Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Exploring Lightweight Federated Learning for Distributed Load Forecasting

Abhishek Duttagupta
Jin Zhao
Shreejith Shanker

Electronic & Electrical Engineering
Trinity College, Dublin

Introduction – Electricity Load Forecasting

- Electricity demand forecasting is a crucial function in the energy industry
  - Allows both consumers and distribution system operators to optimize the usage and distribution of electricity in real-time[1]
  - Aims to balance the supply and demand curve
  - Maintain the stability of the electric grid across the entire network[2]
- Statistical methods like ARIMA etc. have played significant roles in predicting the nature of the load
  - Problem : Non-linear nature of consumer load consumption pattern
- With the advent of Deep Learning (DL), there has been significant breakthrough due to its ability to learn non linear patterns accurately.
  - Various DL[3] and hybrid Models[4], have emerged offering more promising results with minimal error rate

[1] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," International journal of forecasting, vol. 30, no. 4, pp. 1030–1081, 2014.
[2] G. D´ıaz, et al. "Prediction and explanation of the formation of the Spanish day-ahead electricity price through machine learning regression," Applied Energy, vol. 239, pp. 610–625, 2019.
[3] [3] A. R. Khan et al. Khan,"Load forecasting, dynamic pricing and dsm in smart grid: A review, "Renewable and Sustainable Energy Reviews, vol. 54, pp. 1311–1322, 2016.
[4] H. Saxena, et al. "A hybrid machine learning model for forecasting a billing period's peak electric load days," International Journal of Forecasting, vol. 35, no. 4, pp. 1288–1303, 2019.

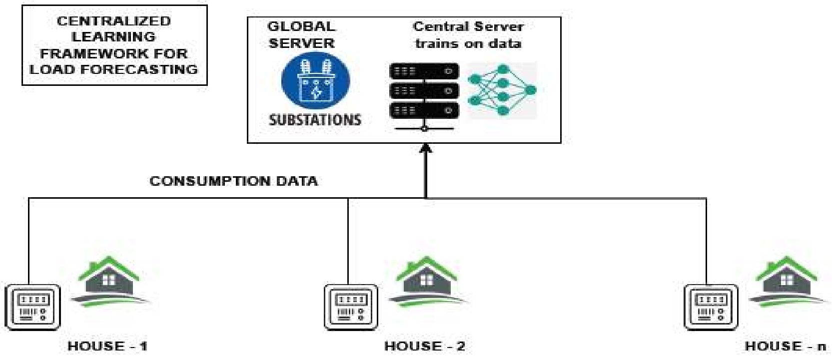
Introduction – Problems associated with DL Models

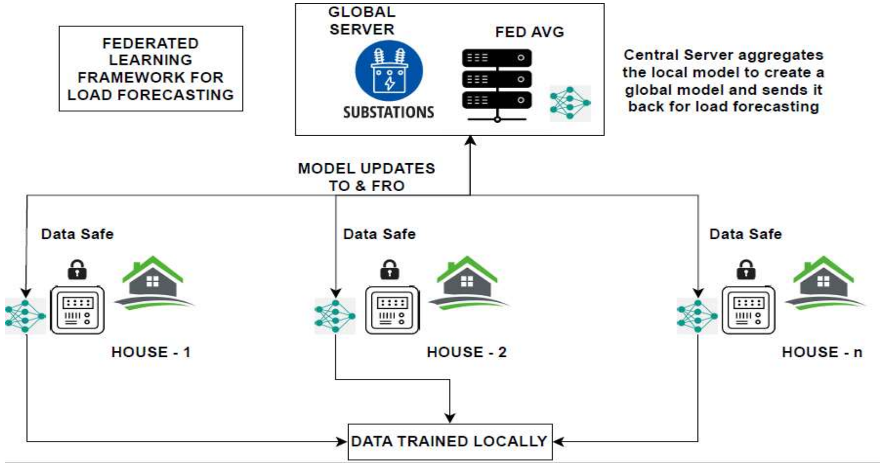
- DL models, while successful in improving accuracy and learning non linear patterns
  - Data Hunger and Complexity
    - Require vast amount of data
    - Models increasingly complex over time
  - Privacy Concerns
    - Accumulation of data at centralized location within distribution grid for model training raises issues
    - Sharing sensitive load usage data has become a concern in context of data security[5]
    - Non-intrusive analysis of granular load curves can provide information about appliances, usage patterns and hence personal details about consumers[6]

[5] G. Eibl and D. Engel, “Influence of data granularity on smart meter privacy,” IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 930–939, 2015.
[6] L. AlAbdulkarim et al. “Impact of privacy concerns on consumers’ acceptance of smart metering in the Netherlands,” in 2011 International Conference on Networking, Sensing and Control, 2011, pp. 287–292.

Background - Why Federated Learning ?

- Federated learning (FL)[7,8] is a promising scheme to address the challenges on sensitive data
  - Learning becomes possible at the edge
  - Only local model updates being aggregated at a global level to improve prediction
  - Eradicates the need to share sensitive user data



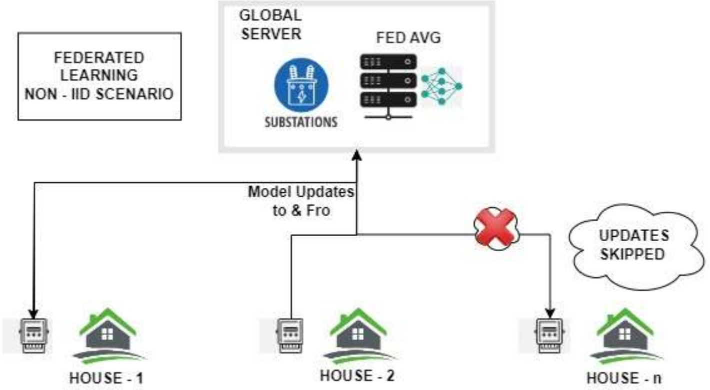


[7] B. Hu, Y. Gao, L. Liu, and H. Ma, “Federated region-learning: An edge computing based framework for urban environment sensing,” in 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018, pp. 1–7.
[8] S. R. Pfohl, A. M. Dai, and K. Heller, “Federated and differentially private learning for electronic health records,” arXiv preprint arXiv:1911.05861, 2019.

Background - Light weighted models for FL

- While FL shows promise, previous works with complex models, have achieved accurate forecasting performance[9,10]
  - The complexity, longer training time of local models as well as their high energy consumption is a barrier to their widespread deployment in constrained devices such as smart meters
  - Non-I.I.D scenario remains untested, where few global updates are skipped, considering connectivity issues between local and global devices
- Use of light weight models becomes ideal, if in a distributed setup they are able to achieve similar or better performance compared to the complex models

APPROACHES	MODEL	PARAMETERS
N. Gholizadeh	LSTM + DNN	11,500
M. Savi	LSTM + DNN	7,500
Our Approach	FNN	5,569



[9] N. Gholizadeh et al, "Federated learning with hyper parameter-based clustering for electrical load forecasting," Internet of Things, vol. 17, p. 100470, 2022.
[10] M. Savi et al, "Short-term energy consumption forecasting at the edge: A federated learning approach," IEEE Access, vol. 9, pp. 95 949–95 969, 2021.

Contributions of the paper

- Can a lightweight feed forward model within an FL framework achieve identical or superior short term load forecasting performance, even while considering limited device connections ?
- Combining the FL flow with clustering to achieve a better global representation within each cluster, leading to improved forecasting performance.
- Evaluate our approach and demonstrate that lightweight models are sufficient to achieve similar load forecasting performance when compared to other complex models in a FL setup.
- Quantify the overhead of the proposed lightweight models on a lightweight microcontroller that mimics a actual smart meter device.

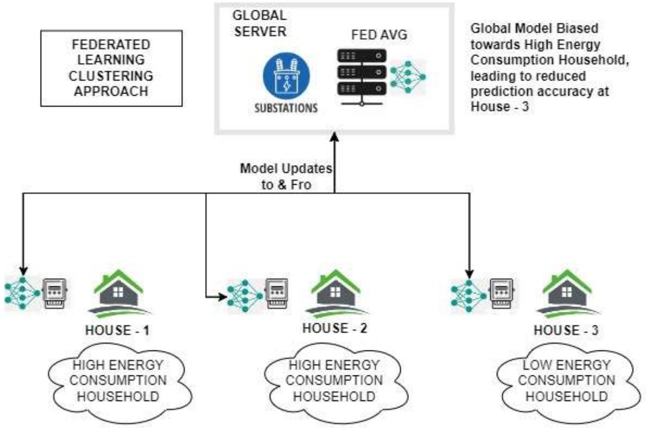
Real World Open Dataset

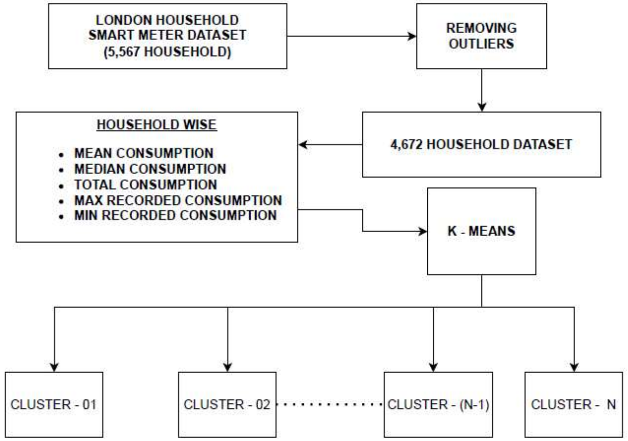
- Dataset used for our simulation and testing : "Smart meter energy consumption dataset from London Households"[11].
  - Energy consumption data recorded from smart meters installed in 5,567 households across London.
  - Consumption coverage period from November 2011 to February 2014
    - Unique Household identifier tag
    - Tariff type (standard or dynamic pricing)
    - Time-stamp
    - Half-hourly electric consumption (in kWh).

[11] UK Power Networks: SmartMeter Energy Consumption Data in London Households, "https://data.london.gov.uk/dataset/smartmeter-energy-usedata-in-london-households," 2014.

Our Approach – Clustering based FL

- With federated models, clustering is particularly effective[12]
  - it allows clients to be organized into smaller clusters based on similar consumption and their locality.
  - improves the aggregation in the substation where local models from similarly performing households have a better probability of achieving an unbiased aggregation to arrive at the new global model.

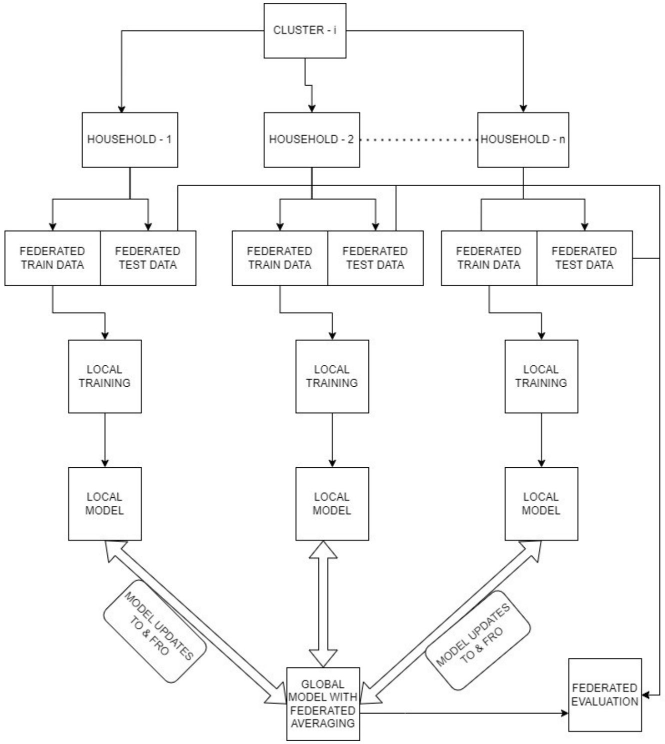




[12] F. L. Quilumba et. al, “Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities,” IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 911–918, 2014.

Evaluation Setup

- Training, Validation and Testing Split : (60:40 Split)
  - Data from Nov 2011 to Feb 2013 : Training and Validation
    Data from Feb 2013 to Feb 2014 : Testing
- A sliding window scheme is used to capture the short history of recent consumption data (336 data points – A week's data)
  - Model : Feed-forward neural networks for their simplicity and computational efficiency
    - A 4-layer model with 2 hidden layers.
    - Model Structure : [16,8,4,1] shape with rectified linear unit (ReLU) activation functions
    - Loss Function : RMSE, MAPE, MAE
    - No of Parameters : 5569



Results – Short term Forecasting

- The table presents the RMSE values averaged across different types of households within a specific cluster from our experiments.

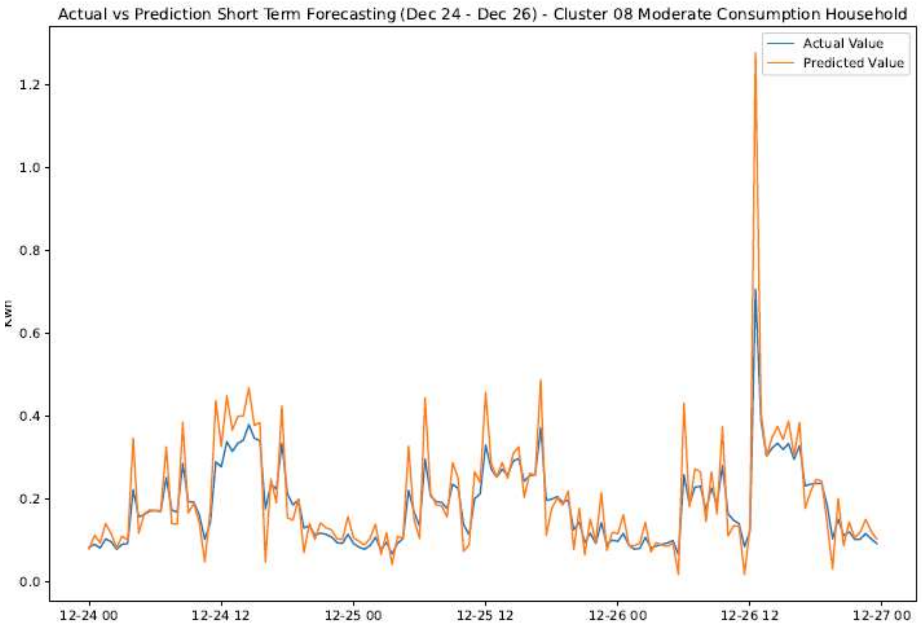
- Model offers better (average) prediction performance for households that received all the updates (i.i.d condition) compared to those that missed out on some of the updates.

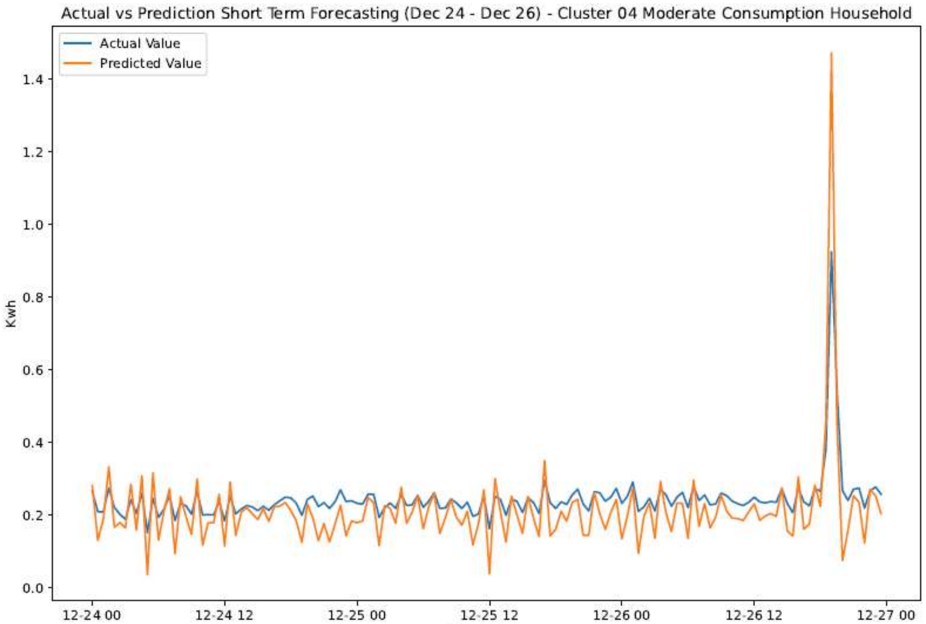
- This demonstrates that in the worse scenario when a house skip updates, the approach is able to provide reasonably accurate predictions.

Month	CLIENT TYPE (RMSE)		
	Mod (All)	**High (Non - i.i.d)**	**Low (Non - i.i.d)**
January	0.0636	0.2507	0.0788
February	0.0633	0.2581	0.0655
March	0.0623	0.2875	0.0760
April	0.0544	0.1531	0.0797
May	0.057	0.0866	0.0740
June	0.0621	0.0533	0.0752
July	0.0638	0.1690	0.7425
August	0.0662	0.2217	0.7119
September	0.0528	0.1417	0.0730
October	0.0611	0.1896	0.0695
November	0.0539	0.2505	0.0684
December	0.065	0.2484	0.0787
Average	**0.0604**	**0.1925**	**0.1828**

Results – Short term Forecasting

We also examined the performance[i] across festive periods where significant uneven consumption is expected.

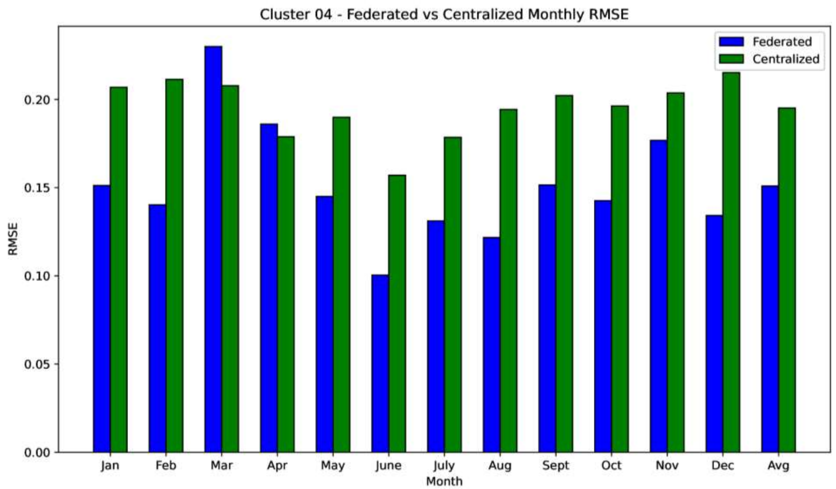




[i] Graphs below denote performances of Moderate consumption Households that missed 25 % of the global updates

Results – Comparison against centralized framework

- Comparing and analysing the effectiveness of distributed FL learning in comparison to a centralized global learning approach
  - A model of similar complexity used
  - Clients participation Ratio kept at a consistent 100 % rate, eliminating the non-I.I.D scenario for the best comparison
  - 20 epochs considered in comparison to 20 Federated Rounds
- The results show that the federated model achieves better performance at the same model complexity.



Results – Comparison against other Complex FL Models

- We compare our model with the FL architecture used in[10], using average monthly RMSE across various clusters.
  - The number of clients contributing fully to the global model was set to be less than 50% than used in the competing method[10] for each cluster.
  - The work[10] uses an LSTM-based model with 7505 parameters compared to our model with only 5569 learnable parameters (lower communication costs during the model aggregation phase).
- Despite using 50% fewer participating nodes in each cluster and 20% fewer computational requirements for the local model, our model is able to achieve an average RMSE of 0.17 compared to 0.14 of the LSTM model.

Month	Our Proposed Model (RMSE) Parameters = 5569	M.Savi Model (RMSE) Parameters = 7505
January	0.2053	0.1463
February	0.1626	0.1530
March	0.2176	0.1535
April	0.1800	0.1445
May	0.1657	0.1304
June	0.1434	0.1259
July	0.1685	0.1457
August	0.1526	0.1406
September	0.1581	0.1302
October	0.1554	0.1335
November	0.1729	0.1471
December	0.1728	0.1328
Average	**0.1712**	**0.1402**

[10] M. Savi and F. Olivadese, “Short-term energy consumption forecasting at the edge: A federated learning approach,” IEEE Access, vol. 9, pp. 95 949–95 969, 2021.

Results – Comparison against other Complex Models

- We also compare our results against the work[13], which similarly uses an LSTM-based model with customized local model improvement at each round, with only 5 federated rounds compared to our 20.
- Our results, in the below table, show that our approach achieves significantly better average prediction accuracy (in terms of MAPE) compared to the results reported in work [13]

Model	Household Ratio	Fed Rounds	Avg MAPE
Taik's Model	36%	5	34.14%
Our Proposed Model	10.50%	20	22.01%

- Additionally, we observe that reducing the number of federated rounds from 20 to 5 increases the average error across all clusters to increase by nearly 4% in our case, which is still better than the results in [13].

[13] A. Taık and S. Cherkaoui, "Electrical load forecasting using edge computing and federated learning," in ICC 2020-2020 IEEE international conference on communications (ICC). IEEE, 2020, pp. 1–6.

Real-world Energy Overhead

- **Hardware Platform:**
  - Implementation on Arduino Uno R4 WiFi platform with an 8-bit processor, coded in Arduino C
- **Evaluation Scenario:**
  - Simulation : Model training on one day's data and predicting half-hourly consumption for the following day
  - Output transfer: Data transferred via serial port to a standard laptop for analysis
- **Monitoring:**
  - Energy consumption measured using a calibrated USB energy monitor
  - Recording energy usage during three key phases: idle, training, and prediction
- **Quantifying overhead:**
  - Assessing the energy overhead incurred during model training and inference at each client (smart meter)
  - Providing insights into the computational cost and energy requirements of the IoT model

Real-world Energy Overhead

- **Minimal Energy Overhead:**
  - Model consumes only 50mWh, averaged over the entire day, in addition to the 440 mWh of idle power consumption on the Arduino platform
  - Resulting in a modest 11% energy overhead during training and inference activities
- **Practical Integration**:
  - Integration of this model is viable with smart meter hardware platforms of similar or higher capabilities
  - The low energy consumption overhead highlights the potential for deploying energy-efficient IoT models on resource-constrained devices, enabling real-world applications

Conclusions

- **Distributed Learning Framework**:
  - Proposed a FL approach for distributed load forecasting using a lightweight fully connected NN
- **Privacy-Preserving Local Learning**:
  - Learning takes place locally at each client's smart meter, with only updated models transferred centrally, mitigating privacy risks associated with aggregating granular energy consumption data.
- **Trade-offs and Performance**:
  - Case I: achieves ~21.4% higher RMSE but significantly lower model parameters (~25.8%)
  - Case II: achieves ~35.6% better aggregate MAPE, highlighting the flexibility of our approach
- **Energy-Efficient Implementation**:
  - Using an Arduino platform, our lightweight model incurs a mere 50 mWh of energy overhead when averaged over multiple cycles, making our approach ideal for decentralized load forecasting applications.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank You