Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v30i19.3649

FEDERATED LEARNING FOR PREDICTION OF ENERGY CONSUMPTION IN EDGE-ENABLED SMART CITIES WITH HEALTHCARE INFRASTRUCTURE CROSSROADS

Jamshaid Iqbal Janjua[1,2*], Areej Fatima[3], Atifa Athar[4], Muhammad Saleem Khan[1], Zia Ul Rehman[5], Rabia Javed[6], Tahir Abbas[6*], Insia Zahra[7]

[1]Computer Sciences, National College of Business Administration & Economics, Lahore, Pakistan.
[2]Al-Khawarizmi Institute of Computer Science, University of Engineering & Technology (UET), Lahore, Pakistan.
[3]Faculty of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan.
[4]Department of Computer Science, Comsats University Islamabad, Lahore Campus, 54000, Pakistan.
[5]Department of Computer Science, Government College University Lahore, 54000, Pakistan.
[6]Department of Computer Science, TIMES Institute, Multan, Pakistan, Pakistan.
[7]Institute of Computer Science and Information Technology, The Women University, Multan, Pakistan.

***Corresponding Author:** Jamshaid Iqbal Janjua, Tahir Abbas
Email: jamshaid.janjua@kics.edu.pk , drtahirabbas@t.edu.pk

Abstract
The smart grid employs large amounts of consumption data to develop advanced machine learning models for various purposes, such as load monitoring and demand response. However, these applications pose security threats and demand a high level of precision. On one hand, the data used is extremely susceptible to privacy concerns. For instance, the comprehensive data collected by a smart meter installed in a consumer's home might offer valuable insights about the precise appliances being utilized and, subsequently, the consumer's behavior at home. Conversely, deep learning models require substantial amounts of varied data to be effectively taught. This study evaluates the deployment of Edge computing and federated learning, a distributed machine learning technique that allows for the increase in both the amount and diversity of data used to train deep learning models, while also preserving privacy. This work introduces the novel application of federated learning to estimate home load, a task that, as far as we know, has not been investigated before. The obtained results are promising. The simulations were performed using Federated on a dataset comprising 300 residential properties. Examining the convergence of smart grid technologies in smart cities with the pharmaceutical market and taking healthcare infrastructure considerations into account holds promise for pioneering solutions that tackle challenges across various sectors. This cooperative strategy has the potential to foster a healthcare system that is more robust, sustainable, and cost-effective, aligning seamlessly with overarching healthcare policy goals.

Keywords: Energy Load Forecasting, Edge Computing, LSTM, Federated Learning, Smart Grid, Pharmaceutical, Healthcare Infrastructure.

INTRODUCTION

Accurate prediction of electricity demand is essential for the progress of the intelligent power distribution system. Infrastructure design necessitates the use of long-term load forecasting, whereas system operations heavily depend on mid-term and short-term load forecasting as essential responsibilities [1].

To ensure the efficient operation of electrical power distribution on a daily basis, it is crucial to accurately predict short-term demand profiles. This estimate is based on the gathering and examination of significant quantities of complex data acquired from homes. However, effectively predicting short-term load forecasting (STLF) for specific instances has been challenging due to the unforeseen attributes of load profiles. The electrical load of a dwelling is significantly impacted by the unpredictable and often challenging behavior of its residents [2, 3].

Load forecasting, also known as energy consumption forecasting, is a crucial undertaking in ensuring the sustainable planning, production, and transportation of electricity in modern power networks [4]. Short-term load forecasting is crucial for the proper functioning of the power system, whilst long-term load forecasting is important for the strategic planning and growth of the system [5]. There is a growing fascination with precisely predicting the immediate electricity usage of individual residential consumers. The interest in this topic derives from the growing trend of generating renewable energy, namely through solar systems, which has resulted in a shift towards a more decentralised energy production. This forecasting technique is considered a valuable tool for households to improve their self-reliance. Residential demand exhibits significant volatility and rapid fluctuations, which introduces intricacy to this procedure in contrast to long-term or aggregated short-term forecasting [6].

In recent times, there has been a specific emphasis on developing Artificial Intelligence (AI) algorithms to maximize the efficiency of short-term residential load forecasting, as indicated by several research [7]. Nevertheless, the most of the suggested AI algorithms inherently necessitate substantial quantities of extensive historical data [8], which must be gathered and stored by the power system in centralised locations to ensure precise model training. The Advanced Metering Infrastructure (AMI) has recently improved the efficiency of this process. The Advanced Metering Infrastructure (AMI), now being implemented in many countries worldwide, utilizes intelligent metres installed at consumers' locations. These sophisticated metres allow for the collection of accurate energy consumption data at intervals of 15 minutes or more.

Recent assessments of state-of-the-art methodologies [9, 10] have shown that deep neural networks possess significant promise as a resolution for the short-term load forecasting (STLF) issue at the household level. This is due to their ability to comprehend complex and non-linear patterns. Neural networks created via artificial means

Outperform other prediction methods, such as Auto Regressive Integrated Moving Average (ARIMA)[11] and Support Vector Regression (SVR). Merely depending on the adoption of deep learning models alone will not lead to significant improvements, considering the inherent characteristics of the models.

Overfitting is a widespread occurrence [12]. An overfitted model is a model that has unduly absorbed the complex details of the training data, including the noise, which impairs its ability to generalize when applied to new data. To tackle this issue, it is recommended to increase the variety and quantity of the data by combining usage statistics from multiple households. In general, the proposed frameworks [13, 14] function under the assumption that each data record is transferred from smart meters to a centralised computer infrastructure using broadband networks in order to train models. Nevertheless, this assumption gives rise to concerns regarding privacy as load profiles reveal substantial sensitive information, including device usage and household occupancy. The transfer of complex information over networks makes it susceptible to malicious interception and manipulation.

The Machine Learning group has created a novel on-device method known as Federated Learning (FL) to address privacy concerns and improve the quantity and diversity of data sets [15]. Federated Learning is a decentralised approach to machine learning where each device collaborates in building a central model without transferring any data. Figure 1 illustrates that the server starts by initializing the model through either randomly assigning variables or by exploiting publicly accessible data. Subsequently, the model is disseminated to a cohort of arbitrarily chosen devices (clients) for localized training utilizing their own data. Each client sends the model weights to the server, which subsequently computes the mean and employs it to update the global model. This technique will continuously iterate until the global model achieves a state of equilibrium.

The advent of technology has offered both consumers and energy providers with novel services, although it has also engendered significant apprehensions around privacy. Both regulatory authorities and consumers have voiced these concerns. Without a doubt, the gathered data has the capability to reveal patterns in consumer behaviour [16], and in certain countries, individuals have the option to decline the installation of a smart metre. Several privacy-enhancing methodologies including data aggregation and/or obfuscation have been suggested to protect privacy while allowing energy suppliers or other entities to access pertinent information [17]. However, these methods are not consistent with any proposed techniques for forecasting short-term energy use in residential areas, as they all require accurate measurements obtained directly from households as input data. Moreover, the current AI-powered solutions require significant computational resources for training the model, and their capacity to effectively handle the vast amount of data generated by millions of smart metres is not simple or straightforward. A different strategy involves training the AI model using a smaller portion of the data, which will inevitably affect the model's ability to generalize.

Edge Computing [18] is a decentralised approach that aims to bring computer operations closer to the end consumers. It expands the reach of Cloud Computing to the periphery of the network, offering significant benefits in terms of velocity, effectiveness, dependability, confidentiality, protection, and scalability. Edge Computing has facilitated the development of numerous innovative applications, particularly in the Internet of Things (IoT) field [19]. This is regarded to be the most efficient approach for addressing the privacy and scalability concerns that occur when utilising an Advanced Metering Infrastructure (AMI) for the collection and analysis of energy usage data. Nevertheless, it is crucial to consider numerous important factors while implementing this approach for short-term distributed load forecasting. The main factors are mostly associated with the core characteristics of current AI methodologies. These strategies guarantee effective model generalization only when trained centrally using data gathered from a diverse set of users. Unlike the fundamental principle of Edge Computing, it is recommended to locally distribute and process sensitive data.

Federated Learning (FL) has emerged as a method to combine Artificial Intelligence (AI) and Edge Computing. Federated Learning is a decentralised machine learning method in which a central entity oversees the training of a collective global model through a federation of participant devices. The uniqueness of the approach lies in the fact that each device trains a separate model using its own data, which is stored locally and never sent. The central entity just receives the model parameters to update the shared global model. Federated Learning enables the complete utilization of Edge Computing for machine learning objectives and has already demonstrated successful implementation in various fields, including human-computer interaction, language modelling, healthcare, transportation, and Industry 4.0. These domains prioritise the criteria of privacy and/or scalability.

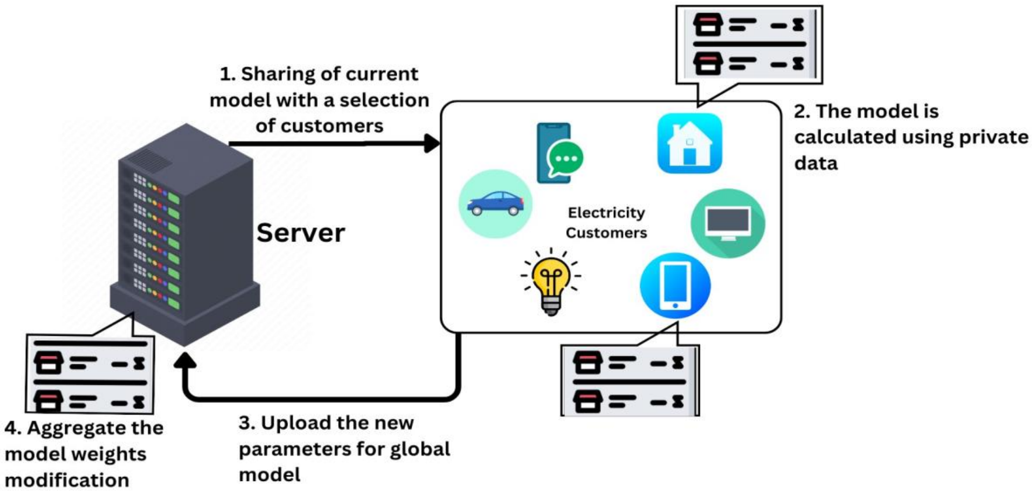

Figure 1. Communication between server and clients in Federated Learning

The objective of this study is to evaluate the use of Edge computing, in combination with the Federated Learning technique, to tackle the problem of Short-Term Load Forecasting (STLF) for residential power consumption. Edge computing is the act of analyzing data at the outermost part of a network, instead of depending on cloud or remote server analysis. We utilize Long-short Term Memory (LSTM), a customized deep neural network specifically developed for predicting future values in time series data. LSTM employs historical data on the electricity consumption of the household to forecast future trends. We are conducting a study on a group of houses that share common characteristics, such as their geographical location and construction type. The assessment is swiftly completed in order to limit the impact of weather fluctuations and seasonal variations. Federated learning takes place on edge devices situated within a grid network of residential areas. Edge equipment is typically located at the endpoint of the electrical distribution network and acts as a smart middleman between the user and the electric power supply. This can take the form of either a smart meter or a more sophisticated gadget. The main contributions of our study are as follows: (1) We propose a framework that enables the use of Edge devices in the smart grid for Federated Learning (FL); (2) We evaluate the improvement in accuracy achieved by FL through simulations; and (3) We assess the reduction in network load through numerical results. Furthermore, we integrate the benefit of heightened privacy attained through decentralisation and Edge computing.

Exploring the intersections between smart grid technologies, the pharmaceutical market, and healthcare infrastructure considerations holds the potential for innovative solutions addressing challenges across multiple sectors. This collaborative approach can enhance operational efficiency in pharmaceutical manufacturing by optimizing electricity usage, leading to cost savings and improved drug production efficiency. Additionally, efficient energy management driven by intelligent modeling and Federated Learning integration may contribute to cost reductions, promoting affordability of medicines in line with health policy goals. The emphasis on sustainability in smart grid technologies aligns with the increasing focus on environmental practices in pharmaceutical manufacturing, aiding companies in meeting regulatory compliance and sustainability requirements outlined in health policies. The integration of Federation Learning in smart grids also contributes to transparent decision-making, crucial for regulatory compliance in the pharmaceutical market. Predictive modeling techniques can be adapted for health policy planning, ensuring a consistent and reliable supply of essential medicines through effective resource allocation. Moreover, the resilient features of smart grid technologies inspire strategies to enhance the resilience of the pharmaceutical supply chain, aligning with health policies aiming to strengthen healthcare infrastructures against disruptions. The research underscores the importance of cross-sector collaboration, suggesting that collaboration between the energy and pharmaceutical sectors can lead to cross-cutting innovations. Policymakers may find value in fostering such collaborations to address shared challenges and align policies for a more sustainable and efficient healthcare ecosystem.

LITERATURE REVIEW / RELATED WORKS

Several recent studies have utilized deep neural networks, namely Long-short term memory (LSTM), to tackle the task of predicting short-term electricity consumption. While benchmarks indicate LSTM's comparative efficiency compared to alternative approaches [20,21], the results do not meet the anticipated level of accuracy in terms of Root Mean Square Error (RMSE) and Mean Average Percentage Error (MAPE). In order to improve the accuracy of predictions, the authors in [22] propose using a modified variant of LSTM called sequence-to-sequence LSTM. This alternative produces improved results when working with data at a one-minute interval, but it does not show any significant improvement for data at a one-hour period compared to the traditional LSTM model. Furthermore, certain authors [9] view the goal of discovering the most effective LSTM network as a challenge of optimizing hyperparameters and utilize the evolutionary approach to address this matter. It is claimed that finding the best combination of window size and number of hidden neurons in each layer is a probabilistic undertaking. Contrary to claims made by specific specialists, the primary cause of the current issue is not primarily attributed to the structure of neural networks. Yet, the main obstacle is in the ability of data-driven forecasting models to generalize. Many proposed models exhibit a decline in accuracy when applied to new datasets [10]. Several studies suggest including additional weather data [23] or data from domestic appliances [2]. The weather has a substantial impact on the total power consumption, but the immediate load is primarily controlled by the activity of the residents [3, 24, 25]. However, the act of collecting data from household devices is both expensive and encroaches upon individuals' privacy.

One alternate approach to improve the training data is to aggregate data from multiple clients. The authors in [13] utilize clustering to classify individuals with similar traits, hence reducing the variability of uncertainty within the groups. The authors in [14] propose a pooling technique to mitigate overfitting by increasing the diversity of the data. Nevertheless, these systems demonstrate a significant level of centralization and are vulnerable to privacy issues. If consumption data is collected or intercepted without legal authority, it can compromise the privacy of its transmission over networks [26]. Diverse procedures were implemented to ensure the protection of users' identities in the smart grid. The authors propose a technique that employs clustering to allocate a common serial number to users who are in close proximity to one another. However, this method has difficulties in delivering customized service to individual clients because of the inherent lack of personal identity. Additional research aims to hide consumption information by utilizing data aggregation as the primary method [27, 28]. Nevertheless, this technique is in conflict with the goals of short-term load forecasting (STLF).

None of the previously stated publications address the issues pertaining to both user privacy and forecast accuracy. The aim of the proposed project is to use Edge Equipment in the Home Area Network (HAN) to carry out tasks including client selection and training neural networks at the Edge using the federated learning approach. Through the utilization of data, it is possible to create a worldwide model while ensuring the protection of the privacy of the persons involved.

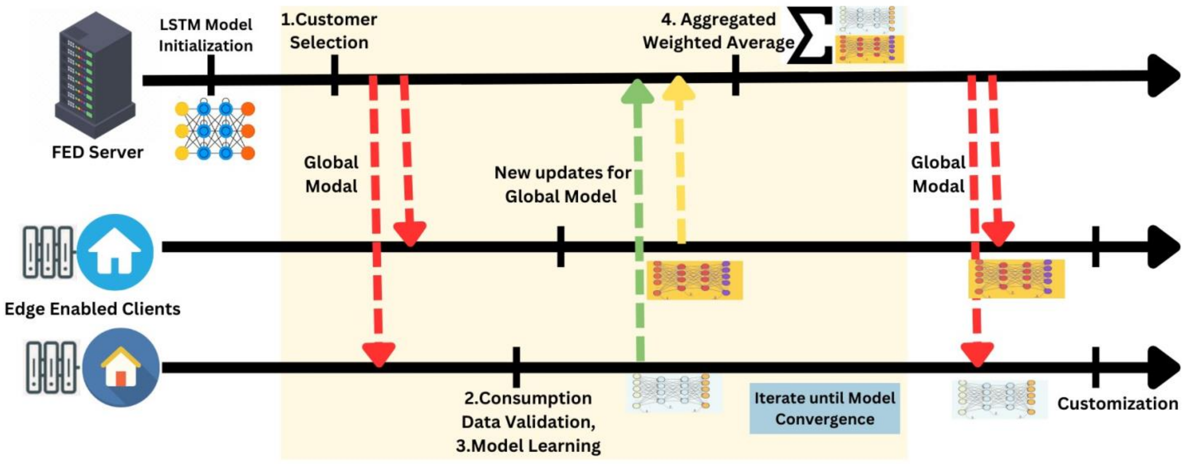

Figure 2. Sequence of the roles in federated learning model

MATERIALS AND METHODS

The network structure illustrated in Figure 2 comprises two main components: a Federated FED server and clients. Clients pertain to residential structures that are furnished with Edge technology, predominantly comprising of intelligent meters and other devices within the Home Area Network. The objective of this work is to employ FL to develop a comprehensive LSTM-based model for Short-Term Load Forecasting (STLF) on a global level. The FED server coordinates the training iterations, which are subsequently executed by the clients utilizing their individual electricity consumption data. This section offers a thorough elucidation of LSTM and its use in predictive modeling, along with FL and its integration into our system architecture.

LSTM based Prediction

Applying Long Short-Term Memory (LSTM) models to forecast time series data is studied by approach to forecast future electrical load in time series analysis. A time series is a set of data points that are uniformly distributed and arranged, illustrating the variations in a specific variable over duration of time. Time series forecasting involves the creation of models that capture the connections between present data points and previous data. Nevertheless, the precision of the predictions is significantly impacted by the chosen model and the caliber of the past data points.

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that is designed to effectively capture and retain long-term dependencies in sequential data.

An LSTM is a type of machine learning model that falls under the field of Artificial Neural Networks (ANNs), more particularly, Recurrent Neural Networks (RNNs). This methodology is commonly utilised to address sequence classification problems, and it is particularly well-suited for forecasting time series data. Prior to providing a more detailed elucidation of LSTM, we will initially present a succinct summary of the architecture of ANNs and RNNs, emphasizing the fundamental distinctions between them.

An artificial neural network (ANN) consists of several neurons arranged in consecutive layers. A neuron functions as the basic component of artificial neural networks (ANNs). The input data is processed through a mathematical operation, usually a weighted sum, and then passed through an activation function, which can be thought of as a threshold function, to obtain the final value. The outcome is subsequently transmitted to further neurons. Each neuron is linked to many weights, denoted as wj (where j represents the index of each weight), that are utilized to compute weighted sums and are associated with the interconnections of neurons. The training phase of an Artificial Neural Network (ANN) is to determine the optimal weights wj for the model, with the objective of maximizing the network's performance in achieving its designated job. In an artificial neural network (ANN), the output of a neuron in a given layer is consistently utilized as input for one or more neurons in the subsequent layer. The feeding technique is commonly known as "feed-forward". An artificial neural network (ANN) consists of three separate layers: an input layer, which receives input data and performs initial processing; one or more hidden layers, where neurons process input from previous layers and transmit the results to subsequent layers; and an output layer, which produces the final result.

A Recurrent Neural Network (RNN) is an Artificial Neural Network (ANN) that allows for connections between neurons in a layer to extend to neurons in previous layers, within the same layer, or even to themselves in a loop. Recurrent neural networks (RNNs) provide the crucial capability to retain and employ an internal state for the purpose of capturing and analyzing temporal variations at a given time t. Nevertheless, conventional Recurrent Neural Networks (RRNs) has limited capacity to retain information over extended durations.

An LSTM, short for Long Short-Term Memory, is a specialized and intricate form of RNN, or Recurrent Neural Network, which employs an LSTM cell as its fundamental component rather than a neuron. An LSTM is distinguished by its ability to retain and retrieve information over long durations. An LSTM cell employs two vectors to encode its internal state: h(t) represents the short-term state, which is always equal to the cell output y(t) at any given time t, while c(t) represents the long-term cell state, which is maintained and modified over time. The LSTM cell consists of three gates that regulate the inclusion or exclusion of information from the cell state c(t) and the calculation of the cell output y(t): The input gate determines which information from the current input x(t) should be retained for computing the current state c(t). The forget gate is responsible for determining which information from the previous state c(t-1) is important and should be retained, while disregarding any irrelevant information, in order to compute c(t). The output gate is responsible for calculating the final output y(t) and the current state c(t), which will be used as the input for the cell in the next time step t + 1.

LSTM can construct temporal links between previous data points and present occurrences, while addressing the problems of vanishing and exploding gradients that are commonly seen in recurrent neural networks (RNNs). Gradient vanishing is the occurrence where the size of the gradient diminishes considerably for elements that are far apart in time, causing the weights to remain unaltered in lower layers. Conversely, gradient bursting is characterised by a significant increase in the amplitude of the gradient [29]. This is achieved by employing two key components: the memory cell, which stores essential states from earlier events, and the gates, which regulate the flow of information. The LSTM model comprises three separate gates: the input gate, the output gate, and the forget gate. During the process of learning, individuals develop the capacity to reset the memory cell linked to unimportant traits. The primary techniques employed to achieve state-of-the-art progress in sequence learning, specifically in language translation and speech recognition, are LSTM and its several iterations. In the context of residential short-term load forecasting (STLF), it is expected that the LSTM network will be able to analyze the consumption profile of residents, remember the specific conditions associated with it, and use this information to predict future consumption.

Federated Learning

Federated learning is a decentralised kind of machine learning where the primary training process takes place on individual devices referred to as clients. Google initially included this feature into mobile device keyboards to anticipate the following word [30]. This approach is well-suited for various situations:

a) When data necessitates stringent privacy safeguards,
b) When the data size is considerably larger than the model updates,
c) In extensively distributed systems with a large number of devices surpassing the number of nodes in a data centre,
d) In supervised training when labels can be inferred directly from the user.

Federated Learning is a decentralised method for training a machine learning model. This approach entails using the data from individual devices to do local training, while a central server combines the locally learned models to generate a global model. Afterwards, this global model is distributed among the devices for additional training iterations. By using this iterative process, an unlimited number of devices can participate in model training without the need to upload collected data to a centralised location. Transmission is solely necessary for locally trained models. Federated Learning has demonstrated efficacy even in scenarios where the remote devices train the model using data that is not independent and identically distributed. Furthermore, studies have shown that this strategy provides advantages in terms of data transfer volume, as compared to a centralised training method that necessitates sending data to a singular place [31].

1) Protocol for Training
Although a comprehensive account of a standard foreign language training approach may be found in [32], we will only emphasize the fundamental elements in this discussion. The progression of time is divided into discrete intervals referred to as rounds, during which two entities, specifically the end devices and the centralised server, participate in the workflow. To provide further clarification, the subsequent operations are executed:

a) The centralised server initiates the training of the machine learning model by selecting the weights of the model in a random manner. Afterwards, it sends the chosen model type (such as ANN, RNN, LSTM) and the specific hyperparameters (such as number of layers, number of neurons) to the terminal devices.
b) The commencement of the training phase. Within each iteration, the subsequent tasks are performed:

i. The server chooses a subset S of end devices.
ii. The server transmits its present global model (i.e., weights) to the chosen end devices.
iii. The edge devices utilize their local data to train the receiving model.
iv. The edge devices transmit the locally trained model to the server by delivering updated weight values.
v. The server aggregates all the local models it receives (i.e., it computes the new weights based on the updated weights received from any end device) and produces an updated global model.

The selection of terminal devices in each iteration is contingent upon the particular application, as extensively examined in [33]. Another crucial factor is the centralised server's aggregation process. In this discussion, we will provide a concise overview of the commonly used aggregation method known as Federated Averaging (FedAVG).

2) Process of Federated Averaging.
FedAVG, as outlined in reference [31], seeks to minimise the collective loss function. FedAVG is a variant of FedSGD, a technique used to combine data, as explained in reference [31]. Federated Stochastic Gradient Descent (FedSGD) involves the chosen end devices doing one iteration of Stochastic Gradient Descent (SDG) in each round and transmitting the resulting model weights to the server. The server calculates the mean of the received weights, taking into account the number of training samples utilized locally. This method can be interpreted as a step in the gradient descent algorithm that is applied on the global model. An inherent limitation of this technique is that

achieving global model training is delayed due to the execution of only one SDG step each round.

FedAVG addresses this constraint by introducing the concepts of local epoch and batch. During each round, multiple local epochs are performed on a batch, which represents a subset of the local data. At every local epoch and batch, a step of stochastic gradient descent (SGD) is executed, resulting in a reduction in the communication between the server and the device. Afterwards, the server calculates the average of the given weights, using the proportion of locally used training samples, using the identical approach as FedSDG.

3) Federated Privacy Aware Approach

Federated learning has proven to be useful in situations when datasets include imbalances or non-identical distributions. During the process of federated learning, iteration goes in the following manner: Firstly, a subset of clients is chosen, and each client is given the latest model. Our customers are using Edge devices that are installed in residential premises, such as smart meters. Chosen customers contribute Stochastic Gradient Descent (SGD) updates on locally stored data, while a server aggregates the client modifications to generate a novel global model. The new model is allocated to a specific group of customers primarily focus on load prediction and privacy considerations. This technique is repeated until the desired degree of prediction accuracy is attained.

The server use the FederatedAveraging algorithm [15] to combine the updates received from the clients. At the beginning, the global model is randomly initialized or pre-trained using publicly available data. During each training round, the server assigns a global model, wr, to a subset K of clients who have enough data records and show significant variability in their consumption load. This is done to optimize the effectiveness of the training data. The objective of this criterion is to guarantee a comprehensive collection of data points that accurately represents the normal consumption habits of the renters. Each client k in the subset utilizes nk instances from its local data. The volume in our case is dictated by the period of data production by the smart meter and the quantity of locally stored data. The dataset used comprises sliding windows that contain a predetermined number of previous steps. Each client k utilizes Stochastic Gradient Descent (SGD) to compute the average gradient gk, using a learning rate η. The updated versions of the models are sent to the server for consolidation.

However, the centralised strategy may not be appropriate for accommodating the energy use of all clients. To effectively tackle this issue, one might employ the strategy of implementing Personalization. The primary focus of several programmes is on customization, which involves understanding user behavior and adapting accordingly. The method involves training a centralised model by including user-specific data to construct an individualized model for each user. To do this, the model can undergo local retraining for a certain number of epochs, using only the user's data [34].

Federated learning presents fewer privacy problems when compared to centralised server storage. Nevertheless, despite the anonymization of data, users' identities are still vulnerable and can be revealed by reverse engineering. The model updates provided by each client are temporary and never stored on the server; weight adjustments are retained in the computer's memory and are deleted after consolidation. The federated learning approach requires that the individual weight submissions be scrutinized or evaluated. The aforementioned approach is more secure than server training due to the network and server's inability to effectively handle complex user data. For billing purposes, it is necessary to supply specific data in a consolidated style, even when this data does not reveal a significant amount of information. Presently, scholars are investigating techniques like as secure aggregation [35] and differential privacy [36] to ensure compliance with trust criteria.

System Architecture

This section presents our federated design for short-term energy consumption forecasting, which is widely recognized as the federated architecture in Figure 3. The suggested framework facilitates precise prediction of energy consumption in residential buildings by means of the cooperative training of a Long Short-Term Memory (LSTM) neural network. The training procedure of local LSTM networks necessitates the utilization of numerous edge devices that generate energy usage measurements. The global LSTM network is trained using centralised model aggregation and redistribution methods, as proposed in Federated Learning. Our architecture considers two separate entities:

1) Energy Company: It provides electricity to residential structures. The improvement of energy supply could depend on the adoption of a Smart Grid [37].
2) Residential customers refer to persons who use power in their homes and receive bills from the energy supplier.

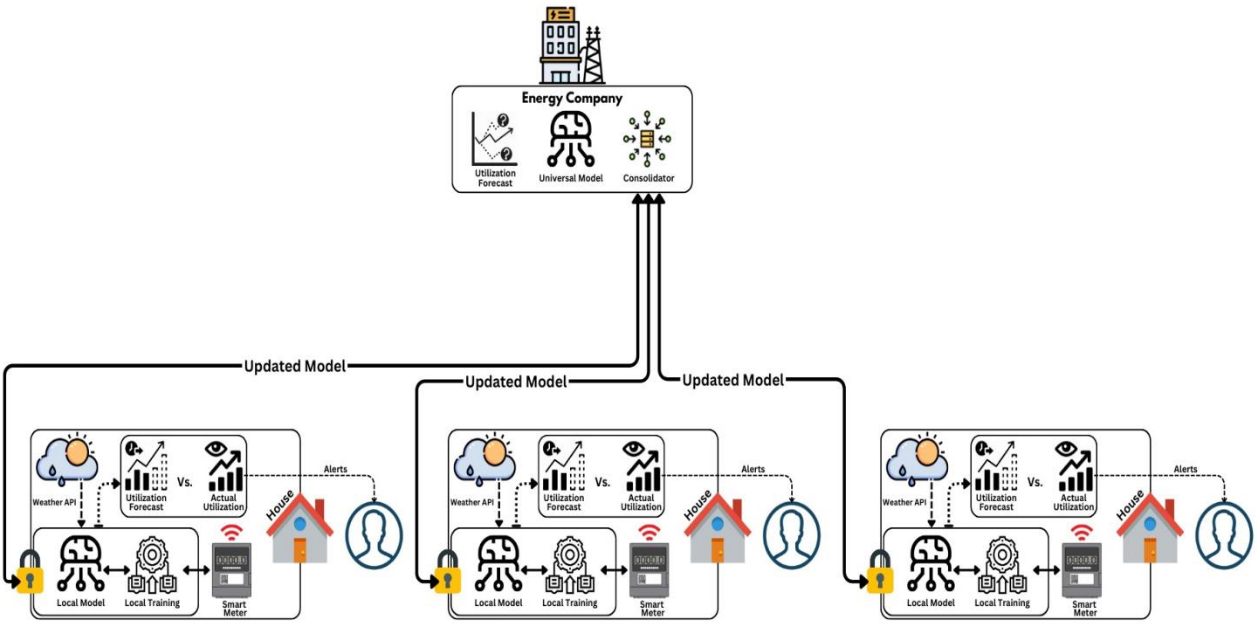

Figure 3. Smart cities energy load prediction based on proposed federated architecture.

Three functional nodes, derived from the prior functional nodes of Federated Learning are combined and engage in mutual interaction:

1) Smart Meter: A device owned by the energy supplier and installed at the premises of a user. The system captures energy consumption data at a detailed level, enabling the collection of one measurement every minute [38].
2) Edge Node: An edge computing node refers to a computing device, such as a GPU board or a PC that can be employed for specific activities or general purposes. The ownership of the energy firm may reside either at the client's facilities or with the customer themselves. The device establishes a connection with the Smart Meter using either a wireless protocol, such as Zigbee, or a wired connection, such as Ethernet. The main objective of this system is to store the energy consumption measurements obtained from the Smart Meter. Furthermore, it has the capability to interface with external sensors that can measure climatic variables such as relative humidity, temperature, wind speed, and other related characteristics. Alternatively, it can acquire less precise data from external meteorological data archives. The primary responsibility of the Edge Computing Node is to perform localized model training. It uses the previously mentioned collected data, in conjunction with calendar information. Additionally, it is linked to the Aggregator to exchange models and can communicate with the customer's end devices through specialized methods. Data visualization apps are commonly used for both data analysis and as a

definitive presenting method for communicating information.

3) Aggregator: An aggregator refers to a centralised server that is under the management of the energy company. The primary objective of the system is to collect local models, which have been trained by various Edge Computing Nodes, at the conclusion of each round and combine them. The result of this method is a comprehensive model that is then deployed on Edge Computing Nodes. The Aggregator establishes secure, uninterrupted connections with the Edge Computing Nodes to ease communication.

The proposed design can be executed with the involvement of alternative institutions, such as a statistical institute or healthcare infrastructure instead of the energy corporation. In this scenario, it is crucial for the consumer to have an energy meter sensor that is linked to their magnetothermic switch. This is because consumers are unable to quickly retrieve the data acquired by the Smart Meter.

RESULTS & DISCUSSION

Pre-processing

Data preprocessing include the activities of purifying, modifying, and organizing unprocessed data in preparation for analysis. Data management involves several methods, including data cleansing, data integration, data transformation, and data reduction. On the contrary, evaluation approaches are employed to appraise the effectiveness and excellence of a model or algorithm. These tactics involve measures such as

The investigation was conducted utilizing data obtained from dataset. An internet-based platform for the transfer and storage of information. Dataport gathers distinct electrical consumption data at the circuit level, with recording intervals ranging from one minute to one second, for around 800 households. Additionally, it offers details regarding the production of electricity through Photovoltaics and the process of charging Electrical Vehicles for a particular subset of these residences [39]. We selected a sample of 300 individuals from this dataset who possess comparable traits. The houses share the same architectural style, especially detached-family homes, and are situated in the same geographical area. The collection consists of recordings obtained at hourly intervals, covering the time span from 1st Jan, 2019 to 31st Mar, 2019. The weather fluctuations throughout this time frame are insignificant, hence the influence of seasonal factors can be disregarded in this study. The data of each client is handled to ensure it is prepared for subsequent analysis. Initially, we convert the data into a normalized range that ranges from 0 to 1. Afterwards, we convert the time series into sliding windows, where each window consists of 12-unit duration for retrospective analysis and 1-unit duration for prospective analysis. Ultimately, the data is partitioned into two distinct subsets: a training subset, comprising 80% of the data, and a testing subset, including the remaining 20%. The clients are divided into two distinct cohorts: 240 customers who actively engage in the federated learning process, and the remaining 60 consumers who are reserved for further evaluation of the model's ability to accommodate nonparticipating clients. The model's performance is assessed by employing the RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) metrics, which objectively measure the accuracy of the predictions. RMSE quantifies the error in terms of energy, while MAPE indicates the error's magnitude as a percentage relative to the real value. The formulas for calculating RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) are as follows:

$$MAPE = \frac{100\%}{F} \sum_{i=1}^{F} \left| \frac{z_i - \hat{z}_i}{z_i} \right|$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{F}(z_i - \hat{z}_i)^2}{N}}$$

Quantitative Outcomes

1) Assessed scenarios:

The many cases that were assessed are condensed in Table 1. As mentioned earlier, only a subset of customers participates in training the model during each iteration. In order to assess the impact of larger subsets, we updated the number of customers in the chosen subset in each round.

Table 1. Scenarios from dataset

Case	Customers in Subset	Iterations
1	10	1
2	25	1
3	10	5
4	25	5
5	10	5

Furthermore, we adjust the number of epochs for local training. The federated learning algorithm was executed for a total of 25 cycles across all situations.

2) Global model outcomes:

The assessed scenarios generated global models through the utilization of the federated learning procedure. The evaluation of these models is based on the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), as specified in Tables 2 and 3. The MAPE computations excluded consumption numbers that were null. Table 2 presents a succinct summary of the outcomes achieved by the participants engaged in various scenarios. The MAPE values reported for different models in Table 2 are considered acceptable, considering that our load forecast is specifically targeted at individual houses and has a narrow time frame of 1 hour. This level of accuracy is expected, given similar findings have been recorded in prior studies [40, 41]. Furthermore, these researches indicate that the accuracy of predictions is frequently restricted when making projections for brief timeframes. An important finding is that the global model is more suitable for certain clients compared to others, due to differences in their characteristics. In addition, it is observed that choosing a larger number of clients in each iteration yields advantages. However, if the cost of sending updates across the network is high, this difference can be offset by increasing the number of local training epochs. The outcomes are comparable when applied to the subgroup of clients who abstained from participating in the programme.

Table 2. Performance for global model of 240 contributing customers

Case over metric	1	2	3	4	5
MAPE	39.30%	38.29%	36.33%	37.81%	39.65%
RSME	0.495	0.468	0.466	0.473	0.455

Table 3. Performance for global model of 60 non-contributing customers

Case over metric	1	2	3	4	5
MAPE	43.88%	39.85%	38.08%	40.05%	41.19%
RSME	0.478	0.440	0.420	0.433	0.418

3) Customization Behavior:

This section is specifically focused on analyzing the impact of personalization on the performance of the models. Firstly, we assess the efficacy of conducting model re-training on each unique customer to determine if it yields superior outcomes. Following that, we use the identical procedure to the group of clients who did not partake in the training. Each client engaged in a process of retraining the models for a combined duration of 5 epochs. The results for the group of customers who participated in the training are shown in Table 4, while the results for the clients who did not take part are given in Table 5. It is clear that there is a widespread improvement in most of the models. Model 1 demonstrates a 4.57% rise in MAPE (Mean Absolute Percentage Error) for the clients who participated and a 4.28% enhancement for the clients who did not participate. Nevertheless, certain clients may not experience any improvement in performance during the process of retraining. The issue at hand, as previously said, is intricately linked to the precision of the historical data points. The utilization of models derived from the spending patterns of these clients leads to a significantly elevated Mean Absolute Percentage Error (MAPE), hence adversely impacting the overall average outcomes. These clients should be considered as anomalies; nonetheless, their analysis is outside the limits of this study.

Table 4. Performance metrics with customization of 240 contributing customers

Case over metric	1	2	3	4	5
MAPE	35.23%	35.29%	33.17%	33.04%	34.02%
RSME	0.440	0.441	0.421	0.421	0.435

Table 5. Performance metrics with customization of 60 contributing customers

Case over metric	1	2	3	4	5
MAPE	39.10%	39.61%	38.39%	37.92%	38.69%
RSME	0.406	0.406	0.406	0.399	0.404

In order to demonstrate the enhanced accuracy in predicting outcomes through personalization, we conducted a random selection of one customer (customer 3202) from the group of participants and another client (customer 7356) from the group of non-participants. We utilized the global model together with its associated individualized models. Figure 4 exhibits the present load profiles with the projected profiles executed as prediction. Both models accurately depict the general patterns identified in the consumption profiles. It is evident that we may develop robust models to comprehend the purchasing habits of a population by utilizing merely a portion of the customers inside it. To meet the precise accuracy demands of specific applications, the model can undergo retraining to create a customized model that closely aligns with the profile's curves. This will lead to more precise predictions. However, the forecasts derived from the global model may serve as a valuable initial point of reference for new clients who do not own sufficient data for customization.

The efficacy of the suggested methodology relies on the edge devices' capability to conduct on-site training. Contemporary Internet of Things (IoT) devices possess ample computational hardware to execute intricate machine learning models. However, the process of training a neural network is expected to have a substantial negative impact on the device's performance. Nevertheless, a limited quantity of nimble machine learning frameworks has surfaced, offering a robust foundation for forthcoming implementations. The precision of the models, even following customization, continues to fluctuate based on the unique characteristics of each user. To enhance the outcomes, it is advisable to integrate neural networks with supplementary techniques, such as utilizing pre-existing customer grouping based on factors that are unrelated to geographical proximity. Further inquiry is required to address the issue of outliers within this particular setting.

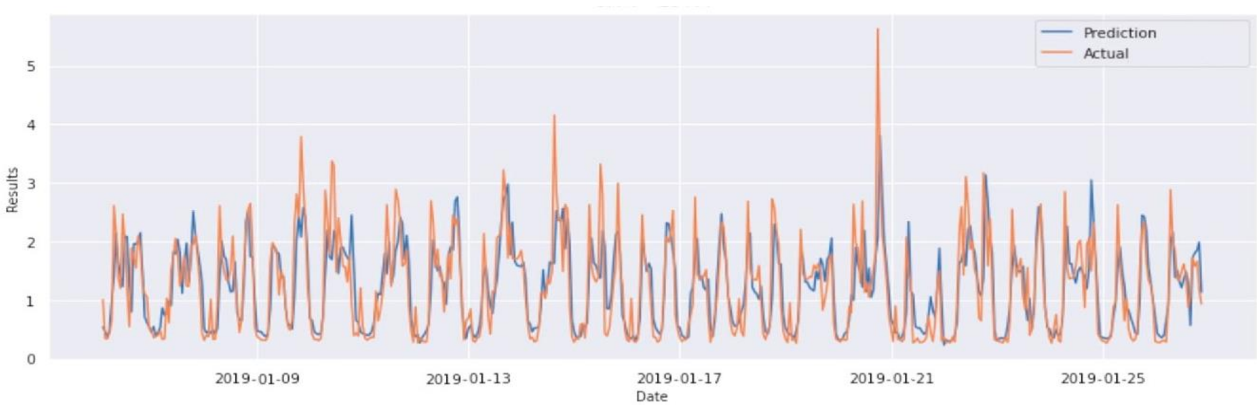

Figure 4. Forecasting for hourly power load in training the global model

Intersection of Smart Grids, Pharmaceutical & Healthcare Infrastructure Framework

By exploring the intersections between smart grid technologies and the pharmaceutical market, as well as health policy considerations, there is potential for innovative solutions that address challenges spanning multiple sectors. This collaborative approach can contribute to a more resilient, sustainable, and affordable healthcare system in line with broader health policy objectives. The research has potential relevance to the pharmaceutical market and health policy in several ways:

Operational Efficiency in Pharmaceutical Manufacturing - The pharmaceutical industry requires substantial energy resources for manufacturing processes. Implementing smart grid technologies, as suggested in the research, can optimize electricity usage in pharmaceutical plants. This could lead to cost savings and improved operational efficiency, indirectly impacting drug production costs.
Cost Reduction and Affordability of Medicines - Efficient energy management, driven by intelligent modeling and Federation Learning integration, may contribute to cost reductions in pharmaceutical production. Lowering operational costs in the pharmaceutical market can potentially lead to more affordable medicines, aligning with health policy goals to improve accessibility to essential drugs.

Environmental Sustainability and Regulatory Compliance - Smart grid technologies emphasize sustainability, which is increasingly becoming a focus in pharmaceutical manufacturing. Aligning electricity consumption with green practices could help pharmaceutical companies meet environmental regulations and sustainability requirements outlined in health policies.

Data Transparency and Regulatory Compliance - The integration of explainable artificial intelligence (Federation Learning) in the smart grid could contribute to transparent decision-making processes. In the pharmaceutical market, transparency is crucial for regulatory compliance. Insights gained from Federation Learning integration may inform strategies for maintaining transparency in pharmaceutical processes, aiding compliance with healthcare policies.

Resource Allocation and Health Policy Planning - Predictive modeling techniques employed in smart grids can be adapted for predicting electricity demands, in line with forecasting pharmaceutical production needs. This predictive capability can be valuable for health policymakers in planning resource allocation for pharmaceuticals, ensuring a consistent and reliable supply of essential medicines.

Resilience in Healthcare Supply Chain - Both pharmaceutical manufacturing and healthcare delivery are integral components of the broader health ecosystem. The resilient features of smart grid technologies can inspire strategies to enhance the resilience of the pharmaceutical supply chain. This is particularly relevant in the context of health policies aiming to strengthen healthcare infrastructures against disruptions.

Cross-Sector Collaboration and Policy Alignment - The research underscores the importance of interdisciplinary collaboration. In a similar vein, collaboration between the energy and pharmaceutical & healthcare sectors can lead to cross-cutting innovations. Policymakers may find value in fostering such collaborations to address shared challenges and align policies for a more sustainable and efficient healthcare ecosystem.

CONCLUSIONS

Forecasting the energy requirements for a short duration for an individual is a difficult undertaking due to the erratic nature of consumption patterns. This study presents a system framework that utilizes Edge computing and federated learning to address privacy and data diversity issues in short-term load forecasting in the smart grid. This research is a pioneering and groundbreaking investigation of federated learning in the smart grid domain. In contrast to centralised approaches, the suggested solution employs federated learning, which leverages edge devices for model training. This effectively mitigates security risks, limiting them exclusively to the device. We conducted a study to assess the efficacy of centralised and personalized models in federated situations. The simulation findings suggest that this strategy holds substantial promise in developing efficient models with decreased networking needs in comparison to a centralised model, all while maintaining the confidentiality of consumption data.

The synergy of smart grids, pharmaceuticals, and healthcare offers transformative potential for cost-effective drug production and improved accessibility. Smart grid technologies optimize electricity usage in pharmaceutical manufacturing, aligning with health policies for sustainable practices. Integrating intelligent modeling and Federation Learning supports cost reductions, contributing to affordable medicines as per health policy goals. Predictive modeling aids resource allocation, ensuring a reliable supply of essential drugs, while smart grid resilience inspires strategies for a robust pharmaceutical supply chain. Cross-sector collaboration is essential for innovative solutions, highlighting the need for joint efforts between energy and pharmaceutical sectors to create a more efficient and sustainable healthcare ecosystem.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

FUNDING

The author received no financial support for the research.

References

1. Elena Mocanu, Phuong H. Nguyen, Madeleine Gibescu, and Wil L. Kling. Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks, 6:91–99, June 2016.
2. W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu. Short-Term Residential Load Forecasting Based on Resident Behaviour Learning. IEEE Transactions on Power Systems, 33(1):1087–1088, January 2018.
3. Dhaou Said et al. Advanced scheduling protocol for electric vehicle home charging with time-of-use pricing. pages 6272–6276, June 2013. ISSN: 1938-1883.
4. A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. A. Khan, ‘‘Load forecasting, dynamic pricing and DSM in smart grid: A review,’’ Renew. Sustain. Energy Rev., vol. 54, pp. 1311–1322, Feb. 2016.
5. S. Fallah, R. Deo, M. Shojafar, M. Conti, and S. Shamshirband, ‘‘Com- putational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions,’’ Energies, vol. 11, no. 3, p. 596, Mar. 2018.

6. Y. Peng, Y. Wang, X. Lu, H. Li, D. Shi, Z. Wang, and J. Li, ‘‘Short-term load forecasting at different aggregation levels with predictability analysis,’’ in Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT Asia), May 2019, pp. 3385–3390.
7. B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, ‘‘Recent advances in the analysis of residential electricity consumption and applications of smart meter data,’’ Appl. Energy, vol. 208, pp. 402–427, Dec. 2017.
8. P. Zhang, X. Wu, X. Wang, and S. Bi, ‘‘Short-term load forecasting based on big data technologies,’’ CSEE J. Power Energy Syst., vol. 1, no. 3, pp. 59–67, Sep. 2015.
9. A. Almalaq and J. J. Zhang. Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings. IEEE Access, 7:1520–1531, 2019.
10. Salah Bouktif, Ali Fiaz, Ali Ouni, and Mohamed Adel Serhani. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †. Energies, 11(7):1636, July 2018.
11. Abderrahime Filali et al. Prediction-Based Switch Migration Scheduling for SDN Load Balancing. In ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pages 1–6, May 2019. ISSN: 1938-1883.
12. Yanbo Huang. Advances in Artificial Neural Networks – Method- ological Development and Application. Algorithms, 2(3):973–1007, September 2009.
13. B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett. Incorporating Practice Theory in Sub-Profile Models for Short Term Aggregated Residential Load Forecasting. IEEE Transactions on Smart Grid, 8(4):1591–1598, July 2017.
14. H. Shi, M. Xu, and R. Li. Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. IEEE Transactions on Smart Grid, 9(5):5271–5280, September 2018.
15. H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agu¨era y Arcas. Communication-efficient learning of deep networks from decentralized data.
16. F. Mármol, C. Sorge, O. Ugus, and G. Pérez, ‘‘Do not snoop my habits: Preserving privacy in the smart grid,’’ IEEE Commun. Mag., vol. 50, no. 5, pp. 166–172, May 2012.
17. M. R. Asghar, G. Dán, D. Miorandi, and I. Chlamtac, ‘‘Smart meter data privacy: A survey,’’ IEEE Commun. Surveys Tuts., vol. 19, no. 4, pp. 2820–2835, Jun. 2017.
18. W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, ‘‘Edge computing: A survey,’’ Future Gener. Comput. Syst., vol. 97, pp. 219–235, Aug. 2019.
19. G. Premsankar, M. D. Francesco, and T. Taleb, ‘‘Edge computing for the Internet of Things: A case study,’’ IEEE Internet Things J., vol. 5, no. 2, pp. 1275–1284, Apr. 2018.
20. Salah Bouktif, Ali Fiaz, Ali Ouni, and Mohamed Adel Serhani. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †. Energies, 11(7):1636, July 2018.
21. Jian Zheng, Cencen Xu, Ziang Zhang, and Xiaohua Li. Electric load forecasting in smart grids using Long-Short-Term-Memory based Recur- rent Neural Network. In 2017 51st Annual Conference on Information Sciences and Systems (CISS), pages 1–6, March 2017.
22. D. L. Marino, K. Amarasinghe, and M. Manic. Building energy load forecasting using Deep Neural Networks. In IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, pages 7046– 7051, October 2016.
23. Guangya Zhu, Tin-Tai Chow, and Norman Tse. Short-term load fore- casting coupled with weather profile generation methodology. Building Services Engineering Research and Technology, 39(3):310–327, May 2018.
24. Dhaou Said et al. Scheduling protocol with load managementfor EV charging. In 2014 IEEE Global Communications Conference, pages 362–367, December 2014. ISSN: 1930-529X.
25. Jihene Rezgui et al. Smart charge scheduling for EVs based on two-way communication. In 2017 IEEE International Conference on Communications (ICC), pages 1–6, May 2017. ISSN: 1938-1883.
26. P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin. Smart Grid Metering Networks:

A Survey on Security, Privacy and Open Research Issues. IEEE Communications Surveys Tutorials, pages 1– 1, 2019.
27. Y. Gong, Y. Cai, Y. Guo, and Y. Fang. A Privacy-Preserving Scheme for Incentive-Based Demand Response in the Smart Grid. IEEE Transactions on Smart Grid, 7(3):1304–1313, May 2016.
28. H. Park, H. Kim, K. Chun, J. Lee, S. Lim, and I. Yie. Untraceability of Group Signature Schemes based on Bilinear Mapping and Their Improvement. In Fourth International Conference on Information Technology (ITNG'07), pages 747–753, April 2007.
29. S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, November 1997.
30. Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Franc¸oise Beaufays, Sean Augenstein, Hubert Eichner, Chloe´ Kiddon, and Daniel Ramage. Federated learning for mobile keyboard prediction.
31. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, ''Communication-efficient learning of deep networks from decentralized data,'' in Proc. Int. Conf. Artif. Intell. Statist., 2017, pp. 1273–1282.
32. K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, H. Brendan McMahan, T. Van Overveldt, D. Petrou, D. Ramage, and J. Roselander, ''Towards federated learning at scale: System design,'' 2019, arXiv:1902.01046. [Online]. Available: http://arxiv.org/abs/1902.01046
33. T. Nishio and R. Yonetani, ''Client selection for federated learning with heterogeneous resources in mobile edge,'' in Proc. IEEE Int. Conf. Com- mun. (ICC), May 2019, pp. 1–7.
34. Khe Chai Sim, Petr Zadrazil, and Franc¸oise Beaufays. An Investiga- tion into On-Device Personalization of End-to-End Automatic Speech Recognition Models. In Interspeech 2019, pages 774–778. ISCA, September 2019.
35. Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone,H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical Secure Aggregation for Federated Learning on User-Held Data. arXiv:1611.04482 [cs, stat], November 2016. arXiv: 1611.04482.
36. Robin C. Geyer, Tassilo Klein, and Moin Nabi. Differentially Private Federated Learning: A Client Level Perspective. arXiv:1712.07557 [cs, stat], December 2017. arXiv: 1712.07557.
37. G. Dileep, ''A survey on smart grid technologies and applications,'' Renew. Energy, vol. 146, pp. 2589–2625, Feb. 2020.
38. Y. Wang, Q. Chen, T. Hong, and C. Kang, ''Review of smart meter data analytics: Applications, methodologies, and challenges,'' IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 3125–3148, May 2019.
39. Pecan street inc. dataport 2019 [online] https://dataport.pecanstreet.org/.
40. W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang. Short- Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. IEEE Transactions on Smart Grid, 10(1):841–851, January 2019.
41. Matthew Rowe, Timur Yunusov, Stephen Haben, William Holderbaum, and Ben Potter. The Real-Time Optimisation of DNO Owned Storage Devices on the LV Network for Peak Reduction. Energies, 7(6):3537– 3560, June 2014.