A Federated Multi-Task Learning Framework with Dual Attention Mechanisms for Smart Buildings

1[st] Mir Hassan
Department of Information Engineering and Computer Science
University of Trento, Italy
mir.hassan@unitn.it

2[nd] Kasim Sinan Yildrim
Department of Information Engineering and Computer Science
University of Trento, Italy
kasim.yildrim@unitn.it

3[rd] Giovanni Iacca
Department of Information Engineering and Computer Science
University of Trento, Italy
giovanni.iacca@unitn.it

***Abstract*—Accurate energy forecasting and occupancy detection are critical for energy management and occupant comfort in smart buildings. Centralized models, for example, based on LSTM and GRU, perform well with homogeneous datasets but face challenges in distributed IoT settings due to privacy concerns and sensor data heterogeneity (e.g., light, temperature, humidity, CO_2). In this paper, we propose Federated Forecasting (FedFor), a privacy-preserving framework integrating Federated Learning (FL), Multi-Task Learning (MTL), and dual attention mechanisms. FedFor simultaneously enhances energy forecasting and occupancy detection by leveraging task-specific and temporal patterns while keeping data localized. We evaluated our approach on two datasets, namely *ThingSpeak* and *Occupancy Detection Data (ODD)*. Results show that FedFor outperforms the baselines, achieving 99.11% accuracy for occupancy detection and reducing Mean Absolute Error (MAE) to 0.0097 for forecasting. Compared to state-of-the-art, our FedFor method is effective in addressing privacy concerns and data heterogeneity.**

***Index Terms*—Federated learning, Time-Series, Environmental Monitoring, Distributed Systems, Sensors**

I. Introduction

The increasing integration of IoT-enabled smart building systems promises substantial advancements in energy efficiency, cost reduction, and occupant comfort [1]. These systems use diverse sensors measuring temperature, CO_2, light, and humidity for applications such as energy usage forecasting and occupancy detection. Accurate forecasting and occupancy detection are critical for optimizing Heating, Ventilation, and Air Conditioning (HVAC) systems and reducing energy waste, especially during periods of partial occupancy [2], [3]. However, achieving these objectives simultaneously remains a significant challenge due to the heterogeneity of IoT data and stringent privacy concerns [4], [5]. Centralized machine learning (ML) approaches, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), are widely used for sequential data processing in these settings but face several limitations [3], [6].

The primary issues with existing solutions lie in privacy risks, model limitations, and task segregation. Centralized systems aggregate sensitive IoT data, such as occupancy patterns, creating vulnerabilities to inference attacks [7]. Furthermore, IoT sensor data is often highly heterogeneous across devices and locations, making it challenging for centralized models to generalize effectively [5], [8]. Furthermore, traditional forecasting models, like LSTM and GRU, struggle to capture long-range temporal dependencies in sequential data, resulting in inaccuracies in energy forecasting and occupancy detection [6], [9]. Finally, existing frameworks usually treat forecasting and occupancy detection as separate tasks, failing to exploit shared patterns between them, which could enhance performance [10], [11].

Federated Learning (FL) offers a promising alternative by decentralizing model training, where raw data remains on local devices, and only model updates are aggregated centrally [4]. This approach preserves privacy and has demonstrated potential in various applications, including power load forecasting and occupancy detection [12], [13]. However, existing FL frameworks lack mechanisms to model task-specific dependencies, such as the interdependencies between forecasting and occupancy detection tasks, which is a critical aspect in Multi-Task Learning (MTL) scenarios [14], [15]. Additionally, FL methods rarely incorporate advanced mechanisms like attention, which are crucial for capturing long-range temporal dependencies and task-specific patterns in heterogeneous data [9], [16].

To address these limitations, we propose the Federated Forecasting (FedFor) framework, which integrates FL with multi-task and dual attention mechanisms to simultaneously address energy forecasting and occupancy detection. Unlike existing methods, FedFor leverages shared patterns between tasks to enhance accuracy while preserving privacy through decentralized data processing. The dual attention mechanism improves the model's ability to focus on long-range temporal dependencies and task-specific features, making it well-suited for heterogeneous IoT data. Evaluations on real-world datasets demonstrate that FedFor achieves superior accuracy compared to state-of-the-art baselines, offering a robust solution for privacy-preserving MTL in smart buildings.

Contributions: The core contributions of this work are:

1) A novel Federated Forecasting (FedFor) framework that integrates FL with a dual attention mechanism, enabling simultaneous energy forecasting and occupancy detection.
2) A privacy-preserving, MTL approach that effectively captures task-specific and temporal patterns in heterogeneous IoT data.

3) Comprehensive evaluation on two real-world datasets, demonstrating superior accuracy, scalability, and robustness compared to state-of-the-art baselines.

The rest of the paper is structured as follows. In Section II, we provide a brief literature review, presenting an overview of the existing research and methodologies specifically in the domain of smart building management and FL. Section III introduces the FedFor framework, detailing the integration of FL with time-series forecasting. Section IV presents the experimental setup, including details on the datasets and the specific metrics used to evaluate the performance of the FedFor algorithm. Additionally, comparisons with existing methods and potential limitations are presented, to evaluate the FedFor algorithm comprehensively. In Section V, we engage in the discussion, analyzing the results, drawing comparisons with existing methods, and highlighting the significance of our findings. Finally, Section VI concludes the paper, summarizing key insights, contributions, and potential avenues for future research in this domain.

II. Related Work

In this section, we provide a brief review of the current literature related to smart city resource management, power load forecasting, and smart building management. We highlight key advancements in FL and related applications and identify the gaps that our proposed FedFor framework aims to address.

A. Smart Cities Resource Management

The deployment of IoT sensors in smart cities is increasing, due to their promise of improving safety and maintainability. In recent years, technology has recognized the enormous potential of artificial intelligence in smart cities, focusing on various sectors such as smart transportation, smart grid, smart governance, smart healthcare, smart management, smart industry, and smart monitoring systems [17]. Hanjri et al. [12] proposed a water consumption prediction system for smart cities using an FL approach that predicts the future measurement of home-based water consumption from past data. Another work [13] proposed a wind power forecasting method that uses an FL approach with a deep recurrent neural network. Their model predicts future wind generation values based on historical wind speed measurements. In addition, Farooq et al. [18] proposed a model that combines locally-trained models from eighteen clients to identify imminent flooding at specific stations, issuing alerts five days in advance. They used a local feed-forward neural network (FFNN) for flood prediction, incorporating multiple regional parameters to estimate water levels. Zhang et al. [19] introduced a framework for traffic speed forecasting using FL for privacy and preservation, called FASTGNN, which integrates a GNN-based model for local training and an FL strategy to protect the shared topological information. In another study [20], authors developed an FL-based spatial-temporal model for freight traffic speed forecasting.

B. Power Load Forecasting

Yang et al. [16] proposed a power load forecasting method based on VMD-FK-SecureBoost, integrating variational mode decomposition (VMD), federated k-means clustering (FK), and SecureBoost to improve prediction accuracy while preserving privacy. VMD decomposes sequences into modes clustered by FK, enabling SecureBoost to perform federated learning with enhanced privacy. Mao et al. [21] introduced CMULA-FL, a communication-efficient approach employing compressed model updates with lazy uploads, bidirectional quantization, and error compensation. CMULA-FL significantly reduces communication costs while maintaining high prediction accuracy and convergence speed. However, its complexity and sensitivity to hyperparameter tuning indicate the need for further theoretical analysis.

C. Smart buildings

Singh et al. [22] introduced occupancy estimation using ML algorithms such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), random forest (RF), and Support Vector Machine (SVM). They used low-cost and non-intrusive sensors measuring: CO_2, temperature, illumination, sound, and motion. Alina et al. [23] introduced a novel hybrid solution in the context of smart buildings that combines Particle Swarm Optimization (PSO) and Levenberg-Marquardt (LM) in an FL framework to enhance model training, reduce bandwidth usage by optimizing data upload processes, and outperform FedAvg [4] both in terms of accuracy and communication efficiency. In [24], the authors proposed an approach for Federated Transfer Learning (FTL) to handle buildings' data without compromising the sensitive information and clustered them according to specific characteristics. In [25], authors introduced a privacy-preserving framework that combines FL and transfer learning to evaluate HVAC system regulation in heterogeneous buildings. Furthermore, Khan et al. [26] presented an FL approach to predict multi-occupancy in buildings, by combining IoT sensors, edge computing, and ML to extract insights into the occupancy patterns.

D. Multi-Task Learning

E. Multi-Task Federated Learning

Hsu et al. [10] proposed a personalized FL algorithm combining multi-task learning (MTL) principles with adaptive clustering for non-IID IoT data. Although improving accuracy, its reliance on static clustering limits adaptability to dynamic environments. Zhang et al. [15] developed a federated MTL framework optimizing updates and clustering for heterogeneous data, but independent client-specific optimization overlooked inter-task correlations. Lu et al. [14] introduced FedHCA^2 to mitigate model incongruence through conflict-averse aggregation, yet its computational overhead hindered scalability for large multi-client systems. Ponomarenko-Timofeev et al. [27] proposed a federated SVM personalization method effective for classification and regression, but convergence was slower in highly heterogeneous networks due to kernel-based optimization inefficiencies. Corinzia and Buhmann [11] introduced VIRTUAL, a variational-inference-based federated MTL algorithm for non-IID data, outperforming baselines but incurring high computational costs and requiring Bayesian optimization, making it unsuitable for resource-constrained edge devices.

F. Limitations of Existing Works

A summary of the aforementioned works is provided in Table I. Despite significant advancements, several limitations remain. In smart city applications, works like Hanjri et al. [12] and Ahmadi et al. [13] effectively address prediction accuracy but struggle with adaptability to dynamic environments due to static clustering and centralized assumptions. Similarly, frameworks such as FASTGNN by Zhang et al. [19] preserve privacy but fail to capture interdependencies between related tasks like traffic forecasting and environmental analysis. In power load forecasting, approaches like CMULA-FL by Mao et al. [21] achieve communication efficiency but exhibit high complexity in hyperparameter tuning, limiting scalability in real-world IoT deployments. In smart building management, methods such as Singh et al. [22] and Wang et al. [25] focus on single-task optimization, overlooking shared patterns between tasks like energy forecasting and occupancy detection. MTL frameworks like FedHCA^2 by Lu et al. [14] and clustering-based FL by Hsu et al. [10] enhance task personalization but face scalability challenges due to static clustering assumptions and computational overhead. Privacy-preserving methods, such as those by Khan et al. [26], address data security but inadequately handle statistical heterogeneity in multi-occupancy scenarios. Advanced techniques like VIRTUAL by Corinzia and Buhmann [11] effectively manage non-convex models but are computationally prohibitive for edge devices with constrained resources. Moreover, methods like Zhang et al. [15] lack mechanisms to model inter-task correlations, limiting their ability to leverage shared dependencies.

These limitations underscore the need for a unified framework that integrates FL and MTL to optimize interdependent tasks like energy forecasting and occupancy detection.

III. Proposed FedFor Framework

A. Federated Learning

The FedFor framework employs an iterative FL process, which involves local training on the nodes and global aggregation on the server. The workflow is detailed as follows:

1. **Global Model Initialization:** The central server initializes the global model parameters θ^0.
2. **Model Distribution:** At the beginning of each communication round t, the server sends the current global model parameters θ^t to a selected subset of nodes.
3. **Local Training at Nodes:** Each selected node k updates its models based on the received parameters θ^t by training on its local dataset D_k for a specified number of epochs E, resulting in updated parameters θ_k^{t+1}.
4. **Model Update Transmission:** Nodes send their updated model parameters θ_k^{t+1} back to the server.
5. **Global Aggregation:** The server aggregates the received updates to form the new global model parameters θ^{t+1}. This is done by Federated Averaging (FedAvg) [4]:

$$\theta^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_k^{t+1}, \tag{1}$$

where n_k is the number of samples on node k, and $n = \sum_{k=1}^{K} n_k$.
6. **Iteration:** Steps 2 to 5 are repeated for multiple rounds until the convergence criteria are met.

This workflow ensures that the global model benefits from the diverse data across all nodes while maintaining data privacy and reducing communication overhead.

B. Neural Network Architecture with Dual Attention

The core of the FedFor framework is a neural network model enhanced with a dual attention mechanism to improve forecasting accuracy and occupancy detection.

1) Shared LSTM Layers: At the heart of the model are LSTM layers that capture temporal dependencies in the time-series data. The LSTM layers process input sequences $X = \{x_1, x_2, \ldots, x_T\}$, producing a sequence of hidden states $H = \{h_1, h_2, \ldots, h_T\}$, where each h_t encapsulates information from time step t.

2) Temporal Attention Mechanism: To enable the model to focus on the most relevant time steps, a temporal attention mechanism is employed, with weights α_t computed as:

$$e_t = v^\top \tanh(W h_t + b), \quad \alpha_t = \frac{\exp(e_t)}{\sum_{t'=1}^{T} \exp(e_{t'})} \tag{2}$$

where W and v are learnable parameters, and e_t represents the relevance score of hidden state h_t. The context vector c is then obtained as:

$$c = \sum_{t=1}^{T} \alpha_t h_t. \tag{3}$$

This mechanism allows the model to weigh different time steps according to their importance for the prediction task.

3) Task-Specific Attention and Output Layers: In the MTL setting, separate attention mechanisms and output layers are used for each task:

- **Occupancy Detection Task:** Uses a sigmoid activation function with Binary Cross-Entropy Loss. The context vector c_{occ} is derived using task-specific attention weights.
- **Energy Forecasting Task:** Uses a linear activation function with Mean Squared Error Loss (MSE). The context vector c_{for} is similarly obtained.

4) Cross-Task Attention Mechanism: The cross-task attention mechanism leverages the interdependence between tasks by integrating contextual information from one task into the attention computation of the other. This is formalized as:

$$\beta_t^{(j)} = \frac{\exp(g(h_t^{(j)}, c^{(i)}))}{\sum_{t'=1}^{T} \exp(g(h_{t'}^{(j)}, c^{(i)}))}, \tag{4}$$

TABLE I: Comparison of FedFor with state-of-the-art works.

Ref.	Application Domain	Data Type	Key Benefits	AM*	Forecasting	Occupancy Detection	Multi-tasking
[10]	IoT Systems	Time series	Adaptive clustering for non-IID data, improved model accuracy.	✗	✗	✗	✓
[15]	Distributed Systems	Time series	Optimizes convergence and accuracy for dynamic environments.	✗	✗	✗	✓
[14]	IoT Systems	Time series	Hyper conflict-averse aggregation enhances heterogeneity handling.	●	✗	✗	✓
[27]	Heterogeneous Networks	Time series	Effective for classification/regression in non-IID settings.	✗	✗	✗	✓
[11]	Distributed Systems	Time series	Variational inference for non-convex optimization.	✗	✗	✗	✓
[13]	Wind Power Forecasting	Time series	Utilizes FL for better generalization.	✓	✓	✗	✗
[12]	Water Management Forecasting	Time series	Effective for monthly household data.	●	✓	✗	✗
[16]	Power Load Forecasting	Time series	Good short-term accuracy.	✗	✓	✗	✗
[18]	Flood Forecasting	Time series	Effective for regional data.	✗	✓	✗	✗
[21]	Power Load Forecasting in IoT	Time series	Communication-efficient.	✗	✓	✗	✗
[28]	Smart Traffic Management	Time series	Captures diverse traffic patterns.	✓	✓	✗	✗
[19]	Traffic Speed Forecasting	Time series	Protects topological information.	✓	✓	✗	✗
[20]	Traffic Speed Forecasting	Time series	High forecasting accuracy and robustness.	✓	✓	✗	✗
[29]	Smart Transportation System	Time series	Enhances GPS positioning accuracy.	✓	✓	✗	✗
[30]	Smart Building	Numerical	Effective with limited historical data.	✗	✗	✓	✗
[24]	Smart Building	Energy consumption	Efficient model training.	✗	✗	✓	✗
[31]	Smart Building	Energy Measurement	Enhances operational efficiency.	✗	✗	✓	✗
[25]	Smart Building	Energy Measurement	Privacy-preserving and efficient.	✗	✗	✓	✗
[26]	Smart Building	Time series	Optimizes HVAC systems.	✗	✗	✓	✗
[22]	Smart Building	Time series	Accurate occupancy prediction.	✗	✗	✓	✗
FedFor	Smart Building	Time series	Multi-task model and dual attention improve prediction accuracy.	✓	✓	✓	✓

*AM (Attention Mechanism): ✓Directly applicable (Yes) ✗Not applicable (No) ●Indirectly applicable (Partial)

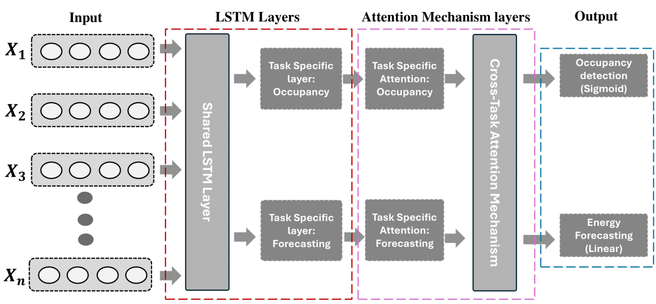

Fig. 1: Multi-task learning architecture with shared LSTM layers and task-specific attention mechanisms. Cross-task attention facilitates shared feature learning, enhancing performance.

where g is a scoring function, $h_t^{(j)}$ is the hidden state for task j, and $c^{(i)}$ is the context vector from task i. This mechanism enhances the model's ability to capture relevant features that are informative across tasks.

Figure 1 illustrates the architecture, showcasing the shared LSTM layers, the task-specific attention mechanisms, and the integration of cross-task attentions.

C. Federated Forecasting Algorithm

The complete FedFor algorithm integrates the FL process described above with the proposed neural network model, ensuring privacy preservation and efficient training. Algorithm 1 outlines the steps involved in the algorithm.

Algorithm 1 FedFor: Federated Forecasting Algorithm

Require: Initial global model parameters θ^0, local datasets D_k, total communication rounds T, number of local epochs E, learning rate η

Ensure: Final global model parameters θ^T

1: **for** each round $t = 0, 1, \ldots, T-1$ **do**
2: Server selects a subset $\mathcal{K}_t$ of nodes
3: Server sends θ^t to all nodes $k \in \mathcal{K}_t$
4: **for** each node $k \in \mathcal{K}_t$ **in parallel do**
5: **Local Training:**
6: Initialize $\theta_k^{t,0} = \theta^t$
7: **for** epoch $e = 1$ to E **do**
8: Compute gradients $\nabla L_k(\theta_k^{t,e-1})$ using dual attention mechanisms
9: Update parameters:

$$\theta_k^{t,e} \leftarrow \theta_k^{t,e-1} - \eta \nabla L_k(\theta_k^{t,e-1})$$

10: **end for**
11: Set $\theta_k^{t+1} = \theta_k^{t,E}$
12: Send θ_k^{t+1} to the server
13: **end for**
14: **Global Aggregation:**
15: Update global model: $\theta^{t+1} = \sum_{k \in \mathcal{K}_t} \frac{n_k}{n} \theta_k^{t+1}$
16: **end for**

IV. Experiments and Results

In this section, we evaluate FedFor and compare it with baseline and state-of-the-art MTL models for energy and occupancy detection.

A. Datasets

In our experiments, we use the *ThingSpeak* [32] and *ODD* (Occupancy Detection Data) [33] datasets. These datasets encompass critical environmental variables, e.g., humidity, temperature, and light intensity, which are integral to smart building applications.

The *ThingSpeak* dataset was acquired through a custom-built IoT weather station, designed to capture real-time environmental data. This dataset includes measurements of humidity (expressed in percentage), temperature (in degrees Celsius), and light intensity (in lux).

The *ODD* dataset, on the other hand, was collected from a controlled office environment, leveraging an array of advanced sensors to record additional features, such as CO_2 levels (measured in parts per million, PPM) and occupancy status, the latter being verified through camera-based labeling.

B. Baselines

To evaluate FedFor, we used GRU and LSTM models, with and without attention mechanisms, as baselines. GRU offers faster training and lower computational overhead due to its simpler architecture [34], while LSTM better models long-term dependencies, although with a higher computational cost [35]. Adding attention mechanisms further improves their focus on relevant input features, enhancing predictive performance. However, these models do not inherently address key FL challenges such as privacy, scalability, and inter-task collaboration.

We also included as baselines Federated Multi-Task Learning (FMTL) [15] and Personalized Federated Learning (PFL) [10] as representatives of state-of-the-art MTL methods. FMTL emphasizes dynamic clustering and task-specific aggregation, and has been implemented with the same LSTM architecture used in FedFor, for consistency. PFL incorporates adaptive clustering for non-IID data, faithfully reproduced following their original methodology.

These baselines ensure a comprehensive and fair comparison, highlighting FedFor's strengths in addressing FL challenges while maintaining competitive performance in time-series forecasting and MTL settings.

C. Experimental Configuration

Experiments were conducted on an Azure Lab Virtual Machine[1] with Intel Xeon E5-2690 v4 CPUs (Architecture: x86_64, 6 cores), TensorFlow-GPU 2.6.0, Keras 2.8.0, Flower Federated Learning[2], and Python 3.9. The computational setup leveraged Keras's modular interface integrated with TensorFlow, simplifying the implementation and integration of complex neural architectures. The loss was computed using Mean Squared Error (MSE), with optimization performed via the Adam optimizer [36]. A sigmoid activation function in the final layer was specifically chosen for time-series data.

D. Evaluation Metrics

To evaluate model performance, we utilized task-specific metrics tailored to forecasting and occupancy detection. For forecasting, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were employed to measure prediction accuracy and emphasize error magnitude. For occupancy detection, the F1-Score was used to balance precision and recall, ensuring robust evaluation in imbalanced scenarios. These metrics provided a precise and comprehensive assessment that is aligned with our objectives.

E. Results

In the following, we present the results of two experiments conducted to evaluate the performance of our models: Experiment 1 focuses on single-task learning, while Experiment 2 examines the MTL setting. In both experiments, we perform energy forecasting and occupancy detection, with distinct configurations for each task.

[1]https://labs.azure.com/virtualmachines

[2]https://flower.dev

1) Experiment 1: Single-Task Learning: In this experiment, we evaluate the performance of the models on two tasks: energy forecasting and occupancy detection, each treated as a separate task. We use the *ThingSpeak* and *ODD* datasets for these evaluations. Detailed results for the single-task learning experiment are provided in Table II. For the LSTM-Attention models, with and without FL, we also report the confusion matrices in Figure 2.

a) Energy Forecasting: For the *ThingSpeak* dataset, the GRU model in its base configuration achieved an MAE of 0.0391 and an RMSE of 0.0546. When enhanced with attention, these metrics improved slightly, particularly for RMSE (0.0499). The most pronounced improvements were observed on the *ODD* dataset, where the attention-equipped GRU reduced the MAE to 0.0096 and the RMSE to 0.0184, significantly outperforming its base counterpart (MAE: 0.0628, RMSE: 0.0885).

Similarly, LSTM models demonstrated substantial gains with the addition of attention mechanisms. On the *ThingSpeak* dataset, the base LSTM achieved an MAE of 0.0177 and an RMSE of 0.0266. Incorporating attention further refined these metrics to an MAE of 0.0185 and RMSE of 0.0274. For the *ODD* dataset, attention-equipped LSTM achieved its best performance with an MAE of 0.0119 and an RMSE of 0.0212.

In the federated setting, the FL GRU and FL LSTM models showed similar trends. The FL GRU model, when augmented with attention, achieved an MAE of 0.0292 and an RMSE of 0.0382 on the *ThingSpeak* dataset, outperforming its base configuration (MAE: 0.0446, RMSE: 0.0577). On the *ODD* dataset, the FL GRU with attention recorded an MAE of 0.0335 and an RMSE of 0.0409. Similarly, the FL LSTM model with attention attained an MAE of 0.0320 and an RMSE of 0.0414 for the *ThingSpeak* dataset, with corresponding values of 0.0400 and 0.0515 for the *ODD* dataset.

b) Occupancy Detection: For occupancy detection, we obtained excellent results on the *ODD* dataset. The non-federated LSTM with attention achieved an accuracy of 98.71%, with precision, recall, and F1 scores of 99.80%, 94.15%, and 97.45%, respectively. This performance demonstrates the model's ability to effectively capture the temporal dependencies necessary for occupancy classification.

In the federated setting, the FL LSTM with attention achieved an accuracy of 90.50%, with precision, recall, and F1 scores of 90.00%, 90.50%, and 90.20%, respectively. While slightly lower than the centralized model, this result validates the feasibility of FL in decentralized environments, where privacy is a concern.

2) Experiment 2: Multi-Task Learning Model: In this experiment, we evaluate the performance of the multi-task model, which simultaneously performs energy forecasting and occupancy detection. This configuration leverages a dual-task learning approach, allowing the model to share information across tasks and potentially improve overall performance.

Looking at the results reported in Table III, it can be seen that the dual-task LSTM model with attention achieved outstanding performance. On the *ODD* dataset, the model

TABLE II: Single-task performance for energy forecasting and occupancy detection on *ThingSpeak* and *ODD* datasets.

Model Type	**Configuration**	**ThingSpeak**		**ODD**			
		MAE	**RMSE**	**MAE**	**RMSE**	**Accuracy (%)**	**F1 Score**
GRU	Base	0.0391 ± 0.0039	0.0546 ± 0.0046	0.0628 ± 0.0073	0.0885 ± 0.0111	83.39 ± 1.01	81.21 ± 0.07
	Attention	0.0377 ± 0.0019	0.0499 ± 0.0023	0.0096 ± 0.0012	0.0184 ± 0.0015	96.53 ± 1.01	96.10 ± 0.03
LSTM	Base	0.0177 ± 0.0019	0.0266 ± 0.0024	0.0215 ± 0.0113	0.0322 ± 0.0154	95.50 ± 0.02	93.25 ± 0.05
	Attention	0.0185 ± 0.0016	0.0274 ± 0.0024	0.0119 ± 0.0014	0.0212 ± 0.0020	98.71 ± 0.01	97.45 ± 0.01
FL GRU	Base	0.0446 ± 0.0307	0.0577 ± 0.0395	0.0483 ± 0.0512	0.0578 ± 0.0603	85.60 ± 2.43	86.71 ± 4.41
	Attention	0.0292 ± 0.0074	0.0382 ± 0.0069	0.0335 ± 0.0089	0.0409 ± 0.0102	87.00 ± 1.00	85.81 ± 1.20
FL LSTM	Base	0.0392 ± 0.0116	0.0485 ± 0.0135	0.0496 ± 0.0061	0.0623 ± 0.0094	89.91 ± 3.12	80.21 ± 4.60
	Attention	0.0320 ± 0.0055	0.0414 ± 0.0057	0.0400 ± 0.0103	0.0515 ± 0.0147	90.50 ± 1.01	90.20 ± 0.14

achieved an MAE of 0.0097 and an RMSE of 0.0134, outperforming both the single-task and baseline state-of-the-art models [15] and [14] in all metrics.

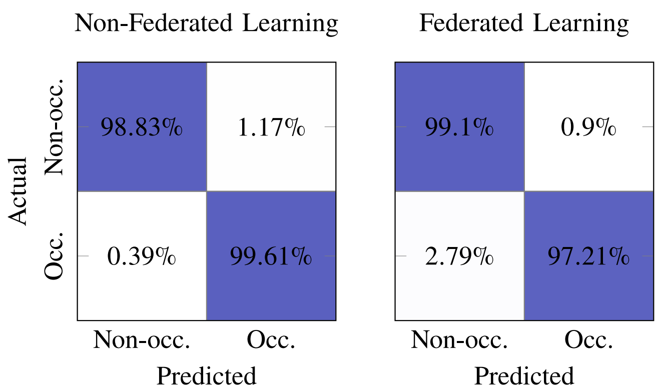

Fig. 2: Confusion matrices using the LSTM-Attention model for Non-Federated and Federated Learning settings.

V. Discussion

The results underscore the transformative potential of FedFor in advancing smart building analytics by integrating FL, MTL, and dual attention. FedFor consistently improves performance across centralized and federated settings. The dual attention mechanism enables precise temporal and task-specific feature extraction, enhancing predictive accuracy for energy forecasting and occupancy detection, especially in dynamic, non-stationary environments. By preserving data privacy and maintaining robust performance, FedFor strikes an effective balance for deployment in privacy-sensitive applications. FedFor sets a new standard in smart building management by leveraging inter-task synergies, improving operational efficiency. Its superior performance, with 99.11% occupancy detection accuracy and an MAE of 0.0097, highlights its effectiveness in addressing forecasting and privacy challenges, outperforming state-of-the-art baselines.

A. Limitations and Challenges

While FedFor demonstrates high accuracy and predictive capability, it faces notable challenges in implementation due to its computational complexity. These challenges are particularly pronounced in resource-constrained environments, such as edge devices or IoT ecosystems. Optimizing the model architecture for efficiency, leveraging distributed edge computing to offload processing, and designing lightweight variants of the models could mitigate these issues. Future efforts should prioritize these strategies to enhance the applicability of the proposed solution and facilitate its broader adoption.

VI. Conclusion

This study introduced the FedFor framework, a novel approach combining attention mechanisms with FL and MTL to address the dual challenges of forecasting and energy management in smart building systems. The proposed framework demonstrated superior performance, achieving up to 99.11% prediction accuracy, thereby validating its effectiveness in decentralized and resource-constrained environments. The integration of attention mechanisms enhanced the model's ability to capture long-range temporal dependencies, enabling precise occupancy and energy usage forecasting. The FL approach preserved data privacy while maintaining high predictive accuracy, positioning FedFor as a scalable solution for real-world applications. By enabling collaborative learning across distributed nodes, and leveraging the knowledge from task interdependencies following the MTL paradigm, the framework can support energy-efficient management strategies.

Acknowledgment

Funded by the European Union | European Innovation Council

Funded by the European Union (project no. 101071179). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EISMEA. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] A. O. Windapo and A. Moghayedi, "Adoption of smart technologies and circular economy performance of buildings," *Built Environment Project and Asset Management*, vol. 10, no. 4, pp. 585–601, 2020.

[2] N. Li, G. Calis, and B. Becerik-Gerber, "Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations," *Automation in Construction*, vol. 24, pp. 89–99, 2012.

TABLE III: Multi-task performance for energy forecasting and occupancy detection on the *ODD* dataset.

Model Type	Configuration	MAE	RMSE	Accuracy (%)	F1 Score
SOTA	Zhang et al. [15] (base)	0.0404 ± 0.0054	0.0608 ± 0.0101	96.71 ± 0.12	96.21 ± 0.16
	Lu et al. [14]	0.0181 ± 0.0022	0.0293 ± 0.0033	98.88 ± 0.00	97.34 ± 0.00
Multi-task (ours)	Dual-Task LSTM (base) + Attention	**0.0108 ± 0.0012**	**0.0156 ± 0.0014**	**98.91 ± 0.16**	**97.45 ± 0.37**
	Dual-Task LSTM (base, federated) + Attention	**0.0097 ± 0.0010**	**0.0134 ± 0.0012**	**99.11 ± 0.07**	**97.88 ± 0.16**

[3] A. S. Telaga, S. R. Aisy, and S. P. Arumi, "Occupancy detection through machine learning and environmental data," in *AIP Conference Proceedings*, vol. 2926. AIP Publishing, 2024.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.

[5] S. Hu, P. Wang, C. Hoare, and J. O'Donnell, "Building Occupancy Detection and Localization Using CCTV Camera and Deep Learning," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 597–608, 2022.

[6] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," *Neurocomputing*, vol. 332, pp. 320–327, 2019.

[7] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for internet of things: Recent advances, taxonomy, and open challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1759–1799, 2021.

[8] R. P. Romansky and I. S. Noninska, "Challenges of the digital age for privacy and personal data protection," *Mathematical Biosciences and Engineering*, vol. 17, no. 5, pp. 5288–5303, 2020.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[10] H.-Y. Hsu, K. H. Keoy, J.-R. Chen, H.-C. Chao, and C.-F. Lai, "Personalized Federated Learning Algorithm with Adaptive Clustering for Non-IID IoT Data Incorporating Multi-Task Learning and Neural Network Model Characteristics," *Sensors*, vol. 23, no. 22, p. 9016, 2023.

[11] L. Corinzia, A. Beuret, and J. M. Buhmann, "Variational federated multi-task learning," 2019, arXiv preprint arXiv:1906.06268.

[12] M. E. Hanjri, H. Kabbaj, A. Kobbane, and A. Abouaomar, "Federated Learning for Water Consumption Forecasting in Smart Cities," 2023, arXiv preprint arXiv: 2301.13036.

[13] A. Ahmadi, M. Talaei, M. Sadipour, A. M. Amani, and M. Jalili, "Deep federated learning-based privacy-preserving wind power forecasting," *IEEE Access*, vol. 11, pp. 39 521–39 530, 2022.

[14] Y. Lu, S. Huang, Y. Yang, S. Sirejiding, Y. Ding, and H. Lu, "FedHCA2: Towards Hetero-Client Federated Multi-Task Learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5599–5609.

[15] H. Zhang, M. Tao, Y. Shi, X. Bi, and K. B. Letaief, "Federated multi-task learning with non-stationary and heterogeneous data in wireless networks," *IEEE Transactions on Wireless Communications*, 2023.

[16] Y. Yang, Z. Wang, S. Zhao, and J. Wu, "An integrated federated learning algorithm for short-term load forecasting," *Electric Power Systems Research*, vol. 214, p. 108830, 2023.

[17] S. Pandya, G. Srivastava, R. Jhaveri, M. R. Babu, S. Bhattacharya, P. K. R. Maddikunta, S. Mastorakis, M. J. Piran, and T. R. Gadekallu, "Federated learning for smart cities: A comprehensive survey," *Sustainable Energy Technologies and Assessments*, vol. 55, p. 102987, 2023.

[18] M. S. Farooq, R. Tehseen, J. N. Qureshi, U. Omer, R. Yaqoob, H. A. Tanweer, and Z. Atal, "FFM: Flood Forecasting Model Using Federated Learning," *IEEE Access*, vol. 11, pp. 24 472–24 483, 2023.

[19] C. Zhang, S. Zhang, J. J. Q. Yu, and S. Yu, "FASTGNN: A Topological Information Protected Federated Learning Approach for Traffic Speed Forecasting," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8464–8474, 2021.

[20] X. Shen, J. Chen, S. Zhu, and R. Yan, "A decentralized federated learning-based spatial–temporal model for freight traffic speed forecasting," *Expert Systems with Applications*, vol. 238, p. 122302, 2024.

[21] Z. Mao, H. Li, Z. Huang, C. Yang, Y. Li, and Z. Zhou, "Communication-Efficient Federated Learning for Power Load Forecasting in Electric IoTs," *IEEE Access*, vol. 11, pp. 47 930–47 939, 2023.

[22] A. P. Singh, V. Jain, S. Chaudhari, F. A. Kraemer, S. Werner, and V. Garg, "Machine Learning-Based Occupancy Estimation Using Multivariate Sensor Nodes," in *IEEE Global Communications Conference Workshops*, 2018, pp. 1–6.

[23] A. Abboud, M.-E.-A. Brahmia, A. Abouaissa, A. Shahin, and R. Mazraani, "A Hybrid Aggregation Approach for Federated Learning to Improve Energy Consumption in Smart Buildings," in *International Wireless Communications and Mobile Computing*, 2023, pp. 854–859.

[24] E. M. Campos, A. G. Vidal, J. L. Hernández Ramos, and A. Skarmeta, "Federated Transfer Learning for Energy Efficiency in Smart Buildings," in *IEEE Conference on Computer Communications Workshops*, 2023, pp. 1–6.

[25] Z. Wang, P. Yu, and H. Zhang, "Privacy-Preserving Regulation Capacity Evaluation for HVAC Systems in Heterogeneous Buildings Based on Federated Learning and Transfer Learning," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3535–3549, 2023.

[26] I. Khan, F. C. Delicato, E. Greco, M. Guarascio, A. Guerrieri, and G. Spezzano, "Occupancy Prediction in Multi-Occupant IoT Environments Leveraging Federated Learning," in *IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress*, 2023, pp. 0036–0043.

[27] A. Ponomarenko-Timofeev, O. Galinina, R. Balakrishnan, N. Himayat, S. Andreev, and Y. Koucheryavy, "Multi-Task Model Personalization for Federated Supervised SVM in Heterogeneous Networks," *arXiv preprint arXiv:2303.10254*, 2023.

[28] C. Zhang, S. Dang, B. Shihada, and M.-S. Alouini, "Dual Attention-Based Federated Learning for Wireless Traffic Prediction," in *IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[29] X. Zhou, Q. Yang, Q. Liu, W. Liang, K. Wang, Z. Liu, J. Ma, and Q. Jin, "Spatial–Temporal Federated Transfer Learning with multi-sensor data fusion for cooperative positioning," *Information Fusion*, vol. 105, p. 102182, 2024.

[30] M. Khalil, S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Transfer learning approach for occupancy prediction in smart buildings," in *International Renewable Engineering Conference*. IEEE, 2021, pp. 1–6.

[31] Y. Zhao, W. Zeiler, G. Boxem, and T. Labeodan, "Virtual occupancy sensors for real-time occupancy information in buildings," *Building and Environment*, vol. 93, pp. 9–20, 2015.

[32] O. Olanrele, O. Ajayi, and O. Adeyemi, "Feeds of humidity, temperature and light intensity downloaded from ThingSpeak," 2022.

[33] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016.

[34] W. Shu, K. Cai, and N. N. Xiong, "A Short-Term Traffic Flow Prediction Model Based on an Improved Gate Recurrent Unit Neural Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16 654–16 665, 2022.

[35] X. Wang, X. Wang, M. He, M. Zhang, and Z. Lu, "Spatial-Temporal Graph Model Based on Attention Mechanism for Anomalous IoT Intrusion Detection," *IEEE Transactions on Industrial Informatics*, 2023.

[36] S. Liu, H. Ji, and M. C. Wang, "Nonpooling convolutional neural network forecasting for seasonal time series with trends," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2879–2888, 2019.