Improving Electric Load Demand Forecasting With Hard Representation Regularization

Paraskevi Nousi
Swiss Data Science Center (SDSC)
ETH Zurich
Zurich, Switzerland
paraskevi.nousi@sdsc.ethz.ch

Maria Tzelepi
Information Technologies Institute (ITI)
Centre of Research and Technology Hellas (CERTH)
Thessaloniki, Greece
mtzelepi@iti.gr

***Abstract*—Electric load demand forecasting is a challenging task which can be of service to downstream tasks and is of interest to various communities, such as power supply companies. In this paper, we tackle this task by introducing a hard representation based regularization technique for neural networks, which can be incorporated into the loss function of any standard network. Specifically, hard samples are first identified by the network by their contribution to the loss during training. Then, the most difficult representations within a batch, which more heavily contribute to the network's loss and in turn hinder the network's performance, are forced to become more similar to their nearest representations in their batch. In this way, we achieve improved generalization ability, and more stable performance. To evaluate the effectiveness of our method and facilitate future research on the task, we curated and publicly shared a dataset with energy demand data from Switzerland. Experimental studies on three datasets show that the proposed method can be effectively applied to the task and provide improvements in terms of performance with minimal hyperparameter tuning.**

***Index Terms*—electric load demand forecasting, hard representation regularization, swiss energy market, time-series forecasting, hard-mining.**

I. Introduction

Electric Load Demand Forecasting (ELDF), i.e., prediction of electricity demand values in the future, is a persisting challenge in today's energy markets due to many factors which may significantly affect the forecasting performance [1], [2]. Typically, historical load data is employed as input to the forecasting models alongside additional data such as weather data and temporal information, such as day of the year, month, etc. The output horizon of the models determines the ELDF category: a) short-term, concerning forecasting of a few hours up to one-day ahead or a week ahead, b) mid-term, concerning longer time periods of up to one year, and c) long-term, concerning time frames of up to several years ahead. In this paper we are particularly concerned with the task of one-day (24 hours) ahead forecasting.

ELDF applications vary widely, from power system planning and operation to energy trading [3]. Achieving an efficient balance between supply and demand is critical for power companies, allowing them to avoid excess reserve of power generation or power interruptions due to load shedding. Such reasons have made research on the field a particularly interesting topic in recent years [4]. Publicly available datasets such as ISO-NE [5] or Spain Energy dataset[1] facilitate research in the field and allow for reproduction of previous results as well as fair comparisons. However, further data could help enrich the research in the field by providing more variance and larger training sets which may aid in the generalization ability of the produced models.

In this paper, we deal with the *Swiss Energy Market* and curate a dataset in a format similar to the aforementioned datasets which we make publicly available in the hopes of facilitating future research in the field. Some pre-existing works on the Swiss Energy market have used similar data and various approaches. For example, in [6] linear regression was used to model the annual electricity consumption of Swiss enterprises. In [7], Switzerland's national electricity demand was modeled using Bayesian hyper-tuned Neural Networks. In [8] a Long Short-Term Memory (LSTM) was used to forecast energy load demand values, applied to Switzerland's power-load data. Our work focuses on a) providing a public dataset to facilitate reproducibility of research, and b) a hard-mining based regularization method that can be combined with any Neural Network based method.

Other prior works in the field of energy load demand forecasting including statistical models [9] and machine learning models [10]. More recently, Deep Learning (DL) models have dominated the field, following their significant accomplishments in various other tasks. More specifically, DL models have been proposed to deal with the ELDF task in [11], achieving remarkable results. In [12] a Deep Belief Network (DBN) was proposed to forecast the hourly load of a power grid, outperforming previous approaches in both daily and weekly predictions at an hourly step. In [13] a method was proposed using a hybrid model of a typical neural network with an LSTM model. Specifically, the method entailed predicting the demand separately for the four seasons of the year, based on the intuition that energy needs might differ significantly during different seasons.

In this work, we propose a hard-representation-mining based regularization technique, which is added as an auxiliary training objective to the standard regression task of neural

[1]https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather

networks, and which first detects difficult samples based on the network's predictions. Then targets are added for the most difficult samples to the objective, specifically they are encouraged to approach samples which the network deems as easier again based on its predictions. Our proposed method is based on the point that some difficult or outlying samples, which more heavily contribute to the network's loss, might hinder the network's performance by forcing it to overfit those samples while forfeiting some generalization ability.

The main contributions of this paper can be summarized as follows:

- We propose a novel hard-mining based regularization method for the task of one-day ahead energy load forecasting. The proposed method is model agnostic and can be added as a plug-and-play module on any state-of-the-art model to increase performance.
- We make a large ELDF dataset complete with historical load data and weather data publicly available, alongside the code used to clean and process it.
- We perform experiments on three publicly available datasets and showcase the effectiveness of the proposed method, without adding any computational overhead during deployment.

The remainder of the manuscript is structured as follows. Section II presents in detail the proposed hard representation regularization method. Subsequently, the experimental evaluation is presented in Section III, including the presentation of the dataset on Swiss Energy Market. Finally, the conclusions are drawn in Section IV.

II. Proposed Method

In this work we propose a novel regularization method for improving the forecasting performance towards the electric load demand forecasting task. The proposed Hard Representation Regularization (HRR) method is grounded on the concept of multitask learning [14], which has been extensively applied as a regularization strategy, considering classification tasks [15]. In this paper we incorporate it to time-series forecasting tasks, and specifically to electric load demand forecasting, developing a novel hard representation regularization method.

Specifically, apart the from the main regression loss employed for training the model for electricity demand prediction task, we introduce an additional regularization loss to the output of the model that forces the hard representations to become more similar to their nearest non-hard representations. That is, during the training we identify the hard representations, i.e., the representations with the maximum main regression loss based on one percentage threshold (e.g., 20% of representations with the maximum regression loss), and we attach an additional auxiliary loss at the space generated by the output layer which forces these hard representations to approach their nearest, i.e., most similar in terms of a similarity metric (e.g., Euclidean distance) non-hard representations (i.e., representations other than the ones deemed as hard in the previous step) inside their batch, defining another percentage threshold (e.g., each hard representation to approach the 20% of nearest non-hard representations).

More specifically, we consider a neural network for electric load demand forecasting $F_{\mathcal{W}}$, parametrized by weights $\mathcal{W}$, with an input space $\mathcal{X} \subseteq \mathbb{R}^D$, and an output space $\mathcal{Y} \subseteq \mathbb{R}^d$. We also consider the input samples $\mathbf{x}_i,\ i = 1, \cdots, N$, their corresponding outputs, $\mathbf{y}_i$, i.e., $\mathbf{y}_i = F_{\mathcal{W}}(\mathbf{x}_i)$ and their ground truth vectors $\mathbf{g}_i \in \mathbb{R}^d$. As it will be presented in the subsequent section, the input samples are built so as to consider previous load and temperature values with respect to the target day. Specifically, $\mathbf{x}_i \in \mathbb{R}^{171}$, while since we aim to predict the load demand for next day on an hourly basis, $\mathbf{g}_i, \mathbf{y}_i \in \mathbb{R}^{24}$. The model is trained using a regressions loss, ℓ_r. Several regression losses can be utilized (e.g., L1, L2), in this paper we utilize Smooth-L1, which acts both as L1 when errors are large and as L2 when errors decrease, providing robust performance, since we have seen in preliminary experiments that performs better, while it has also been utilized in the relevant literature [16]. Smooth-L1 is formulated as follows for a sample i:

$$\ell_r(\mathbf{g}_i, \mathbf{y}_i) = \begin{cases} 0.5(\mathbf{g}_i - \mathbf{y}_i)^2, & \text{if } |\mathbf{g}_i - \mathbf{y}_i| < 1 \\ |\mathbf{g}_i - \mathbf{y}_i| - 0.5, & \text{otherwise.} \end{cases} \tag{1}$$

During the training process, we define a threshold, T_1, for the percentage of representations with the maximum regression loss inside the batch, in order to identify the hard representations. That is, considering each output representation $\mathbf{y}_i = F_{\mathcal{W}}(\mathbf{x}_i)$, associated with its loss value $\ell_r(\mathbf{x}_i, \mathbf{y}_i)$, we sort the representations based on their loss values, and we define the hard representations, as: $\mathbf{y}_j^h,\ j = 1, \cdots, N_h$, where $N_h = T_1 \times batch\_size/100$. Then, we introduce the additional regularization objective that encourages the hard representations to approach their nearest non-hard representations. To do so, we define a second threshold T_2 for the percentage of the nearest non-hard representations that each hard representation is forced to approach to. That is, each hard representation $\mathbf{y}_j^h$ is associated with a subset of non-hard nearest representations $\mathbf{y}_k^n,\ k = 1, \cdots, N_n$, where $N_n = T_2 \times (batch\_size - N_h)/100$. Therefore, our goal is to minimize the regression loss, ℓ_r between each hard representation $\mathbf{y}_j^h$ and the mean value, $\mathbf{m}_j$ of nearest non-hard representations. The total for training the regularized forecasting model is formulated as:

$$\ell_{\text{total}} = \ell_r(\mathbf{g}, \mathbf{y}) + \lambda \ell_r(\mathbf{m}, \mathbf{y}^h), \tag{2}$$

where the parameter λ controls the contribution of the two loss terms. Thus, training the forecasting model with both the main and the additional regularization loss leads to advanced generalization ability and more stable performance. During the test, the test samples are simply propagated to the trained model and the load demand predictions for the target day are produced.

III. Experimental Evaluation

In this section we present the experiments conducted in order to validate the proposed method. We first present the utilized datasets, including the novel dataset constructed for the electric load demand forecasting task on Swiss energy data. Subsequently, we present the evaluation metrics and the utilized model, followed by the implementation details. Finally the experimental results are discussed.

A. Datasets

The main target of this paper is to address the electric load demand forecasting on Swiss Energy Market. However, in order to further validate the effectiveness of the proposed method, we also perform experiments on two additional datasets.

1) Swiss Energy Market: We have collected historical load data from SwissGrid[2] and complemented them with weather data from Open-Meteo[3], specifically temperature values from Zurich, the largest city in Switzerland, to match the format of other existing datasets in the field. In total, the data spans 16 years, specifically starting from 2009 and up to 2024. This is the largest of the datasets used in this works and, to the best of our knowledge, one of the largest datasets designed for this task in terms of years covered. Figure 1 shows the curated electricity load data, where it is evident that energy demand data exhibits multiple periodicities. In Figure 2 the daily energy demand is shown as a function of the day of the year, where the difference in demand between weekdays and weekends can be seen clearly. This behaviour influences the choice of features we choose to use as input to the model, as will be detailed in Section III-C.

Preprocessing of the dataset included fixing some missing values, namely by interpolating values that were close to 0 as well as values greater than 10^7 via cubic interpolation. For the weather data, there were no missing or outlier values in the collected data. Data from years 2009-2022 are used for training, 2023 is used for validation, and finally data from 2024 are used for testing. We make this dataset as well as the code used to generate it publicly available to facilitate further research on the field. The dataset can be found at this repository[4].

2) ISONE: ISO-NE[5] consists in historical load and weather data (i.e., temperature) from New England, collected from eight years. Specifically, for training we utilized data from years 2007-2012, data from 2013 are utilized for validation, and finally data from 2018 are used for testing.

3) Spain Energy Market: The dataset of Spain Energy Market consists in historical load data, provided by ENTSO-E Transparency Platform, and weather information (i.e., temperature), acquired from OpenWeather. Data from four years are used. Specifically, for training we utilized data from years 2015 - 2017, data from 2017 are used for validation, and finally data from 2018 are used for testing.

[2]https://www.swissgrid.ch
[3]https://open-meteo.com
[4]https://github.com/vivinousi/energy_demand_ch
[5]https://github.com/yalickj/load-forecasting-resnet

B. Implementation Details

All the models are trained using Adam optimizer with an initial learning rate of 0.003. The mini-batch consists of 64 samples, and the models are trained for 1,000 epochs. T_1 is set to 75%, T_2 is set to 75%, while the parameter λ in eq. (2) for controlling the relative importance of the two contributed losses is set to 0.001. All the experiments conducted on an NVIDIA GeForce RTX 3050 with 4GB of GPU memory. The proposed method was implemented using the Pytorch framework.

C. Model Architecture

In this work, a simple Multilayer Perceptron (MLP) is employed for evaluating the effectiveness of the proposed HRR method. The model consists of one hidden layer with 32 neurons, while the output layer consists of 24 neurons, since our goal is to predict the load demand for each of the 24 hours of the next day. The input features are built according to [17]. Specifically, the input consists of a total of 171 features that include the load of the day one day prior the target day (24 features), the load one week prior the target day (24 features), and the load one month, i.e., 28 day, prior the target day (24 features). The corresponding temperatures for the aforementioned days are also included (72 features), as well as the temperature of the target day (24 features). Finally, two binary indicators are included, of the target day being holiday and weekend, as well as another indicator of which day of the week is the target day (3 features in total).

D. Evaluation Metrics

We use a common metric considering time-series forecasting tasks, i.e., Mean Absolute Percentage Error (MAPE), in order to evaluate the effectiveness of the proposed method. MAPE considering a set of N_t test samples is formulated as follows:

$$\text{MAPE} = \frac{100\%}{N_t} \sum_{t=1}^{N_t} \left| \frac{\mathbf{r}_t - \hat{\mathbf{r_t}}}{\mathbf{r}_t} \right|, \tag{3}$$

where $\mathbf{r}_t$ corresponds to the ground truth, and $\hat{\mathbf{r_t}}$ to the model's prediction. We execute each experiment five times, and we report the mean value of MAPE and the standard deviation. We also provide the percentage of improvement achieved by the proposed method over the baseline.

E. Experimental Results

First, in Table I we provide the experimental results of the proposed HRR method against baseline, i.e., training without the proposed regularization objective, on the three considered datasets, in terms of test MAPE. We also provide the percentages of improvement for each case. Best performance is printed in bold. As it can be observed the proposed method considerably improves the baseline forecasting performance on all the utilized datasets. Specifically, HRR provides improvements up to 6.85%. It can also be noticed that in two out of three datasets the standard deviation of baseline training is large, while the regularized HRR training is stable, indicated

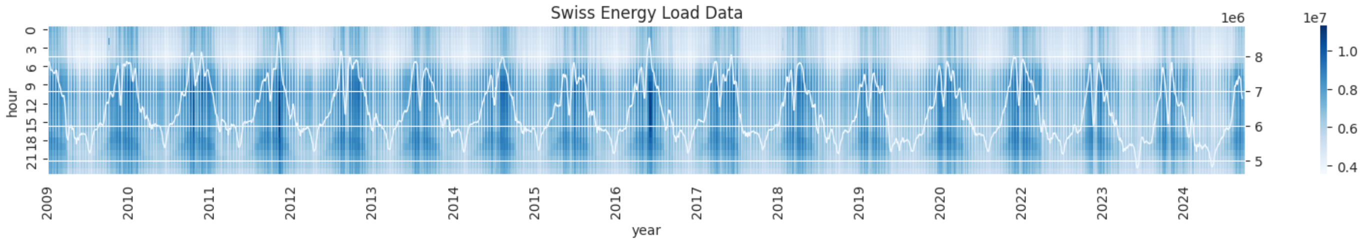

Fig. 1. Swiss electricity demand data, with hours shown on the y-axis and days shown on the x-axis, over the span of 16 years (mean daily demand value shown in light blue).

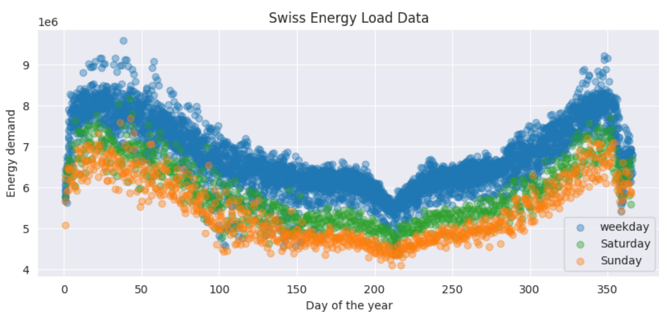

Fig. 2. Distribution of the proposed dataset per day of the year, with colors highlighting whether each day is a weekday or a day of the weekend.

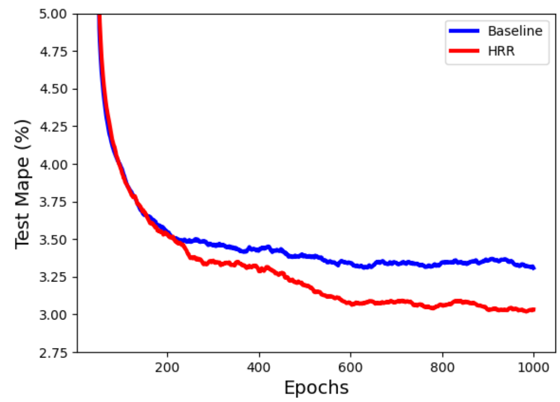

Fig. 3. Test MAPE (%) throughout the training epochs for the proposed method against baseline on Swiss Energy Market.

by the low standard deviation on all the utilized datasets. Correspondingly, in Fig.3, we provide the mean test MAPE (%) of the five executions throughout the training epochs for the proposed method against baseline on Swiss Energy Market, where the steadily improved forecasting performance of HRR is demonstrated.

TABLE I
EVALUATION OF THE PROPOSED HRR METHOD AGAINST BASELINE IN TERMS OF TEST MAPE (%) ON THE THREE CONSIDERED DATASETS.

Method	Swiss Energy Market	ISONE	Spain Energy Market
Baseline	2.82 ± 0.10	4.67 ± 0.38	6.92 ± 0.01
HRR (Proposed)	$\mathbf{2.68 \pm 0.01}$	$\mathbf{4.35 \pm 0.01}$	$\mathbf{6.86 \pm 0.01}$
Improvement (%)	$\downarrow$ 4.96	$\downarrow$ 6.85	$\downarrow$ 0.86

Subsequently, in Table II we present the evaluation of the proposed method against baseline on the main dataset of this paper, for various combinations of the percentage thresholds T_1 and T_2, i.e., the percentage threshold for defining the hard representations inside each batch, and the percentage of the nearest representations that each hard representation is forced to approach to, respectively. Best results considering the comparison of the proposed method against is printed in bold, while the best performance considering the combinations of the percentage thresholds is also underlined. Furthermore, for better comprehension, we visualize the aforementioned results using a heatmap in Fig. 4. The color gradient represents the forecasting performance in terms of test MAPE, allowing for readily identifying optimal and suboptimal combinations of the parameters T_1 and T_2. From the demonstrated results several remarks can be drawn. First, we can notice that the proposed method significantly improves the forecasting performance (up to 4.96%) for all the considered combinations of thresholds, apart from one case ($T_1 = 25\%$ and $T_2 = 25\%$). Furthermore, we can observe that for lower percentage of T_1 the HRR method performs worse in general, for all the combinations with T_2. This is marginally violated in the combinations of $T_1 = 50\% - T_2 = 50\%$ and $T_1 = 50\% - T_2 = 75\%$, where the lower percentage of T_1 performs better. Moreover, we can observe that as T_2 percentage increases, regardless of the T_1 percentage, the forecasting performance improves, except for a single case, i.e., $T_1 = 50\% - T_2 = 50\%$ against $T_1 = 50\% - T_2 = 75\%$, where the latter combination provides slightly worse performance. Finally, from the demonstrated results it is evident that we can achieve the best performance for the combination of maximum values of the two percentage thresholds.

Finally, as mentioned previously, the proposed HRR method is orthogonal to current state-of-the-art models and methods for electric load demand forecasting, and hence it could be combined with them for further improving their forecasting performance. To validate this claim, we also perform experiments applying the proposed HRR method in combination with the recent state-of-the-art AFORE method [16] on our main dataset, using the same experimental setup. The exper-

TABLE II
EVALUATION OF THE PROPOSED HRR METHOD FOR VARIOUS COMBINATIONS OF THE PERCENTAGE THRESHOLDS T_1 AND T_2 IN TERMS OF TEST MAPE (%) ON THE SWISS ENERGY MARKET

Method	$\mathbf{T_1}$ (%)	$\mathbf{T_2}$ (%)	Test MAPE (%)	Improvement (%)
Baseline	-	-	2.82 ± 0.10	-
HRR	25	25	2.86 ± 0.01	↑ 1.41
HRR	25	50	**2.75 ± 0.01**	↓ 2.48
HRR	25	75	**2.70 ± 0.01**	↓ 4.25
HRR	50	25	**2.81 ± 0.01**	↓ 0.35
HRR	50	50	**2.70 ± 0.10**	↓ 4.25
HRR	50	75	**2.72 ± 0.01**	↓ 3.54
HRR	75	25	**2.76 ± 0.10**	↓ 2.12
HRR	75	50	**2.75 ± 0.01**	↓ 2.48
HRR	75	75	**2.68 ± 0.01**	↓ 4.96

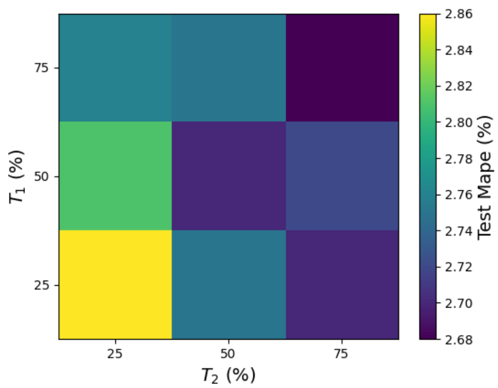

Fig. 4. Heatmap of various combinations of the percentage thresholds T_1 and T_2 of the proposed HRR method on Swiss Energy Market.

imental results are provided in Table III. From the provided results, we can first observe that the proposed HRR method performs marginally better compared to the AFORE method applied on the Swiss Energy Market, providing an improvement of 4.96% against 4.25%. In addition we can observe that the proposed method can indeed be combined with current state-of-the-art methods providing further improvements for both of the combined methods.

TABLE III
EVALUATION OF THE PROPOSED HRR METHOD APPLIED IN COMBINATION WITH THE AFORE METHOD IN TERMS OF TEST MAPE (%) ON THE SWISS ENERGY MARKET.

Method	Swiss Energy Market	Improvement (%)
Baseline	2.82 ± 0.10	-
HRR	**2.68 ± 0.01**	↓ 4.96
AFORE [16]	**2.70 ± 0.01**	↓ 4.25
AFORE & HRR	**2.66 ± 0.01**	↓ 5.67

IV. CONCLUSIONS

In this paper, we have investigated the problem of electric load demand forecasting using DL models and proposed a hard-mining based regularization technique. The proposed method can be integrated as part of the objective function of any neural network model that uses regression loss to forecast predictions. The method's effectiveness was evaluated on three datasets: two pre-existing ones and a newly curated dataset from the Swiss Energy Market, which is publicly shared to aid future researchers in the field in training more robust models and making comparisons. Results on all three cases showed that improvements in forecasting accuracy can be achieved with the proposed method for the 24-hour ahead forecasting problem.

REFERENCES

[1] Tao Hong and Shu Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[2] Naqash Ahmad, Yazeed Ghadi, Muhammad Adnan, and Mansoor Ali, "Load forecasting techniques for power system: Research challenges and survey," *IEEE Access*, vol. 10, pp. 71054–71090, 2022.

[3] Maria Jacob, Cláudia Neves, and Danica Vukadinović Greetham, *Forecasting and assessing risk of individual electricity peaks*, Springer Nature, 2020.

[4] Isaac Kofi Nti, Moses Teimeh, Owusu Nyarko-Boateng, and Adebayo Felix Adekoya, "Electricity load forecasting: a systematic review," *Journal of Electrical Systems and Information Technology*, vol. 7, no. 1, pp. 1–19, 2020.

[5] Kunjin Chen, Kunlong Chen, Qin Wang, Ziyu He, Jun Hu, and Jinliang He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, 2018.

[6] Carlo Stingl, Konstantin Hopf, and Thorsten Staake, "Explaining and predicting annual electricity demand of enterprises–a case study from switzerland," *Energy Informatics*, vol. 1, pp. 149–166, 2018.

[7] Ashkan Safari, Hamed Kharrati, and Afshin Rahimi, "Voltavistaman: Energy dynamics intelligent predictive analysis utilizing bayesian hyper-tuned neural networks–a case study on switzerland's national electricity demand," in *2024 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2024, pp. 316–322.

[8] Yangyang Mu, Ming Wang, Xuehan Zheng, and He Gao, "An improved lstm-seq2seq-based forecasting method for electricity load," *Frontiers in Energy Research*, vol. 10, pp. 1093667, 2023.

[9] James W Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of the Operational Research Society*, vol. 54, no. 8, pp. 799–805, 2003.

[10] Anthony Setiawan, Irena Koprinska, and Vassilios G Agelidis, "Very short-term electricity load demand forecasting using support vector regression," in *2009 International Joint Conference on Neural Networks*. IEEE, 2009, pp. 2888–2894.

[11] Kasun Amarasinghe, Daniel L Marino, and Milos Manic, "Deep neural networks for energy load forecasting," in *Proceedings of the IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 2017, pp. 1483–1488.

[12] Yusen He, Jiahao Deng, and Huajin Li, "Short-term power load forecasting with deep belief network and copula models," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2017, vol. 1, pp. 191–194.

[13] Shams Forruque Ahmed et al., "Short-term electrical load demand forecasting based on lstm and rnn deep neural networks," *Mathematical Problems in Engineering*, vol. 2022, 2022.

[14] Rich Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[15] Maria Tzelepi and Anastasios Tefas, "Improving the performance of lightweight cnns for binary classification using quadratic mutual information regularization," *Pattern Recognition*, vol. 106, pp. 107407, 2020.

[16] Maria Tzelepi, Paraskevi Nousi, and Anastasios Tefas, "Improving electric load demand forecasting with anchor-based forecasting method," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[17] Nikitas Maragkos, Maria Tzelepi, Nikolaos Passalis, Apostolos Adamakos, and Anastasios Tefas, "Electric load demand forecasting on greek energy market using lightweight neural networks," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2022, pp. 1–5.