Unsupervised, Federated and Privacy-Preserving Detection of Anomalous Electricity Consumption in Real-World Scenarios

Joppe W. Bos
NXP Semiconductors, Leuven, Belgium
joppe.bos@nxp.com

Michiel De Vis
Sirris, Brussels, Belgium
michiel.devis@sirris.be

Charles Faes
Engie Lab Laborelec, Linkebeek, Belgium
charles.faes@engie.com

Nicolás González-Deleito
Sirris, Brussels, Belgium
nicolas.gonzalez@sirris.be

Anna Hristoskova
Sirris, Brussels, Belgium
anna.hristoskova@sirris.be

Sarah Klein
Sirris, Brussels, Belgium
sarah.klein@sirris.be

Sreeraj Rajendran
Sirris, Brussels, Belgium
sreeraj.rajendran@sirris.be

***Abstract*—Being able to detect anomalous electricity consumption events in residential buildings is one of the main steps when reducing unnecessary energy usage. In this paper, we present an unsupervised approach for detecting such kind of events locally at individual households, while leveraging the knowledge from other households and ensuring the privacy of the consumption. This is realized by learning a model locally on the households' edge devices for on-the-edge anomaly detection, using federated learning to leverage the knowledge from other households and ensure the privacy of the training data, and using a secure aggregation approach to handle model weight updates, protecting against a malicious central server and avoiding the risk of model inversion attacks. The approach has as advantages that it respects privacy, does not require annotated data, circumvents the cold-start problem, and is flexible with respect to detecting different types of occurring anomalies. We demonstrate its effectiveness on a real-world, highly granular electricity consumption dataset comprising 14 Belgian households during a period of 36 months, and show effectiveness by installing malfunctioning devices in a lab environment.**

I. Introduction

Just about every new (IoT) device entering the market is in one way or another connected to its environment. These devices generate vast amounts of data which are in turn used for a plethora of use cases with the help of advanced machine learning techniques. Over the past years, this data collection has risen and so do the privacy concerns about this information, especially if the data was collected in a centralized storage in order to use it for the training or inference of machine learning models.

An interesting approach to limit the amount of raw and possibly privacy-sensitive (or confidential) data needing to be treated centrally, while reducing bandwidth requirements and enabling cross-device learning in highly distributed environments is Federated Learning [1]–[4]. In this approach, instead of collecting the data in a centralized fashion, a model is trained locally on each device and only the model parameters are shared with a central server. That server combines the individual model parameters to derive a global model, whose parameters can be further pushed back to the different devices. Subsequently, the local models can be further updated, the model updates can be shared with the server to update the global model, and this model further shared with the different devices. However, it is important to stress that model updates can also leak some amount of private data which could be exploited by an attacker (cf. [5], [6]). Further, when using federated learning in real-world use cases, the central server has access to all possibly privacy-sensitive model updates which implies some level of trust.

All privacy concerns related to sensitive user-data can be mitigated if the model updates are encrypted by the data owner before they are sent to an external party. This ensures that only the legitimate data owner can access the data by decrypting it using its private decryption key. However, this form of encryption limits the possibility to outsource *computations* since the external party will not have access to the decryption key. Performing such computational tasks is often crucial to the business value of cloud services.

Such privacy concerns often result in the slow wide-scale adoption and acceptance of these new data processing technologies but provide opportunities for advances in the research area of privacy-enhancing technologies. In this paper, we present a privacy-preserving federated learning technique applied to a real-world electricity consumption data. With our approach, we are able to detect malfunctioning appliances in private households, without sharing any privacy-related data: neither during the model training nor the inference phases. This is realized by handling the model weight updates in a privacy-sensitive manner using a secure aggregation approach [7]. Furthermore, our approach overcomes the so-called cold-start problem [8]: in order to detect anomalies, only a few weeks of data is needed compared to long-time training

This research was conducted within the SunRISE project labeled by PENTA and funded by VLAIO, the Flanders Innovation & Entrepreneurship Agency, under grant agreement HBC.2018.2157. Author list in alphabetical order; see https://www.ams.org/profession/leaders/culture/CultureStatement04.pdf.

requirements from single-source machine learning models.

II. Preliminaries and Related Work

Optimization of electricity consumption in public or corporate buildings is an active research field due to the rising interest in low or even zero-energy buildings. This led to an increased interest for the detection of malfunctioning electrical devices since these often waste energy. An extensive review by Himeur et al. [9] describes the most commonly used techniques for detecting abnormal electricity consumption data. While most studies on anomaly detection focus on energy theft of defective meters, Rashid et al. [10] propose the *Rimor* approach that focuses on detecting contextual anomalies by comparing a forecasted consumption to the actual one. They validate their approach on public aggregated consumption data with additional weather information about temperature and humidity.

Predicting future electricity consumption and warning on significant deviation is one way of detecting anomalies. Forecasting the short-term electricity load at individual households is a well-known problem that is extensively studied in the smart grid domain: see e.g. the overview by Lopez-Martin et al. [11]. The use of federated learning approaches [1]–[4] enabling to process the individual consumption values locally at each household's smart meter and resulting in less bandwidth-intensive and more privacy-friendly solutions is very recent. Taïk and Cherkaoui [12] are among the first ones to investigate such a setting, for which they consider a Long Short-Term Memory (LSTM) neural network [13]. Interestingly, Savi and Olivadese [14] propose a similar approach, but, in addition to using the previous electricity consumption data for a household, they also consider in their LSTM model the consumption data from similar days at the same hour for that household and other calendar and weather-related features. To that end, the authors use a publicly available dataset with (after preprocessing) 1500+ households in London, from Nov. 2012 to Feb. 2014, enriched with weather information and including demographic characteristics of the considered households. Individual models are built for each demographic category present in the data, as well as for similar types of households derived in a data-driven fashion. These models are evaluated against a model derived from a fully centralized approach. The federated and centralized models perform similarly when considering the demographic categories, and the federated models perform slightly better when considering the data-driven categories.

Finally, the problem of computing meaningful operations on encrypted data in order to address privacy concerns was already introduced in the late 1970s by Rivest et al. [15]. However, it took until 2009 before a concrete instantiation was found by Gentry [16] in the form of a *fully homomorphic encryption* (FHE) scheme. Such a scheme allows an untrusted party to carry out arbitrary computations on *encrypted data* without learning anything about the content of this data. The key to allowing arbitrary computations is that an FHE scheme allows both homomorphic addition and multiplication operations on the encrypted data. Previous homomorphic encryption schemes were *partially homomorphic*, i.e. they only provided one of the two operations and therefore were not fully homomorphic. Although these generic techniques allow one to compute *any* function in a privacy-preserving manner, they are often too slow in practice. Dedicated techniques to compute one specific algorithm or function are often crafted to make the privacy-preserving solution significantly faster.

III. Detecting Abnormal Electricity Consumption

Electricity consumption data with high temporal granularity is not only highly personal but also very context dependent. It is influenced by the time people get up and go to bed, and might change dramatically over time, for example when a newborn is in the house or a whole society goes into lockdown. Nevertheless, common patterns can be observed across different households and by applying federated learning, these patterns can be captured in a shared model.

A. Data Characteristics and Context

In order to outline, test and verify the presented approach, we use a real-world dataset provided by Engie Laborelec. This data is composed of per minute electricity consumption for 14 Belgian households during a period of 36 months, from Sept. 2017 until Sept. 2020. This data originates from the smart meters installed at those households, reporting on the overall electricity consumption in each household separately. The smart meters are connected to a Raspberry Pi acting as a gateway and being used for research purposes. In the context of the present research, each Raspberry Pi is seen as an edge node with the capability of performing some limited computations.

B. Data Processing

In order to ensure a reliable anomaly detection, three preprocessing steps are performed on the edge nodes: (1) discretization in a streaming fashion, (2) mode decomposition, and (3) normalization across households in a streaming fashion.

Firstly, as the electricity consumption levels of the H different households can vary significantly, the electricity data used is normalized in order to train a common model across those households. Let us denote the total electricity consumption data for a specific household h from time t' to t as $E_h(t - t')$. In a first step, a temporal binning is used such that one can sum the data over an overlapping window of length w with stride w_s; when using ℓ data points, this results in $M = \lfloor (\ell - w)/(w_s) \rfloor + 1$ temporal bins.

Secondly, electricity consumption of private households shows temporal variations at different scales. For instance, one can observe strong seasonalities as the overall consumption is typically higher in winter than in summer. Similarly, one can also observe daily patterns reflecting the personal situation of the inhabitants. Additionally, several appliances like refrigerators or heat pumps exhibit specific repeating consumption fingerprints. While these patterns are predictable, stochastic

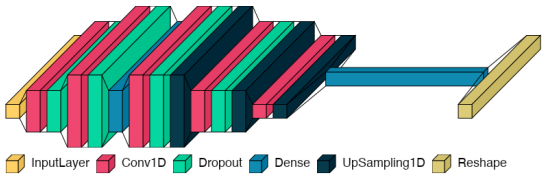

Fig. 1. Architecture of the reconstruction local model with two convolutional layers visualized with [21]. The discretized, decomposed and normalized input data X of dimension daily temporal bins M × modes K is reconstructed to $\hat{X}$ with the same dimension and reconstruction loss ϵ.

contributions occur on the short-term due to spontaneous actions.

In order to model these different layers of periodic and stochastic contributions, we split the binned consumption signal in different modes by using Variational Mode Decomposition (VMD) [17]. This has been successfully used already in different contexts, e.g. for fault detection and diagnosis in wind power gearboxes [18] and for harmonic detection in power grids [19]. The decomposition is applied to separate binned daily consumption data $E_h(m)$, $0 \leq m < M$, into $K > 1$ modes $\tilde{E}_h(m,k)$, $1 \leq k \leq K$ representing repeating patterns at different time scales, such that $E_h(m) \approx \sum_{k=1}^{K} \tilde{E}_h(m,k)$, for $0 \leq m < M$.

Thirdly, in order to make the temporally binned and decomposed data comparable across different households, a minmax-normalization, mapping the maximal value to 1 and the minimal value to 0, is performed. This is applied in a streaming fashion by calculating over a window $W \gg w$, the maximal and minimal consumption per mode k, denoted as $\tilde{E}_h^{W,\max}(k) = \max_{W \leq m \leq t} \tilde{E}_h(m,k)$ and analogously for $\tilde{E}_h^{W,\min}(k)$. This allows us to define the discretized, decomposed and normalized consumption per household as $E_h^W(t-w,k) = \frac{\tilde{E}_h(t-w,k) - \tilde{E}_h^{W,\min}(k)}{\tilde{E}_h^{W,\max}(k) - \tilde{E}_h^{W,\min}(k)}$, which is used as input for our reconstruction model.

C. Local Learning

The daily consumption patterns are reconstructed using a convolutional autoencoder as illustrated in Fig. 1. Such an autoencoder transforms an input signal $X \in [0,1]^{M \times K}$, where M is the number of daily temporal bins and K the number of modes, to a reconstructed copy $\hat{X}$ with reconstruction error ϵ with respect to a predefined loss function $l(X,\hat{X}) = \epsilon$. The input signal X is first encoded and compressed using two convolutional layers per mode to dense layer. Next, this is upsampled via two additional convolutional decoder layers to the reconstructed signal $\hat{X}$. A general overview of convolutional autoencoders can be found in [20].

D. Global Learning

The overall model architecture for the global, federated model is the same as the one described above in Section III-C for the local model. Instead of training the model weights on the locally available data only, a federated training approach is used. The updates are performed in a federated fashion similarly to the `FedSGD` approach proposed in [22]. First, the model is initialized at the central server with random weights and these weights are shared with the local devices at the H households. On each local device, the model is trained but only for one epoch. The local weights and their gradients per household are communicated to the central server. There, the local weights of the encoding part of the autoencoder up to the dense layer are averaged, while for the remaining layers for decoding the signal from the dense layer to the final output signal, the gradients of the weights are averaged. Then, the averaged weights and gradients are sent back to the local device, where the new model weights for the decoding layers are calculated. With this approach, the encoding layers are trained in common while for the decoding layers, device specific contributions are better reflected.

While such a federated learning approach has the benefit of not requiring to share the raw electricity consumption values, it is still possible to derive some private information through a model inversion attack [5], [6]. For this reason, we further add an additional security layer to our approach as discussed later in section III-F: the goal is to protect the local gradients per household such that these model inversion attacks on individual households are not possible anymore.

E. Anomaly detection

Once the model is trained for a certain time period pattern the abnormal electricity consumption is detected by computing $(\hat{E}_h^W(0,k), \ldots, (\hat{E}_h^W(24-w,k))$. Next, the reconstruction error, given by $l(X,\hat{X}) = (X-\hat{X})^2$, is calculated. The daily pattern is considered to be an outlier or anomaly if for any of the K modes the reconstruction error is bigger than the rolling mean plus three times the standard deviation of the reconstruction error. The practical parameters and results applied to the electricity consumption data are discussed in Section IV-B.

F. Privacy-Preserving Aggregation using Shares

A conceptually simple yet efficient method for privacy-preserving aggregation was recently presented in the scope of demand side management of residential loads [7]. This approach circumvents the usage of generic but computationally expensive privacy-preserving techniques. The idea behind it stems from the cryptographic research community and consists in working with multiple shares [23] in order to split-up and obfuscate the data, and recombine it when needed. All but one of the shares are communicated in plain text while the last share is encrypted. This allows certain computations on the plain shares without revealing information as long as these computations preserve the way the data is obfuscated. Assume we have $H > 1$ households that would like to communicate the gradients of their model weights g_i^h per epoch i to a central server, but the server should only know the aggregated value $g_i = \sum_h g_i^h$ (as described in section III-D) and should not learn any information about the individual values g_i^h. First, the server generates in a setup phase a public/private key pair; this public key is provisioned securely to each individual client. Next, each client masks its privacy-sensitive

gradient of the model weights g_i^h by sampling a uniform random value r_i^h, of the same size of the domain where the gradient g_i^h is from, and additively splitting g_i^h into a pair of shares $(p_i^{1,h}, p_i^{2,h}) = (r_i^h, g_i^h - r_i^h)$. This means that the gradient g_i^h can be reconstructed by summing the two shares $p_i^{1,h} + p_i^{2,h} = g_i^h$. Moreover, a single share provides no useful information on g_i^h and is information-theoretically secure. One of the shares (say, $p_i^{2,h}$) is encrypted using the pre-provisioned public key, resulting in $c_i^h = \text{ENC}(p_i^{2,h})$, and the pair $(p_i^{1,h}, c_i^h)$ is communicated to an untrusted third party which we will denote by U. This introduction of an additional party is a disadvantage of this approach. However, the only requirement of this third party is that it does not collude with the central server that computes the final result and issues the cryptographic keys. When U receives the H pairs $(p_i^{1,h}, c_i^h)$, for $h \in [1, H]$, it cannot compute g_i from these pairs. It can, however, sum the plain shares $p_i^{1,h}$ and forward the result and the H encrypted shares to the aggregation server, i.e. $(s = \sum_h p_i^{1,h}, c_i^1, c_i^2, \ldots, c_i^H)$. Finally, that server can decrypt the individual c_i^h values but cannot learn anything from them since $\text{DEC}(c_i^h) = p_i^{2,h} = g_i^h - r_i^h$, for $h \in [1, H]$, which means the original values g_i^h are masked. However, adding all the decrypted values to s results in $s + \sum_{h=1}^{H} \text{DEC}(c_i^h) = \sum_{h=1}^{H} r_i^h + \sum_{h=1}^{H} (g_i^h - r_i^h) = \sum_{h=1}^{H} g_i^h$, as desired. This allows us to aggregate the gradients of the model weights when training the global model in both a privacy-preserving and federated fashion.

IV. Benchmark and Discussion

For both the local and global training scenarios, we use $K = 3$ different modes (see Section III-A) in order to capture behaviors which occur with varying frequency, as discussed in section III-B. Moreover, we use the data for one day as input. This means $\ell = 1440$ data points, corresponding to the minutes per day. Using a rolling window of size $w = 3$ with stride $w_s = 1$, this results in $M = 767$ temporal bins for each of the $K = 3$ modes.

A. Training and Evaluation Data

We train and evaluate our approach on a real-world dataset provided by Engie Laborelec as described in Section III-A. For this, the discretized, decomposed and normalized data of the households is split into a training set and a test set of two years and one year, respectively. The data of one of the households was used in order to simulate defects and malfunctioning devices. For this, the normal consumption data of two consecutive days during the test year was replayed in Engie Laborelec's Home Lab. This Home Lab is a true-scale testing facility built to represent a range of residential units, with three-phase and mono-phase power supplies. These functional replicas of actual homes are equipped with standard electrical cabinets and head metering features, both conventional and smart, coupled to production and load emulators. The electrical cabinets are a plug and play feature where different equipment can be connected. It is possible to replay a consumption profile through the use a generator acting as a load for simulated residential unit. In addition, other equipment can be added in parallel to the simulated consumption profile to generate an anomaly.

Fig. 2. Logarithmic distribution of daily mean squared reconstruction error of local (left) and federated (right) models for a test data set for the two modes (from top to bottom).

Different type of anomalies were generated: the addition of a low consuming defective equipment adding a small constant load as well as the addition of a defective boiler. This defective boiler dissipates heat faster than it should, therefore increasing the heating cycles of its water. On the selected data we validate our federated and unsupervised approach for detecting abnormal electricity consumption. In that case, before preprocessing the data, the actual consumption of the two days is replaced by the data created with defect boiler or heating device by replaying the energy consumption in the Home Lab and adding the anomalies and recording the new energy consumption profile.

B. Local Results

In the local setting every household uses only its own local data for training the model. Inherently, this means there are no privacy concerns since neither the training data nor the inference data leave the edge node. Using the corresponding test dataset consisting of valid data for each household we evaluate how well the model is reconstructing the daily patterns by calculating the mean squared error between the original patterns and the reconstruction for each mode. These local results for the first and second VMD mode are displayed in the left part Figure 2. For our further analysis, we will concentrate on these two modes as they allow us drawing conclusions on either global electricity increase or electricity usage of devices with very regular short-term pattern. This figure clearly shows that the vast majority of the local test data is reconstructed correctly (note the logarithmic scale on the y-axis). However, a non-negligible part of the test data set has a mean squared error > 0.02 in the two modes.

C. Global Results

The data from multiple households is used to train the models using the federated learning approach while com-

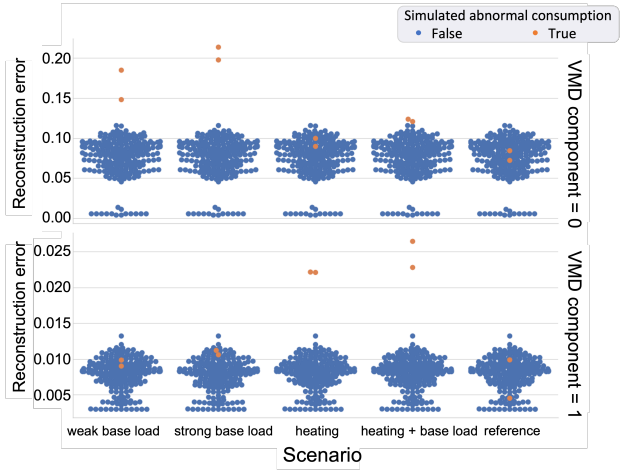

Fig. 3. The reconstruction error when using data with abnormal behavior in the global evaluation setting for the two modes.

municating the per household local gradients in a privacy-preserving manner. The right part of Figure 2 shows the reconstruction error for the same household as used to illustrate the local approach. The results are significantly better. For the two modes, virtually all data is reconstructed with a mean squared error significantly smaller than 0.02, demonstrating the added value of using the global setting. Figure 3 shows the results for the two modes in the global setting on the abnormal data. This figure shows the reconstruction error for the various scenarios for the slowly changing (top) and higher frequency (bottom) mode, respectively. One can see how the base load influences the slowly changing VMD mode with the reconstruction error being significantly higher for these cases (left). For the malfunctioning heating the reconstruction error in the higher frequency mode exceeds the threshold (middle and second from right). For the reference curve (right), the actual consumption for these two days, we see that reconstruction error falls nicely into the distribution of all other errors.

V. Conclusions

In this paper we showed that combining electricity consumption data from multiple households improves the accuracy of detecting malfunctioning home equipment, and showed how to overcome potential privacy implications when training these models. For this, we used federated learning approaches where the models are trained locally and which use a technique based on multiple shares for aggregating the model information centrally in a privacy-preserving fashion. We combine all these elements into a practical and privacy-preserving approach for detecting anomalous electricity consumption events in residential buildings. In addition, we showed that the federated models capture consumption patterns better than the local models (as shown in Fig. 2), and that for newly installed smart meters (cold start problem) our approach enables to reliably detect abnormal electricity consumption for heating as well as for weak and strong base loads.

References

[1] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *ACM CCS 2015*, I. Ray, N. Li, and C. Kruegel, Eds. ACM Press, Oct. 2015, pp. 1310–1321.

[2] J. Konečný, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *CoRR*, vol. abs/1511.03575, 2015.

[3] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *AISTATS*, ser. Proceedings of Machine Learning Research, A. Singh and X. J. Zhu, Eds., vol. 54. PMLR, 2017, pp. 1273–1282.

[5] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart, "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *USENIX Security 2014*, K. Fu and J. Jung, Eds. USENIX Association, Aug. 2014, pp. 17–32.

[6] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?" *Advances in Neural Information Processing Systems*, vol. 33, pp. 16 937–16 947, 2020.

[7] E. J. Palacios-Garcia, J. W. Bos, X. Carpent, and G. Deconinck, "A privacy-friendly aggregation algorithm for demand side management of residential loads," in *ISGT Europe*. IEEE, 2021, pp. 1–5.

[8] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *Research and Development in Information Retrieval*, ser. SIGIR '02. Association for Computing Machinery, 2002, p. 253–260.

[9] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," *Applied Energy*, vol. 287, p. 116601, 2021.

[10] H. Rashid, N. Batra, and P. Singh, "Rimor: Towards identifying anomalous appliances in buildings," in *Systems for Built Environments*, 2018, pp. 33–42.

[11] M. Lopez-Martin, A. Sanchez-Esguevillas, L. Hernandez-Callejo, J. I. Arribas, and B. Carro, "Novel data-driven models applied to short-term electric load forecasting," *Applied Sciences*, vol. 11, no. 12, p. 5708, 2021.

[12] A. Taïk and S. Cherkaoui, "Electrical load forecasting using edge computing and federated learning," in *ICC*, 2020, pp. 1–6.

[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997.

[14] 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.

[15] R. L. Rivest, L. Adleman, and M. L. Dertouzos, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.

[16] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *41st ACM STOC*, M. Mitzenmacher, Ed. ACM Press, May / Jun. 2009, pp. 169–178.

[17] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[18] Z. Wang, G. He, W. Du, J. Zhou, X. Han, J. Wang, H. He, X. Guo, J. Wang, and Y. Kou, "Application of parameter optimized variational mode decomposition method in fault diagnosis of gearbox," *IEEE Access*, vol. 7, pp. 44 871–44 882, 2019.

[19] G. Cai, L. Wang, D. Yang, Z. Sun, and B. Wang, "Harmonic detection for power grids using adaptive variational mode decomposition," *Energies*, vol. 12, no. 2, p. 232, 2019.

[20] Y. Zhang, "A better autoencoder for image: Convolutional autoencoder," in *ICONIP17-DCEC*, 2018.

[21] P. Gavrikov, "visualkeras," https://github.com/paulgavrikov/visualkeras, 2020.

[22] 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.

[23] S. Chari, C. S. Jutla, J. R. Rao, and P. Rohatgi, "Towards sound approaches to counteract power-analysis attacks," in *CRYPTO*, ser. LNCS, M. J. Wiener, Ed., vol. 1666. Springer, 1999, pp. 398–412.