RESEARCH ARTICLE

Direct Short-Term Net Load Forecasting Based on Machine Learning Principles for Solar-Integrated Microgrids

GEORGIOS TZIOLIS[iD], ANDREAS LIVERA[iD], JESUS MONTES-ROMERO[iD], SPYROS THEOCHARIDES[iD], GEORGE MAKRIDES, (Member, IEEE), AND GEORGE E. GEORGHIOU

PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, Department of Electrical and Computer Engineering,
University of Cyprus, 1678 Nicosia, Cyprus

Corresponding author: Georgios Tziolis (tziolis.georgios@ucy.ac.cy)

ABSTRACT Accurate net load forecasting is a cost-effective technique, crucial for the planning, stability, reliability, and integration of variable solar photovoltaic (PV) systems in modern power systems. This work presents a direct short-term net load forecasting (STNLF) methodology for solar-integrated microgrids by leveraging machine learning (ML) principles. The proposed data-driven method comprises of an initial input feature engineering and filtering step, construction of forecasting model using Bayesian neural networks, and an optimization stage. The performance of the proposed model was validated on historical net load data obtained from a university campus solar-powered microgrid. The results demonstrated the effectiveness of the model for providing accurate and robust STNLF. Specifically, the optimally constructed model yielded a normalized root mean square error of 3.98% when benchmarked using a 1-year historical microgrid data. The k-fold cross-validation method was then used and proved the stability of the forecasting model. Finally, the obtained ML-based forecasts demonstrated improvements of 17.77% when compared against forecasts of a baseline naïve persistence model. To this end, this work provides insights on how to construct high-performance STNLF models for solar-integrated microgrids. Such insights on the development of accurate STNLF architectures can have positive implications in actual microgrid decision-making by utilities/operators.

INDEX TERMS Bayesian neural networks, machine learning, microgrid, net load forecasting, photovoltaic.

NOMENCLATURE

ANFIS	Adaptive neuro fuzzy inference system.
ANN	Artificial neural network.
BNN	Bayesian neural network.
BTM	Behind-the-meter.
CNN	Convolutional neural network.
CV	Cross-validation.
DNN	Deep neural network.
DPT	Dew point temperature.
DQA	Data quality assessment.
D_{week}	Day of the week.
ETR	Extraterrestrial radiation.
FPSe2Q	Fully parameterized sequence to quantile.
GA	Genetic algorithm.
GHI	Global horizontal irradiance.
HNL	Historical net load.
K_t	Clearness index.
LM	Levenberg-Marquardt.
LSTM	Long short-term memory.
MAE	Mean absolute error.
$MAPE$	Mean absolute percentage error.

MI	Mutual information.
MISO	Multi-input single-output.
ML	Machine learning.
M_{year}	Month of the year.
N_h	Number of hidden nodes.
n_i	Number of input features.
NLF	Net load forecasting.
n_o	Number of output nodes.
NPM	Naïve persistence model.
nRMSE	Normalized root mean square error.
NWP	Numerical weather prediction.
PCC	Point of common coupling.
P_{max}	Maximum measured net load power.
PV	Photovoltaic.
Q-Q	Quantile-quantile.
RES	Renewable energy sources.
RF	Real feel.
RH	Relative humidity.
RMSE	Root mean square error.
$RMSE_{baseline}$	Root mean square error of the baseline model (naïve persistence model).
$RMSE_{forecasted}$	Root mean square error of the forecasted model (Bayesian neural network model).
RNN	Recurrent neural network.
SS	Skill score.
SSA	Singular spectrum analysis.
SSO	Shark smell optimization.
STLF	Short-term load forecasting.
STNLF	Short-term net load forecasting.
SVR	Support vector regression.
T_{amb}	Ambient temperature.
T_{day}	Time of the day.
UCY	University of Cyprus.
VIS	Visibility.
WRF	Weather research and forecasting.
WS	Wind speed.
y_{actual}	Actual net load.
$y_{forecasted}$	Forecasted net load.
ρ	Pearson correlation coefficient.

I. INTRODUCTION

The growing integration of variable renewable energy sources (RES) poses new challenges and uncertainties to the operational management of modern power systems [1]. Forecasting the net load (i.e., the difference of aggregated load and RES production) is a required to safeguard the efficient management of energy in solar-powered grids and to optimally balance energy supply and demand [2]. In this domain, the field of direct and indirect net load forecasting (NLF) has drawn attention recently. Research efforts focus on improving the NLF accuracy by merging the load and generation forecasts. The direct NLF strategy consists of a single forecast of the net load. The indirect NLF strategy involves forecasting the load and renewable production separately, and then calculating their difference [3].

Over the past years, various short-term load forecasting (STLF) approaches have been proposed in the literature [4], [5], [6], [7], [8], [9], [10], [11]. The STLF performance in a grid-connected microgrid was analyzed in [5] by constructing artificial neural networks (ANNs) model. The ANNs model demonstrated higher accuracies when compared to other methods. Likewise, the performance of five different ANNs models was evaluated for STLF in a building with small-scale loads [6]. The forecasts revealed a relative error of 7% when the loads were utilized on a regular basis [6]. A recurrent neural network (RNN) was also developed to mitigate the uncertainty for residential load forecasting [7]. The implemented model outperformed empirical and machine learning (ML) approaches. A STLF framework for residential users based on deep learning method was presented in [8]. The proposed methodology outperformed other existing methods (i.e., deep neural network, autoregressive moving average, and extreme learning machine). A multi-scale convolutional neural network (CNN) with time-cognition for multi-step STLF was developed in [9]. The CNN outperformed recursive multi-step long short-term memory (LSTM), direct multi-step multi-scale convolution and multi-step gated convolution-based models. The authors in [10] proposed a hybrid STLF with signal decomposition and correlation analysis. The proposed model achieved reduced errors compared to traditional empirical model decomposition. A STLF using neural networks with pattern similarity-based error weights was proposed in [11]. The proposed approach achieved accurate modeling of the target function.

The increased integration of RES in modern electricity systems necessitates the need for renewable production forecasting. Specifically, the high shares of solar photovoltaic (PV) systems (entering the Terawatt era in 2022 [12]), renders PV forecasting imperative for addressing the uncertainty associated with the RES generation [13]. Numerous PV power generation forecasting methodologies have been proposed by utilizing ML principles, statistics, and weather classification methods [14], [15], [16], [17], [18], [19], [20], [21], [22]. ML-based PV power generation forecasting was proposed in [14]. The random forest regression achieved the lowest errors given by mean absolute error (*MAE*) of 0.0098. A day-ahead PV power production forecasting approach using ANNs was presented in [15]. It was validated on a dataset containing weather and PV operational data acquired from a large-scale PV power plant. The results showed that the implemented data-driven network achieved mean absolute percentage error (*MAPE*) in the range of 1.42% to 2.10% for days with different clearness indexes. Another day-ahead PV power production forecasting model was proposed in [16], using ANNs and statistical post-processing. The resulting model yielded *MAPE* of 4.7% and a normalized root mean square error (*nRMSE*) of 6.11%. Moreover, a CNN model and variational mode decomposition was also proposed for PV power forecasts [17]. The model provided a root mean square error (*RMSE*) of 0.30 kW, 0.36 kW, and 0.42 kW

for forecast horizons of 1-, 2-, and 3-hours, respectively. The authors in [18] proposed a hybrid approach for day-ahead PV power generation forecasting at the distribution system of Cyprus. The approach yielded a *nRMSE* and a *MAPE* of 10.29% and 9.11%, respectively. Overall, research efforts on day-ahead PV power forecasting demonstrated relative *RMSE* accuracies higher than 10% [19]. LSTM neural network and synthetic weather forecast were employed for short-term PV power forecasting [20]. The proposed LSTM neural network achieved a *MAPE* of 22.31%, a *MAE* of 0.36 MW, and a *RMSE* of 0.71 MW, outperforming other ML methods (i.e., RNN, generalized regression neural network, and extreme learning machine). In [21], a hybrid deep learning approach based on CNN and LSTM was proposed for PV power generation forecasting. The lowest forecasting errors were obtained for 15-minute ahead horizons (*MAE* 1.028 MW and *RMSE* 2.095 MW) in comparison to persistence and radial basis function neural network models. A hybrid ensemble deep learning framework comprising of multiple LSTM models was also proposed for short-term PV power forecasting. Its forecasts were compared to the outputs of other existing models (i.e., persistence, autoregressive integrated moving average with exogenous variable, multi-layer perceptron, and traditional LSTM) [22]. The model yielded reduced *MAE* (0.80 kW to 1.47 kW), *MAPE* (24.65% to 37.82%), and *RMSE* (1.39 to 2.09 kW) for all investigated horizons (i.e., 7.5-, 15-, 30-, and 60-minute ahead).

Major part of the research was concerned with forecasting either load (i.e., [4], [5], [6], [7], [8], [9], [10], [11]), or renewable generation (i.e., [14], [15], [16], [17], [18], [19], [20], [21], [22]), or load and RES generation separately (i.e., [23]). In the last few years, research has shifted to direct approaches of forecasting the net load. The direct NLF utilizes net load data, which are widely available in contrast with individual load and RES generation data. In addition, for the direct NLF strategy a single model is trained, offering a computational advantage over the indirect NLF strategy. Previous short-term net load forecasting (STNLF) methodologies attempted to enhance forecasting accuracies by leveraging hybrid and ML data-driven principles [2], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39]. A NLF for a PV-assisted charging station was proposed in [24]. The integrated model achieved the lowest errors of 14.65 kW (*MAE*) and 18.85 kW (*RMSE*). The authors in [25] presented a fully parameterized sequence to quantile (FPSe2Q) model for STNLF at the distribution level, that achieved *nRMSE* of 7%. A STNLF with singular spectrum analysis (SSA) and LSTM was proposed in [26]. The proposed model yielded *MAPE* in the range of 3.17% to 5.55% for the examined cases. Sun et al. [27] suggested a probabilistic day-ahead NLF method based on Bayesian deep learning. The model surpassed other state-of-the-art methods (i.e., quantile regression, multiple linear regression, quantile random forests, support vector quantile regression, and gradient boosting quantile regression). Furthermore, an ANN

model was proposed by Kobylinski et al. [28] to predict the short-term micro-scale domestic net load profiles at a 15-minute resolution. The forecasting results showed *MAE* of 5.4%. A multi-energy forecasting framework for integrated local energy systems was proposed by Zhou et al. [29] for the prediction of gas, thermal, and electrical net load. The framework was based on deep learning methods and the findings confirmed that the proposed tool was capable of: (a) minimizing forecasting complexity and cost, (b) enhancing prediction accuracy, and (c) providing lower errors compared to other forecasting methods. An ultra STNLF model based on phase space reconstruction and a deep neural network (DNN) was proposed in [30]. The DNN achieved *MAPE* values in the range of 4% to 15% for different day categories (i.e., sunny and cloudy days). The authors in [31] proposed NLF models for solar-integrated utility-scale feeders at different forecasting horizons ranging from 10- to 30-minutes. The ANN model significantly surpassed the reference persistence forecasts for most cases, exhibiting forecasting skills up to 40%. In addition, the developed ANN slightly outperformed the support vector regression (SVR) model for the majority of the investigated feeders and forecasting horizons. A forecasting engine comprising of three block cascade neural networks utilizing the shark smell optimization (SSO) algorithm was proposed for both direct and indirect STNLF applied to the power system of Ireland [32]. The forecasting models achieved *MAPE* values of 1.12% and 1.31% for direct and indirect forecasting, respectively. In a more recent study [33], the following three strategies were presented for STNLF: (a) aggregated, (b) partially aggregated, and (c) disaggregated. The results proved that the partially aggregated strategy exhibited the highest forecasting performance. A feed-forward neural network, trained by using Levenberg-Marquardt (LM) was also employed for both direct and indirect STNLF in power grids with solar and wind generation [34]. The direct NLF approach outperformed the implemented indirect model. A novel direct STNLF methodology using Bayesian neural networks (BNNs) and statistical post-processing demonstrated *nRMSE* up to 1.3% for different feeders with integrated PV systems [35]. In [36], an hour-ahead NLF methodology was presented based on adaptive neuro fuzzy inference system (ANFIS). The ANFIS provided *nRMSE* values from 0.92% to 2.12% for different time periods. A DNNs model reinforced with a genetic algorithm (GA) for electric NLF, presented forecasting *RMSE* of 1.42% [37]. A comparative analysis was performed in [38] between ANN, RNN, extreme gradient boosting, random forest, *k*-nearest neighbors, and SVR models for STNLF. Random forest outperformed all other models, achieving a *nRMSE* of 4.32%. A multi-input single-output (MISO) LSTM model and an online LSTM model were developed in [39] for short-term net energy forecasting. The implemented online LSTM improved the net energy forecasts at the household level by 7.3%, while the MISO LSTM was more efficient at the aggregated level (13.2% improvement).

Even though the above-mentioned studies provide promising results for forecasting the net load, the literature lacks a direct STNLF method applicable and demonstrated at utility-scale solar-integrated microgrids. Such method is of utmost importance as it uses only net load data. In addition, direct STNLF methods do not require knowledge of system characteristics, topologies, and physical parameters, which are rarely available for PV systems installed behind-the-meter (BTM). This renders the utilization of direct NLF approaches imperative since the constructed data-driven models can capture accurately net load profiles and improve accuracies beyond the state-of-the-art.

The purpose of this study is to advance the field of NLF by proposing a direct STNLF methodology that is applicable to solar-integrated microgrids. The paper expands on a previous study [40], where BNN was proposed for direct STNLF for two buildings located within the University of Cyprus (UCY) microgrid. In this work, the analysis is expanded by investigating different supervised learning regimes and benchmarking the performance of the proposed direct STNLF at additional buildings and the entire microgrid. The impact of net load profiles to the model during a typical week and under different irradiance conditions was also assessed. Moreover, a cross-validation (CV) method and an unseen dataset were used to validate the robustness of the model. The forecasting model development process involved the utilization of historical aggregated net load data obtained from the entire microgrid and three buildings (with and without PV systems), located within microgrid. Lastly, the model's performance was validated using common performance metrics and benchmarked against a baseline naïve persistence model (NPM).

Overall, the main contribution of this paper is the implementation of a direct STNLF using an optimized BNN model applicable for utility-scale microgrids with varying levels of PV penetration. The novel methodology followed in this study can serve as a reference guidance tool to the research/industry community for constructing high-performance data-driven STNLF models that are applicable to actual-environment microgrids. Such insights and information on the development of accurate performing and robust STNLF architectures can have positive implications in the decision-making of solar-integrated microgrids.

The rest of this paper is structured as follows: Section II analyzes the methodology followed for the development of a direct STNLF model, while Section III summarizes the results when applying the optimal constructed model to data from buildings and UCY microgrid. Finally, Section IV concludes the paper.

II. METHODOLOGY

The methodology followed to construct the direct STNLF model is presented in Fig. 1. It consists of five sequential steps: (a) experimental setup and data acquisition, (b) data quality assessment and identification of input features,
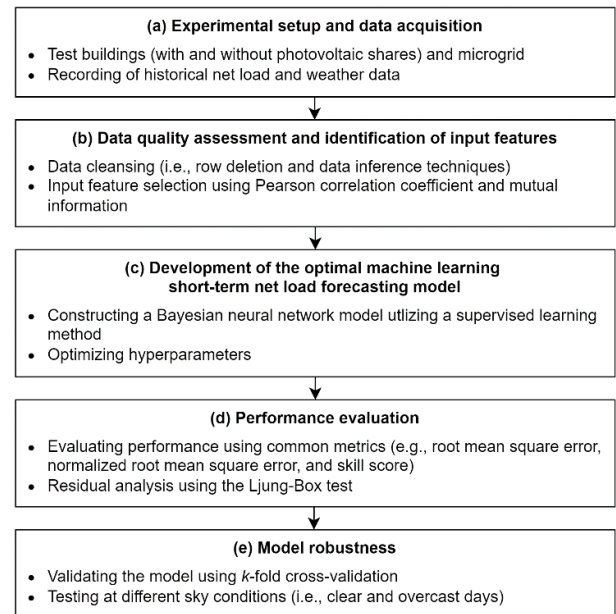
FIGURE 1. Methodology for developing and validating the proposed direct STNLF model.

(c) development of the optimal ML STNLF model, (d) performance evaluation, and (e) assessment of model's robustness.

A. EXPERIMENTAL SETUP AND DATA ACQUISITION

The developed direct STNLF model was constructed utilizing a 1-year historical dataset. The data were acquired from weather sensors and smart meters installed at UCY microgrid in Nicosia, Cyprus. More specifically, net load measurements (at a resolution of 1-second and aggregation of 60-minute recordings) were acquired from three buildings (located within the campus) and the point of common coupling (PCC) of the entire microgrid. Table 1 provides more details about the consumption and production capacities of the investigated buildings and UCY microgrid.

Over the evaluation period, historical numerical weather prediction (NWP) data (useful for energy forecasting [41]) were also calculated for the three buildings and the UCY microgrid by utilizing the weather research and forecasting (WRF) model [42]. The forecasted NWP parameters included values of global horizontal irradiance (GHI), real feel (RF), relative humidity (RH), dew point temperature (DPT), ambient temperature (T_{amb}), wind speed (WS), and visibility (VIS). In addition, historical net load (HNL) at a lag period of a week and time-related parameters such as month of the year (M_{year}), day of the week (D_{week}), and time of the day (T_{day}) were also included in the dataset [40].

B. DATA QUALITY ASSESSMENT AND IDENTIFICATION OF INPUT FEATURES

The initial stage of the methodology included the application of data quality assessment (DQA) routines to ensure high fidelity time series data [43]. Different filtering stages, detection methods, data deletion, and inference techniques

TABLE 1. Consumption and production capacity information of the buildings and microgrid.

| Test setup	Maximum measured net load power (kW)	Total PV capacity (kW$_p$)	PV penetration (%)
Site A (building with PV shares)	307.75	148.32	48.19
Site B (building without PV shares – low consumption)	199.92	0	0
Site C (building without PV shares – high consumption)	378.56	0	0
UCY microgrid (with PV shares)	1,645.77	434.80	26.42

were thus applied to the available dataset. For invalid data (i.e., erroneous and missing measurements) less than 10% of the entire dataset, the row deletion technique [43] was applied to remove those values. Alternatively, data inference/imputation techniques were used to back-fill the missing measurements [43].

Afterwards, the mutual information (*MI*) and Pearson correlation coefficient (ρ) were employed to identify the highly dependent and correlated input data features. Specifically, ρ ranges from -1 (perfect negative correlation) to $+1$ (perfect positive correlation), with zero indicating no correlation [44]. Similarly, the *MI* takes values between zero (i.e., independent variables) and infinity (fully dependent features) [45]. Threshold limits were set for $\rho \geq |0.20|$ and $MI \geq 0.25$ [40]. The input features were chosen so that at least one of ρ or *MI* value fell within the predefined threshold levels. These features were then used to construct the ML model for direct STNLF.

C. SHORT-TERM NET LOAD FORECASTING MODELS

BNNs were selected among different ML and statistical models due to their high accuracy and proven convergence in forecasting applications [46]. In the field of load forecasting, BNNs provided high-precision forecasts, while also estimating automatically the forecasting error. It was also shown that low execution time is required for their training process [18]. Other advantages of BNNs include the simplicity of managing hyperparameters, low sensitivity to small datasets and prevention of overfitting (through an automated procedure for adjusting the regularization term) [18].

BNNs are categorized under the umbrella of artificial intelligence and more specifically of deep learning networks technology with the ability to learn from the data. Such models combine neural network with Bayesian inference (i.e., a method of estimating the posterior probability). To this end, BNNs use probabilistic layers that capture uncertainty (in the predictive output) over weights and activations [47]. The development of such models involves the integration of stochastic components into the network architecture, that consist of a set of interconnected Bayesian neurons (or nodes) [18]. The nodes comprising the network are
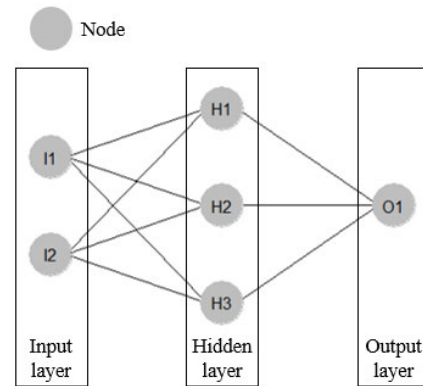
FIGURE 2. Illustration of a typical BNN architecture, consisting of 3 layers (input, hidden, and output layers) and 6 nodes (I1-I2: 2 input nodes, H1-H3: 3 hidden nodes, and O1: 1 output node).

organized in three layers, namely the input, hidden, and output layers (as depicted in Fig. 2). The input layer represents the input data to the network (for instance, 2 input features correspond to 2 input nodes), while the output layer is the desired forecasting parameter generated by the network (i.e., the net load in this study). A rule of thumb method was then used to determine the appropriate number of nodes for the hidden layer of the BNN forecasting model. According to this rule, the number of hidden nodes (N_h) should be 2/3 the size of the input layer, plus the size of the output layer. It is estimated mathematically as [48]:

$$N_h = \frac{2}{3} \cdot n_i + n_o \quad (1)$$

where n_i and n_o is the number of input features and output nodes, respectively. In the case that the value of N_h was not a whole number (e.g., a number with a fractional part), the value was rounded up to the nearest whole number.

To derive the optimal BNN model for STNLF applications, a supervised learning method was followed. Specifically, the 1-year UCY microgrid dataset was divided into three different train subsets (30%, 50%, and 70% portion of the entire dataset), while the test subset remained constant (30% portion of the entire dataset) for validation purposes. The train subsets were used to examine the impact of training set size on the model's performance and to identify the optimal training duration. During this procedure, the test subset was used for evaluating the model's forecasting performance and it was kept constant to allow a fair comparison between the results. In parallel, sequential and random dataset split approaches were also investigated to determine their impact on the forecasting accuracy of the model. In particular, successive samples (e.g., using data from consecutive months) were used for the sequential approach, while the random approach included the selection of randomly obtained samples from the entire dataset (e.g., using data from different months/seasons).

Finally, the forecasts provided by the optimal derived BNN model were benchmarked against the forecasts of a baseline NPM. The simplistic NPM uses historical net load data from

the same day and time of the previous week (i.e., one-week persistence) [49]. It operates on the basis that forecasted values remain the same as previous time-step values.

D. PERFORMANCE EVALUATION

The forecasting accuracies of the models were evaluated using the *RMSE*, *nRMSE*, and skill score (*SS*) performance metrics. The *RMSE* is the standard deviation of the prediction errors. The *nRMSE* is the *RMSE* normalized to the maximum measured net load power of the building/microgrid. The *SS* compares the score obtained by a forecast with the score of a standard forecast. A *SS* = 0% indicates no improvement of the BNN model compared to the NPM. A *SS* = 100% indicates the best possible score by the ML model. These metrics are computed as follows [16]:

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n}(y_{\text{actual,i}} - y_{\text{forecasted,i}})^2} \quad (2)$$

$$nRMSE = \frac{100}{P_{\text{max}}} \cdot \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n}(y_{\text{actual,i}} - y_{\text{forecasted,i}})^2} \quad (3)$$

$$SS = 100 \cdot \left(1 - \frac{RMSE_{\text{forecasted}}}{RMSE_{\text{baseline}}}\right) \quad (4)$$

where $y_{\text{actual,i}}$ and $y_{\text{forecasted,i}}$ are the actual and forecasted net load, respectively, n is the total number of observations, P_{max} is the maximum measured net load power of the building/microgrid, $RMSE_{\text{forecasted}}$ and $RMSE_{\text{baseline}}$ are the *RMSE* of the forecasted BNN model and baseline NPM, respectively.

Furthermore, residual analysis of the net load was performed to examine for Gaussian white noise properties [50]. The Ljung-Box test was also performed to investigate for significant correlation in the residuals [51]. The test evaluates whether the p-value is ≤ 0.05, indicating statistical significance (i.e., the null hypothesis is rejected). For p-value > 0.05, the result is not statistically significant (i.e., the null hypothesis is not rejected).

E. EVALUATING MODEL ROBUSTNESS

The k-fold CV method was employed to evaluate the robustness and stability of the optimal BNN model under different training and testing conditions [16]. The dataset was separated into k equal folds and the BNN model was trained and tested k times. For each learning sequence, k-1 different folds were used for training, while the model was tested on the remaining fold [40].

In addition, three months of unseen data were utilized to verify the robustness of the model. Lastly, to gain more insights about the performance of the optimal BNN model under various sky conditions, the daily average clearness index (K_t) was calculated. Specifically, the K_t ranges from 0 to 1, with K_t = 0 indicates overcast sky and

K_t = 1 indicates clear sky, and calculated as [52]:

$$K_t = \frac{\sum_{i=1}^{24} GHI_i}{\sum_{i=1}^{24} ETR_i} \quad (5)$$

where GHI_i and ETR_i are the global horizontal irradiance and the extraterrestrial radiation at hour i, respectively. In this study, ETR was simulated using solar positioning calculators [53].

The K_t was thus calculated for the unseen dataset to provide a thorough performance investigation of the optimal BNN model on unseen data and different sky conditions.

III. RESULTS

Section III describes the obtained performance verification results from the application of the direct STNLF model to the historical yearly dataset of the three buildings and the entire microgrid.

A. INPUT FEATURE ENGINEERING

An initial investigation pertaining to the implementation of an optimal BNN forecasting model was performed. The analysis presented useful information on the dependency of the different input variables of *HNL*, M_{year}, D_{week}, *DPT*, *RF*, T_{amb}, T_{day}, *RH*, *VIS*, *GHI*, and *WS* to the net load output. Table 2 summarizes the evaluation results of the input feature dependency and correlation to the output. It can be observed that the *HNL* input feature exhibited the highest ρ and *MI* value. Moreover, the ρ values for M_{year}, D_{week}, *DPT*, *RF*, T_{amb}, and T_{day} were greater than the predetermined threshold. Likewise, M_{year}, T_{day}, and *GHI* also yielded values within the *MI* threshold. For *RH*, *VIS*, and *WS*, ρ and *MI* values below the predetermined threshold limits were obtained and thus these input features were not selected.

TABLE 2. Input feature dependency and correlation to the net load output.

Input feature	ρ	MI
HNL	0.87	1.08
M_{year}	0.43	0.31
D_{week}	-0.36	0.21
DPT	0.33	0.12
RF	0.30	0.17
T_{amb}	0.27	0.16
T_{day}	0.21	0.34
RH	-0.06	0.07
VIS	0.04	0.01
GHI	0.03	0.25
WS	0.01	0.02

B. SUPERVISED MACHINE LEARNING MODEL AND OPTIMIZATION

Five models based on different dataset split training approaches (i.e., sequential and random) and different train set durations (i.e., 30%, 50%, and 70% portion of the entire dataset) were constructed. Table 3 depicts the results (given by *nRMSE*) of the devised models with different input

features, when applied to the entire UCY microgrid. The models exhibited *nRMSE* values ranging from 10.47% to 25.72% when trained with a 30% sequential train subset. Lower errors (ranging from 4.03% to 4.67%) were obtained when training at a 30% portion random train set. Furthermore, for 50% train set portion, the *nRMSE* ranged from 6.97% to 12.78% and from 4.08% to 4.66% for sequential and random train set split, respectively. Finally, for 70% train set portion, the *nRMSE* of the models ranged from 4.63% to 5.21% and from 3.98% to 4.66% for sequential and random training, respectively. In summary, the results showed that higher accuracies were obtained when using random training and larger train set duration (i.e., 70% portion of the entire dataset). Overall, the results showed that when training the model with larger amount of data over the yearly period, it was capable to capture the underlying consumption and production profiles and provide improved forecasts.

Ultimately, the optimal BNN model yielded a *nRMSE* of 3.98% when using 70:30% random training and testing (see Table 3). The network interconnection diagram of the model is shown in Fig. 3. The model comprises of 8 input features (HNL, M_{year}, D_{week}, DPT, RF, T_{amb}, T_{day}, and GHI), 7 hidden nodes, and 1 output node (net load).
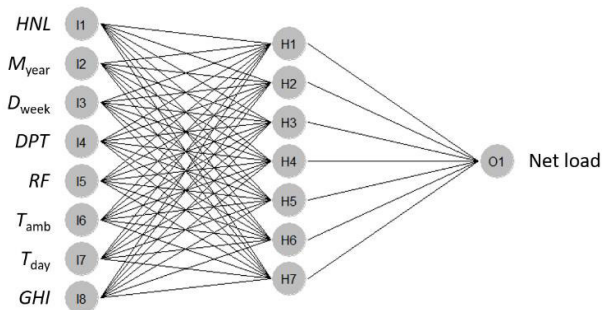

FIGURE 3. Input features, hidden, and output nodes of the devised forecasting model.

C. PERFORMANCE VALIDATION OF BNN MODEL
Fig. 4 presents the performance of the optimal direct STNLF model (constructed using a supervised 70:30% random training and testing approach) when applied to the three buildings and the entire microgrid, over the test set period. For sites A, B, and C, the direct STNLF model yielded a daily mean *nRMSE* of 4.81%, 5.35%, and 4.75%, respectively. The low absolute error deviation of 0.54% obtained when comparing the exhibited *nRMSE* values of a site with PV (site A) and without PV (site B), provided evidence that the accuracy of the BNN model was not affected by the levels of PV penetration at the investigated sites. Accordingly, the model achieved a *nRMSE* of 3.98% when applied to the entire microgrid.

The optimal direct STNLF model achieved *nRMSE* values below 10% for sites A and B (as shown in Fig. 4a and Fig. 4b) for 95% of the test set period days. Similarly, for site C the optimal BNN model yielded *nRMSE* values below 10% for
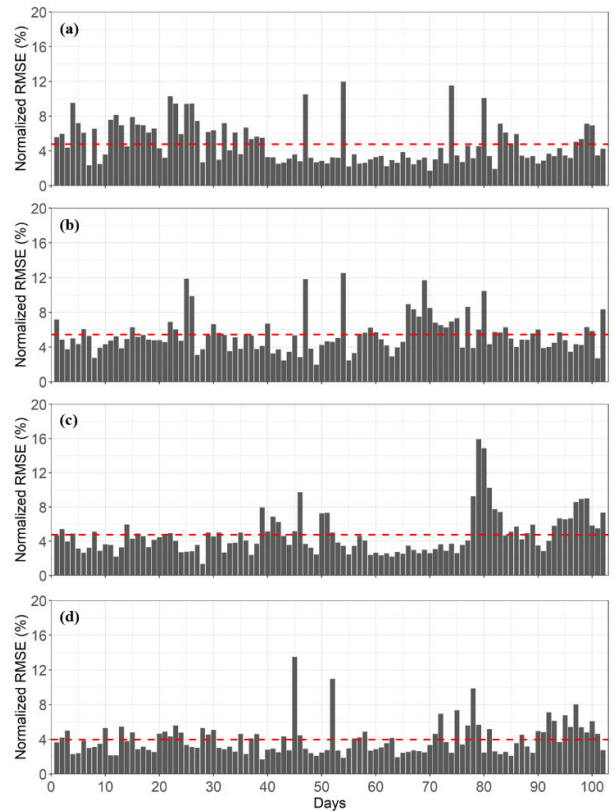

FIGURE 4. Daily *nRMSE* of the optimal BNN model for (a) Site A, (b) Site B, (c) Site C, and (d) UCY microgrid over the test set period. The red dashed line demonstrates the mean *nRMSE*.

97% of the test set period days (see Fig. 4c). Lastly, for the entire microgrid, 98% of the daily *nRMSE* values were below 10% (see Fig. 4d). The low *nRMSE* values (ranging from 3.98% to 5.35%) indicated the applicability of the proposed BNN model on the three sites and microgrid level regardless of the PV shares. The obtained low forecasting errors are attributed to the BNN's ability to capture the non-linear behavior of load and PV generation. It is worth noting that the days when daily *nRMSE* values exceeded 10% (for all three sites and the entire microgrid) were attributed to maintenance events and extreme consumption conditions. The extreme consumption values were caused by unexpected/extreme weather conditions (e.g., heat wave) that affect both the generation and consumption, and hence net load profiles. For such instances, the mode's performance ability was reduced and thus increased *nRMSE* values were obtained due to unexpected net load profiles (caused by the sudden consumption variation).

To assess the quality of the optimal direct STNLF model, residual analysis was conducted over the test set period for the UCY microgrid dataset. Specifically, the p-value obtained from the Ljung-Box test was greater than 0.05, thus failing to reject the null hypothesis and demonstrating the randomness of residuals. Moreover, the residual distribution profile (depicted in Fig. 5a) provided evidence that the residuals were normally distributed, indicating non-seasonality and lack of

TABLE 3. Input features, hidden nodes, and *nRMSE* of the developed models for different training dataset split approaches and durations for UCY microgrid.

Model No.	Input features	Number of hidden nodes	nRMSE (%)					
			Sequential training 30%	Sequential training 50%	Sequential training 70%	Random training 30%	Random training 50%	Random training 70%
1	HNL, M_{year}, GHI	3	13.54	7.21	4.99	4.67	4.66	4.66
2	$HNL, M_{year}, D_{week}, GHI$	4	10.47	6.97	4.70	4.54	4.41	4.29
3	$HNL, M_{year}, D_{week}, DPT, T_{day}, GHI$	5	13.04	7.11	4.63	4.17	4.23	4.02
4	$HNL, M_{year}, D_{week}, DPT, RF, T_{day}, GHI$	6	25.72	8.79	5.13	4.03	4.11	4.02
5	$HNL, M_{year}, D_{week}, DPT, RF, T_{amb}, T_{day}, GHI$	7	23.71	12.78	5.21	4.04	4.08	3.98
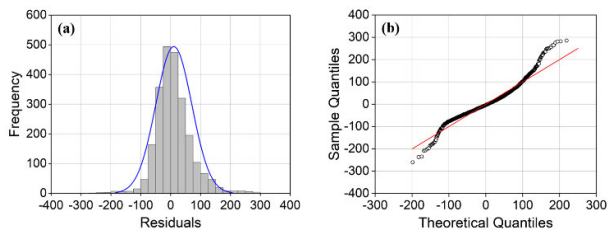
FIGURE 5. (a) Distribution and (b) Q-Q plots of the residuals derived from the application of the BNN model to the UCY microgrid test set. The blue line demonstrates the density curve of a normal distribution, while the red line demonstrates the 45-degree reference line.

FIGURE 7. Daily *nRMSE* of the optimal BNN model when applied to the UCY microgrid (unseen dataset). The red dashed line demonstrates the mean *nRMSE,* while the blue color code shows the daily average clearness index (K_t).

autocorrelation. In addition, the quantile-quantile (Q-Q) plot verified the normal distribution of the residuals since the data points closely followed the reference line (see Fig. 5b).
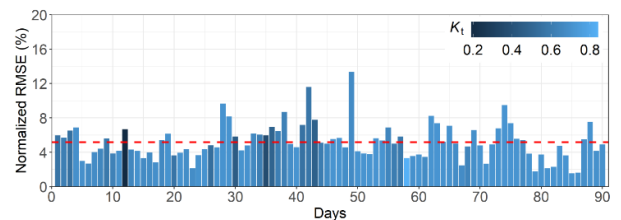
D. ROBUSTNESS VALIDATION

1) CROSS-VALIDATION EVALUATION

The *k*-fold CV results obtained when the BNN model was applied to 10 constructed folds of the entire UCY microgrid data, demonstrated *nRMSE* values ranging from 3.70% to 6.24%, with an average *nRMSE* of 5.13% (see Fig. 6). The low *nRMSE* values exhibited for all folds proved that the proposed BNN model can provide reliable forecasts for solar-integrated microgrids. The accurate forecasts obtained for all folds verified the robustness of the BNN model for capturing time-varying behaviors in a short duration dataset.

FIGURE 6. *nRMSE* obtained by the optimal BNN model during the *k*-fold CV a the microgrid level [40]. The blue dashed line demonstrates the average *nRMSE* obtained for all folds.

2) VERIFICATION ON UNSEEN DATASET

The optimal BNN model was initially validated through simulation, prior to its application to the unseen data of the microgrid. Specifically, the model was trained using a 70% random train subset. Its performance was then validated on a

new three-month unseen dataset of the entire UCY microgrid. The validation results are presented in Fig. 7 for the 90 days of the unseen dataset and at various sky conditions. The optimal BNN model yielded daily *nRMSE* values ranging from 1.55% to 13.35%. Specifically, for 2 days (i.e., 2.22%) of the unseen dataset the daily *nRMSE* values exceeded 10% due to extreme consumption conditions. Furthermore, a daily mean *nRMSE* of 5.16% was obtained when applying the model to the unseen dataset. The obtained low *nRMSE* proved model's robustness for new datasets. Also, the performance of the proposed BNN model was not affected by irradiance conditions since low errors were obtained for both clear and overcast days (indicated by the blue color code in Fig. 7). This indicates BNN model's ability to capture efficiently the net load profiles.

E. BENCHMARKING OF THE MACHINE LEARNING MODEL AGAINST THE NAÏVE APPROACH

The performance of the BNN model was further benchmarked against the forecasts of a baseline NPM. Fig. 8 presents the daily *nRMSE* of the NPM over the test set period for the three buildings and the UCY microgrid. The NPM achieved daily mean *nRMSE* values of 5.47%, 6.38%, and 5.65% for site A, B, and C, respectively. Additionally, the NPM achieved a daily mean *nRMSE* of 4.75% for the entire microgrid.

Several high error values (reaching up to 17% for some days) were obtained for the NPM over the test set period. More specifically, daily *nRMSE* values higher than 10% were obtained for 14% (see Fig. 8a), 13% (see Fig. 8b), 10% (see Fig. 8c), and 4% (see Fig. 8d) of the days in the test set. The high percentage of the daily *nRMSE* values greater than 10% demonstrated the weakness of NPM to capture the
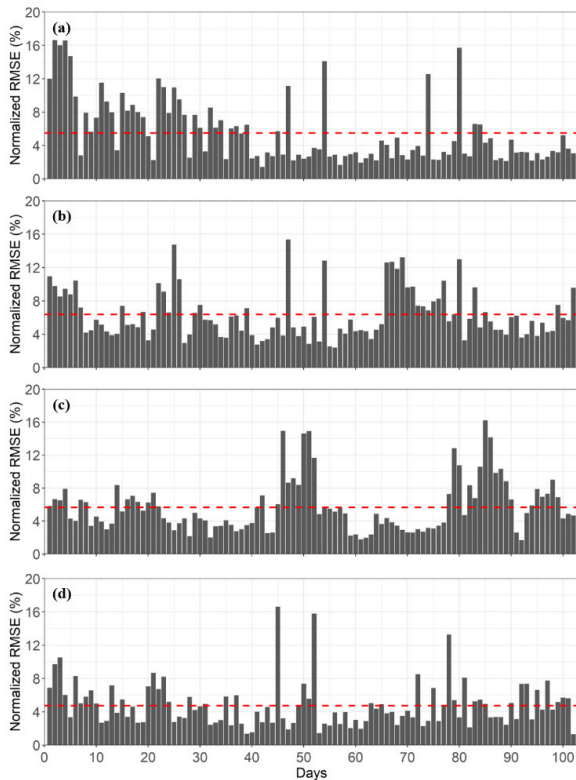
FIGURE 8. Daily *nRMSE* of the NPM for (a) Site A, (b) Site B, (c) Site C, and (d) UCY microgrid over the test set period. The red dashed line demonstrates the mean *nRMSE*.

TABLE 4. Forecasting error of NPM and BNN for the buildings and microgrid under study over the test set period.

| Test setup | NPM | | BNN | | |
	RMSE (kW)	nRMSE (%)	RMSE (kW)	nRMSE (%)	SS (%)
Site A (building with PV shares)	20.31	5.47	16.53	4.81	18.61
Site B (building without PV shares – low consumption)	13.91	6.38	11.39	5.35	18.12
Site C (building without PV shares – high consumption)	24.50	5.65	19.99	4.75	18.41
UCY microgrid (with PV shares)	76.87	4.75	63.21	3.98	17.77

non-linear behavior of load and PV generation. Hence, the NPM cannot provide accurate forecasts for many of the test days due to its simplistic nature; it assumes that future values will remain the same as current values. Despite its limitations, the NPM is a simplistic method that can be used as a baseline model for performance comparisons.

F. OVERALL COMPARISON

Table 4 summarizes the forecasting errors obtained for the optimally constructed direct STNLF model when applied to the buildings and microgrid under study over the test

set period. The NPM forecasts are also provided for comparative purposes. The optimally constructed model yielded *nRMSE* values from 4.75% up to 5.35% for the three buildings, whereas for the entire microgrid the *nRMSE* was 3.98%. The results (*nRMSE* values less than 5.35%) proved that the BNN model is consistent and adaptable for buildings with and without PV shares (irrespective of the consumption level), and for solar-integrated microgrids.

The NPM reported *nRMSE* values ranging from 5.47% to 6.38% for the three buildings, while a *nRMSE* of 4.75% was obtained for the entire microgrid. Comparing the simplistic NPM with the BNN model, *RMSE* differences of up to 4.51 kW and 13.66 kW were obtained for the three buildings and the microgrid, respectively, in favor of the BNN model. In addition, the ML model outperformed the baseline NPM, presenting lower *nRMSE* values (differences up to 1.03% for the sites A to C, and 0.77% for the entire microgrid). Furthermore, *SS* values of approximately 19% were acquired for the three buildings and 17.77% for the entire microgrid, demonstrating the superiority of the BNN model for providing improved forecasts.

Furthermore, Fig. 9 shows the plots of the actual and forecasted net load profiles obtained by the BNN model and NPM for the UCY microgrid over a typical week. The exhibited profiles demonstrated that the BNN model achieved forecasts that were in close agreement to the actual net load. This indicates the BNN's ability to capture non-linear net load profiles. In contrast, the NPM failed to capture the net load behavior throughout the week. Hence, more sophisticated algorithms (such as ML-based) are required to forecast complex net load patterns.

Fig. 10a depicts the performance of the optimal BNN model and NPM applied to the UCY microgrid for low irradiance conditions (*GHI* < 200 W/m^2). The BNN model outperformed the NPM (daily *nRMSE* of 5.17%), achieving a lower daily *nRMSE* (i.e., 4.74%). Similarly, Fig. 10b, shows the net load profiles obtained from the two STNLF models for high irradiance conditions (reaching *GHI* values up to 885 W/m^2 during the day). It can be observed that the BNN model outperformed the forecasts provided by NPM (daily *nRMSE* of 9.73%), achieving a daily *nRMSE* of approximately 4%. High errors were observed by the NPM during daytime hours in contrast to the forecasts of the BNN. The results provided evidence that the BNN forecasts were improved for higher solar irradiation during clear sky days. This is attributed to the fact that the direct BNN model was developed to provide improved STNLF accuracies for solar-integrated microgrids, thus, high and stable solar irradiance profiles (without ramps) favor the capability of the model to achieve accurate forecasts.

The findings of this study are aligned with the ones reported in the literature, where hybrid and ML-based approaches outperformed the simplistic persistence models (with the latter showing a *nRMSE* of 7.04% [38]). The developed hybrid and ML NLF models reporte *nRMSE* values ranging from 0.92% to 7% [25], [35], [36], [38]. Specifically,
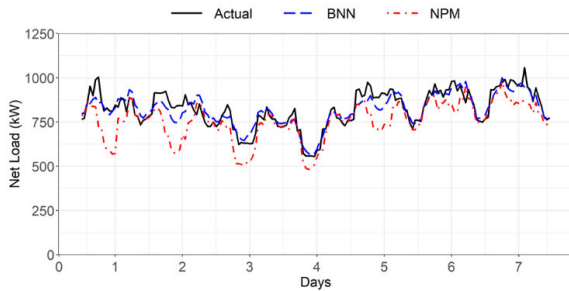
FIGURE 9. Net load profiles given by the optimal BNN model and NPM applied to the UCY microgrid for a typical week.
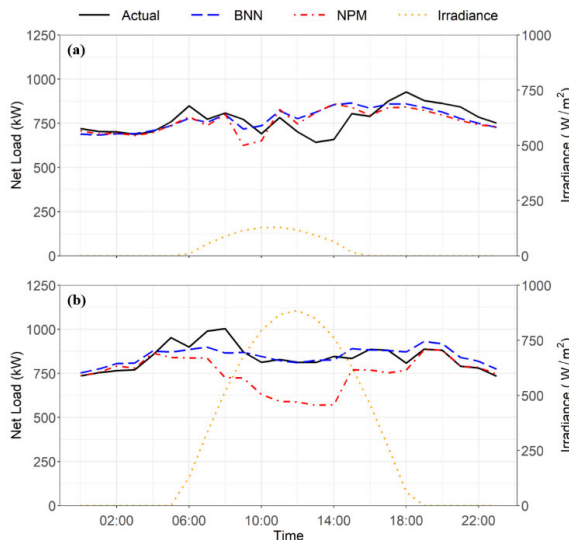
FIGURE 10. Performance of the optimal BNN model and NPM applied to the UCY microgrid for days with (a) low and (b) high irradiance conditions.

an ANFIS model achieved *nRMSE* values between 0.92% to 2.12% when forecasting the net load of a power system [36]. A BNN with statistical post processing reported *nRMSE* values from 1.02% to 1.29% for distribution feeders [35]. In addition, a random forest model yielded a *nRMSE* of 4.32% for a renewable integrated microgrid [38]. Moreover, a hybrid FPSe2Q model showed a *nRMSE* of 7% at the distribution level [25]. Lastly, the proposed BNN achieved a *nRMSE* of 3.98% for an entire utility-scale microgrid with PV shares.

Although the literature revealed the above-mentioned forecasting errors, the NLF models were evaluated under different conditions (e.g., location, datasets, application levels, systems, sizes, etc.). Therefore, no consensus regarding the optimal model can be derived, since the models have never been extensively compared "side-by-side" and/or under the same conditions.

IV. CONCLUSION

The increasing penetration of RES in microgrids as well as the uncertainty linked with their production present new challenges in the management and operation of power systems. A major challenge in optimally planning and scheduling the operation of renewable energy integrated systems is to accurately forecast net load using data-driven approaches.

A direct STNLF methodology was proposed in this work, that is applicable to renewable-powered microgrids. The constructed BNN model provides improved forecasts, and it was evaluated on buildings and microgrid level using net load historical data.

The optimal ML model was developed through an input feature selection and hyperparameter determination stage. The results from this analysis indicated 8 input features (HNL, M_{year}, D_{week}, DPT, RF, T_{amb}, T_{day}, and GHI) and 7 hidden nodes for the optimal BNN forecasting model. It was also shown that its accuracy was affected by the train set data split approach (i.e., random and sequential) and duration (i.e., 30%, 50%, and 70%). Training the model with a 70% random train subset achieved the lowest daily mean *nRMSE* value of 3.98%. For 70% sequential training the error was increased to 5.21%. Also, for lower random training duration subsets, the model yielded a *nRMSE* of 4.04% and of 4.08% for 30% and 50% portion of the entire dataset, respectively.

In addition, the evaluation of the BNN model's robustness (utilizing the k-fold CV technique on the UCY microgrid) showed low *nRMSE* variations (ranging from 3.70% to 6.24% with an average of 5.13%). Moreover, a low daily mean *nRMSE* value of 5.16% was obtained by the model (regardless of the daily average clearness index) when applied to unseen data of the entire microgrid, proving its reliability and robustness.

The forecasts of the BNN model were further compared against the ones acquired from a baseline persistence model. The NPM yielded a *nRMSE* of 4.75% (the error was increased by 0.77% when compared to the BNN) for the entire microgrid. *SS* values ranging from 17.77% to 18.61% were obtained for the microgrid and the buildings under investigation. To this end, the analysis showed that the optimal BNN model achieved improved forecasts when compared to the baseline model at low and high irradiance conditions.

Finally, the proposed direct STNLF methodology can be used to effectively manage and control existing solar-integrated microgrids with varying PV penetration shares. Future research will focus on validating the proposed direct STNLF method in higher capacity microgrids from different locations to verify its scalability, replicability, and location independence. Furthermore, benchmarking on hybrid microgrids that integrate various RES types is another important future activity to ensure full validation and transferability of the model.

REFERENCES

[1] V. S. Tabar and V. Abbasi, "Energy management in microgrid with considering high penetration of renewable resources and surplus power generation problem," *Energy*, vol. 189, Dec. 2019, Art. no. 116264, doi: 10.1016/j.energy.2019.116264.

[2] S. Sreekumar, K. C. Sharma, and R. Bhakar, "Grey system theory based net load forecasting for high renewable penetrated power systems," *Technol. Econ. Smart Grids Sustain. Energy*, vol. 5, no. 1, pp. 1–14, Dec. 2020, doi: 10.1007/S40866-020-00094-4.

[3] H. Shaker, H. Chitsaz, H. Zareipour, and D. Wood, "On comparison of two strategies in net demand forecasting using Wavelet Neural Network," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2014, pp. 1–6, doi: 10.1109/NAPS.2014.6965360.

[4] G. Gross and F. D. Galiana, "Short-term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, Dec. 1987, doi: 10.1109/PROC.1987.13927.

[5] J. Izzatillaev and Z. Yusupov, "Short-term load forecasting in grid-connected microgrid," in *Proc. 7th Int. Istanbul Smart Grids Cities Congr. Fair (ICSG)*, Apr. 2019, pp. 71–75, doi: 10.1109/SGCF.2019.8782424.

[6] S.-J. Baek and S.-G. Yoon, "Short-term load forecasting for campus building with small-scale loads by types using artificial neural network," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2019, doi: 10.1109/ISGT.2019.8791674.

[7] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.

[8] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, "A deep learning method for short-term residential load forecasting in smart grid," *IEEE Access*, vol. 8, pp. 55785–55797, 2020, doi: 10.1109/ACCESS.2020.2981817.

[9] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, pp. 88058–88071, 2019, doi: 10.1109/ACCESS.2019.2926137.

[10] Md. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," *IEEE Access*, vol. 7, pp. 125413–125423, 2019, doi: 10.1109/ACCESS.2019.2937222.

[11] G. Dudek, "Short-term load forecasting using neural networks with pattern similarity-based error weights," *Energies*, vol. 14, no. 11, p. 3224, May 2021, doi: 10.3390/EN14113224.

[12] International Energy Agency (IEA). (2022). *Solar PV*. [Online]. Available: https://www.iea.org/reports/solar-pv

[13] A. Bracale, G. Carpinelli, and P. De Falco, "A probabilistic competitive ensemble method for short-term photovoltaic power forecasting," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 551–560, Apr. 2017, doi: 10.1109/TSTE.2016.2610523.

[14] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam, and D. Mathur, "Machine learning based PV power generation forecasting in Alice Springs," *IEEE Access*, vol. 9, pp. 46117–46128, 2021, doi: 10.1109/ACCESS.2021.3066494.

[15] A. Mellit, A. Massi Pavan, and V. Lughi, "Short-term forecasting of power production in a large-scale photovoltaic plant," *Sol. Energy*, vol. 105, pp. 401–413, Jul. 2014, doi: 10.1016/J.SOLENER.2014.03.018.

[16] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, Jun. 2020, Art. no. 115023, doi: 10.1016/J.APENERGY.2020.115023.

[17] D. Korkmaz, "SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting," *Appl. Energy*, vol. 300, Oct. 2021, Art. no. 117410, doi: 10.1016/J.APENERGY.2021.117410.

[18] S. Theocharides, C. Spanias, I. Papageorgiou, G. Makrides, S. Stavrinos, V. Efthymiou, and G. E. Georghiou, "A hybrid methodology for distribution level photovoltaic power production forecasting verified at the distribution system of Cyprus," *IET Renew. Power Gener.*, vol. 16, no. 1, pp. 19–32, Jan. 2022, doi: 10.1049/RPG2.12296.

[19] J. Kleissl, *Solar Energy Forecasting and Resource Assessment*. Cambridge, MA, USA: Academic, 2013.

[20] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast," *IEEE Access*, vol. 8, pp. 172524–172533, 2020, doi: 10.1109/ACCESS.2020.3024901.

[21] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, and S. Du, "Photovoltaic power forecasting with a hybrid deep learning approach," *IEEE Access*, vol. 8, pp. 175871–175880, 2020, doi: 10.1109/ACCESS.2020.3025860.

[22] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019, doi: 10.1109/ACCESS.2019.2923006.

[23] S. E. Haupt, S. Dettling, J. K. Williams, J. Pearson, T. Jensen, T. Brummet, B. Kosovic, G. Wiener, T. McCandless, and C. Burghardt, "Blending distributed photovoltaic and demand load forecasts," *Sol. Energy*, vol. 157, pp. 542–551, Nov. 2017, doi: 10.1016/J.SOLENER.2017.08.049.

[24] J. Shi, N. Liu, Y. Huang, and L. Ma, "An edge computing-oriented net power forecasting for PV-assisted charging station: Model complexity and forecasting accuracy trade-off," *Appl. Energy*, vol. 310, Mar. 2022,

Art. no. 118456, doi: 10.1016/J.APENERGY.2021.118456.

[25] A. Faustine and L. Pereira, "FPSeq2Q: Fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2440–2451, May 2022, doi: 10.1109/TSG.2022.3148699.

[26] A. Stratigakos, A. Bachoumis, V. Vita, and E. Zafiropoulos, "Short-term net load forecasting with singular spectrum analysis and LSTM neural networks," *Energies*, vol. 14, no. 14, p. 4107, Jul. 2021, doi: 10.3390/EN14144107.

[27] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2020, doi: 10.1109/TPWRS.2019.2924294.

[28] P. Kobylinski, M. Wierzbowski, and K. Piotrowski, "High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources," *Int. J. Electr. Power Energy Syst.*, vol. 117, May 2020, Art. no. 105635, doi: 10.1016/J.IJEPES.2019.105635.

[29] B. Zhou, Y. Meng, W. Huang, H. Wang, L. Deng, S. Huang, and J. Wei, "Multi-energy net load forecasting for integrated local energy systems with heterogeneous prosumers," *Int. J. Electr. Power Energy Syst.*, vol. 126, Mar. 2021, Art. no. 106542, doi: 10.1016/J.IJEPES.2020.106542.

[30] F. Mei, Q. Wu, T. Shi, J. Lu, Y. Pan, and J. Zheng, "An ultrashort-term net load forecasting model based on phase space reconstruction and deep neural network," *Appl. Sci.*, vol. 9, no. 7, p. 1487, Apr. 2019, doi: 10.3390/APP9071487.

[31] Y. Chu, H. T. C. Pedro, A. Kaur, J. Kleissl, and C. F. M. Coimbra, "Net load forecasts for solar-integrated operational grid feeders," *Sol. Energy*, vol. 158, pp. 236–246, Dec. 2017, doi: 10.1016/j.solener.2017.09.052.

[32] M. Bagheri, K. Suieubek, O. Abedinia, M. S. Naderi, and M. S. Naderi, "Direct and indirect prediction of net demand in power systems based on syntactic forecast engine," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/I&CPS Europe)*, Jun. 2018, pp. 1–6, doi: 10.1109/EEEIC.2018.8493990.

[33] M. Beichter, K. Phipps, M. M. Frysztacki, R. Mikut, V. Hagenmeyer, and N. Ludwig, "Net load forecasting using different aggregation levels," *Energy Informat.*, vol. 5, no. S1, pp. 1–21, Sep. 2022, doi: 10.1186/S42162-022-00213-8.

[34] G. Aburiyana, H. Aly, and T. Little, "Net load forecasting model for a power system grid with wind and solar power penetration," in *Proc. Global Conf. Eng. Res.*, Jun. 2021, pp. 525–531.

[35] G. Tziolis, C. Spanias, M. Theodoride, S. Theocharides, J. Lopez-Lorente, A. Livera, G. Makrides, and G. E. Georghiou, "Short-term electric net load forecasting for solar-integrated distribution systems based on Bayesian neural networks and statistical post-processing," *Energy*, vol. 271, May 2023, Art. no. 127018, doi: 10.1016/J.ENERGY.2023.127018.

[36] G. Aburiyana, H. Aly, and T. Little, "Direct net load forecasting using adaptive neuro fuzzy inference system," in *Proc. IEEE Electr. Power Energy Conf. (EPEC)*, Oct. 2021, pp. 131–136, doi: 10.1109/EPEC52095.2021.9621457.

[37] C. Zheng, M. Eskandari, M. Li, and Z. Sun, "GA–reinforced deep neural network for net electric load forecasting in microgrids with renewable energy resources for scheduling battery energy storage systems," *Algorithms*, vol. 15, no. 10, p. 338, Sep. 2022, doi: 10.3390/A15100338.

[38] G. Tziolis, J. Lopez-Lorente, M.-I. Baka, A. Koumis, A. Livera, S. Theocharides, G. Makrides, and G. E. Georghiou, "Comparative analysis of machine learning models for short-term net load forecasting in renewable integrated microgrids," in *Proc. 2nd Int. Conf. Energy Transition Medit. Area (SyNERGY MED)*, Oct. 2022, pp. 1–5, doi: 10.1109/SYNERGYMED55767.2022.9941378.

[39] S. E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D. B. Smith, and M. Minakshi, "From load to net energy forecasting: Short-term residential forecasting for the blend of load and PV behind the meter," *IEEE Access*, vol. 8, pp. 224343–224353, 2020, doi: 10.1109/ACCESS.2020.3044307.

[40] G. Tziolis, A. Koumis, S. Theocharides, A. Livera, J. Lopez-Lorente, G. Makrides, and G. E. Georghiou, "Advanced short-term net load forecasting for renewable-based microgrids," in *Proc. IEEE Int. Smart Cities Conf. (ISC)*, Sep. 2022, pp. 1–6, doi: 10.1109/ISC255366.2022.9922157.

[41] D. Yang, W. Wang, and T. Hong, "A historical weather forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for energy forecasting," *Sol. Energy*, vol. 232, pp. 263–274, Jan. 2022, doi: 10.1016/J.SOLENER.2021.12.011.

[42] J. G. Powers et al., "The weather research and forecasting model: Overview, system efforts, and future directions," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 8, pp. 1717–1737, Aug. 2017, doi: 10.1175/BAMS-D-15-00308.1.

[43] A. Livera, M. Theristis, E. Koumpli, S. Theocharides, G. Makrides, J. Sutterlueti, J. S. Stein, and G. E. Georghiou, "Data processing and quality verification for improved photovoltaic performance and reliability analytics," *Prog. Photovolt., Res. Appl.*, vol. 29, no. 2, pp. 143–158, Feb. 2021, doi: 10.1002/PIP.3349.

[44] I. Jebli, F.-Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by Pearson correlation using machine learning," *Energy*, vol. 224, Jun. 2021, Art. no. 120109, doi: 10.1016/J.ENERGY.2021.120109.

[45] A. Al-Ani and M. Deriche, "Feature selection using a mutual information based measure," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2002, pp. 82–85, doi: 10.1109/ICPR.2002.1047405.

[46] R. Yacef, M. Benghanem, and A. Mellit, "Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study," *Renew. Energy*, vol. 48, pp. 146–154, Dec. 2012, doi: 10.1016/J.RENENE.2012.04.036.

[47] L. Xu, M. Hu, and C. Fan, "Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks," *J. Building Eng.*, vol. 46, Apr. 2022, Art. no. 103853, doi: 10.1016/J.JOBE.2021.103853.

[48] S. Karsoliya, "Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture," *Int. J. Eng. Trends Technol.*, vol. 3, no. 6, pp. 714–717, 2012.

[49] H. S. Hippert, D. W. Bunn, and R. C. Souza, "Large neural networks for electricity load forecasting: Are they overfitted?" *Int. J. Forecasting*, vol. 21, no. 3, pp. 425–434, Jul. 2005, doi: 10.1016/J.IJFORECAST.2004.12.004.

[50] S. Theocharides, M. Theristis, G. Makrides, M. Kynigos, C. Spanias, and G. E. Georghiou, "Comparative analysis of machine learning models for day-ahead photovoltaic power production forecasting," *Energies*, vol. 14, no. 4, p. 1081, Feb. 2021, doi: 10.3390/EN14041081.

[51] H. Hassani and M. R. Yeganegi, "Selecting optimal lag order in Ljung–Box test," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 123700, doi: 10.1016/J.PHYSA.2019.123700.

[52] A. Livera, M. Theristis, G. Makrides, S. Ransome, J. Sutterlueti, and G. E. Georghiou, "Optimal development of location and technology independent machine learning photovoltaic performance predictive models," in *Proc. IEEE 46th Photovoltaic Specialists Conf. (PVSC)*, Jun. 2019, pp. 1270–1275, doi: 10.1109/PVSC40753.2019.8980474.

[53] National Renewable Energy Laboratory (NREL). *Solar and Lunar Position Calculators*. Accessed: May 2, 2023. [Online]. Available: https://midcdmz.nrel.gov/solpos/

GEORGIOS TZIOLIS received the B.Eng. degree in electrical and electronic engineering and the M.Sc. degree in electrical power engineering from Northumbria University, Newcastle, U.K., in 2018 and 2019, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with a scholarship with the University of Cyprus (UCY).

He is a Special Scientist with the PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, UCY. Prior to this, he was an Electrical Engineer, from 2019 to 2021. His research interests include photovoltaic performance data analysis and application of machine learning techniques for net load forecasting.

ANDREAS LIVERA received the B.Sc. degree in electrical engineering from the University of Cyprus (UCY), in 2015, the M.Sc. degree in sustainable energy futures from Imperial College London, U.K., in 2016, and the Ph.D. degree in electrical engineering from UCY, in 2023.

He is currently a Research Scientist with the PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, UCY. His research interests include the photovoltaic performance modeling and monitoring, reliability, failure diagnosis, performance loss quantification, and optimization of field operation and maintenance activities.

JESUS MONTES-ROMERO received the degree in electronic engineering and the Ph.D. degree in renewable energies from the University of Jaén, in 2013 and 2020, respectively.

He is currently a Researcher with the AdPVTech Research Group and also collaborating with the PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, University of Cyprus (UCY). His research interests include test, analysis, and modeling of PV technology behavior, developing monitoring electronics for photovoltaic technology and building set-ups for photovoltaic related monitoring systems.

SPYROS THEOCHARIDES received the B.Sc. degree in computer engineering and the Ph.D. degree in electrical engineering from the University of Cyprus (UCY), in 2016 and 2022, respectively.

He is currently a Postdoctoral Researcher with the PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, UCY. His research interests include linear and dynamic algorithms, machine learning algorithms, data analysis, and PV power forecasting. His exceptional work in the undergraduate senior project, titled "PV Prediction of Power Production and Live Monitoring Tool for PV Systems in Cyprus'', earned him the second place in the EUREC awards for renewable energy.

GEORGE MAKRIDES (Member, IEEE) received the B.Eng. degree in electrical and electronic engineering from the Queen Mary University of London, London, U.K., in 2003, the M.Phil. degree in engineering from the University of Cambridge, Cambridge, U.K., in 2004, and the Ph.D. degree in electrical engineering from the University of Cyprus (UCY), Nicosia, Cyprus, in 2012.

He is currently the Head of the Renewable and Grid Integration Group, FOSS Research Centre for Sustainable Energy. He has published more than 100 papers in international journals and conference proceedings and has participated successfully in various local and European research funded projects. His research interests include outdoor performance of PV technologies, grid integration of variable renewable sources (particularly PV), smart grids, machine learning, and data-driven analytics.

GEORGE E. GEORGHIOU received the B.A., M.Eng., M.A. degrees (Hons.) and Ph.D. degree from the University of Cambridge, Cambridge, U.K., in 1996 and 1999, respectively.

He is currently a Professor and a Founding Member of the FOSS Research Centre for Sustainable Energy, University of Cyprus (UCY). Prior to this, he was a Lecturer and the Undergraduate Course Leader of electrical engineering with the University of Southampton, and a Research Fellow with the Electricity Utilization Group, University of Cambridge, from 1999 to 2002. He has published more than 400 papers in international journals and conference proceedings.

Prof. Georghiou's awards and honors include six outstanding paper awards for the most significant technical scientific contributions and two innovation prizes.

● ● ●