and F. Yang, "An occupancy prediction model for campus buildings based on the diversity of occupancy patterns," *Sustain. Cities Soc.*, vol. 64, 2021, Art. no. 102533.

[11] H. Zou, Y. Zhou, J. Yang, and C. Spanos, "Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT," *Energy Buildings*, vol. 174, pp. 309–322, 2018.

[12] Z. Liu, R. Yuan, Y. Yuan, Y. Yang, and X. Guan, "A sensor-free crowd counting framework for indoor environments based on channel state information," *IEEE Sensors J.*, vol. 22, no. 6, pp. 6062–6071, Mar. 2022.

[13] R. Nawaratne, D. Alahakoon, D. D. Silva, H. Kumara, and X. Yu, "Hierarchical two-stream growing self-organizing maps with transience for human activity recognition," *IEEE Trans. Ind. Inform.*, vol. 16, no. 12, pp. 7756–7764, Dec. 2022.

[14] R. Nawaratne, T. Bandaragoda, A. Adikari, D. Alahakoon, D. D. Silva, and X. Yu, "Incremental knowledge acquisition and self-learning for autonomous video surveillance," in *Proc. IECON -43rd Annu. Conf. IEEE Ind. Electron. Soc.*, 2017, pp. 4790–4795.

[15] D. Nallaperuma, D. D. Silva, D. Alahakoon, and X. Yu, "Intelligent detection of driver behavior changes for effective coordination between autonomous and human driven vehicles," in *Proc. IECON-44th Annu. Conf. IEEE Ind. Electron. Soc.*, 2018, pp. 3120–3125.

[16] S. Zhang, B. Deng, and D. Yang, "CrowdTelescope: Wi-Fi-positioning-based multi-grained spatiotemporal crowd flow prediction for smart campus," *CCF Trans. Pervasive Comput. Interaction*, vol. 31, pp. 31–44, 2022.

[17] U. Singh, J. Determe, F. Horlin, and P. D. Doncker, "Crowd forecasting based on WiFi sensors and LSTM neural networks," *IEEE Trans.Instrum. Meas.*, vol. 69, no. 9, pp. 6121–6131, Sep. 2020.

[18] R. Shumway and D. Stoffer, *Time Series Analysis and its Applications*. Berlin, Germany: Springer, 2000.

[19] V. Erickson, M. Carreira-Perpiñán, and A. Cerpa, "OBSERVE: Occupancy-based system for efficient reduction of HVAC energy," in *Proc. ACM/IEEE 10th Int. Conf. Inf. Process. Sensor Netw.*, 2011, pp. 258–269.

[20] S. Ryu and H. Moon, "Development of an occupancy prediction model using indoor environmental data based on machine learning techniques," *Building Environ.*, vol. 107, pp. 1–9, 2016.

[21] Y. Zhang et al., "Sequential click prediction for sponsored search with recurrent neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2014, pp. 1369–1375.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997.

[23] H. Jaeger, "Echo state network," *Scholarpedia*, vol. 2, 2007, Art. no. 2330.

[24] D. Kleyko, E. Osipov, D. D. Silva, U. Wiklund, and D. Alahakoon, "Integer self-organizing maps for digital hardware," in *Proc. Int. Joint Conf. Neural Netw.*, 2019, pp. 1–8.

[25] S. Leroux and P. Simoens, "Sparse random neural networks for online anomaly detection on sensor nodes," *Future Gener. Comput. Syst.*, vol. 144, pp. 327–343, 2022.

[26] B. Anzengruber, D. Pianini, J. Nieminen, and A. Ferscha, "Predicting social density in mass events to prevent crowd disasters," in *Proc. Social Inform.: 5th Int. Conf.*, 2013, pp. 206–215.

[27] M. Xu, M. Han, T. Qiu, and H. Lin, "Hybrid regularized echo state network for multivariate chaotic time series prediction," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2305–2315, Jun. 2019.

[28] R. Dwivedi et al., "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surveys*, vol. 55, no. 9, pp. 1–33, 2023.

[29] M. E. Eddin et al., "Fine-tuned RNN-based detector for electricity theft attacks in smart grid generation domain," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 733–750, 2022.

[30] A. Keliris, "Automated formulation of attack vectors for industrial control systems security assessment," Ph.D. dissertation, New York University Tandon School of Engineering, Brooklyn, NY, USA, 2019.

[31] N. Magaia, R. Fonseca, K. Muhammad, A. H. F. N. Segundo, A. V. L. Neto, and V. H. C. d. Albuquerque, "Industrial Internet-of-Things security enhanced with deep learning approaches for smart cities," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6393–6405, Apr. 2021.

[32] Y. Liu, Z. Liang, and X. Li, "Enhancing short-term power load forecasting for industrial and commercial buildings: A hybrid approach using time GAN, CNN, and LSTM," *IEEE Open J. Ind. Electron. Soc.*, vol. 4, pp. 451–462, 2023.

[33] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J.-S. Oh, "Semisupervised deep reinforcement learning in support of IoT and smart city services," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 624–635, Apr. 2018.

[34] S. Ameer et al., "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, pp. 128325–128338, 2019.

[35] P. Bellavista, P. Chatzimisios, L. Foschini, M. Paradisioti, and D. Scotece, "A support infrastructure for machine learning at the edge in smart city surveillance," in *Proc. IEEE Symp. Comput. Commun.*, 2019, pp. 1189–1194.

[36] I. Tan, O. Yaik, and O. Sheng, "Predicting shopper volume using ARIMA on public Wi-Fi signals," *Int. Inf. Inst.(Tokyo). Inf.*, vol. 19, 2016, Art. no. 3295.

[37] Y. Wu, S. Gu, T. Yu, and X. Xu, "APS-PBW: The analysis and prediction system of customer flow data based on WiFi probes," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.*, 2018, pp. 477–488.

[38] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Assoc. Comput. Linguistics*, 2014. [Online]. Available: https://aclanthology.org/D14-1179.pdf

[39] P. Yu, L. Miao, and G. Jia, "Clustered complex echo state networks for traffic forecasting with prior knowledge," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2011, pp. 1–5.

[40] O. Adeleke, "Echo-state networks for network traffic prediction," in *Proc. IEEE 10th Annu. Inf. Technol., Electron. Mobile Commun. Conf.*, 2019, pp. 0202–0206.

[41] D. D. Silva and D. Alahakoon, "An artificial intelligence life cycle: From conception to production," *Patterns*, vol. 3, no. 3, 2022, Art. no. 100489.

[42] J. D. Ser, I. Laña, M. Bilbao, and E. Vlahogianni, "Road traffic forecasting using stacking ensembles of echo state networks," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 2591–2597.

[43] Q. Zhang, H. Qian, Y. Chen, and D. Lei, "A short-term traffic forecasting model based on echo state network optimized by improved fruit fly optimization algorithm," *Neurocomputing*, vol. 416, pp. 117–124, 2020.

[44] R. Godahewa, K. Bandara, G. Webb, S. Smyl, and C. Bergmeir, "Ensembles of localised models for time series forecasting," *Knowl.-Based Syst.*, vol. 233, 2021, Art. no. 107518.

[45] P. Montero-Manso and R. Hyndman, "Principles and algorithms for forecasting groups of time series: Locality and globality," *Int. J. Forecasting*, vol. 37, pp. 1632–1653, 2021.

[46] P. Rathnayaka, H. Moraliyage, N. Mills, D. D. Silva, and A. Jennings, "Specialist vs generalist: A transformer architecture for global forecasting energy time series," in *Proc. IEEE 15th Int. Conf. Hum. Syst. Interaction*, 2022, pp. 1–5.

[47] C. Molnar, *Interpretable Machine Learning*. Morrisville, NC, USA: Lulu.com, 2020.

[48] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, pp. 1189–1232, 2001.

[49] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *J. Comput. Graphical Statist.*, vol. 24, pp. 44–65, 2015.

[50] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?" Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 1135–1144.

[51] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[52] L. Grigoryeva and J. Ortega, "Echo state networks are universal," *Neural Netw.*, vol. 108, pp. 495–508, 2018.

[53] P. Steiner, A. Jalalvand, and P. Birkholz, "Cluster-based input weight initialization for echo state networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7648–7659, Oct. 2023.

[54] N. Mills, P. Rathnayaka, H. Moraliyage, D. D. Silva, and A. Jennings, "Cloud edge architecture leveraging artificial intelligence and analytics for microgrid energy optimisation and net zero carbon emissions," in *Proc. IEEE 15th Int. Conf. Hum. System Interaction*, 2022 pp. 1–7.

[55] H. Moraliyage, N. Mills, P. Rathnayake, D. D. Silva, and A. Jennings, "UNICON: An open dataset of electricity, gas and water consumption in a large multi-campus university setting," in *Proc. IEEE 15th Int. Conf. Hum. System Interaction*, 2022 pp. 1–8.

[56] CDAC-lab, "CrowdForecastingData - github repository." [Online]. Available: https://github.com/CDAC-lab/CrowdForecastingData.git

[57] T. Fel, D. Vigouroux, R. Cadène, and T. Serre, "How good is your explanation? Algorithmic stability measures to assess the quality of explanations for deep neural networks," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, pp. 720–730.