

Article

Transfer Learning for Day-Ahead Load Forecasting: A Case Study on European National Electricity Demand Time Series

Alexandros Menelaos Tzortzis ^{1,*}, Sotiris Pelekis ¹, Evangelos Spiliotis ², Evangelos Karakolis ¹, Spiros Mouzakis ¹, John Psarras ¹ and Dimitris Askounis ¹

¹ Decision Support Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, 157 72 Athens, Greece; spelekis@epu.ntua.gr (S.P.); vkarakolis@epu.ntua.gr (E.K.); smouzakis@epu.ntua.gr (S.M.); john@epu.ntua.gr (J.P.); askous@epu.ntua.gr (D.A.)

² Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, 157 72 Athens, Greece; spiliotis@fsu.gr

* Correspondence: atzortzis@epu.ntua.gr

Abstract: Short-term load forecasting (STLF) is crucial for the daily operation of power grids. However, the non-linearity, non-stationarity, and randomness characterizing electricity demand time series renders STLF a challenging task. Various forecasting approaches have been proposed for improving STLF, including neural network (NN) models which are trained using data from multiple electricity demand series that may not necessarily include the target series. In the present study, we investigate the performance of a special case of STLF, namely transfer learning (TL), by considering a set of 27 time series that represent the national day-ahead electricity demand of indicative European countries. We employ a popular and easy-to-implement feed-forward NN model and perform a clustering analysis to identify similar patterns among the load series and enhance TL. In this context, two different TL approaches, with and without the clustering step, are compiled and compared against each other as well as a typical NN training setup. Our results demonstrate that TL can outperform the conventional approach, especially when clustering techniques are considered.

Keywords: short-term load forecasting; multi-layer perceptron; national energy demand; deep learning; transfer learning; time series forecasting; ensembling

MSC: 68T07



Citation: Tzortzis, A.M.; Pelekis, S.; Spiliotis, E.; Karakolis, E.; Mouzakis, S.; Psarras, J.; Askounis, D. Transfer Learning for Day-Ahead Load Forecasting: A Case Study on European National Electricity Demand Time Series. *Mathematics* **2024**, *12*, 19. <https://doi.org/10.3390/math12010019>

Academic Editor: Jianping Gou

Received: 24 October 2023

Revised: 15 December 2023

Accepted: 17 December 2023

Published: 21 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

From generation to transmission and distribution, the operation of electricity systems has to be carefully designed and planned. In this context, load forecasting is critical to properly control the operation of electricity systems and support decisions about fuel allocation, personnel management, and voltage regulation, among others. The negative implications of forecast inaccuracies in such settings have been widely explored and discussed in the literature [1–3]. Therefore, there is an increasing need for highly accurate load forecasting models which has resulted in numerous forecasting approaches through the years.

The present study focuses on short-term load forecasting (STLF) which is most relevant for the day-to-day operation of electricity systems but also supports the decisions of utilities and companies participating in electricity markets. STLF involves day- to week-ahead forecasts, while the resolution of said forecasts ranges from 15 min to one hour. STLF is considered a challenging task as electricity load time series are characterized by non-linearity, non-stationarity, randomness, and multiple seasonal patterns [4]. This is because electrical power demand originates from various electrical loads that, in turn, depend on numerous external variables, including weather and calendar factors. Additionally, the necessity for higher penetration of renewable energy sources demands more accurate

forecasts that effectively adapt to the flexibility and demand response requirements posed within transmission and distribution systems [5]. As a result, STLF has been investigated in a variety of recent research studies [6] and projects [7–9].

Conventional STLF approaches involve the selection of a proper forecasting model and its training using historical data of the target time series (e.g., national electricity demand), often accompanied by key explanatory variables [10–13]. Although such models are theoretically focused on a single time series, in practice, they may also be capable of accurately forecasting other series, provided that said series exhibit a sufficient level of similarity with the ones used for originally training the models. This process of “transferring” the knowledge gained from solving a certain learning problem (source task) to another, usually related problem (target task) [14], is commonly referred to in the literature as transfer learning (TL). The fact that TL enables the development of reusable, generalized models that can be used directly and with a minimal cost for forecasting multiple new series has popularized the research and development of TL. This has been particularly true in applications that involve the utilization of computationally expensive models, such as deep neural networks (NNs), or the prediction of relatively small data sets. In the first case, TL can be exploited to democratize the use of sophisticated models by making them widely available to the public, while in the second case to support forecasting tasks where limited data availability would have introduced several modeling challenges. Effectively, the application of TL in the STLF domain offers the opportunity to generate accurate forecasts using the concept of pre-trained, global forecasting models [15] that, with little or no retraining, can be adapted to the needs of a previously unseen (not included in the training set—zero-shot forecasting) or slightly seen (included in the training set among many other time series—few-shot forecasting) time series.

Although the research done on TL in STLF applications has been relatively limited, the concept of TL has been extensively studied in the fields of computer vision and image recognition [16–18]. The main objective of TL in such settings has been the transfer of existing knowledge, representations, or patterns learned from one context and applying that knowledge to improve performance in a different context [19]. In contrast to traditional machine learning (ML) techniques, which require past (training) and future (testing) data to have the same domain, TL allows learning from different tasks, domains, or feature spaces. This element of TL offers immense flexibility when it comes to the source, nature, and context of the source/target data. There are various TL approaches [20] that can be distinct based on the nature of knowledge that is being transferred, namely, (a) instance transfer, (b) feature representation transfer, (c) parameter transfer, and (d) relational knowledge transfer. Depending on the approach, different techniques have been proposed in the literature: from simply reusing data from the source domain as part of the target domain in a process called “re-weighting” (instance transfer) [21] to more complicated ones, such as “warm-start” (parameter transfer) [22–25], “freezing/fine-tuning” (feature representation) [23,24], and “head replacement” (parameter transfer) [26]. Note that TL techniques are not mutually exclusive. In fact, they can be blended, addressing different aspects of the forecasting problem.

1.2. Motivation and Goal of the Study

Given the benefits of TL and the limited attention it has attracted in the area of STLF, this study focuses on TL approaches for day-ahead forecasting on 27 net aggregated load time series that are available at the ENTSO-E transparency platform [27]. Our main objective is to investigate the potential accuracy improvements that knowledge transfer can offer to conventional NNs for STLF currently used by transmission system operators (TSOs) in Europe for forecasting the national day-ahead electricity demand, especially within countries that share similar load patterns. This is of high interest for electrical power and energy system stakeholders as it can enable better regulation of the balance between electricity production and consumption, reduce operational costs, and enhance the safety and robustness of the systems.

We apply TL by considering an approach which combines the warm-start and fine-tuning techniques. The former technique involves copying the weights and biases from the source task model and using them to initialize the weights of the target task's models, while the latter involves adjusting said weights and biases on the target task. Following this approach, the time and target data needed to train the respective forecasting models are significantly decreased, since a significant part of the solution has already been established from the source task's solution. To assess the performance of the proposed TL approach, we consider three different experimental setups. In the first setup, the forecasting models are optimized, trained, and evaluated for each country individually, i.e., in a conventional fashion and without using any TL operation, thus serving as a baseline. The second setup involves transferring information from a model trained on the 26 countries of the data set to a model tasked with generating forecasts for the respective remaining country. Finally, the third setup applies the same TL approach to the above on subsets of the data set that have been constructed using a clustering algorithm so that information is exchanged within countries of similar electricity demand patterns.

1.3. Related Work

There has been abundant research on NNs and their variations in energy forecasting applications. Multi-layer perceptron (MLP) architectures are frequently employed for solving STLF problems [28–31], especially when it comes to new searching algorithms for model training such as adaptive learning [28], genetic algorithms with particle swarm optimization [32], and modified harmony search algorithm [33]. Graphical Neural Networks (GNNs) have also been utilized [34] to examine the spatial correlation that data sources have with each other, while Residual Neural Network (ResNet) models [35] have been proposed as an organic way to map complex relationships between energy loads of different regions. Feed-forward NNs have also been the topic of research concerning hyperparameter tuning and optimization [36–38], aiming to increase forecast accuracy and decrease the computational costs of said models. Additionally, studies like those of De Felice Matteo and Yao Xin [39] and Vesa Andreea Valeria et al. [40] have introduced ensemble techniques as a means of reducing parameter uncertainty in NN models, leading to more consistent load forecasts.

Focusing on deep learning (DL) NNs for STLF, recent research has put a particular emphasis on more sophisticated model architectures, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks [41–43]. Many variants of RNNs and hybrid DL models [44,45] have been presented in several studies, providing forecasts concerning energy consumption at regional, national, or even building levels. More specifically, Lee and Cho [46] have done notable work implementing LSTM-based models for national-level forecasting using Korean data sets. Yuan et al. [47] have developed a real-time short-term load adjustment tool using data from the Taiwan Power Company. Wang et al. [48] made a comparative study between their proposed LSTM-based model, MLP, and SVM, using weather and schedule data from a large office located in Alaska, United States. The LSTM-based method proposed by Memarzadeh and Keynia [49] has been validated successfully on load and price data collected from the Pennsylvania–New Jersey–Maryland (PJM) and Spain electricity markets. Similarly, advanced NN architectures have gained popularity [4,13,50], introducing innovative deep STLF approaches based on feed-forward [51,52], convolutional [53], or transformer [54,55] setups.

Among the various forecasting approaches used, TL has been a key topic of research with applications in multiple sectors. There is an extensive review conducted by Iman et al. [56], presenting studies involving TL in various topics including but not limited to medical imaging, psychology, natural language processing (NLP), and quantum mechanics. TL techniques have also been employed to solve problems in the energy sector and especially STLF. Following the trend of aforementioned complicated architectures, models such as LSTM, ResNet, and TCN have been used to process data from multiple meters at the regional, national, and international levels, all achieving considerable forecasting

performance. Jung et al. [57] proposed a novel energy load forecasting scheme, collecting calendar, population, and weather data from 25 districts in Seoul. Lee and Rhee [38] proposed a two-layer transfer-learning-based short-term load forecasting (STLF) architecture using data from ISO England and GEFCOM 2012, spanning multiple US cities and states. Cai et al. [58] attempted to increase model performance to adjust to future needs for electricity consumption brought on by the rapid urbanization in Nigeria. Abdulrahman et al. [59], Zhao et al. [60], and Zhang et al. [61] focused on developing models for high-precision short-term load forecasting during COVID-19, collecting data from Switzerland, France, Germany, and Italy as well as metropolitan US cities. Also in the context of TL, clustering techniques have been employed, aiming to identify distribution nodes that have similar trends based on the energy consumption of smart grids [62,63]. Ultimately, ensemble techniques have become popular in the context of TL [34,64], once more demonstrating the importance of NN ensembles within time series forecasting applications.

Among the numerous comparative studies related to DL architectures [65–67], we can distinguish several cases where MLP architectures achieve higher accuracy than other NN models [68], as well as studies demonstrating that MLPs, aided by TL, outperform LSTM architectures [67]. Another study [38] compares TL and meta-learning techniques on national data sets of Korea and Portugal, respectively, highlighting the importance of hyperparameter optimization in the process. Additionally, depending on the nature of the use case [69,70], MLPs can be preferred over a recurrent neural network, including LSTMs. Considering these findings, alongside the numerous applications of MLP-based TL in the energy sector [57,71–73], MLPs clearly demonstrate their usefulness as a base model within the contemporary TL landscape.

Summing up, TL applications have been widely researched in the STLF domain. In terms of model architecture, modern research mostly comprises CNN (e.g., ResNet, TCN) and RNN (LSTM) architectures. However, MLPs are still used for benchmarking and occasionally preferred when relative changes over a few time steps away can be ignored. With regards to regional TL, existing research focuses on applications where TL is implemented on data sets, spanning from multi-building to multi-district and even multi-city (intra-national) levels. However, studies that employ TL at an international level usually expand at a small scale as aforementioned, involving a limited number of countries.

1.4. Contribution

This study aims to investigate the value added by TL to STLF at an international level when applied to forecast the national aggregated electricity load of different countries. In this regard, the contribution of our study can be summarized as follows:

1. While previous research has examined TL for STLF using national [38,58,59] or small-scale international transfer [60,61], the present study extends the scale of international TL to 27 EU countries, covering the entire European continent. To this end, we use load consumption data sets available at ENTSO-E [74]. We argue that global forecasting models can be used as a starting point to provide more accurate load forecasts by taking advantage of knowledge transfer opportunities among different, yet possibly similar electricity demand time series. This is a major contribution to the electrical power and energy system domain that can be of significant interest for TSOs and utilities, at both the European and international levels.
2. While many researchers have chosen to experiment with complex DL architectures [28,32–34,58], in this work we focus on MLPs, a relatively simple NN architecture that is well established in the time series forecasting domain and has been proven to often outperform more sophisticated models. Apart from the fact that MLPs allow for greater flexibility when it comes to parameter adjusting and retraining, which is a feature that makes them a widely accepted benchmarking option in the context of TL, the former are usually faster to compute and easier to implement and replicate. Said characteristics constitute a significant contribution to the research community and potential future users of the present study.

3. Different countries may exhibit different electricity demand patterns due to key latent variables (e.g., geography, population size, and economic state). Therefore, we introduce a time series clustering approach with the objective of identifying countries with similar demand patterns and facilitating TL. Although similar clustering approaches have been examined in the literature [62,63], they were applied on low voltage distribution and building-level data rather than country-level electricity demand time series. To this end, we introduce an additional TL setup in our study, where we apply TL among the countries of a given cluster rather than the complete set of countries available in the data set.

1.5. Structure of the Paper

The rest of the paper is structured as follows. Section 2 covers the methodological steps of our study including the following stages: (i) data pre-processing (sanitization of missing values, duplicates, and outliers), (ii) an exploratory data analysis that provides useful insights about the data set, (iii) the clustering process used to form groups of countries with similar load patterns, (iv) the selected model architecture, (v) the TL methodology and the respective examined setups, and (vi) the model training, validation, and evaluation procedures. Section 3 follows with the results derived from the investigated TL setups, followed by a brief discussion of the results in Section 4. Finally, in Section 5 we conclude our work and present potential future perspectives.

2. Methodology

To effectively address our TL problem, the experimental process took place using an automated machine learning operations (MLOps) pipeline developed with MLflow [75], building up to the one described by Pelekis et al. [13]. A flowchart displaying the basic concepts and stages of the pipeline can be found in Figure A1. The methodological steps of our approach are structured as described in the following sections.

2.1. Data Collection and Curation

The data set used in the present study contains the time series of the national net aggregated electricity demand of 27 different European countries (it includes the electricity demands of all EU countries at an hourly resolution, excluding Luxembourg and Cyprus, with the addition of Switzerland and Norway). The data set is available online and was obtained by the ENTSO-E transparency platform [27]. The time series, containing a total of 1,577,954 data points, spans from 2015 up to 2021, as this was the most recent version of the data collection containing full calendar years at the time when the experiments were carried out.

Prior to model training, data wrangling had to be carried out on the electrical load time series. To do so, we extended the methodology proposed in [4] to all countries included in the data set. Specifically, we proceeded with the following operations:

1. **Removal of duplicate entries:** Data may become duplicated as a result of storage or measurement errors; therefore, cleaning our data of such duplicates is necessary to avoid costly mistakes (e.g., skewed prediction results).
2. **Removal of outliers:** After calculating the mean and standard deviation of the load for each month in the data set, we removed any values that arithmetically deviated from the mean more than a certain amount, aiming to exclude any outliers. More specifically, we set the maximum distance from the mean as 4.5 times the standard deviation, so as to maintain data credibility and remove only extreme outliers. This process detected 233 different outliers, with many of them forming groups in segments of our data set as shown in Figure 1a,b.
3. **Conversion from native UTC to local time of each country:** Electricity demand time series can exhibit different patterns, depending on the time of the day (e.g., due to daylight hours). To accommodate that and align the international time series patterns through time, for each time series we inferred the timezone based on the country it referred to, changing it to local time rather than UTC which was the initial setting.

- Missing data imputation:** In the cases where imputation was required, we employed a hybrid method between weighted average of historical data and simple linear interpolation, similar to Peppanen et al. [76] and Pelekis et al. [4]. The weights of each method depend exponentially on the distance of the missing value in question from the nearest timestamp that has a value according to the formula:

$$w = e^{\alpha \times d_i}, r = w \times L + (1 - w) \times H$$

where α (empirically set to 0.3) is a positive weight parameter that depicts how rapidly the weight of simple interpolation drops as the distance to the closest non-empty value increases, d_i is the (positive) distance (in samples) to the closest (preceding or succeeding) available sample, L is the imputation suggestion by linear interpolation, H is the imputation suggestion for historical data, and r is the resulting imputed value. Therefore, as d_i increases, the contribution of the linear interpolation suggestion decreases and the contribution of the historical data suggestion increases exponentially. Said algorithm resulted in the imputation of 13,290 data points within our data set; an illustrative example of the process is demonstrated in Figure 1c,d.

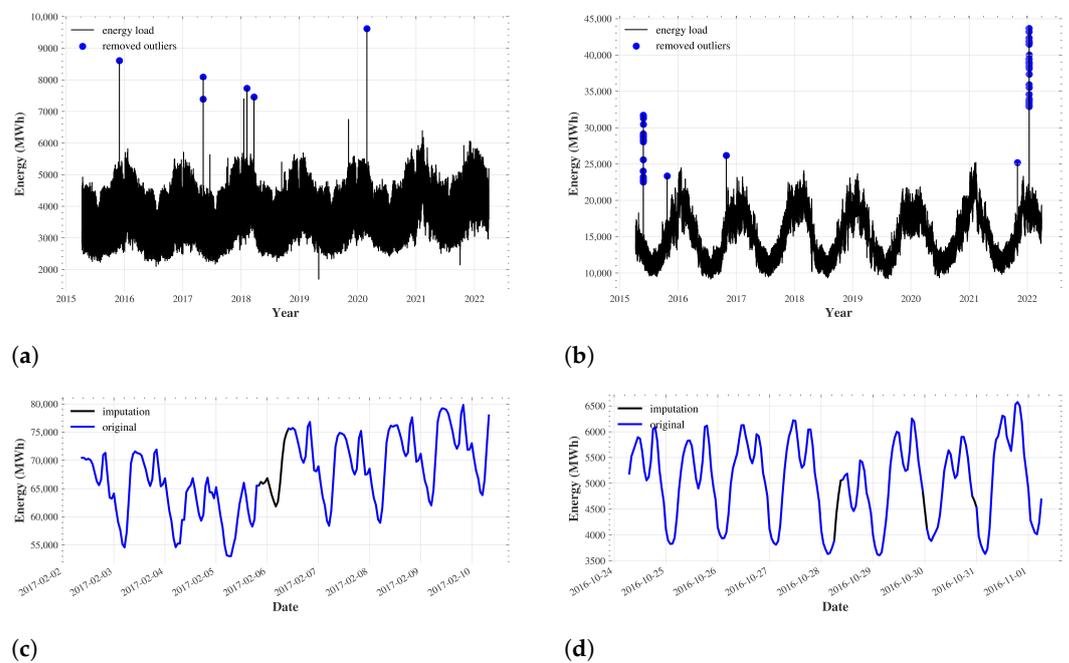


Figure 1. Visual representation of the outlier removal and missing value imputation procedures that were performed during the pre-processing of the data set. (a) Extreme outliers at random data points (Denmark). (b) Outliers forming grouping at specific data points (Norway). (c) Imputation matching complicated pattern (France). (d) Imputation on multiple close proximity data points (Greece).

2.2. Load Profiling

Proceeding with load profiling, we initially calculated the average load profiles of the entire load time series data set, grouped by and aggregated on selected time units (hour, weekday, and month). Specifically, Figure 2 illustrates the graphs of the daily (Figure 2a), weekly (Figure 2b), and yearly (Figure 2c) load profiles of the countries of the data set. Subsequently, several key patterns can be noted with respect to our data, as follows:

- Daily load profile:** In general, most countries exhibit a steep increase starting from sunrise, reaching a peak at noon, with a small decrease during the noon break time, followed by a steady decrease towards night hours (Figure 2a). This can be attributed to the increased energy demand during working hours, and confirmed by the decrease during noon break/lunch time when working activity is decreased. However, several

countries (e.g., Switzerland, France) exhibit differentiating patterns leading to the need for further investigation.

- **Weekly load profile:** We observe a steady energy demand during the working days and a steep decrease as we reach the weekend days (Figure 2b). This is attributed to the fact that commercial and industrial activity is decreased at the end of the week (Saturday and Sunday).
- **Yearly load profile:** For the vast majority of countries, the summer months have on average the lowest energy requirements and they steadily increase as we move towards winter. After that, the energy demand follows the reverse pattern, steadily decreasing until it reaches the lowest point again at summer (Figure 2c). This behavior can be attributed to the increased demand of heating loads during the winter. However, the opposite trend can be observed for southern European countries with much warmer climates (Greece, Spain, Italy, Croatia, and Portugal) which demonstrate higher energy load during summer months due to the increased cooling demand.

Taking the above into consideration, it becomes evident that most countries follow common patterns during the span of any time profile given. However, as several differences and sub-classifications can be observed—especially within the yearly (Figure 2c) and daily profiles (Figure 2a)—we are prompted to compare their energy loads and seek further classifications and dissimilarities through the clustering procedure of the following section.

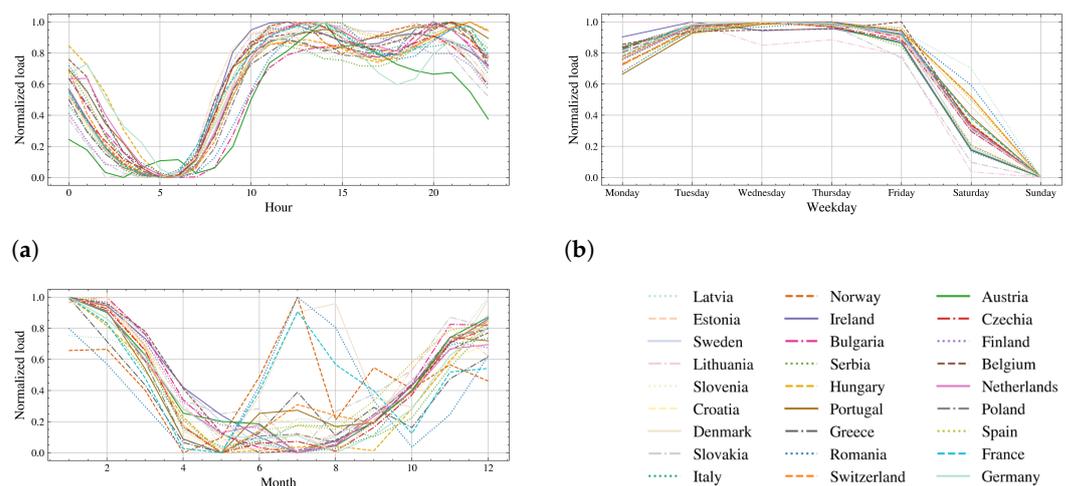


Figure 2. Average load profiles of the entire load time series data set aggregated by selected time intervals (hour, weekday, and month) for the time period of 2015 to 2021. (a) Average daily load profile (hourly resolution). (b) Average weekly load profile (daily resolution). (c) Average yearly load profile (monthly resolution).

2.3. Clustering

Motivated by the insights of the previous section regarding the international load profile patterns, we are prompted towards grouping countries together in clusters that will act as source domains during TL, aiming at higher forecasting performance. A hierarchical clustering approach using Ward’s linkage has been identified as the best option for clustering functional data that exhibit periodic trends [77,78]. To measure the distance between clusters (and by extension, countries), our approach was focused on the countries’ timely load profiles. In this direction, each country is represented as a vector consisting of 4 sub-vectors, each one containing its daily, weekly, and yearly profiles, as depicted in Figure 3. The vectors have been normalized to (i) ensure a clustering process of high quality that only depends on load shapes rather than magnitudes, as the latter are highly variable among countries, and (ii) enhance the computational efficiency of the algorithm. Note that longer-term load profiles have been ignored as we focus on STLTF (the contribution of such profiles is considered negligible for the examined forecasting horizon).

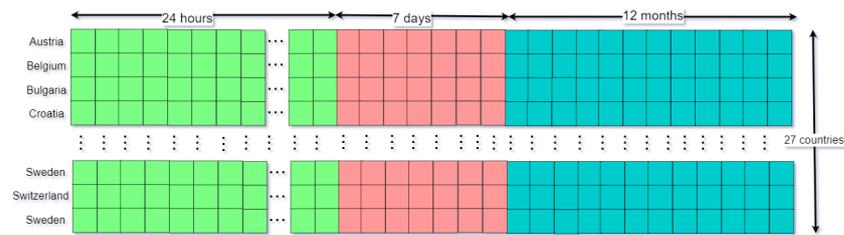


Figure 3. An illustration of the derived load profile vectors containing daily and yearly profiles for each country of the data set.

The results of the clustering process are illustrated in Figure 4, where we can observe four basic clusters: (i) Mediterranean countries (cluster 1), (ii) central European countries (cluster 2), (iii) Eastern European countries (cluster 3), (iv) Scandinavian and Baltic countries (cluster 4). Given these results, which imply a certain correlation among the countries of the same cluster, we additionally came up with several geographical and socio-economic facts that can be used to validate the results of the derived grouping by the clustering algorithm:

1. Mediterranean sea (cluster 1): Spain, Greece, Italy, and Croatia share many connections due to their unique geographical position. To develop a more integrated energy market in the Mediterranean region, many of these nations work together on electricity and gas interconnections. They seek to increase cross-border gas and electricity trade to guarantee a steady supply of energy.
2. Visegrad Group (cluster 2): Poland, Slovakia, and Hungary have made significant progress at building transnational energy pipelines, electrical connections, and transportation networks.
3. Benelux (cluster 2): Belgium and the Netherlands share a commitment to free trade as well as a history of economic cooperation, while they are also renowned for their close proximity and highly advanced logistics and transportation networks.
4. Baltic states (cluster 4): Estonia, Latvia, and Lithuania share historical ties and experiences among the Baltic states, particularly their time under Soviet authority. Since attaining their independence, they have worked to improve their collaboration.
5. Scandinavia (cluster 4): Sweden, Norway, and Denmark adopt a similar social structure, are located in close proximity to one another, and have a comparable cultural history. They work together on a variety of local concerns and have close trading relations (energy policies, resource management, and development of RES, among others).

The aforementioned classifications and details on current affairs imply that the clustering algorithm has intuitively and sufficiently grouped the observed load patterns of the European countries involved in our study.

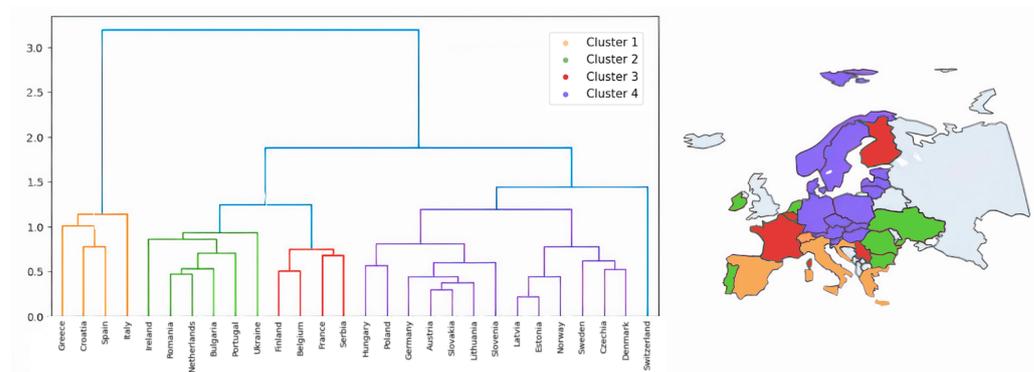


Figure 4. Dendrogram and related choropleth based on clustering performed on daily, weekly, and yearly profiles of the countries.

2.4. Selected Model

With respect to NN selection, we have opted for the MLP for several reasons. Compared to other popular NN architectures, MLPs are relatively simpler, more flexible, and similarly skillful in learning how to map inputs to outputs within a wide range of forecasting applications. Their structure allows examination with high precision and they do not require the time-consuming network design search process needed for deeper and more state-of-the-art architectures, therefore allowing for the rapid deployment of the experimental approach and interpretable comparisons. This trait also results in a more straightforward application of TL techniques, enabling the execution of a wider variety of tests compared to larger and usually computationally intensive architectures.

The MLP consists exclusively of perceptrons (or neurons), which are organized in one or more layers. Each perceptron of a layer is fed with the outputs of each perceptron of the previous layer and feeds, in turn, each perceptron of the next layer (forward pass). Equation (1) is used to calculate a given neuron's output.

$$y = f\left(\sum_{i=1}^n (w_i x_i) + b\right) \quad (1)$$

The weighted (with weights w_i) sum of the outputs (x_i) of all the neurons in the previous layer plus a bias term (b) is computed first. Then, the activation function f is applied to this result. The purpose of this function is to enable the model to capture non-linear relationships among the inputs (lags in our case) and outputs (forecasts). Various activation functions have been used in the literature, such as the sigmoid, hyperbolic tangent (tanh), and sigmoid linear units (SiLU), as well as the rectified linear unit (ReLU), with the latter being among the most popular (see Equation (2)).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

During the training process of an MLP, the back-propagation algorithm is used. The weights and biases of the MLP are iteratively updated with the use of optimizers to ensure the optimal values for its weights and biases. Among multiple optimizer options (e.g., Gradient Descent, Adagrad, Adadelta) our choice was the adaptive moment estimation algorithm (ADAM) since it is one of the most widely adopted options in current research.

2.5. Transfer Learning Setup

We define TL as the extension of ML that uses knowledge previously gained from solving a task in order to improve performance on a subsequent task (T). Given the popular study by Pan and Yang [21] to define TL, we first need to formalize the following:

Let D be a domain that includes (i) a feature space (X) and (ii) a marginal probability distribution $P(X)$ for said X . Each task T (denoted by $T = \{Y, f(\cdot)\}$) consists of (i) a label space Y and (ii) a function $f(\cdot)$, which is unobserved and can only be inferred from the training data ($\{x_i, y_i\} \mid x_i \in X$ and $y_i \in Y$), and $f(\cdot)$ is used to predict the label $f(x)$ of a new instance x . A general-purpose definition of TL is stated as follows:

Definition 1. Given a source domain D_S and a learning task T_S , a target domain D_T and learning task T_T , TL aims to help improve the learning of the target predictive function $f(\cdot)$ in D_T using the knowledge in D_S and T_S , where $D_S \neq D_T$ or $T_S \neq T_T$.

Aiming to adapt this definition to our TL use case, we can come up with the following statements for each one of our TL experiments:

- **Domains (D_S/D_T):** The feature space is the sequence of historical values used to train the model. Specifically, the length of this sequence directly depends on the optimal look-back window (l) of each trained NN. Mathematically, this leads to the following source and target feature spaces, respectively: $X_S, X_T \in \mathbb{R}^l$. Therefore, the source

and target feature spaces are the same within the TL workflow of a given experiment. In addition, the source and target marginal probability distributions for the predictor variable (future energy demand) are different ($P_S(X) \neq P_T(X)$) within the same experiment, since they depend on the countries included during the training (source training) and fine-tuning (target training) procedures. Given the above, the source and target domains differ within our TL studies ($D_S \neq D_T$).

- **Tasks (T_S/T_T):** The label space consists of the range of the possible energy demand forecasts, which are directly dependent on the forecast horizon used. The forecast horizon is constant (24 data points) within our day-ahead forecasting setting, causing source and target label spaces to be the same ($Y_S, Y_T \in \mathbb{R}^{24}$). With regards to the $f(\cdot)$ function, it corresponds to our trained NN, whose parameters (weights and biases) are used for the knowledge transfer procedure. Since $f(\cdot)$ is determined by the provided training data, and each experiment contains a different set of countries, thereby source and target $f(\cdot)$ functions are different ($f_S(\cdot) \neq f_T(\cdot)$). Given the above, the source and target tasks differ within our TL studies ($T_S \neq T_T$).

Considering these factors, our TL setups fall within the scope of homogeneous ($Y_S = Y_T$ and $X_S = X_T$) inductive transfer learning [19,79], which can be defined as follows:

Definition 2. (*Inductive Transfer Learning*). Given a source domain D_S and a learning task T_S , and a target domain D_T and learning task T_T , inductive transfer learning aims to help improve the learning of the target predictive function $f(\cdot)$ in D_T using the knowledge in D_S and T_S , where $T_S \neq T_T$.

Regarding the specifics of TL sub-techniques, within our study, we apply warm-start paired with fine-tuning. Warm-start utilizes previously trained weights or parameters. Specifically, warm-start refers to models being trained on the source domain/task and their parameters being subsequently used as initial values for the target model's parameters (parameter transfer). Note that warm-start applies only an initialization of all models' parameters, in contrast to head replacement which initializes all but the final layers. Warm-start initially involves copying the entire solution from the source task to the target task. This is where fine-tuning comes into play to adjust our source task's solution to the target task. Taking into consideration that (i) normally, NNs' weights and biases are typically initialized using random values, which does not offer any initial advantageous/disadvantageous bias for the parameters regarding the target task, and that (ii) the source and target are similar in nature (task, domain, or feature space), the warm-start technique allows us to start training the target models from a beneficial ("warmer") position on the loss surface, leading to faster convergence and decreased training time, as the optimal solution has already been significantly approached during the training of the source model.

Aiming to evaluate the added value of TL with or without clustering on the day-ahead national load forecasting of European countries, 3 main setups were implemented to support our comparisons:

1. **Baseline:** We model each country included in the data set individually. In this respect, each country has a unique, individual forecasting model which is trained, optimized, and tested using historical data from said country alone. These models are used as a baseline to evaluate the potential improvements of the two TL setups described below.
2. **All-but-One (AbO):** A TL setup according to which, given a certain country, a model is first pre-trained on the data of all the other countries included in the data set and then fine-tuned using the data of the selected country. As shown in Figure 5, our data set contains data from 27 different countries. Therefore, in this setup, we perform experiments equal to the number of countries, where each time a different country is set as the target domain, while the remaining 26 are set as the source domain. A model (f_S) is trained in the source domain and its parameters are used (via the warm-start

technique) to develop a new model (f_s) in order to generate forecasts in the target domain.

3. **Cluster-but-One (CbO):** As stated in Section 2.3, countries with similar geographical, climatic, and socio-economic characteristics may also share similarities in their electricity demand. In this context, this TL setup is developed to examine the performance of TL between countries pertaining to the same cluster. The approach is similar to the AbO setup, with the exception that in a given country’s experiment, the source domain comprises countries that belong to the same cluster, rather than the entire data set.

For the purpose of completeness and objective assessment, we also introduce a seasonally naive model of weekly seasonality, namely, seasonally naive with a look-back window of 168 values, namely sNaive(168). This naive approach uses the observed values of the same day of the previous week as day-ahead forecasts for each of the countries and is expected to produce moderate results (daily and weekly seasonalities are incorporated), thus supporting the objective benchmarking of the TL and baseline approaches [4]. Note that in addition to the seasonal naive model, in our experiments we also considered functional mean time series models since they are capable of identifying both intra-week and intra-year seasonal patterns. However, since the accuracy of said models was lower than that of the sNaive(168) model, we omit reporting their results in the rest of our analysis for reasons of brevity.

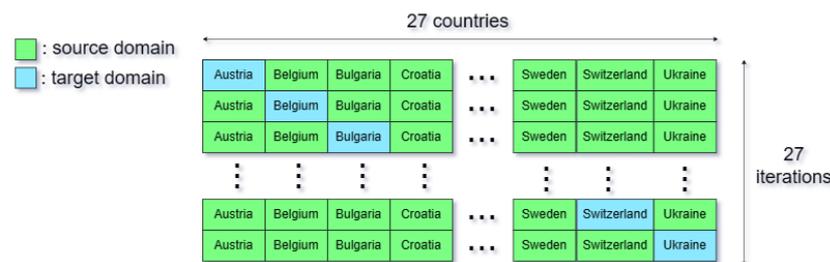


Figure 5. An illustration of the separation process of source and target domains for each experiment in the AbO setup. Each row of cells represents a separate execution of the entire ML pipeline. Source countries in each row are depicted in green, while the corresponding target domain country is depicted in blue. The same approach has been followed in the CbO setup as well, with the exception that this separation takes place internally within the countries of each cluster, excluding those that do not pertain to it.

2.6. Model Training and Validation Pipeline

In Section 2.2, we described the MLP architecture selected for our models. As mentioned, during the model training and validation procedure, all models have been optimized to minimize the L2 loss (MSE), using the ADAM optimizer. Additionally, early stopping was applied with a patience of 10 epochs.

Our data set contains data from 27 different countries. Therefore, depending on the setup and target domain, each model might be subject to different (but mostly overlapping) data points. In this context, the training process can be summarized as follows:

- Each model is trained on a training set spanning from 2015 to 2019, that is, $N_s - N_t$ unique data points, where N_s is the number of observations pertaining to the countries of the source domain and N_t is the number of observations belonging to the target country. Note that N_s can vary depending on (i) the historical data availability of each country and (ii) the TL setup. For example, the AbO setup results in much larger training data sets than CbO. The N_t variable is affected by the fact that not all countries’ observations date back to 2015, which leads to varying population sizes among target domains. This is not the case for the validation and test sets, as all time series have available observations for the years 2020 and 2021.

- Each model is optimized on the validation set (year 2020: 239,018 unique data points) to identify appropriate hyperparameter values.
- Each model is evaluated on the remaining, and previously unseen, test set (year 2021: 243,541 unique data points) without retraining the model on the full data set (union of the training and validation sets).

Note that the aforementioned data set splitting process allowed the inclusion of complete calendar years within each data set, thus maintaining the symmetry of seasonal patterns.

For each experiment, the hyperparameters of the MLPs were tuned using a multi-dimensional grid of hyperparameter values and more specifically the tree-structured Parzen estimator (TPE) optimization approach, as described by Bergstra et al. [80] and implemented in Python (version 3.8) programming language in Optuna optimization library (version 2.10) [81], similar to Pelekis et al. [4]. Additionally, the successive halving algorithm—a combination of random search and early stopping—was internally applied as a pruning criterion for TPE trials. A crucial hyperparameter that is common among all architectures is the look-back window (l), namely, the number of historical time series values (lags) that the model looks back at when being trained to produce forecasts. The l directly determines the input layer size of the trained neural network. The specific details of the optimization process for each architecture are presented in Table 1. A total of 100 trials were executed for the optimization of each NN architecture.

Following the identification of the best model architecture in terms of hyperparameter values, 20 separate networks sharing the same hyperparameter values but different pseudo-random initialization of neural weights were trained for each NN architecture. Then, an ensemble of these models was used to produce the final forecasts. To improve the robustness of the results, the final forecast of our ensemble model was derived as the average of the predictions of all models. The ensemble method was performed using Ensemble-PyTorch (version 0.1.9) [82], a unified framework for ensembling in PyTorch, with a user-friendly API for ensemble training/evaluation, as well as high efficiency in training with parallelism.

During the TL step, hyperparameter tuning was bypassed, as the hyperparameters and trainable parameters of the given source model were passed, as they were, for the initialization of the target model (warm-start technique). The retraining of the target model followed, being fine-tuned until its validation loss plateaued out (early stopping).

With respect to computing resources, the training, validation, and testing of the models were executed on an Ubuntu 22.04 virtual machine with an NVIDIA Tesla V100 GPU, 32 CPU cores, and 64 GB of RAM.

Table 1. The hyperparameters optimized per NN architecture and the respective search spaces.

Arguments	Value Range	Type
number of layers	{2, ..., 6}	Discrete
layer sizes	{128, 256, 512, 1024, 2048}	Discrete
l	{168, 336, 504, 672}	Discrete
forecast horizon	24	Fixed
learning rate	$\{1 \times 10^{-5}, \dots, 1 \times 10^{-4}\}$	Continuous
batch size	{256, 512, 1024}	Discrete

2.7. Model Evaluation

Various forecasting performance measures are common in the literature for evaluating forecasting accuracy, such as the mean absolute error (MAE), the mean squared error (MSE), the root mean square error (RMSE), the mean absolute percentage error (MAPE), the symmetric mean absolute percentage error (sMAPE), and the mean absolute scaled error (MASE) Hyndman and Koehler [83]. Within the present study, MAPE was selected as it is a widely accepted choice in STLF applications and results in interpretable measurements.

Consequently, MAPE, defined as below, served as our main evaluation measure and was used for optimizing the hyperparameters of the models:

$$MAPE = \frac{1}{m} \sum_{t=1}^m \left| \frac{Y_t - F_t}{Y_t} \right| \cdot 100(\%), \quad (3)$$

where m represents the number of samples, while Y_t and F_t stand for the actual values and the forecasts at time t , respectively.

We should note that although MAPE is frequently judged for becoming infinite, undefined, or heavily skewed when the actual values are (close to) zero, and also puts a different penalty on positive errors than on negative errors [84], it still serves as a useful measure in the STLF literature since it enables intuitive comparisons between countries of different magnitudes of electricity consumption. Moreover, given that in such applications the actual values are much greater than zero, the disadvantages of MAPE become less relevant in practice. Nevertheless, to confirm that our key findings are not affected by the selected accuracy measure, in the appendix we also report our results according to RMSE.

3. Results

Proceeding to the results of our experiments, the error metrics for each scenario have been collected in Table 2. The last row lists the average improvement (average MAPE difference) of each model compared to its baseline model. The last column contains the MAPE of the best-performing model alongside its name. The final performance of each architecture is displayed in Figures 6 and 7 (derived from Tables 2 and 3, respectively), taking into account the ensembling modeling technique that followed the hyperparameter optimization process. More details on the optimal hyperparameters and training times of the best-performing models can be found in Table A1 of the Appendix A. Each line of the table represents a country, containing (a) the name of its best-performing setup, (b) the training time of the target model (fine-tuning process), (c) the training time of the baseline model, and (d) hyperparameters of the best model. As an initial observation, sNaive(168) is heavily outperformed in all cases by all NN models (AbO, CbO, baseline). This is but an expected outcome, since naive models do not entail any kind of learning process; they only repeat past values of the time series. In fact, this model was selected as a hard baseline, helping to ensure the validity of our setups rather than to actually benchmark them. For the sake of completeness, Appendix B lists the evaluation results with respect to RMSE as well. Said results are in general alignment with those of Table 2. More precisely, 70% of the countries in our RMSE metrics support the findings of our MAPE metrics. Nonetheless, note here that the hyperparameter tuning performed during our experimental process was based on optimal MAPE, therefore justifying the presence of certain small inconsistencies.

Focusing on the effectiveness of TL, Figure 7 demonstrates that within each cluster, at least one of two TL setups achieve, on average, a lower forecasting error than the baseline. Therefore, it becomes apparent that TL, as a whole, outperforms in all cases the traditional NN approach that is represented by our baseline model. Note in Table 2 that AbO or CbO are the only alternating values, demonstrating that there is always a TL setup that outperforms the baseline NN. In this context, the last cell of the table corresponds to the average improvement of TL (0.34 %), in general, compared to the NN baseline model. Additionally, note that TL enables the advantageous initialization of target models from their respective source models, leading to much faster convergence and therefore a significant decrease in the target pipeline's (fine-tuning) execution compared to the respective baseline model execution. Said decrease in execution time ranges from 12.09% up to 66.87% (average 47%) depending on the experiment (target country). More details regarding the exact training times per experiment can be sought in Table A1 of Appendix A.

Table 2. Accuracy (MAPE) of the examined forecasting approaches (Baseline, AbO, CbO) alongside the naive model sNaive(168). The bold fonts correspond to the lowest error.

Country	Cluster No.	MAPE %				
		Baseline	AbO	CbO	sNaive(168)	Best Model
Italy	1	2.72	2.37	2.47	5.37	2.37 (AbO)
Croatia	1	3.35	3.00	2.86	6.13	2.86 (CbO)
Spain	1	2.00	1.95	2.28	4.51	1.95 (AbO)
Greece	1	3.66	3.60	3.39	7.97	3.39 (CbO)
Serbia	2	2.82	3.27	2.46	6.41	2.46 (CbO)
Portugal	2	2.24	2.79	2.23	4.31	2.23 (CbO)
Belgium	2	2.55	2.56	2.50	4.58	2.50 (CbO)
Ireland	2	2.15	2.03	2.09	3.32	2.03 (AbO)
Netherlands	2	4.21	4.16	4.26	5.59	4.16 (AbO)
France	2	4.53	2.31	2.22	7.15	2.22 (CbO)
Romania	2	2.54	2.09	2.33	4.36	2.09 (AbO)
Bulgaria	2	2.80	2.67	3.32	6.87	2.67 (AbO)
Finland	2	2.26	2.16	2.08	5.75	2.08 (CbO)
Hungary	3	2.96	3.26	2.88	5.64	2.88 (CbO)
Germany	3	2.76	3.17	2.42	4.26	2.42 (CbO)
Slovakia	3	2.06	1.94	2.17	3.95	1.94 (AbO)
Austria	3	3.07	3.02	3.04	5.32	3.02 (AbO)
Slovenia	3	3.56	3.58	3.49	6.44	3.49 (CbO)
Poland	3	2.18	2.40	2.05	4.39	2.05(CbO)
Lithuania	3	2.77	2.47	2.35	4.95	2.35 (CbO)
Switzerland	4	4.71	4.01	3.97	6.25	3.97 (CbO)
Norway	4	2.37	2.33	2.06	5.56	2.33 (CbO)
Denmark	4	3.87	2.80	2.79	5.46	2.79 (CbO)
Estonia	4	3.52	3.36	3.72	6.94	3.36 (AbO)
Czechia	4	2.21	1.83	1.96	4.97	1.83 (AbO)
Latvia	4	2.31	2.11	2.23	4.29	2.11 (AbO)
Sweden	4	3.13	2.84	3.19	6.85	2.84 (AbO)
Average error		2.94	2.74	2.70	5.47	2.61 (TL)
Average improvement (over Baseline)		-	0.20	0.24	-2.53	0.33 (TL)

Table 3. Average MAPE(%) per cluster and training setup. The bold fonts indicate the minimum error per cluster.

Cluster No.	Baseline	AbO	CbO	sNaive(168)	Best Setup
1	2.93	2.73	2.75	5.99	AbO
2	2.85	2.67	2.61	5.37	CbO
3	2.77	2.83	2.63	4.99	CbO
4	3.16	2.75	2.85	5.75	AbO

Considering the analytical results of Table 2 and Figure 6, it is notable that the AbO setup results in an average error reduction of 0.16% compared to the baseline in terms of absolute MAPE percentage. Therefore, this setup outperforms the baseline on average, even if this is not the case for all the countries in our data set. In the same direction, CbO leads to a higher average error reduction of 0.24% compared to the baseline, therefore confirming the value added by the clustering scheme proposed within our study.

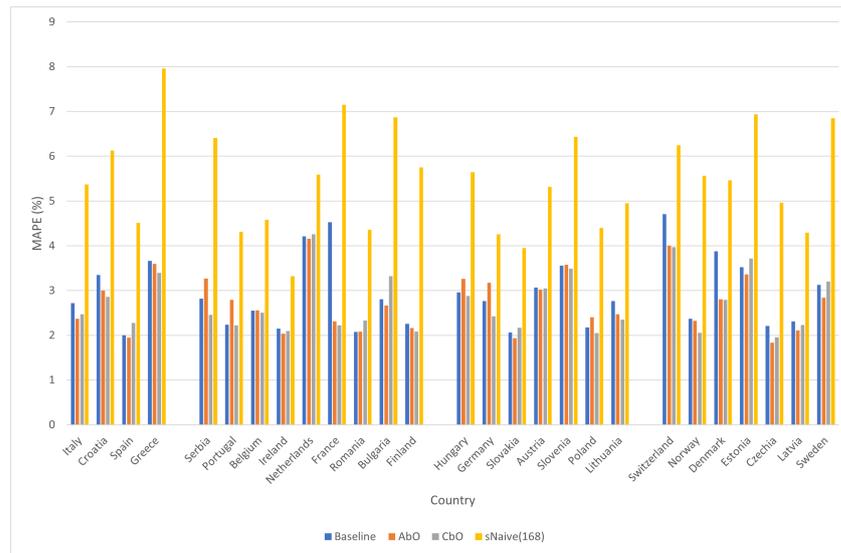


Figure 6. Side-by-side barplots depicting the MAPE(%) for each TL setup and baseline for each country.

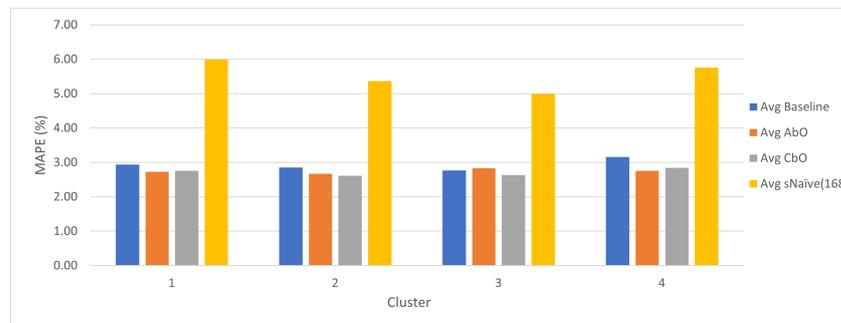


Figure 7. Side-by-side barplots depicting the average MAPE(%) for each TL setup and the baseline for each cluster.

4. Discussion

4.1. A Note on the Best-Performing TL Setup Overall

With respect to CbO, its average improvement (0.24%) compared to the baseline is higher than AbO (0.20%) and therefore has already been considered as the best-performing TL setup. To further reinforce this statement, it can be noted that CbO also holds the majority of countries where it achieved better results than its TL competitor (AbO). More specifically, among the 27 European countries that were examined in total, 12 (44.5%) of them demonstrate better results with the AbO setup compared to CbO, with the opposite being true for the other 15 (55.5%). This outcome can be attributed to the stratified training data selection proposed by our clustering methodology. In this direction, by providing a specific, highly correlated group of countries as its training set, the target model can take advantage of the similarities among countries and transfer the required knowledge in an optimal fashion. Despite the fact that CbO proved to be quite effective in most of the countries, the shortcomings of said method cannot be neglected. Countries like Spain, Romania, and Bulgaria exhibit reduced performance with CbO compared to the respective baseline MLP.

Focusing on Table 3 and Figure 7, an interesting observation lies in the fact that cluster 3 is the only case where AbO is significantly outperformed, even by the baseline model. This finding further validates the superiority of the clustering scheme, confirming that CbO is the most consistent TL setup. Regarding the poor performance of AbO in cluster 3, it should be noted however that its forecast error is not always worse than the baseline NN, being higher mainly in four (Germany, Poland, and Hungary) of the seven countries included in the cluster.

4.2. A Note on Selecting the Optimal TL Setup Based on the Target Country

Another interesting discussion topic relates to the selection between AbO and CbO depending on the target country of interest. Our results suggest that said selection should be evaluated individually per use case, and therefore specific caution is advised during the selection process. However, a good indicator for clustering could, by definition, be the degree of similarity within the energy demand patterns observed among the countries of each cluster. This can be further supported by the significant overlap between countries, where clustering was the most effective for countries that indeed have several interconnections with each other. In this direction, several interesting insights can be extracted from the official ENTSO-E transmission system map [85], as follows:

1. **Cluster 1:** Croatia, Italy, and Greece have made a lot of progress developing energy interconnections between them. There is an established 380-400 kV Transmission Line between Redipuglia (Italy) and Melina (Croatia) as well as a 400KV HVDC link between Galatina (Italy) and Arachthos (Greece)
2. **Cluster 2:** Belgium and the Netherlands are closely related in matters of energy and have created transmission lines such as the Van Eyck–Maasbracht interconnection. Additionally, France and Belgium have also developed cross-border interconnections between Mastaing–Avelgem and Avelin–Avelgem.
3. **Cluster 3:** Austria is connected by several interconnections with Hungary, Germany, Slovenia, and Slovakia (e.g., Kainarkdal–Cirkovce, Silz–Oberbrumm, etc.). Poland and Lithuania are also connected through various electricity interconnections, including the LitPol Link, a high-voltage direct current (HVDC) interconnection that allows for electricity exchange between the two countries.
4. **Cluster 4:** Sweden has a strong energy connection with Norway, primarily through hydroelectric power generation, as well as Denmark through electricity transmission lines and substations. Estonia and Latvia are also connected through electricity interconnections (e.g., Valmiera–Trisguliina), promoting regional energy cooperation and grid stability.

An interesting observation lies in cluster 1. Croatia, Italy, and Greece have developed energy interconnections with each other, while, on the contrary, Spain has not—at least at the time of writing this paper. In this context, Spain is the only country in its cluster where CbO was outperformed by the NN baseline (and AbO by extension). Despite the appeal of this theory, there are a few exceptions, such as the case of Sweden which did not respond as expected to the CbO approach.

5. Conclusions and Future Work

This study conducted a comparative analysis of TL approaches used for day-ahead national net aggregated electricity load forecasting. A feed-forward NN architecture was selected to implement the baseline forecasting models, followed by proper hyperparameter tuning and ensembling. The warm-start technique, accompanied by fine-tuning, was employed as a widely used and streamlined TL technique. Within our initial approach (AbO), we examined the value of straightforward TL. We developed a separate model for each country of the data set, aiming at forecasting its energy load (target domain). Each model draws its initial parameter values from another unique model, pre-trained using the data from the rest of the countries included in the data set (source domain). Moreover, motivated by an exploratory data analysis which revealed a certain correlation among the energy load patterns of different countries—due to similarities in their geographical and socio-economic conditions—we investigated whether the identified correlations could be exploited to further boost forecasting performance. In this direction, we developed a hierarchical clustering algorithm to derive clusters of countries and then used them to form the proper source domains for the TL pipeline by pre-training the respective global model using only the time series of the countries pertaining to the same cluster with the target country (CbO setup).

Overall, the experimental observations emerging from our research can be summarized as follows:

- Time series analysis reveals that countries exhibit a consistent pattern in their daily, weekly, and yearly load profiles.
- The process of country clustering suggests that countries sharing similar geographical and/or socio-economic factors exhibit a strong correlation and are hence clustered accordingly.
- The performance of the baseline NN model is consistently outperformed by at least one of the examined TL approaches, either AbO or CbO. Specifically, the clustering-based TL model (CbO) outperforms the rest of the models, improving the average MAPE by 0.04% compared to AbO and by 0.24% when compared to the baseline.
- Despite its overall predominance, CbO often exhibited worse performance, especially for countries with outlying socio-economic profiles and/or load patterns and magnitudes. In such cases, the AbO model is recommended to counter the negative bias that clustering entails.

The methodology and results of our study can be used by researchers and forecasting practitioners alongside certain stakeholders of the energy sector, such as transmission and distribution system operators, to develop unified TL-based models for previously unseen national load time series. Said models combine both increased accuracy and an average training acceleration of 47% compared to their respective, conventional NN approach.

Regarding future perspectives, it would be of high interest to explore the clusters and investigate their pertaining countries more thoroughly with respect to political and socio-economic factors, such as energy policies, energy providers, and cross-country grid interconnections. Such an analysis could allow the selection of more appropriate clustering algorithms or even the manual classification of countries aiming for an optimized CbO performance. Furthermore, a possible future next step of this research could involve the experimentation and TL approach with more complex NN architectures as baseline models, such as LSTM, CNN, or transformers. With respect to the specifics of TL, the use of additional TL techniques (e.g., freezing, head replacement/feature extraction) is also highly recommended in an attempt to optimize the TL pipeline, leading to a further increase in both the accuracy and computational efficiency of the models. With respect to available data sets, further performance increases could be potentially achieved by (i) enriching the utilized electricity demand data set with non-European countries to further extend the scope and scale of the study and (ii) incorporating weather forecasts or energy price fluctuations within the training procedure. Moreover, future research could involve the comparison of global forecasting models with the existing TL models (AbO, CbO) alongside their zero-shot versions, i.e., approaches that omit the fine-tuning stage. Such a comparison would lead to interesting results regarding the cases where (a) target data sets are (global) or are not (TL) available during the model training process and (b) fine-tuning is hard to conduct after handing the pre-trained model (zero-shot TL case). Finally, given that probabilistic forecasts are particularly useful for supporting the operation of electricity systems, future research could expand our work on point forecasts in the field of probabilistic forecasting, using methods that either produce quantile/distribution forecasts directly or estimate the uncertainty around the point forecasts empirically.

Author Contributions: Conceptualization, A.M.T., S.P. and E.S.; Methodology, A.M.T., S.P. and E.S.; Software, A.M.T.; Validation, S.P. and E.S.; Formal analysis, A.M.T.; Investigation, A.M.T.; Resources, S.P., E.S. and E.K.; Data curation, A.M.T.; Writing—original draft, A.M.T. and S.P.; Writing—review & editing, A.M.T., S.P., E.S. and E.K.; Visualization, A.M.T.; Supervision, S.P., S.M., J.P. and D.A.; Project administration, S.M., D.A. and J.P.; Funding acquisition, S.M. and J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by the European Union’s Horizon 2020 research and innovation program under the I-ENERGY project, grant agreement No. 101016508. Additionally, the HPC resources utilized for training and optimizing the required ML models in this study have been provided by the EGI-ACE project, which also receives funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 101017567.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://transparency.entsoe.eu>. The source code is publicly available at: <https://github.com/pelekhs/transfer-learning-forecasting>.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AbO	All-but-One
ADAM	ADaptive Moment estimation algorithm
CbO	Cluster-but-One
DL	Deep Learning
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLOps	Machine Learning Operations
MLP	Multi-Layer Perceptron
NN	Neural Network
RMSE	Root Mean Squared Error
sNaive	Seasonally Naive
STLF	Short-Term Load Forecasting
TL	Transfer Learning
TPE	Tree-structured Parzen Estimator
TSO	Transmission System Operator

Appendix A. Table of Optimal Hyperparameter Setups

Table A1. Training time and hyperparameters of the best-performing setups in terms of MAPE.

Country	Best Performing Setup	Pipeline Duration (Minutes)				Best hyperparameters			
		Source	Target	Baseline	Lookback Window	Learning Rate	Layer Number	Layer Sizes	Batch Size
Italy	AbO	402	20	51	168	0.000206	4	{1024,256,2048,1024}	1024
Croatia	CbO	90	17	40	168	0.000162	3	{256,256,1024}	256
Spain	AbO	402	16	50	168	0.000224	2	{256,2048}	512
Greece	CbO	72	24	58	504	0.000313	5	{512,256,512,128,2048}	1024
Serbia	CbO	126	21	38	168	0.000617	3	{2048,1024,1024}	256
Portugal	CbO	168	29	44	672	0.000541	4	{1024,2048,512,1024}	512
Belgium	CbO	114	21	47	168	0.000839	2	{128,512}	256
Ireland	AbO	546	29	33	336	0.000116	4	{512,256,512,2048}	256
Netherlands	AbO	492	23	55	168	0.000511	3	{1024,2048,128}	256
France	CbO	138	25	46	168	0.000428	3	{512,2048,2048}	256
Romania	AbO	534	25	46	336	0.000213	3	{512,256,256}	256
Bulgaria	AbO	1134	32	47	336	0.000169	4	{2048,256,1024,256}	256
Finland	CbO	156	23	46	168	0.000307	3	{512,1024,256}	256
Hungary	CbO	132	26	46	168	0.000928	4	{128,512,512,128}	256
Germany	CbO	156	14	31	504	0.000810	4	{128,256,128,2048}	256
Slovakia	AbO	600	26	66	168	0.000130	3	{2048,256,512}	256
Austria	AbO	348	19	49	168	0.000644	4	{2048,128,512,256}	512
Slovenia	CbO	168	29	36	168	0.000385	5	{512,256,128,128,128}	256
Poland	CbO	150	25	54	504	0.000368	4	{128,2048,1024,256}	512
Lithuania	CbO	174	25	32	504	0.000240	2	{1024,1024}	256
Switzerland	CbO	162	28	54	336	0.000136	5	{1024,256,2048,512,512}	256
Norway	CbO	150	22	51	168	0.000188	2	{128,1024}	256
Denmark	CbO	126	20	44	168	0.000140	3	{2048,1024,512}	512
Estonia	AbO	546	25	58	336	0.000117	4	{512,1024,512,2048}	256
Czechia	AbO	498	24	44	336	0.000165	2	{1024,2048}	256
Latvia	AbO	1584	40	46	168	0.000124	4	{1024,256,2048,512}	256
Sweden	AbO	570	30	60	168	0.000278	4	{256,512,256,128}	256

Appendix B. Evaluation Results in Terms of RMSE

Table A2. Accuracy (RMSE) of the examined forecasting approaches (Baseline, AbO, CbO) alongside the naive model sNaive(168).

Country	Cluster No.	RMSE %				Best Model
		Baseline	AbO	CbO	sNaive(168)	
Italy	1	1099.00	1275.70	1276.80	2851.57	1099.00 (Baseline)
Croatia	1	95.57	88.99	86.53	175.85	86.53 (CbO)
Spain	1	684.20	839.90	915.60	1992.61	684.20 (Baseline)
Greece	1	293.20	297.20	280.50	686.37	280.50 (CbO)
Serbia	2	175.70	165.00	136.00	337.47	136.00 (CbO)
Portugal	2	194.20	228.80	191.40	404.31	191.40 (CbO)
Belgium	2	335.00	346.50	341.70	633.81	335.00 (Baseline)
Ireland	2	127.60	122.40	128.90	217.57	122.40 (AbO)
Netherlands	2	703.50	713.90	717.90	1002.51	703.50 (Baseline)
France	2	3430.40	1849.50	1803.00	5857.20	1803.00 (CbO)
Romania	2	194.90	204.10	225.30	435.16	194.90 (Baseline)
Bulgaria	2	172.50	166.50	198.80	435.59	166.50 (AbO)
Finland	2	303.70	292.90	289.10	791.46	289.10 (CbO)
Hungary	3	212.20	227.60	206.10	399.04	206.10 (CbO)
Germany	3	2272.50	2528.00	2013.70	3718.17	2013.70(CbO)
Slovakia	3	89.53	101.1	107.20	196.07	89.53 (Baseline)
Austria	3	309.40	322.60	318.40	552.92	309.40 (Baseline)
Slovenia	3	112.10	120.40	112.30	174.46	112.10 (Baseline)
Poland	3	681.20	758.90	645.20	1392.36	645.20 (CbO)
Lithuania	3	57.28	52.74	50.07	100.65	50.07 (CbO)
Switzerland	4	483.20	369.90	365.30	586.48	365.30 (CbO)
Norway	4	516.00	491.20	442.90	1286.17	442.90 (CbO)
Denmark	4	217.60	158.00	158.60	304.92	158.00 (AbO)
Estonia	4	44.64	43.32	47.51	89.93	43.32 (AbO)
Czechia	4	245.30	208.70	220.20	558.88	208.70 (AbO)
Latvia	4	28.38	26.85	27.20	51.94	26.85 (AbO)
Sweden	4	652.70	603.40	677.10	1496.47	603.40 (AbO)
Average Accuracy		508.57	480.88	443.83	990.00	420.98 (TL)
Average Improvement (<i>over Baseline</i>)		-	41.76	64.75	-481.42	73.31

Appendix C. ML Pipeline Flowchart

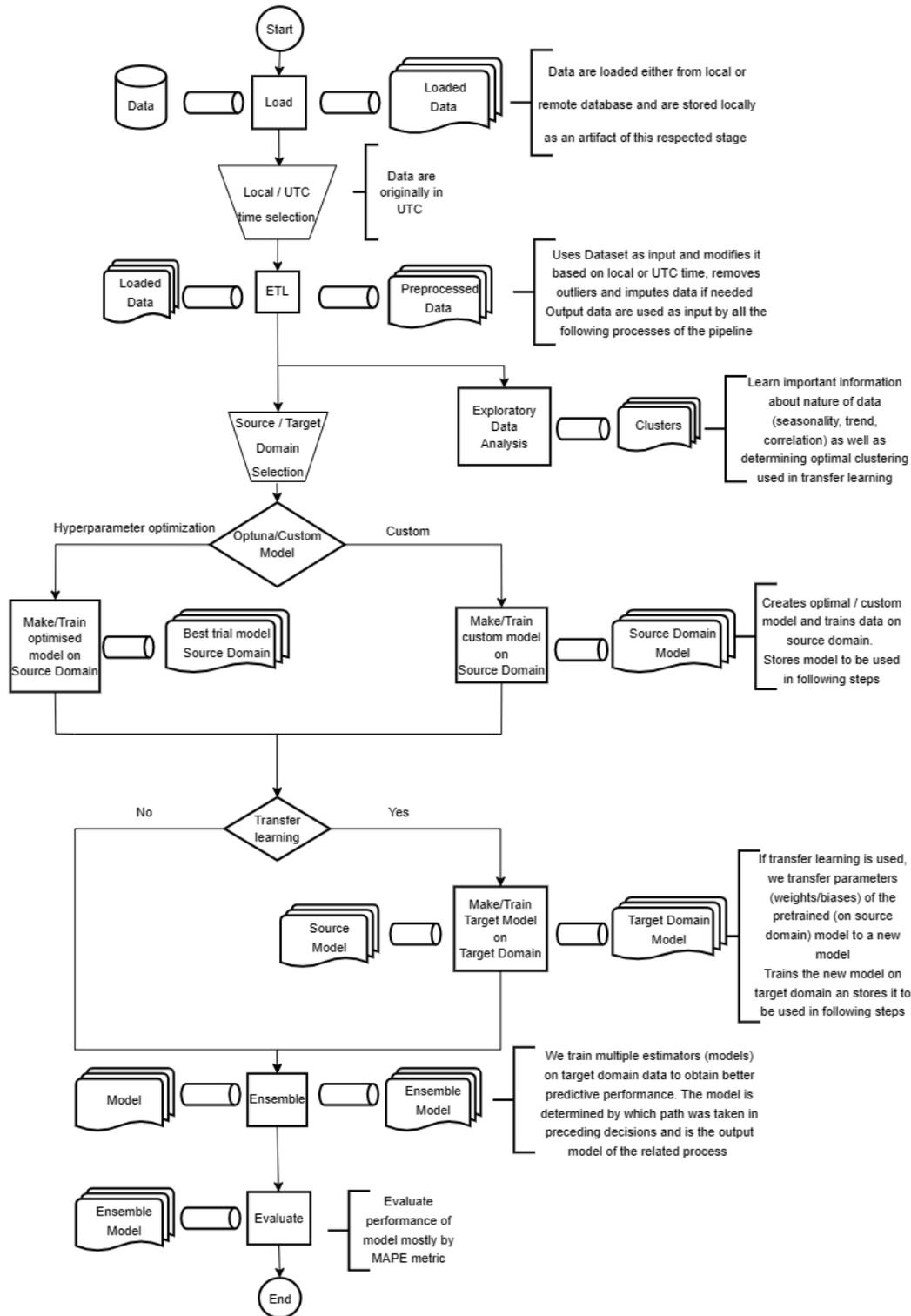


Figure A1. Illustration of the pipeline’s execution flow for the machine learning life-cycle stages.

References

1. Feinberg, E.A.; Genethliou, D. Load Forecasting. In *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence*; Chow Joe, H., Wu Felix, F., Momoh, J., Eds.; Springer: Boston, MA, USA, 2005; pp. 269–285. [\[CrossRef\]](#)
2. Alfares, H.K.; Nazeeruddin, M. Electric load forecasting: Literature survey and classification of methods. *Int. J. Syst. Sci.* **2002**, *33*, 23–34. [\[CrossRef\]](#)
3. Hahn, H.; Meyer-Nieberg, S.; Pickl, S. Electric load forecasting methods: Tools for decision making. *Eur. J. Oper. Res.* **2009**, *199*, 902–907. [\[CrossRef\]](#)
4. Pelekis, S.; Seisopoulos, I.K.; Spiliotis, E.; Pountridis, T.; Karakolis, E.; Mouzakitis, S.; Askounis, D. A comparative assessment of deep learning models for day-ahead load forecasting: Investigating key accuracy drivers. *Sustain. Energy Grids Netw.* **2023**, *36*, 101171. [\[CrossRef\]](#)
5. Pelekis, S.; Phipergias, A.; Karakolis, E.; Mouzakitis, S.; Santori, F.; Ghoreishi, M.; Askounis, D. Targeted demand response for flexible energy communities using clustering techniques. *Sustain. Energy Grids Netw.* **2023**, *36*, 101134. [\[CrossRef\]](#)
6. Bahrami, S.; Chen, Y.C.; Wong, V.W. Deep Reinforcement Learning for Demand Response in Distribution Networks. *IEEE Trans. Smart Grid* **2021**, *12*, 1496–1506. [\[CrossRef\]](#)
7. Karakolis, E.; Pelekis, S.; Mouzakitis, S.; Markaki, O.; Papapostolou, K.; Korbakis, G.; Psarras, J. Artificial Intelligence for Next Generation Energy Services Across Europe—The I-ENERGY Project. In Proceedings of the ES 2021: 19th International Conference e-Society 2021, Lisbon, Portugal, 3–5 March 2021; pp. 61–68.
8. Wehrmeister, K.A.; Bothos, E.; Marinakis, V.; Magoutas, B.; Pastor, A.; Carreras, L.; Monti, A. The BD4NRG Reference Architecture for Big Data Driven Energy Applications. In Proceedings of the 13th International Conference on Information, Intelligence, Systems and Applications, IISA 2022, Corfu, Greece, 18–20 July 2022. [\[CrossRef\]](#)
9. Pelekis, S.; Karakolis, E.; Pountridis, T.; Kormpakis, G.; Lampropoulos, G.; Mouzakits, S.; Askounis, D. DeepTSF: Codeless machine learning operations for time series forecasting. *arXiv* **2023**, arXiv:2308.00709v2.
10. Moghaddas-Tafreshi, S.M.; Farhadi, M. A linear regression-based study for temperature sensitivity analysis of iran electrical load. In Proceedings of the IEEE International Conference on Industrial Technology, Chengdu, China, 21–24 April 2008. [\[CrossRef\]](#)
11. Cui, H.; Peng, X. Short-Term City Electric Load Forecasting with Considering Temperature Effects: An Improved ARIMAX Model. *Math. Probl. Eng.* **2015**, *2015*. [\[CrossRef\]](#)
12. Haben, S.; Giasemidis, G.; Ziel, F.; Arora, S. Short term load forecasting and the effect of temperature at the low voltage level. *Int. J. Forecast.* **2019**, *35*, 1469–1484. [\[CrossRef\]](#)
13. Pelekis, S.; Karakolis, E.; Silva, F.; Schoinas, V.; Mouzakitis, S.; Kormpakis, G.; Amaro, N.; Psarras, J. In Search of Deep Learning Architectures for Load Forecasting: A Comparative Analysis and the Impact of the Covid-19 Pandemic on Model Performance. In Proceedings of the 2022 13th International Conference on Information, Intelligence, Systems and Applications (IISA), Corfu, Greece, 18–20 June 2022; pp. 1–8. [\[CrossRef\]](#)
14. Torrey, L.; Shavlik, J. Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2010; pp. 242–264. [\[CrossRef\]](#)
15. Semoglou, A.A.; Spiliotis, E.; Makridakis, S.; Assimakopoulos, V. Investigating the accuracy of cross-learning time series forecasting methods. *Int. J. Forecast.* **2021**, *37*, 1072–1084. [\[CrossRef\]](#)
16. Chang, J.; Yu, J.; Han, T.; Chang, H.j.; Park, E. A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer. In Proceedings of the 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), Dalian, China, 12–15 October 2017; pp. 1–4.
17. Gopalakrishnan, K.; Khaitan, S.K.; Choudhary, A.; Agrawal, A. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* **2017**, *157*, 322–330. [\[CrossRef\]](#)
18. Kentsch, S.; Lopez Caceres, M.L.; Serrano, D.; Roure, F.; Diez, Y. Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study. *Remote Sens.* **2020**, *12*, 1287. [\[CrossRef\]](#)
19. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [\[CrossRef\]](#)
20. Ribani, R.; Marengoni, M. A Survey of Transfer Learning for Convolutional Neural Networks. In Proceedings of the 32nd Conference on Graphics, Patterns and Images Tutorials, SIBGRAPI-T 2019, Rio de Janeiro, Brazil, 28–31 October 2019; pp. 47–57. [\[CrossRef\]](#)
21. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [\[CrossRef\]](#)
22. Künzel, S.R.; Stadie, B.C.; Vemuri, N.; Ramakrishnan, V.; Sekhon, J.S.; Abbeel, P. Transfer Learning for Estimating Causal Effects using Neural Networks. *arXiv* **2018**, arXiv:1808.07804.
23. Shafahi, A.; Saadatpanah, P.; Zhu, C.; Ghiasi, A.; Studer, C.; Jacobs, D.; Goldstein, T. Adversarially robust transfer learning. In Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, 26–30 April 2020.
24. Mitra, S.; Warushavithana, M.; Arabi, M.; Breidt, J.; Pallickara, S.; Pallickara, S. Alleviating Resource Requirements for Spatial Deep Learning Workloads. In Proceedings of the 22nd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing, CCGrid 2022, Taormina, Italy, 16–19 May 2022; pp. 452–462. [\[CrossRef\]](#)

25. Gunduz, S.; Ugurlu, U.; Oksuz, I. Transfer learning for electricity price forecasting. *Sustain. Energy Grids Netw.* **2023**, *34*, 100996. [CrossRef]
26. Gao, Y.; Mosalam, K.M. Deep Transfer Learning for Image-Based Structural Damage Recognition. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 748–768. [CrossRef]
27. ENTSO-E. ENTSO-E Transparency Platform. Available online: <https://transparency.entsoe.eu/> (accessed on 15 May 2023).
28. Ho, K.; Hsu, Y.Y.; Yang, C.C. Short term load forecasting using a multilayer neural network with an adaptive learning algorithm. *IEEE Trans. Power Syst.* **1992**, *7*, 141–149. [CrossRef]
29. Kandil, N.; Wamkeue, R.; Saad, M.; Georges, S. An efficient approach for short term load forecasting using artificial neural networks. *Int. J. Electr. Power Energy Syst.* **2006**, *28*, 525–530. [CrossRef]
30. Hayati, M.; Shirvany, Y. Artificial neural network approach for short term load forecasting for Illam region. *World Acad. Sci. Eng. Technol.* **2007**, *28*, 280–284.
31. Arvanitidis, A.I.; Bargiotas, D.; Daskalopulu, A.; Laitos, V.M.; Tsoukalas, L.H. Enhanced Short-Term Load Forecasting Using Artificial Neural Networks. *Energies* **2021**, *14*, 7788. [CrossRef]
32. Mishra, S.; Patra, S.K. Short term load forecasting using neural network trained with genetic algorithm particle swarm optimization. In Proceedings of the 1st International Conference on Emerging Trends in Engineering and Technology, ICETET 2008, Maharashtra, India, 16–18 December 2008; pp. 606–611. [CrossRef]
33. Amjady, N.; Keynia, F. A New Neural Network Approach to Short Term Load Forecasting of Electrical Power Systems. *Energies* **2011**, *4*, 488–503. [CrossRef]
34. Wu, D.; Lin, W. Efficient Residential Electric Load Forecasting via Transfer Learning and Graph Neural Networks. *IEEE Trans. Smart Grid* **2022**, *14*, 2423–2431. [CrossRef]
35. Zhang, Z.; Zhao, P.; Wang, P.; Lee, W.J. Transfer Learning Featured Combining Short-Term Load Forecast with Small-Sample Conditions. In Proceedings of the Conference Record—IAS Annual Meeting (IEEE Industry Applications Society), Detroit, MI, USA, 9–14 October 2021. [CrossRef]
36. Hernandez, L.; Baladron, C.; Aguiar, J.; Carro, B.; Sanchez-Esguevillas, A.; Lloret, J.; Chinarro, D.; Gomez-Sanz, J.; Cook, D. A multi-agent system architecture for smart grid management and forecasting of energy demand in virtual power plants. *IEEE Commun. Mag.* **2013**, *51*, 106–113. [CrossRef]
37. Farsi, B.; Amayri, M.; Bouguila, N.; Eicker, U. On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach. *IEEE Access* **2021**, *9*, 31191–31212. [CrossRef]
38. Lee, E.; Rhee, W. Individualized short-term electric load forecasting with deep neural network based transfer learning and meta learning. *IEEE Access* **2021**, *9*, 15413–15425. [CrossRef]
39. De Felice, M.; Yao, X. Short-Term Load Forecasting with Neural Network Ensembles: A Comparative Study [Application Notes]. *IEEE Comput. Intell. Mag.* **2011**, *6*, 47–56. [CrossRef]
40. Vesa, A.V.; Cioara, T.; Anghel, I.; Antal, M.; Pop, C.; Iancu, B.; Salomie, I.; Dadarlat, V.T. Energy flexibility prediction for data center engagement in demand response programs. *Sustainability* **2020**, *12*, 1417. [CrossRef]
41. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [CrossRef]
42. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †. *Energies* **2018**, *11*, 1636. [CrossRef]
43. Kwon, B.S.; Park, R.J.; Song, K.B. Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer. *J. Electr. Eng. Technol.* **2020**, *15*, 1501–1509. [CrossRef]
44. Sajjad, M.; Khan, Z.A.; Ullah, A.; Hussain, T.; Ullah, W.; Lee, M.Y.; Baik, S.W. A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. *IEEE Access* **2020**, *8*, 143759–143768. [CrossRef]
45. Rafi, S.H.; Al-Masood, N.; Deeba, S.R.; Hossain, E. A short-term load forecasting method using integrated CNN and LSTM network. *IEEE Access* **2021**, *9*, 32436–32448. [CrossRef]
46. Lee, J.; Cho, Y. National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? *Energy* **2022**, *239*, 122366. [CrossRef]
47. Yuan, T.L.; Jiang, D.S.; Huang, S.Y.; Hsu, Y.Y.; Yeh, H.C.; Huang, M.N.L.; Lu, C.N. Recurrent Neural Network Based Short-Term Load Forecast with Spline Bases and Real-Time Adaptation. *Appl. Sci.* **2021**, *11*, 5930. [CrossRef]
48. Wang, X.; Fang, F.; Zhang, X.; Liu, Y.; Wei, L.; Shi, Y. LSTM-based Short-term Load Forecasting for Building Electricity Consumption. In Proceedings of the 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, BC, Canada, 12–14 June 2019; pp. 1418–1423. [CrossRef]
49. Memarzadeh, G.; Keynia, F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electr. Power Syst. Res.* **2021**, *192*, 106995. [CrossRef]
50. Zhao, W.; Li, T.; Xu, D.; Wang, Z. A global forecasting method of heterogeneous household short-term load based on pre-trained autoencoder and deep-LSTM model. *Ann. Oper. Res.* **2022**. [CrossRef]
51. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.

52. Singh, N.P.; Joshi, A.R.; Alam, M.N. Short-Term Forecasting in Smart Electric Grid Using N-BEATS. In Proceedings of the ICPC2T 2022—2nd International Conference on Power, Control and Computing Technologies, Proceedings, Raipur, India, 1–3 March 2022. [CrossRef]
53. Yin, L.; Xie, J. Multi-temporal-spatial-scale temporal convolution network for short-term load forecasting of power systems. *Appl. Energy* **2021**, *283*, 116328. [CrossRef]
54. Huy, P.C.; Minh, N.Q.; Tien, N.D.; Anh, T.T.Q. Short-Term Electricity Load Forecasting Based on Temporal Fusion Transformer Model. *IEEE Access* **2022**, *10*, 106296–106304. [CrossRef]
55. Giacomazzi, E.; Haag, F.; Hopf, K. Short-Term Electricity Load Forecasting Using the Temporal Fusion Transformer: Effect of Grid Hierarchies and Data Sources. *arXiv* **2023**, arXiv:2305.10559v1. <https://doi.org/10.1145/10.1145/3575813.3597345>.
56. Iman, M.; Arabnia, H.R.; Rasheed, K. A Review of Deep Transfer Learning and Recent Advancements. *Technologies* **2023**, *11*, 40. [CrossRef]
57. Jung, S.M.; Park, S.; Jung, S.W.; Hwang, E. Monthly Electric Load Forecasting Using Transfer Learning for Smart Cities. *Sustainability* **2020**, *12*, 6364. [CrossRef]
58. Cai, L.; Gu, J.; Jin, Z. Two-Layer Transfer-Learning-Based Architecture for Short-Term Load Forecasting. *IEEE Trans. Ind. Inform.* **2020**, *16*, 1722–1732. [CrossRef]
59. Abdulrahman, M.L.; Gital, A.Y.u.; Ibrahim, K.M.; Zambuk, F.U.; Umar, I.M.; Yakubu, Z.I. Predicting Electricity Consumption in Residential Building's Using Deep Transfer Learning with Attention-LSTM. In Proceedings of the 2022 International Mobile and Embedded Technology Conference, MECON 2022, Noida, India, 10–11 March 2022; pp. 540–546. [CrossRef]
60. Zhao, P.; Cao, D.; Wang, Y.; Chen, Z.; Hu, W. Gaussian Process-Aided Transfer Learning for Probabilistic Load Forecasting Against Anomalous Events. *IEEE Trans. Power Syst.* **2023**, *38*, 2962–2965. [CrossRef]
61. Zhang, Z.; Liu, J.; Pang, S.; Shi, M.; Goh, H.H.; Zhang, Y.; Zhang, D. General short-term load forecasting based on multi-task temporal convolutional network in COVID-19. *Int. J. Electr. Power Energy Syst.* **2023**, *147*, 108811. [CrossRef]
62. Syed, D.; Zainab, A.; Refaat, S.S.; Abu-Rub, H.; Bouhali, O.; Ghrayeb, A.; Houchati, M.; Bañales, S. Inductive Transfer and Deep Neural Network Learning-Based Cross-Model Method for Short-Term Load Forecasting in Smarts Grids. *IEEE Can. J. Electr. Comput. Eng.* **2023**, *46*, 157–169. [CrossRef]
63. Campos, E.M.; Vidal, A.G.; Hernández Ramos, J.L.; Skarmeta, A. Federated Transfer Learning for Energy Efficiency in Smart Buildings. In Proceedings of the IEEE INFOCOM 2023—Conference on Computer Communications Workshops, INFOCOM WKSHPS 2023, New York, NY, USA, 17–20 May 2023. [CrossRef]
64. Tan, M.; Yuan, S.; Li, S.; Su, Y.; Li, H.; He, F.H. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. *IEEE Trans. Power Syst.* **2020**, *35*, 2937–2948. [CrossRef]
65. Hu, W.; Luo, Y.; Lu, Z.; Wen, Y. Heterogeneous transfer learning for thermal comfort modeling. In Proceedings of the BuildSys 2019—6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA 13–14 November 2019; pp. 61–70. [CrossRef]
66. Ma, J.; Cheng, J.C.; Jiang, F.; Chen, W.; Wang, M.; Zhai, C. A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. *Energy Build.* **2020**, *216*, 109941. [CrossRef]
67. Khalil, M.; McGough, S.; Pourmirza, Z.; Pazhooesh, M.; Walker, S. Transfer Learning Approach for Occupancy Prediction in Smart Buildings. In Proceedings of the 2021 12th International Renewable Engineering Conference, IREC 2021, Amman, Jordan, 14–15 April 2021. [CrossRef]
68. Feng, C.; Mehmani, A.; Zhang, J. Deep Learning-Based Real-Time Building Occupancy Detection Using AMI Data. *IEEE Trans. Smart Grid* **2020**, *11*, 4490–4501. [CrossRef]
69. Chen, Y.; Zheng, Y.; Samuelson, H. Fast Adaptation of Thermal Dynamics Model for Predictive Control of HVAC and Natural Ventilation Using Transfer Learning with Deep Neural Networks. In Proceedings of the American Control Conference, Denver, CO, USA, 1–3 July 2020; pp. 2345–2350. [CrossRef]
70. Chen, Y.; Tong, Z.; Zheng, Y.; Samuelson, H.; Norford, L. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. *J. Clean. Prod.* **2020**, *254*, 119866. [CrossRef]
71. Demianenko, M.; De Gaetani, C.I. A Procedure for Automating Energy Analyses in the BIM Context Exploiting Artificial Neural Networks and Transfer Learning Technique. *Energies* **2021**, *14*, 2956. [CrossRef]
72. Kazmi, H.; Suykens, J.; Driesen, J. Large-scale transfer learning for data-driven modelling of hot water systems. *Build. Simul. Conf. Proc.* **2019**, *4*, 2611–2618. [CrossRef]
73. Wu, D.; Wang, B.; Precup, D.; Boulet, B. Multiple Kernel Learning-Based Transfer Regression for Electric Load Forecasting. *IEEE Trans. Smart Grid* **2020**, *11*, 1183–1192. [CrossRef]
74. ENTSO-E. ENTSO-E Website. Available online: <https://www.entsoe.eu/> (accessed on 15 May 2023).
75. Alla, S.; Adari, S.K. *Beginning MLOps with MLFlow*; Apress: New York, NY, USA, 2021. [CrossRef]
76. Peppanen, J.; Zhang, X.; Grijalva, S.; Reno, M.J. Handling bad or missing smart meter data through advanced data imputation. In Proceedings of the 2016 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2016, Minneapolis, MN, USA, 6–9 September 2016. [CrossRef]
77. Ferreira, L.; Hitchcock, D.B. A Comparison of Hierarchical Methods for Clustering Functional Data. *Commun.-Stat.-Simul. Comput.* **2009**, *38*, 1925–1949. [CrossRef]

78. Vijaya, V.; Sharma, S.; Batra, N. Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. In Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019, Faridabad, India, 14–16 February 2019; pp. 568–573. [CrossRef]
79. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1–40. [CrossRef]
80. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for Hyper-Parameter Optimization. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 2546–2554.
81. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631. [CrossRef]
82. Ensemble PyTorch. Ensemble PyTorch | Documentation. Available online: <https://ensemble-pytorch.readthedocs.io/en/latest/> (accessed on 15 May 2023).
83. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
84. Koutsandreas, D.; Spiliotis, E.; Petropoulos, F.; Assimakopoulos, V. On the selection of forecasting accuracy measures. *J. Oper. Res. Soc.* **2022**, *73*, 937–954. [CrossRef]
85. Grid Map. Available online: <https://www.entsoe.eu/data/map/> (accessed on 15 May 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.