Received 5 December 2022, accepted 20 December 2022, date of publication 26 December 2022, date of current version 26 April 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3232475

RESEARCH ARTICLE

Deep Federated Learning-Based Privacy-Preserving Wind Power Forecasting

AMIRHOSSEIN AHMADI[1], (Student Member, IEEE), MOHAMMAD TALAEI[2], MASOD SADIPOUR[3], ALI MORADI AMANI[4], (Member, IEEE), AND MAHDI JALILI[4], (Senior Member, IEEE)

[1]Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[2]Department of Energy Engineering, Sharif University of Technology, Tehran 11155-8639, Iran
[3]Department of Mechanical Engineering, University of Denver, Denver, CO 80208, USA
[4]School of Engineering, Royal Melbourne Institute of Technology, Melbourne, VIC 3001, Australia

Corresponding author: Amirhossein Ahmadi (amirhossein.ahmadi@ucalgary.ca)

ABSTRACT Given the growing installed capacity, wind energy will exert a profound impact on the flexibility of modern energy systems. Wind power forecasting is a practical solution for dealing with the attributed variations and uncertainties, balancing supply and demand, and improving the reliability of the system. To achieve more accurate and generalizable forecast models, comprehensive data sets, supplied by multiple wind farms owing to their spatio-temporal dependencies, are required. In addition, data aggregation/collaboration across many wind farms scattered around a country is difficult, if not impossible, due to complex administrative processes, industry competition, and data privacy and security concerns. This article offers federated learning-based wind energy forecasting as a novel decentralized collaborative modeling method capable of training a single model on data from many wind farms without jeopardizing the privacy or security of data. To this end, rather than sending private data across sites, local model parameters are securely transmitted. A comparison between the proposed private distributed model and non-private centralized and fully private localized models indicates the high performance of the proposed federated learning-based wind power forecasting with 87.96% accuracy. Enjoying the smoothing effect, the higher generalizability of the proposed model with 83.63% accuracy is also substantiated in comparison to localized and centralized approaches while the privacy of the underlying data is preserved.

INDEX TERMS Wind power forecasting, federated learning, deep learning, distributed collaborative learning, data privacy and security.

I. INTRODUCTION

Wind energy is now a prominent renewable energy source and an essential alternative energy solution for energy development due to rising energy demand, dwindling global resource reserves, and environmental protection concerns [1]. It is clear that the utilization of wind energy has increased dramatically in recent years, thereby exerting a profound impact on the flexibility of modern energy systems. According to [2], the total capacity of all erected wind turbines globally reached 837 GW by the end of 2021 indicated in Fig. 1. Despite all of the benefits associated with wind energy, various issues such as variability, internal instability, and uncertainty limit its high penetration into energy systems [3]. To overcome these difficulties and mitigate existing uncertainties, accurate forecasting of wind power has been offered as a dependable and low-cost solution [4].

The proposed forecasting techniques, based on methodology, are categorized into four broad groups: physical, statistical, intelligent, and hybrid models [5]. The physical models focus on numerical weather forecasting and use various meteorological data collected from observation systems to forecast wind speed. Although useful for long-term prediction horizons, physical models require additional factors, such

The associate editor coordinating the review of this manuscript and approving it for publication was Xianzhi Wang.

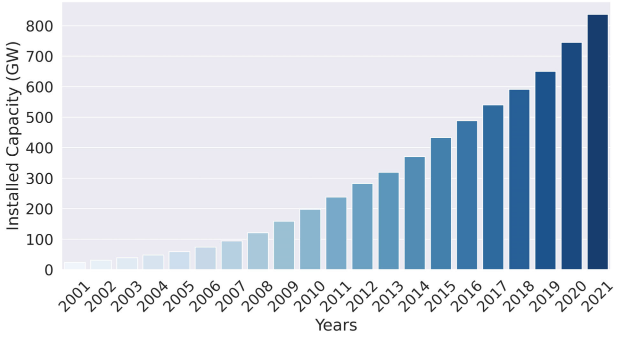

FIGURE 1. **Total installed wind power from 2005 to 2021 worldwide.**

as geographic and geomorphic conditions, temperature, and pressure. Additionally, these techniques necessitate the use of a lot of measuring sensors, which are not necessarily economical [6]. Statistical and intelligent models use past observations to extract time-varying relationships in time-series [7]. Various statistical models for wind speed forecasting have been introduced, including Kalman filter [8], Box. Jenkins models (AR, ARIMA models, etc.) [9], and Particle Swarm Optimization [10]. While statistical techniques perform well when estimating basic time series, they are incapable of handling nonlinear data and perform poorly when processing datasets with complex behavior. To address these issues, intelligent models are adapted because of their great capacity for learning volatility and nonlinearity. Data mining methods such as artificial neural networks (ANNs) [11], machine learning models [12], and deep learning [13] are used to create those intelligent models. These models have been widely utilized in recent years for a range of energy and power system applications and have consistently outperformed other models for wind forecasting applications [14].

It has long been observed that the combined (relative) variability of multiple wind generators (or solar generators) installed in a wider area is less than the variability experienced by a single system [15]. Additionally, intelligent models are susceptible to overfitting [16], limiting their capacity to generalize when deployed to new datasets. Uncorrelated locations represent a smoothing effect that can reduce variability associated with wind turbines, and therefore, improve the accuracy and generalizability of deterministic forecasts [17], [18]. Typically, suggested frameworks assume that all data records from smart meters are transmitted over broadband networks to a centralized computing infrastructure for model training. Nonetheless, this assumption creates privacy issues, since data profiles disclose a wealth of sensitive data, such as the connection of wind turbines and control centers, the wind farm network, and the turbine itself. Sending such sensitive data across networks exposes it to hostile interception and exploitation. Thus, the primary drawback of both conventional and intelligence methods used in previous forecasting models is the need for centralized data. The centralized data is very sensitive since it may readily be utilized to infer critical/private information or conduct cybersecurity attacks [19], [20]. For example, reference [21] examined the effect of data integrity attacks on the physical system of a wind farm. As such, companies are becoming more worried that their information is being utilized (or worse, misused) without their knowledge or consent. Under this landscape, collecting and exchanging data across various energy companies becomes more difficult, if not unfeasible, while the value of collaboration over data exchange is not immediately apparent.

The preservation of data privacy in centralized databases has been the subject of numerous studies in recent years. For example, methods for securing multi-client decision trees with vertically partitioned data were presented in [22] and [23]. Following their work, Vaidya and Clifton developed secure association mining methods [24], Naive Bayes classifier [25], and secure k-means [26]. Private Support Vector Machine methods have been developed for both vertically and horizontally partitioned data [27]. Secure methods for multi-group linear regression and classification were suggested in reference [28]. Using homomorphic encryption, the authors of [29] devised a privacy-preserving linear regression technique for horizontally partitioned data. Aono et al. [30] pioneered the use of homomorphic encryption to secure logistic regression. Shokri and Shmatikov [31] suggested training neural networks with updated parameters for horizontally partitioned data. With recent advancements in deep learning, privacy-preserving neural network inference has garnered considerable academic attention [32], [33].

Despite the efforts made in previous literature, the privacy issues associated with forecasting systems with multiple clients have remained a persistent challenge. To solve such a challenge while expanding the amount and diversity of data sets, the machine learning community has suggested Federated Learning (FL) [34]. FL is a decentralized collaborative approach to machine learning in which each device contributes to the training of a central model without providing any data. As shown in Fig. 2, the server initially initializes the model randomly or using publicly accessible data. The model is then sent to a randomly chosen group of devices (clients) for local training using their data. Each client updates the model's weights on the server, which are then averaged and utilized to update the global model. This procedure will be continued until the global model reaches a state of equilibrium. FL-based frameworks have been proposed for other applications such as traffic flow forecasting [35], load forecasting [36], renewable scenario generation [37], behind-the-meter solar generation disaggregation [38] and solar irradiation forecasting [39]. Given the importance of developing accurate yet generalizable forecasting algorithms based on data from multiple parties by maintaining data privacy, to our knowledge, there is no relevant study in the existing literature that addresses this issue explicitly in wind power generation applications.

Based on the above discussion, it can be seen that prior studies have relied mainly on the idea of providing a more

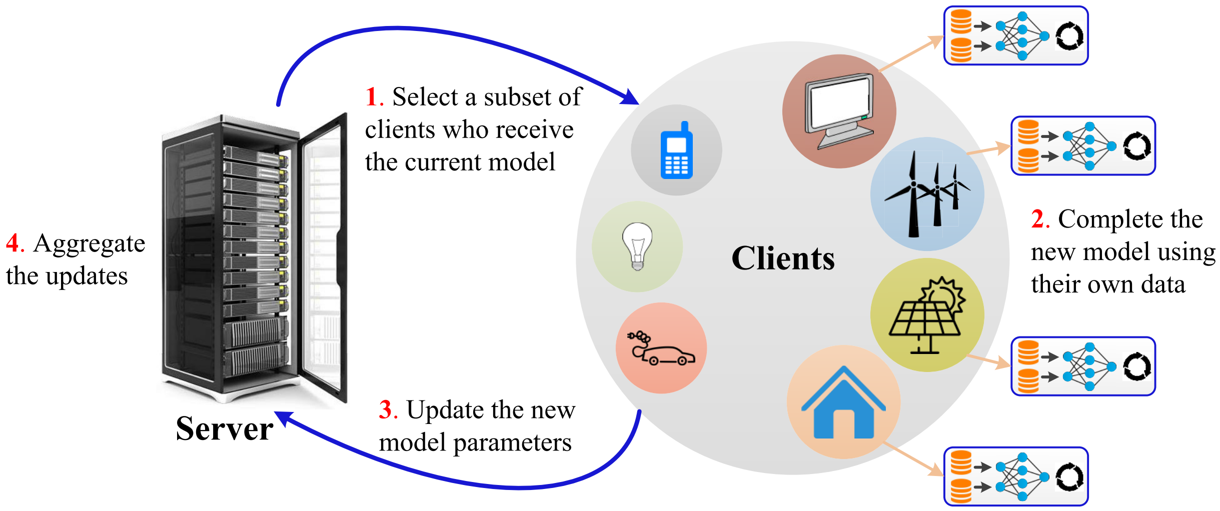
1. Select a subset of clients who receive the current model
2. Complete the new model using their own data
3. Update the new model parameters
4. Aggregate the updates
Server
Clients

FIGURE 2. Client-server iterative interactions within federated learning.

accurate model for forecasting by combining different models of machine learning and deep learning algorithms. Most suggested techniques have significant flaws since the little emphasis is placed on preserving the privacy of data associated with wind farms and meteorological stations. In addition, the proposed methods are only able to perform forecasting operations for specific areas where forecasting models are trained with data related to that area and cannot generalize forecasting for adjacent areas. Furthermore, combining machine learning algorithms with numerical weather predictions and terrain-specific conditions, while can increase the accuracy of the forecasting, adds to the complexity of the models, and requires much higher computational time. This paper proposes the use of FL to train a privacy-preserved wind power forecasting model. We use Long Short Term Memory (LSTM), a deep recurrent neural network for predicting time series, which makes use of historical measurements of the wind speed to anticipate future values of wind generation. Our contributions to this study are as follows:

- For the first time, an FL-based wind power forecasting scheme is proposed to offer a secure method of protecting data privacy by training the forecasting models locally and avoiding the exchange of raw data across various wind farms. The proposed scheme enables forecasting to benefit from the improved performance provided by global model aggregation in the absence of data exchange (Section II).
- A comparison between the proposed private distributed model, non-private centralized, and fully private localized models is conducted indicating the high accuracy of the proposed federated learning-based wind power forecasting using real-world datasets (Section III).
- Enjoying the smoothing effect, the higher generalizability of the proposed model is also substantiated while the privacy of the underlying data is preserved (Section III).

Lastly, Section IV concludes the paper and elaborates on some future research directions.

II. FEDERATED LEARNING

Modern energy systems have recognized the enormous potential of artificial intelligence as a result of the emergence and advancement of industry 4.0, and have begun to anticipate more complex, creative algorithms in a variety of applications, including forecasting. However, except for a few industries, others have only restricted and/or poor-quality data, thereby limiting the full potential of artificial intelligence. Data privacy and security, on the other hand, have recently become a global concern. Federated learning is a novel modeling technique that allows a single model to be trained on data from many sources without jeopardizing data privacy or security. It can unleash the full potential of artificial intelligence with promising applications where data is decentralized, typically unbalanced, and not identically distributed.

As previously stated, many factors contribute to the issue of large amounts of data required to train joint machine learning models. Thus, it is logical to explore methods for developing machine learning architectures that do not rely on accumulating all data in a single storage place for model training. A possible approach is to build a model at each location where a data source is situated, and then to allow those locations to communicate their unique models in order to reach a consensus on a global model. To ensure client data security and privacy, the communication mechanism is meticulously designed to prevent any site from interfering with the private data of another site. Simultaneously, the model is constructed as though the data sources were merged. Therefore, rather than sending data across sites, model parameters are securely transmitted, ensuring that third parties would not second guess the contents of another party's data.

FL aims to train a model on decentralized data $D_1, \ldots, D_m$ that is usually imbalanced and not identically distributed. A centralized approach is to assemble all data as $D = D_1 \cup \ldots \cup D_N$ to train a model (M_{cen}). However, to preserve data privacy, federated learning considers collaborative training of a model (M_{fed}), such that its accuracy (A_{fed}) satisfies

$$|A_{fed} - A_{cen}| < \delta \tag{1}$$

where δ is a non-negative real number and A_{cen} is the accuracy of M_{cen}. This equation conveys the intuition that the joint model resulted from performs roughly the same as when all data sources are combined. Because FL data providers would not disclose their data to a centralized server and other clients, we enable the FL system to perform somewhat worse than a joint model. This extra security and privacy assurance is worth much more than the loss in accuracy for many applications including wind power forecasting.

For wind power forecasting, we suggest the FL system utilize a central coordinating server, which is utilized to further create the joint model. The FL architecture may alternatively be built in a peer-to-peer fashion, eliminating the need for a coordinator; however, this will result in increased computational load. Fig. 3 illustrates the proposed FL coordinator scheme in a wind power forecasting system. The coordinator in this scenario is a central aggregation server (parameter server), which distributes an initial model to the local data owners 1–M (clients or participants). Each one of the data owners 1–M trains a local model with their own dataset and updates the model's weights through the aggregating server. The aggregation server then combines the model updates received from the clients (e.g., through federated averaging [40]) and sends them back. This procedure is repeated until the convergence criterion is satisfied or the maximum number of iterations is exceeded. Under this architecture, the original data of the individual providers never leaves the possession of the local data owners. This method not only protects user privacy and data security but also eliminates the communication cost associated with raw data transmission. To avoid data leakage, communication between the coordinating server and clients may be encrypted (e.g., utilizing homomorphic encryption [41]).

This is a horizontal federated learning technique in which several users with the same feature space but different samples train a model jointly on a server. Algorithm 1 details the step-by-step procedure proposed as the FL wind power forecasting. To begin, a small set of randomly chosen participants, referred to as a mini-batch, computes model parameters locally, encrypts them, and sends them to the server encrypted. The server then performs secure aggregation without jeopardizing any participant's privacy and returns to participants the aggregated parameters. Aggregation is a well-known approach that is based on stochastic gradient descent and is used in many different applications [42]. Finally, participants use the decrypted parameters to update their own models. This approach is repeated until the loss function converges, at which point the training phase is terminated.

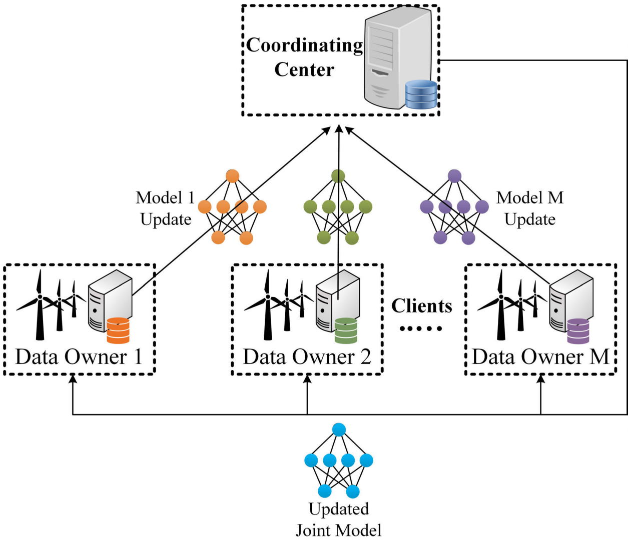

FIGURE 3. Client-server model of the proposed forecasting model.

Algorithm 1: Federated Averaging

```
Define minibatch size $B$, number of clients $m$ and epochs
  $E$, the rate of learning $\xi$ and global model $w_g$.

[Client $i$]
ClientTraining($i$,$w_g^t$):
for each epoch $j \in [1, E]$ do
    for batch $\kappa \in B$ do
        $w_i^t \leftarrow w_i^t - \xi \nabla \psi(w_i^t, \kappa)$
    end
end

[Server $i$]
Initialize $w_g^0$
for each round $t \in [1, T]$ do
    Select a random set $\mathcal{S}_t$ of $m$ clients from $\mathcal{N}$
    for each client $i \in \mathcal{S}_t$ parallelly do
        $w_i^{t+1} \leftarrow$ClientTraining($i$,$w_g^t$)
    end
    $w_g^t = \frac{1}{\sum_{i \in \mathcal{N}} D_i} \sum_{i=1}^{N} D_i w_i^t$    (Averaging aggregation)
end
```

A. LONG SHORT TERM MEMORY

As a major shortcoming in dealing with sequential data, traditional neural networks suffers from a lack of memory to reflect temporal dependencies. Recurrent neural networks (RNNs) are proposed to address this issue by allowing information to persist through a recursive network. As illustrated in Fig. 4, the network acts as a memory allowing information to be passed from one step of the network to the

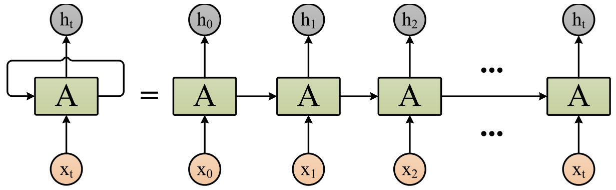

FIGURE 4. An unrolled recurrent neural network.

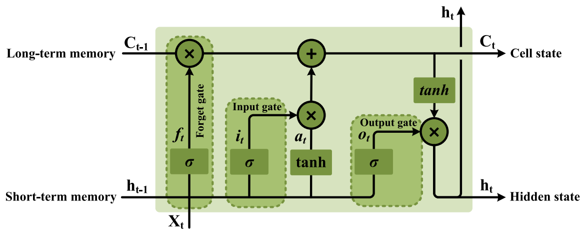

FIGURE 5. Principle architecture of the LSTM memory cell.

next. However, RNNs are not suitable for tackling long-term dependencies due to their short-term memory originating from the vanishing gradient problem, during which the gradient shrinks as it back propagates through time [43]. The LSTM networks are deep RNNs enabling learning long-term dependencies through cell state [44]. They are able to regulate the flow of information, i.e. remove or add information to the cell state, through internal mechanisms called gates. As depicted in Fig. 5, an LSTM consists of three consecutive gates including forget, input and output gates. First, the forget gate decides whether the information coming from the previous time stamp is to be remembered or is irrelevant and can be thrown away from the cell state (forget). Next, the input gate decides what new information should be added to the cell state. Finally, the output gate decides what parts of the updated cell state should be passed from the current timestamp to the next time stamp.

Each of these gates has unique computational relationships and functions, the process of calculating each variable at time t is shown as follows

$$f_t = \sigma\left(W_{lf} l_t + W_{mf} h_{t-1} + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_{li} l_t + W_{mi} h_{t-1} + b_i\right) \tag{3}$$

$$o_t = \sigma\left(W_{lo} l_t + W_{mo} h_{t-1} + b_o\right) \tag{4}$$

$$a_t = \tanh\left(W_{la} l_t + W_{ma} h_{t-1} + b_a\right) \tag{5}$$

$$c_t = c_{t-1} * f_t + i_t * a_t \tag{6}$$

$$m_t = o_t * \tanh c_t \tag{7}$$

where σ is the logistic sigmoid function, f_t, i_t, o_t, c_t, and a_t denotes forget gate, input gate, output gate, memory cell, and hidden vector respectively. $W_{l*} = \left(W_{lf}, W_{li}, W_{la}, W_{lo}\right)$ and $W_{m*} = \left(W_{mf}, W_{mi}, W_{ma}, W_{mo}\right)$ represents trainable weights of the respective gates while b_f, b_i, b_o, and b_a are output biases. Lastly, operator $*$ defines the Hadamard product.

B. PERFORMANCE METRICS

To assess the effectiveness of the proposed models, we use various performance indices with respect to accuracy. The following paragraphs introduce those performance indices.

1) MEAN ABSOLUTE ERROR (MAE)

MAE, which evaluates the mean absolute difference between predictions and observations, is expressed in (8) as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right|. \tag{8}$$

It is worth mentioning that because MAE does not have a differentiable function, most ML algorithms that use gradient descent have a hard time incorporating MAE as the evaluation metric. To compensate for this problem, other performance metrics should be considered.

2) ROOT MEAN SQUARE ERROR (RMSE)

RMSE, as expressed in (9), can consider the error's direction by measuring the root of the mean of the distance between predictions and observations.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{9}$$

To make the RMSE metric more sensible when it is used in RESs models, normalized RMSE (nRMSE) is often proposed, whose formula is depicted in (10).

$$\mathrm{nRMSE} = \frac{1}{P_{inst}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{10}$$

where P_{inst} is the installed capacity of the wind power plant, which is 1 MW in selected wind farms.

III. SIMULATION RESULTS

This section evaluates the FL forecasting method's performance using real-world datasets, and the findings are compared to centralized models operating under non-shared data situations. Additionally, the influence of the participation ratio on FL accuracy is examined using ten clients. The results indicate that the proposed FL scheme delivers competitive performance while ensuring data privacy.

A. DATA ANALYTICS

Geographically, Iran is located in a mountainous region with great potential for wind power generation. As illustrated in Table 1, nine different wind farms scattered around the country are considered here as: Abadan, Chabahar, Kahrizak, Khaf, Zahedan, Mahshahr, Neyshabur, Nikouyeh, Sonqor, and Tabriz. Datasets with a 10-min sampling measured at the height of 40 m were collected from these wind farms, whose statistical information is provided in Table 2. Moreover, Fig. 6 depicts the Weibull distribution of these wind farms for data measured at the height of 40 m. As can be seen, multiple wind farms with different profiles can provide representative data for the country due to their spatio-temporal dependencies.

TABLE 1. Selected wind farms located in Iran.

Location	Longitude	Latitude
Kahrizak	51°21’36.45"E	35°31’2.68"N
Tabriz	38° 7’ 86" N	46° 28’ 8" E
Sonqor	45° 1’ 04" N	70° 28’ 29" E
Abadan	30°20’21.12"N	48°18’15.48"E
Mahshar	49°13’ E	30° 36’ N
Chabahar	25° 29’ 27" N	60° 64’ 96" N
Zahedan	29° 1’ 28" N	60° 1’ 53" N
Khaf	35° 32’ 36.6"E	-75° 12’ 21.6"E
Neyshabour	45° 01’ 04"E	70° 28’ 29"N

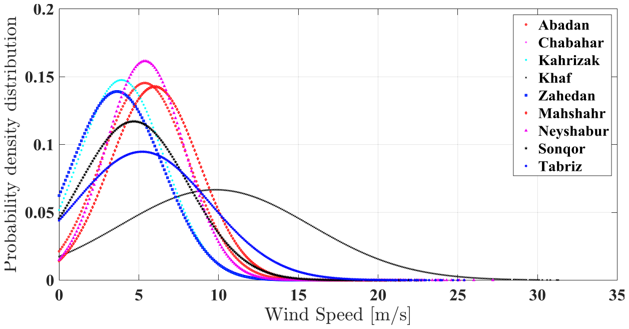

FIGURE 6. Weibull distribution for selected wind farms.

B. PREPROCESSING

Data preprocessing is a critical component of machine learning, as it prepares data for knowledge discovery by cleaning, integrating, reducing, transforming, and discretizing it. Data cleaning tries to fill in missing values, smooth out noise, discover and eliminate outliers, and resolve data inconsistencies. Data integration attempts to resolve issues such as entity identification, tuple duplication, data value conflict, and redundancy and correlation. By compressing data and lowering its dimensionality and numerosity, data reduction aims to produce a reduced representation of the data. Through data normalization, aggregation, and generalization, data transformation assists in the translation of data into a suitable format, whereas data discretization replaces raw data values with ranges. The preprocessing stage is described in this article as follows: Missing values are replaced with the median using the Simple-Imputer function, outliers are found and eliminated using the Z-score metric, duplicates are simply deleted, and data is normalized using the MinMaxScaler function.

C. CASE STUDY SCENARIOS

We offer comparisons against centralized learning, localized learning, and federated-based cases to evaluate the efficacy of using FL to wind power forecasting. Table 3 summarizes these various scenarios.

To begin, we created a centralized, non-distributed learning method that is most often used in situations when data privacy is not a significant issue during training. This case consolidates individual wind farm information and conducts training in a single place. Also, centralized training establishes a baseline for the capabilities of a single, collaborative forecasting model in a non-private environment. We train models for 35 epochs with early stopping depending on the lowest error obtained on the validation set.

The second scenario is an entirely private localized learning environment in which each dataset is trained independently, and the training process is isolated from every other wind farm. This technique results in forecasting models that are specifically customized to each wind farm and cannot profit from the information contained in other wind farms’ data. In accordance with the centralized learning method, training was performed for a maximum of 35 epochs. It is essential to emphasize that in localized situations, individual datasets are private and unobservable to other data owners.

Next, we offer an FL-based approach with the same objective as centralized learning: to train a single, joint model that generalizes well enough to give accurate predictions for all individual wind farms. However, FL provides advantages that exceed a localized learning environment as FL has higher generalizability. Unlike centralized learning, the training data from each wind farm is not pooled in the FL. The training data, on the other hand, is kept private by each local client.

The only way to determine how effectively a model generalizes to new situations is to test it on unseen data. In this regard, we hold out client data to assess the generalizability of the algorithms (centralized, localized, and federated) when they are exposed to completely new data. We use one client for testing the model and other clients for training the models. After the machine learning model has been trained and verified, a holdout subset is employed to give a final estimate of its performance. Using client data as a held-out subset allows to build generalizable models that are applicable to future data collection, rather than only the data used to train the model.

D. FORECASTING RESULTS OF DIFFERENT SCENARIOS

To evaluate the performance of the representative approaches, we report the R^2, RMSE, MSE, and MAE metrics obtained on the test set for each of the 9 wind farms for different case studies. The average performance indices are also overall clients (in FL and distributed methods) or over validation sets (in centralized approaches). The performance results of Case 1, Case 2, and Case 3 are detailed in Table 4, Table 5, and Table 6, respectively.

The centralized approach provides access to all databases gathered from the various clients. As a result, model accuracy is anticipated to be higher in comparison to alternative approaches that utilize far fewer data. Using the root mean square error as an example, the centralized method performs 24.97 percent better than the localized approach and 5.35 percent higher than the FL. This demonstrates that centralized models are capable of effective predictions, although with a high data need and a trade-off in privacy. While centralized models may allow for the learning of collective behaviors, they also risk the privacy of the energy facilities since data must be collected in a single place.

The localized learning method involves training a model for each client separately, utilizing just the data that is

TABLE 2. Statistics of datasets measured at the height of 40 m.

Wind farm	Time interval (10 min)	Samples	Min (m/s)	Mean (m/s)	Max (m/s)	Std
Kahrizak	25/08/2015 08:30-03/01/2017 13:30	58000	0	3.9119	22.5	2.7024
Tabriz	02/11/2015 09:40-07/01/2017 14:10	50491	0	5.2114	25.4	4.2109
Sonqor	05/01/2015 18:30-09/12/2017 09:20	109514	0	4.6829	22.7	3.4069
Abadan	27/09/2015 20:10-31/12/2016 23:50	58704	0	5.3765	21.3	2.743
Mahshahr	27/09/2015 11:10-01/01/2017 01:20	58110	0	6.0038	22.4	2.7947
Chabahar	23/07/2016 12:40-18/12/2017 15:40	73296	0	5.3839	17.6	2.4691
Zahedan	22/10/2015 18:10-22/02/2017 17:20	57021	0	3.9362	27.2	3.1672
Khaf	07/06/2015 12:40-25/10/2017 09:40	78452	0	9.8017	31.3	5.9815
Neyshabur	11/04/2014 13:40-27/06/2017 18:50	125297	0	7.1062	31.0	3.9829

TABLE 3. Case study scenarios.

Case Number	Model	Privacy of data	Generalizability
1	Centralized	×	✓
2	Loacalized	✓	×
3	FL	✓	✓
4	Centralized with holdout	×	✓
5	Loacalized with holdout	✓	×
6	FL with holdout	✓	✓

TABLE 4. forecasting performance values for the 9 databases and overall average—Centralized (case 1) without privacy.

	R^2	RMSE	MSE	MAE
Client 1	0.913340	0.046200	0.002134	0.022918
Client 2	0.929181	0.044452	0.001976	0.017802
Client 3	0.854238	0.083068	0.006900	0.043897
Client 4	0.929228	0.102861	0.010580	0.059779
Client 5	0.820338	0.094974	0.009020	0.052766
Client 6	0.939103	0.042611	0.001816	0.025924
Client 7	0.889405	0.058199	0.003387	0.027474
Client 8	0.920639	0.084182	0.007087	0.050290
Client 9	0.871659	0.076947	0.005921	0.041675
Avergae	**0.836947**	**0.070388**	**0.005424**	**0.038058**

TABLE 5. forecasting performance values for the 9 databases and overall average—Localized (case 2) with privacy.

	R^2	RMSE	MSE	MAE
Client 1	0.908720	0.047415	0.002248	0.023576
Client 2	0.758238	0.082132	0.006746	0.054363
Client 3	0.845407	0.085547	0.007318	0.045549
Client 4	0.928566	0.103341	0.010679	0.062976
Client 5	0.842676	0.088874	0.007899	0.046552
Client 6	0.938765	0.042729	0.001826	0.027266
Client 7	0.891274	0.057706	0.003330	0.025978
Client 8	0.918712	0.085198	0.007259	0.049155
Client 9	0.884263	0.073070	0.005339	0.039869
Avergae	**0.879624**	**0.0740013**	**0.005849**	**0.041698**

TABLE 6. forecasting performance values for the 9 databases and overall average—Federated (case 3) with privacy.

	R^2	RMSE	MSE	MAE
Client 1	0.790412	0.101105	0.010222	0.046193
Client 3	0.889205	0.070981	0.005038	0.035849
Client 5	0.821530	0.108309	0.011731	0.056959
Client 8	0.569000	0.127282	0.016201	0.079118
Avergae	**0.767536**	**0.101919**	**0.010798**	**0.0545297**

accessible to that particular wind farm. The high value of R^2 score along with lower values of MAE, MSE, and RMSE indicate a reasonable performance for the localized model. This means that the LSTM architecture is adequate to learn complex generation profiles that are specific to each client. Also, the localized model maintains privacy as there is no sharing of data between clients. However, this approach lacks generalizability because the training samples are limited to one specific location, and new and unseen data might result in poor performance of the localized models.

The FL method uses iterative communication between a supermodel and each client for each round of training. A subset of clients is selected, each of which trains its own local data separately for a limited number of epochs. As a result, a pool of local models is created that can be used to further update the supermodel. The selected clients update/co-train the supermodel by sending the parameters associated with the local models. Because of such a training procedure, the federated model can preserve privacy in contrast to the centralized model. Additionally, the FL outperforms the localized model by 7.42%, as measured by the R^2 score. The average values of RMSE, MSE, and MAE are also 2.67%, 6.9%, and 32.37% less than the localized model, respectively.

For scenarios involving a held-out subset with a localized approach (Case 4, Case 5, and Case 6), we perform the experiments by holding out the worst-performing client model with the highest errors on the validation set. As such, Client 1, Client 3, Client 5, and Client 8 are not involved in the training phase, and we only use them to assess the generalizability of the representative models during the testing phase. Table 7, Table 8, and Table 9 show the details of the obtained performances over the held-out clients as well as the average metric scores for different scenarios.

As it is shown, when the models are exposed to previously unobserved data, their overall performance suffers a degradation. Nonetheless, the centralized method outperforms the localized approach by a significant margin (e.g., 19.57% higher R^2 and 21% lower RMSE), although at the expense of a large amount of data and a reduction in privacy. This time, however, the FL shows higher performance compared to the localized and centralized approaches. For example, the

TABLE 7. Projected performance metrics for held-out clients and overall average scores–Centralized (case 1) without privacy.

	R^2	RMSE	MSE	MAE
Client 1	0.790412	0.101105	0.010222	0.046193
Client 3	0.889205	0.070981	0.005038	0.035849
Client 5	0.821530	0.108309	0.011731	0.056959
Client 8	0.569000	0.127282	0.016201	0.079118
Avergae	**0.767536**	**0.101919**	**0.010798**	**0.0545297**

TABLE 8. Projected performance metrics for held-out clients and overall average scores–Localized (case 2) with privacy.

	R^2	RMSE	MSE	MAE
Client 1	0.664218	0.127973	0.016377	0.075146
Client 3	0.771841	0.101859	0.010375	0.059884
Client 5	0.652427	0.151148	0.022846	0.097586
Client 8	0.479060	0.139933	0.019581	0.108098
Avergae	**0.641886**	**0.130228**	**0.017294**	**0.085179**

TABLE 9. Projected performance metrics for held-out clients and overall average scores–Federated (case 3) with privacy.

	R^2	RMSE	MSE	MAE
Client 1	0.807401	0.096921	0.009394	0.043524
Client 3	0.898757	0.067852	0.004604	0.032132
Client 5	0.811393	0.111342	0.012397	0.056825
Client 8	0.827760	0.080463	0.006474	0.047635
Avergae	**0.836327**	**0.089144**	**0.008217**	**0.045029**

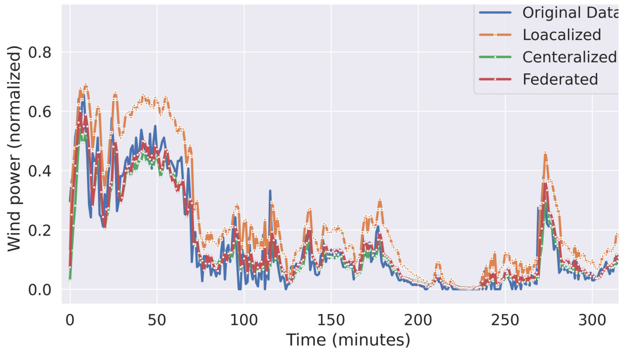

FIGURE 7. Predictions and ground data of wind power generation for Client 1 on February 22, 2018 between 10AM–3PM.

R^2 is 8.96% and 23.24% more than those of centralized and localized methods, respectively. Additionally, the RMSE is 14.33% and 46.08% lower, respectively, than the centralized and localized methods. The projected values for a 5-hour forecasting horizon of Client 1 are displayed in Fig. 7 to help comprehend the capabilities of federated learning in comparison to centralized and localized techniques.

E. COMPARISON STUDY

This section will compare the suggested strategy to numerous cutting-edge machine learning methods. The purpose of this study is to get a better understanding of the advantages and limits of the decomposition-based model in contrast to powerful and timely techniques such as support vector machine (SVM), random forest (RF), and multi-layer perceptron (MLP). There are several machine learning models, each with its own characteristics and uses. These models were chosen as representative of the most popular and effective supervised learning techniques. These algorithms provide very precise, consistent, and interpretable prediction models. Nonetheless, the proposed method is applicable to the remaining machine learning models. This section begins with an overview of the typical ML algorithms. Following the findings comes the debate. Table 10 displays the performance of the recommended forecasting models for a six-hour-ahead forecasting horizon with varied evaluation criteria for various data sets (Client1, Client3, Client5, Client8, and average performance).

TABLE 10. Forecasting performance of all representative ML and proposed models for different clients.

	Metric	SVM	RF	MLP	LSTM
Client1	RMSE	14.8	11.7	13.7	13.2
	MAE	18.1	14.2	28.2	21.3
	MAPE	9.6	15.4	8.4	7.4
	R^2	0.751	0.882	0.834	0.845
Client3	RMSE	26.3	23.3	14.7	19.0
	MAE	26.1	31.4	36.24	44.4
	MAPE	13.9	18.2	20.15	12.6
	R^2	0.783	0.734	0.824	0.895
Client5	RMSE	36.4	33.7	24.6	41.7
	MAE	22.6	21.4	19.4	21.5
	MAPE	14.1	13.2	8.1	8.2
	R^2	0.588	0.736	0.627	0.769
Client 8	RMSE	17.3	19.6	13.7	13.9
	MAE	33.2	23.8	14.6	23.5
	MAPE	8.7	8.2	6.2	4.6
	R^2	0.789	0.722	0.956	0.873
Average	RMSE	24.2	22.0	16.1	22.3
	MAE	25.2	24.4	26.9	34.7
	MAPE	11.9	14.6	12.6	8.6
	R^2	0.739	0.762	0.773	0.745

As expected, diverse algorithms display a variety of traits and performance characteristics. While certain algorithms, such as RF, perform better for some customers, they may be surpassed by other models in other contexts and on average. On the contrary, MLP did well in all circumstances. Both MLP and LSTM are capable of projecting wind power rather well, with LSTM beating MLP in simulations on average. Nevertheless, the suggested federated strategy has shown consistent, high-level performance across all stations, as indicated by the mean result. The worse performance of the ML algorithms (in comparison to the suggested models) is attributable to their inability to account for the non-stationarity and variability of wind profile data. Although ML models are capable of learning data, they are unable to capture the time-dependent characteristics of the wind series. LSTM, on the other hand, may match datasets better since it maintains temporal relationships. Using the MAPE as an example, the suggested method performs 3.4% better than MLP, 9.4% better than RF, and 6.7% better than SVM on average.

IV. CONCLUSION

Collective wind energy forecasting is a difficult task, given the privacy concerns surrounding wind farm data. Here, we proposed a privacy-preserving wind power predictor system by federating the training of machine learning models between several wind farms. To our understanding, this is one of the first studies that examine federated learning in the context of learning-based wind energy prediction. By using a federated learning method, we may substantially decrease the overall communication between clients and the central server as server-client data transmission is no longer required. Because the server does not gather data from individual wind farms, data privacy is preserved. Federated learning outperforms localized models in our trials and performs rather well when compared to centralized approaches. When exposed to unseen data, federated learning shows higher genralizeability compared to its counterparts.

REFERENCES

[1] M. Fan, Z. Li, T. Ding, L. Huang, F. Dong, Z. Ren, and C. Liu, "Uncertainty evaluation algorithm in power system dynamic analysis with correlated renewable energy sources," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5602–5611, Nov. 2021.

[2] B. Park, Z. Zhou, A. Botterud, and P. Thimmapuram, "Probabilistic zonal reserve requirements for improved energy management and deliverability with wind power uncertainty," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4324–4334, Nov. 2020.

[3] A. Ahmadi, M. Nabipour, B. Mohammadi-Ivatloo, A. M. Amani, S. Rho, and M. J. Piran, "Long-term wind power forecasting using tree-based learning algorithms," *IEEE Access*, vol. 8, pp. 151511–151522, 2020.

[4] S. Taheri, P. Hosseini, and A. Razban, "Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: A state-of-the-art review," *J. Building Eng.*, vol. 60, Nov. 2022, Art. no. 105067.

[5] A. Arjomandi-Nezhad, A. Ahmadi, S. Taheri, M. Fotuhi-Firuzabad, M. Moeini-Aghtaie, and M. Lehtonen, "Pandemic-aware day-ahead demand forecasting using ensemble learning," *IEEE Access*, vol. 10, pp. 7098–7106, 2022.

[6] Z. Liang, H. Chen, S. Chen, Y. Wang, C. Zhang, and C. Kang, "Robust transmission expansion planning based on adaptive uncertainty set optimization under high-penetration wind power generation," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 2798–2814, Jul. 2021.

[7] S. Taheri, M. Jooshaki, and M. Moeini-Aghtaie, "Long-term planning of integrated local energy systems using deep learning algorithms," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106855.

[8] M. A. González-Cagigal, J. A. Rosendo-Macias, and A. Gómez-Expósito, "Parameter estimation of wind turbines with PMSM using cubature Kalman filters," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 1796–1804, May 2020.

[9] H. Liu, H.-Q. Tian, and Y.-F. Li, "Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction," *Appl. Energy*, vol. 98, pp. 415–424, Oct. 2012.

[10] L. Tan, J. Han, and H. Zhang, "Ultra-short-term wind power prediction by Salp swarm algorithm-based optimizing extreme learning machine," *IEEE Access*, vol. 8, pp. 44470–44484, 2020.

[11] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303–315, Feb. 2014.

[12] R. Ak, O. Fink, and E. Zio, "Two machine learning approaches for short-term wind speed time-series prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1734–1747, Aug. 2016.

[13] A. Omidi, A. Heydarian, A. Mohammadshahi, B. A. Beirami, and F. Haddadi, "An embedded deep learning-based package for traffic law enforcement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 262–271.

[14] J. Zhao, J. Wang, Z. Guo, Y. Guo, W. Lin, and Y. Lin, "Multi-step wind speed forecasting based on numerical simulations and an optimized stochastic ensemble method," *Appl. Energy*, vol. 255, Dec. 2019, Art. no. 113833.

[15] R. Perez, P. Lauret, M. Perez, M. David, T. E. Hoff, and S. Kivalov, "Solar resource variability," in *Wind Field and Solar Radiation Characterization and Forecasting*, vol. 6. Springer, pp. 149–170, 2018.

[16] C. Zhang, O. Vinyals, R. Munos, and S. Bengio, "A study on overfitting in deep reinforcement learning," 2018, *arXiv:1804.06893*.

[17] D. W. Van Der Meer, J. Munkhammar, and J. Widén, "Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals," *Sol. Energy*, vol. 171, pp. 397–413, Sep. 2018.

[18] J. Lee, W. Wang, F. Harrou, and Y. Sun, "Wind power prediction using ensemble learning-based models," *IEEE Access*, vol. 8, pp. 61517–61527, 2020.

[19] A. Ahmadi, M. Nabipour, S. Taheri, B. Mohammadi-Ivatloo, and V. Vahidinasab, "A new false data injection attack detection model for cyberattack resilient energy forecasting," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 371–381, Jan. 2023.

[20] A. Ahmadi, M. Nabipour, B. Mohammadi-Ivatloo, and V. Vahidinasab, "Ensemble learning-based dynamic line rating forecasting under cyberattacks," *IEEE Trans. Power Del.*, vol. 37, no. 1, pp. 230–238, Feb. 2022.

[21] S. Sridhar and G. Manimaran, "Data integrity attacks and their impacts on SCADA control system," in *Proc. IEEE PES Gen. Meeting*, Jul. 2010, pp. 1–6.

[22] R. K. Tai, J. P. Ma, Y. Zhao, and S. S. Chow, "Privacy-preserving decision trees evaluation via linear functions," in *Proc. Eur. Symp. Res. Comput. Secur.* Cham, Switzerland: Springer, 2017, pp. 494–512.

[23] J. Vaidya and C. Clifton, "Privacy-preserving decision trees over vertically partitioned data," in *Proc. IFIP Annu. Conf. Data Appl. Secur. Privacy*. Cham, Switzerland: Springer, 2005, pp. 139–152.

[24] J. Vaidya and C. Clifton, "Secure set intersection cardinality with application to association rule mining," *J. Comput. Secur.*, vol. 13, no. 4, pp. 593–622, 2005.

[25] J. Vaidya and C. Clifton, "Privacy preserving Naïve Bayes classifier for vertically partitioned data," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2004, pp. 522–526.

[26] J. Vaidya and C. Clifton, "Privacy-preserving K-means clustering over vertically partitioned data," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2003, pp. 206–215.

[27] Y. Rahulamathavan, R. C.-W. Phan, S. Veluru, K. Cumanan, and M. Rajarajan, "Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud," *IEEE Trans. Dependable Secure Comput.*, vol. 11, no. 5, pp. 467–479, Sep. 2014.

[28] M. D. Cock, R. Dowsley, A. C. A. Nascimento, and S. C. Newman, "Fast, privacy preserving linear regression over distributed datasets based on pre-distributed data," in *Proc. 8th ACM Workshop Artif. Intell. Secur.*, Oct. 2015, pp. 3–14.

[29] V. Nikolaenko, U. Weinsberg, S. Ioannidis, M. Joye, D. Boneh, and N. Taft, "Privacy-preserving ridge regression on hundreds of millions of records," in *Proc. IEEE Symp. Secur. Privacy*, May 2013, pp. 334–348.

[30] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1333–1345, May 2018.

[31] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2015, pp. 1310–1321.

[32] T. Ryffel, A. Trask, M. Dahl, B. Wagner, J. Mancuso, D. Rueckert, and J. Passerat-Palmbach, "A generic framework for privacy preserving deep learning," 2018, *arXiv:1811.04017*.

[33] P. Li, J. Li, Z. Huang, T. Li, C.-Z. Gao, S.-M. Yiu, and K. Chen, "Multi-key privacy-preserving deep learning in cloud computing," *Future Generat. Comput. Syst.* vol. 74, pp. 76–85, Sep. 2017.

[34] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[35] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7751–7763, Aug. 2020.

[36] N. Gholizadeh and P. Musilek, "Federated learning with hyperparameter-based clustering for electrical load forecasting," *Internet Things*, vol. 17, Mar. 2022, Art. no. 100470.

[37] Y. Li, J. Li, and Y. Wang, "Privacy-preserving spatiotemporal scenario generation of renewable energies: A federated deep generative learning approach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2310–2320, Apr. 2022.

[38] J. Lin, J. Ma, and J. Zhu, "A privacy-preserving federated learning method for probabilistic community-level behind-the-meter solar generation disaggregation," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 268–279, Jan. 2022.

[39] X. Zhang, F. Fang, and J. Wang, "Probabilistic solar irradiation forecasting based on variational Bayesian inference with secure federated learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7849–7859, Nov. 2021.
[40] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.
[41] D. Liu, T. Miller, R. Sayeed, and K. D. Mandl, "FADL: Federated-autonomous deep learning for distributed electronic health record," 2018, *arXiv:1811.11400*.
[42] S. Taheri, A. Ahmadi, B. Mohammadi-Ivatloo, and S. Asadi, "Fault detection diagnostic for HVAC systems via deep learning algorithms," *Energy Buildings*, vol. 250, Nov. 2021, Art. no. 111275.
[43] Y. Hu, A. Huber, J. Anumula, and S.-C. Liu, "Overcoming the vanishing gradient problem in plain recurrent networks," 2018, *arXiv:1801.06105*.
[44] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3168–3176, May 2020, doi: 10.1109/TII.2019.2902129.

AMIRHOSSEIN AHMADI (Student Member, IEEE) received the B.Sc. degree in electrical engineering and control systems from the University of Tabriz, Tabriz, Iran, in 2012, and the M.Sc. degree from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2015. He is currently pursuing the Ph.D. degree in electrical and software engineering with the University of Calgary, Calgary, Canada. He was an Instructor with the Industrial Control Laboratory, Amirkabir University of Technology, and a Research Assistant with the Department of Electrical Engineering, University of Tabriz. His main research interests include renewable energy integration, storage systems, electrical vehicles, forecasting, and deep learning. He serves as a Reviewer for IEEE Transactions on Power Systems and IEEE Transactions on Industrial Informatics.

MOHAMMAD TALAEI received the B.Sc. degree in mechanical engineering from Bu-Ali Sina University, Iran, in 2018, and the M.S. degree in energy systems engineering from the Sharif University of Technology, Iran, in 2021. He is currently an Instructor of data science course at the Sharif University of Technology. He is also the Co-Founder of Fitech1 in the field of time series and machine learning, where he and his colleagues work on vertical solutions for tabular time series problems. His research interests include time series, machine learning, optimization, and deep learning. He serves as a Reviewer for the IEEE Internet of Things Journal.

MASOD SADIPOUR received the B.Sc. degree in mechanical engineering from the Ferdowsi University of Mashhad, Iran, in 2016, and the M.S. degree from the University of Tehran, Iran, in 2019. He is currently pursuing the Ph.D. degree with the Department of Mechanical and Material Engineering, Ritchie School of Engineering, University of Denver, Denver, CO, USA. His research interests include biomedical engineering, CFD, FSI, machine learning, and optimization.

ALI MORADI AMANI (Member, IEEE) received the bachelor's and master's degrees in electrical engineering (control systems). He is currently with the School of Engineering, Royal Melbourne Institute of Technology (RMIT University), Melbourne, Australia. His research interests include control of future power systems, EV integration into power grids, fault-tolerant control systems, and control of complex networks.

MAHDI JALILI (Senior Member, IEEE) received the Ph.D. degree from the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, in 2009. He was a Faculty Member at the Sharif University of Technology, Tehran, Iran, from 2009 to 2014. In 2014, he joined RMIT University, Melbourne, Australia, as an Australian Research Council DECRA Fellow and a Vice-Chancellor's Research Fellow. He is currently an Associate Professor of AI and electrical engineering at RMIT University and leads the Electric Vehicle Living Laboratory. His main research interests include complex systems, network science, machine learning, and their applications in energy systems.