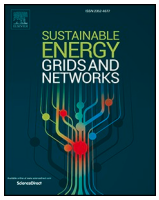
Sustainable Energy, Grids and Networks 37 (2024) 101256

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan



Direct short-term net load forecasting in renewable integrated microgrids using machine learning: A comparative assessment

Georgios Tziolis [a,*], Javier Lopez-Lorente [b], Maria-Iro Baka [c], Anastasios Koumis [a], Andreas Livera [a], Spyros Theocharides [a], George Makrides [a], George E. Georghiou [a]

[a] PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus
[b] DNV, Arnhem 6812 AR, the Netherlands
[c] QUE Technologies, 62 Papandreou Andrea Street, Chalandri GR 15232, Greece

ARTICLE INFO

Keywords:
Machine learning
Microgrid
Net load forecasting
Photovoltaic
Renewable energy sources

ABSTRACT

Modern microgrids require accurate net load forecasting (NLF) for optimal operation and management at high shares of renewable energy sources. Machine learning (ML) principles can be used to develop precise and reliable NLF models. This paper evaluates the performance of different ML models, that are optimally trained using supervised learning regimes, for direct short-term net load forecasting (STNLF) in renewable microgrids. Different categories of ML models, such as neural network, ensemble, linear regression, nearest neighbor, and support vector machine were used. The comparative assessment was conducted utilizing historical net load, meteorological, and time-related categorical data acquired from the renewable integrated microgrid of the University of Cyprus in Nicosia, Cyprus. The results showed that all STNLF ML models achieved normalized root mean square error (*nRMSE*) values below 10%. Amongst the investigated models, the Bayesian neural network (BNN) presented the highest forecasting accuracy, exhibiting a daily average error of 3.58%. In addition, the BNN model yielded robust forecasts regardless of the season and weather conditions. Finally, the results demonstrated that optimally constructed ML models can be applied to provide STNLF in renewable integrated microgrids, which can be used by microgrid operators to efficiently control and manage their assets.

Nomenclature

ANFIS	Adaptive neuro-fuzzy inference system.
ANN	Artificial neural network.
ARIMAX	Autoregressive integrated moving average with exogenous variables.
ARMA	Autoregressive moving average.
ARX	Autoregressive with exogenous variables.
BNN	Bayesian neural network.
BTM	Behind-the-meter.
DPT	Dew point temperature.
DT	Decision tree.
D_{week}	Day of the week.
GBM	Gradient boosting machine.
GHI	Global horizontal irradiance.
HI	Heat index.
HNL_{day}	Daily time-lagged historical net load.
HNL_{week}	Weekly time-lagged historical net load.
IEA	International Energy Agency.
KNN	*k*-nearest neighbor.
LR	Linear regression.
LSTM	Long short-term memory.
MAE	Mean absolute error.
MAPE	Mean absolute percentage error.
MISO	Multi-input single-output.
ML	Machine learning.
MLP	Multi-layer perceptron.
MLR	Multiple linear regression.
M_{year}	Month of the year.
n	Total number of observations.
NLF	Net load forecasting.
NPM	Naïve persistence model.
NPM_{day}	Naïve persistence model using daily time-lagged data.
NPM_{week}	Naïve persistence model using weekly time-lagged data.

* Corresponding author.
E-mail address: tziolis.georgios@ucy.ac.cy (G. Tziolis).

https://doi.org/10.1016/j.segan.2023.101256
Received 12 June 2023; Received in revised form 24 October 2023; Accepted 10 December 2023
Available online 14 December 2023

nRMSE	Normalized root mean square error.
OLS	Ordinary least squares.
P_{max}	Maximum measured net load power.
PV	Photovoltaic.
P_v	Water vapor pressure.
$P_{v,sat}$	Water vapor saturation pressure.
RES	Renewable energy sources.
RF	Random forest.
RH	Relative humidity.
RMSE	Root mean square error.
RNN	Recurrent neural network.
SD	Standard deviation.
STLF	Short-term load forecasting.
STNLF	Short-term net load forecasting.
SVM	Support vector machine.
SVR	Support vector regression.
SVRX	Support vector regression with exogenous inputs.
T_{amb}	Ambient temperature.
T_{day}	Time of the day.
UCY	University of Cyprus.
WS	Wind speed.
XGBoost	Extreme gradient boosting.
y_{actual}	Actual net load.
$y_{forecasted}$	Forecasted net load.

1. Introduction

The International Energy Agency (IEA) reported that the annual renewable capacity additions will reach 460 GW in 2027, with solar photovoltaic (PV) and wind constituting the highest shares [1]. Current research efforts aim to develop accurate net load forecasting (NLF) models that effectively mitigate the variability and uncertainty issues arising due to the increasing penetration of renewable energy sources (RES) in modern power grids [2].

Net load is defined as the difference between consumption and renewable energy generation [3]. NLF can be achieved directly (i.e., a single forecast of the net load) or indirectly (i.e., by calculating the difference between the load and the RES generation forecasts) [4]. Direct NLF has become increasingly important lately due to its computational advantage over indirect NLF and the availability of net load data [5]. Moreover, high shares of PV systems are installed behind-the-meter (BTM) and their contribution is not observable to grid operators [6], [7]. In this domain, the use of data-driven techniques (i.e., machine learning algorithms) that do not require additional knowledge of system characteristics becomes imperative for the construction of high-performing NLF models.

Load forecasting is essential to energy suppliers that strive to maintain balance in electricity networks [8–17]. Specifically, the load forecasting techniques presented in [9], [10], were applied to residential households, while the forecasting models of [11], [12] focused on commercial-scale buildings. In addition, [13], [14] performed load forecasting in microgrids, whereas distribution system level forecasting was presented in [15], [16]. In [17] load forecasting for low-voltage electricity networks was applied. Prior research studies compared the performance of various machine learning (ML) models for short-term load forecasting (STLF) applications [18–25]. Specifically, a previous study evaluated the performance of different ML models based on artificial neural networks (ANNs), *k*-nearest neighbors (KNNs), random forest (RF), recurrent neural networks (RNNs), and support vector regression (SVR) [18]. The analysis was performed on several datasets at different load aggregation levels (e.g., low, medium, and high). The findings demonstrated that the RF was the best performing STLF model for all datasets. Another comparative STLF study presented the performance of adaptive neuro-fuzzy inference system (ANFIS), ANN, multiple linear regression (MLR), and SVR models [19]. The results proved that the ANN was the most reliable and accurate model, with a prediction error of 1.67% when applied to electricity load data. In [20], decision tree (DT), KNN, linear regression (LR), long short-term memory (LSTM), RF, and SVR models were benchmarked for STLF. The results showed that the RF and SVR algorithms exhibited the highest forecasting accuracies amongst the investigated models. The authors in [21] evaluated the performance of extreme gradient boosting (XGBoost), LSTM, and RF models for STLF in smart buildings. In terms of accuracy and execution time efficiency, the XGBoost algorithm outperformed all other models. The XGBoost yielded a mean absolute percentage error (*MAPE*) of 2.01%. A comparison between STLF methods for several types of buildings (residential, schools, and shopping centers) was carried out in [22]. The investigated techniques comprised of KNN, LSTM, multi-layer perceptron (MLP), MLR, RF, and SVR. The KNN model yielded the lowest mean normalized root mean square error (*nRMSE*) of 1.01% for the residential dataset. Conversely, the MLP model achieved the lowest error for the school and shopping center loads, with *nRMSE* of 0.28% and 0.18%, respectively. Another STLF study presented the MLR application on power system data at two distinct seasons. The constructed MLR model provided *MAPE* values of 3.52% and 4.34% for the dry and rainy seasons, respectively [23]. In [24], MLR was proposed for STLF, yielding a *MAPE* of 3.99%. The authors in [25] proposed a STLF method using exponential smoothing and gradient boosting machine (GBM). The proposed method achieved *MAPE* values in the range of 2.08% to 2.62% for the two test sets.

Even though prior studies focused on STLF, the net effects of increased RES penetration to the performance of NLF algorithms remain a field of main concern, especially for microgrids. Along this context, the implementation and actual-life demonstration of novel short-term net load forecasting (STNLF) methodologies for the construction of accurately performing models have attracted considerable attention in microgrids with RES penetration. Several studies employed statistical and ML principles for STNLF [26–35]. An ANN-based model was proposed for STNLF in micro-neighborhoods with high RES penetration [28]. The ANN model outperformed an autoregressive integrated moving average with exogenous variables (ARIMAX) model, yielding an average mean absolute error (*MAE*) equivalent to 5.4% of the maximum measured net load. Direct and indirect STNLF methodologies were compared for a distribution substation utilizing MLP and autoregressive with exogenous variables (ARX) models [29]. The MLP model outperformed the ARX model, achieving *MAPE* of 9.33% and 8.94% for the direct and indirect method, respectively. In [30], direct and indirect NLF methods based on the LSTM model were compared for systems with disaggregated BTM PV generation. The comparison proved that the indirect NLF achieved a root mean square error (*RMSE*) reduction of 9.98% compared to the direct NLF. Multi-input single-output (MISO) LSTM, batch LSTM, online LSTM, autoregressive moving average (ARMA), and persistence models were compared for indirect residential STNLF [31]. The results demonstrated the superiority of the online LSTM model at the household level (7.3% improvement) and the MISO LSTM at the aggregate level (13.2% improvement). The authors in [32], conducted a comparison between direct and indirect NLF for a microgrid with high renewable energy penetration. The results provided evidence that the direct strategy outperformed the indirect approach. Specifically, the ARX model achieved a reduced *MAPE* (4.60%), while the support vector regression with exogenous inputs (SVRX) model achieved lower *MAE* (54.74 kW) and *RMSE* (82.59 kW). A direct STNLF methodology based on the Bayesian neural network (BNN) model for renewable-based microgrids was proposed in [33], [34]. The devised model outperformed the naïve persistence model (NPM), yielding *nRMSE* values between 3.98% and 5.35% when applied to individual buildings (with and without PV shares) and to the solar-integrated microgrid. A direct STNLF model based on BNN and statistical post-processing was proposed in [35]. The results showed that the BNN model with post-processing outperformed the simplistic NPM achieving *nRMSE* values between 1.02% and 1.29% for all three solar-integrated distribution feeders.

Table 1
Taxonomy of ML models used for STLF and STNLF in the literature.

Category	Model	Forecasting	
		Load	Net load
Neural network	ANN	[18],[19]	[28]
	BNN	-	[33]–[35]
	RNN (simple and LSTM)	[18],[20]–[22]	[30], [31]
	MLP	[22]	[29]
Ensemble	RF	[18],[20]–[22]	-
	XGBoost	[21]	-
	GBM	[25]	-
Linear regression	LR	[20]	-
	MLR	[19],[22]–[24]	-
Neural network and fuzzy logic	ANFIS	[19]	-
Decision tree	DT	[20]	-
Nearest neighbor	KNN	[18],[20], [22]	-
Support vector machine	SVR	[18]–[20], [22]	[32]

Table 1 summarizes the categorical taxonomy of previous STLF and STNLF studies based on ML models.

Despite the fact that numerous ML models utilized and compared different STLF models (e.g., [18]–[22]), a thorough comparison of optimized methods to develop direct STNLF models leveraging ML principles applicable to utility-scale renewable microgrids remains an unexplored area. Moreover, the pressuring needs for highly accurate and robust STNLF models is another area of increasing research interest for future power systems. The purpose of this work is to bridge this gap while meeting the needs of the electric power industry by presenting and analyzing various ML models applied to a renewable integrated microgrid capable of facilitating its management and operation. This paper expands on a previous work [36], where six ML algorithms (ANN, XGBoost, KNN, RF, RNN, and SVR) were benchmarked for direct STNLF in a renewable integrated microgrid. These ML models were chosen based on their high accuracies demonstrated in prior load forecasting literature [18]–[22]. In this paper, the analysis is extended by developing and tuning additional ML models, extensively investigating the daily performance for all STNLF models, and evaluating the daily net load profiles for the best performing model over different meteorological seasons and weather conditions. The ML investigated models included ANN, BNN, XGBoost, GBM, KNN, MLP, MLR, RF, RNN, and SVR. The BNN model was added due to its promising results when applied to STNLF methodologies applicable to renewable-based microgrids [33], [34] and solar-integrated distribution systems [35]. The MLP model was also selected since it provided highly accurate forecasts for different building types [22], and STNLF for a distribution substation [29]. In addition, the GBM and MLR models were employed due to the high STLF accuracies achieved in prior studies [23]–[25]. All ML models were developed utilizing historical net load, weather, and categorical data from the microgrid of the University of Cyprus (UCY) in Nicosia, Cyprus. The performance of each model was assessed using common performance metrics and against two NPMs, which served as baseline (reference) models.

Overall, the contributions of this work are the following:

- Introduction of new data-driven NLF concepts applied to utility-scale renewable microgrids, therefore presenting new knowledge on the usage of historical data and data-driven approaches for direct NLF.
- Implementation of a novel direct STNLF methodology for the development of accurate and robust direct forecasting ML models. The proposed methodology presents a unique pipeline of stages applicable to renewable-based microgrids at all scales.
- Comparison and actual-life verification of different models (ten ML and two NPMs) for direct STNLF, thereby presenting models with accuracies that exceed the state-of-the-art. The best performing model can achieve accurate and reliable direct STNLF that facilitates decision-making by microgrid operators.
- Determination of the best performing model for STNLF by looking at the hourly and daily performance of the day-ahead forecasting horizon. This presents important information to the research community on novel approaches utilized to develop optimized STNLF models.
- Robustness and adaptability verification of the best performing model in different seasons and weather conditions to provide insights about the impact of external weather factors on forecasting performance.

The rest of this paper is organized as follows: Section 2 describes the methodology and models used in this study, Section 3 discusses the results emanating from this work, while Section 4 highlights the conclusions of this work.

2. Methodology

The methodology used to develop and evaluate the performance of the optimally performing STNLF models, leveraging ML principles, consists of four steps (as shown in Fig. 1): (a) experimental setup and input feature acquisition, (b) data pre-processing, (c) training and tuning of ML models, and creation of baseline models, and (d) performance evaluation.

The detailed procedure for forecasting the net load output is shown in Fig. 2. In particular, the input features that were used for the training stage of the ML models comprised of categorical and numerical (computed and measured) parameters. The dataset comprised of data acquired at a granularity of a second and recorded at hourly intervals. The time of the day (T_{day}), day of the week (D_{week}), and month of the year (M_{year}), were the categorical data. The computed data were the heat index (*HI*), while the historically measured data were the global horizontal irradiance (*GHI*), relative humidity (*RH*), ambient temperature (T_{amb}), wind speed (*WS*), daily time-lagged historical net load (HNL_{day}), and weekly time-lagged historical net load (HNL_{week}). Data pre-processing was then performed to clean the raw data and construct a high-fidelity cleansed dataset over a three-year period. The cleansed dataset was divided into 70% train set and 30% test set. The ML forecasting models (i.e., ANN, BNN, XGBoost, GBM, KNN, MLP, MLR, RF, RNN, and SVR) were thus developed and tuned using the train set. The naïve persistence model using daily time-lagged data (NPM_{day}) and the naïve persistence model using weekly time-lagged data (NPM_{week}) employed the HNL_{day} and HNL_{week} input parameters, respectively. The STNLF models were then used to forecast the net load at hourly intervals. The performance of the models was evaluated using the test set. Specifically, the forecasted net load data (derived from the STNLF models) were compared with the actual net load data obtained from the test set (at hourly resolution). In addition, the forecasted and actual net load data were also aggregated into daily intervals. Moreover, the forecasted and actual net load was compared using the aggregated daily data in order to rank the STNLF models (by performing day-by-day evaluation). Finally, common performance metrics were employed for the performance evaluation.

2.1. Experimental setup

A three-year historical dataset of hourly average net load data from UCY microgrid was used to develop day-ahead direct STNLF models (developed using the Python programming language) by leveraging ML principles. The dataset contained net load data obtained from an advanced metering infrastructure consisting of a network of smart

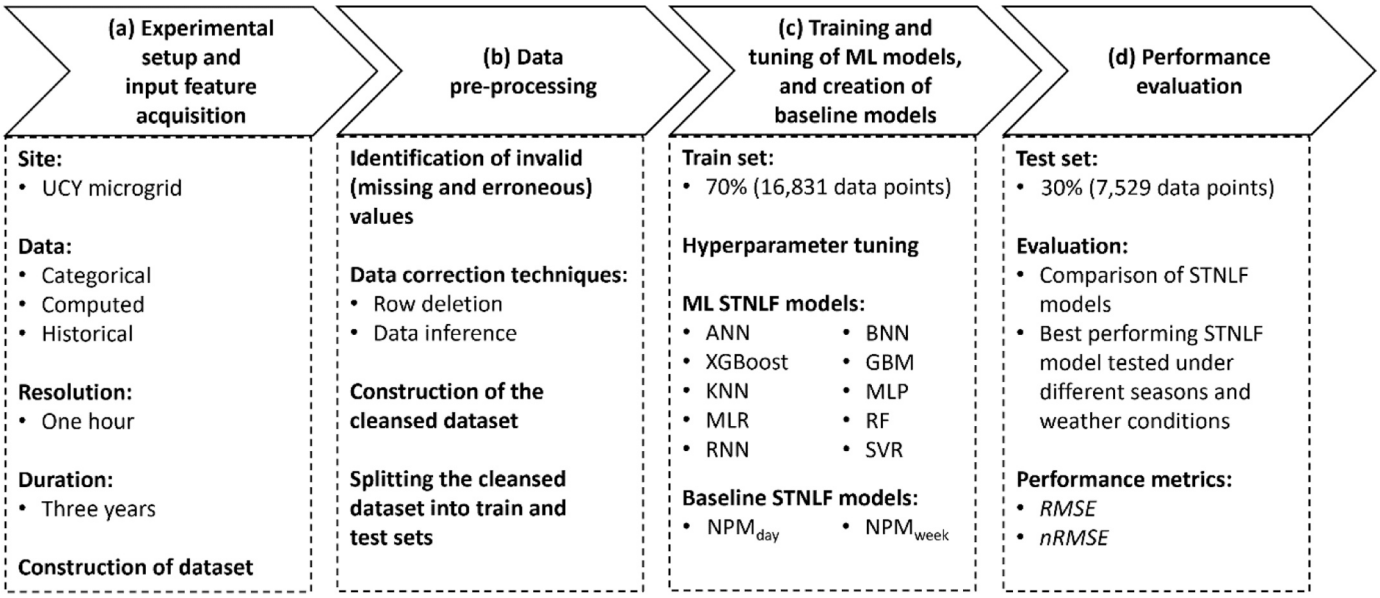


Fig. 1. Flowchart of the proposed methodology for developing and evaluating data-driven direct STNLF models.

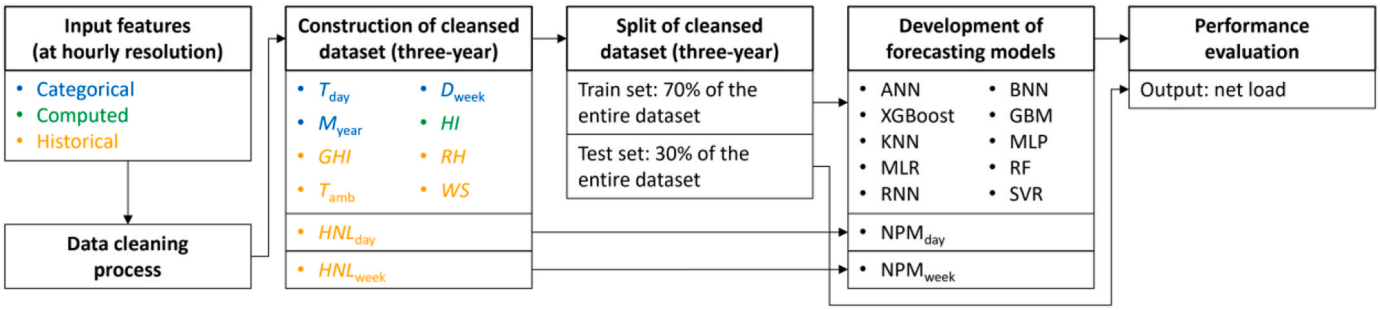


Fig. 2. Flowchart detailing the training and testing procedure of the STNLF models.

meters installed at key positions of the low-voltage side (400 V AC) of the microgrid. Fig. 3 provides a satellite imagery of the utility-scale UCY microgrid comprising of 17 buildings and distributed generation (PV systems).

The net load characteristics of UCY microgrid are depicted in Fig. 4. Fig. 4a shows the net load duration curve over the three-year period, while Fig. 4b illustrates the hourly mean net load histogram. The maximum measured net load power of the microgrid is 3.996 MW, while the total PV capacity is 434.80 kW_p (equivalent to 10.88% PV penetration).

Fig. 3. Satellite imagery of UCY microgrid (utility-scale microgrid).

2.2. Input feature acquisition and data pre-processing

Categorical (i.e., T_{day}, D_{week}, and M_{year}) and weather-related variables were employed as input features for the development of ML models. The weather parameters include *GHI*, *RH*, T_{amb}, *WS*, and *HI*, which captures the real feel. The weather-related variables, except for *HI*, were historical on-site measurements.

The *HI* was computed using Eq. (1) as a function of T_{amb} and dew point temperature (*DPT*). Due to lack of *DPT* data, Eqs. (2) - (4) were used to estimate this parameter. Using Eq. (2), the partial water vapor pressure (P_v) can be estimated as a function of *RH* (available from field measurements) and water vapor saturation pressure ($P_{v,sat}$) [37], [38].

$$HI = T_{amb} - 1.0799 \cdot e^{0.03755 T_{amb}} \left[1 - e^{0.0801 (DPT - 14)}\right] \quad (1)$$

$$DPT = \frac{243.5 \log(P_v / 6.112)}{17.67 - \log(P_v / 6.112)} \quad (2)$$

$$P_v = RH \cdot P_{v,sat} \quad (3)$$

$$P_{v,sat} = 6.112 \cdot e^{\frac{17.67 T_{amb}}{T_{amb} + 243.5}} \quad (4)$$

Furthermore, two additional features were used: HNL_{day} and HNL_{week}. Table 2 summarizes the input features employed by utilizing categorical data, historical measurements, and computed variables.

Data pre-processing was performed prior to model development through a series of filtering stages to identify missing data and remove

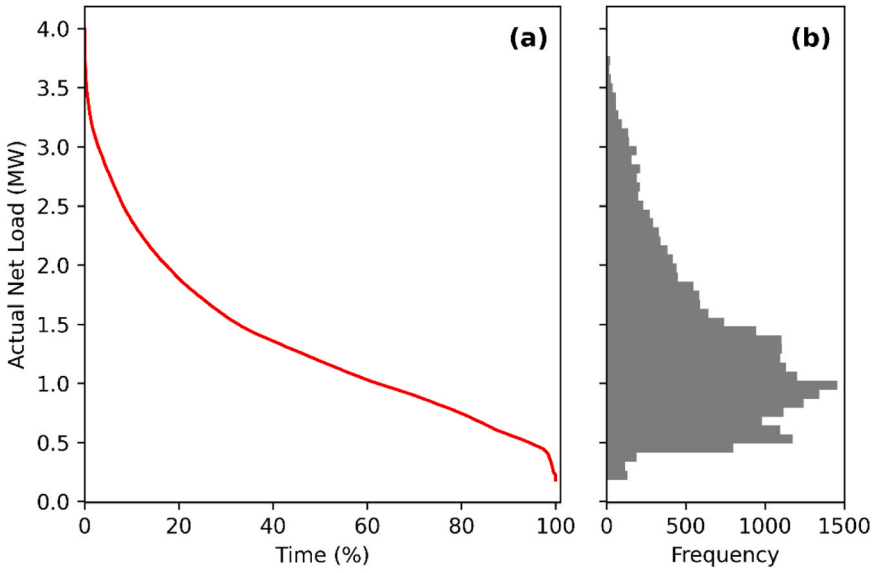


Fig. 4. Net load characteristics of the microgrid: (a) net load duration curve and (b) histogram of hourly mean net load.

Table 2
Input features used in the ML models.

Input feature	Type of data	Unit
T_{day}	Categorical	Numeric
D_{week}	Categorical	Numeric
M_{year}	Categorical	Numeric
GHI	Historical	W/m^2
RH	Historical	%
T_{amb}	Historical	°C
WS	Historical	m/s
HI	Computed from historical data using Eqs. (1)-(4)	°C
HNL_{day}	Historical	W
HNL_{week}	Historical	W

any erroneous entries to ensure high quality data. Specifically, row deletion was performed for less than 10% invalid data points of the entire dataset. Otherwise, data inference techniques were applied to reconstruct the dataset [39].

2.3. Short-term net load forecasting models

Ten supervised ML models were implemented for direct STNLF and demonstrated at the investigated renewable integrated microgrid. Furthermore, NPMs were utilized as baseline models to compare the behavior of more sophisticated ML models with that of simplistic ones. The developed models are briefly described in this section.

2.3.1. Artificial neural network

ANN is a powerful computational tool for dealing with complex problems with both linear and non-linear relationships. ANN-based models can have an adaptive behavior to noisy data [40]. The ANN architecture consists of a number of hidden layers and nodes (neurons). The ANN tuning was conducted by assessing various activation functions, learning rate coefficients, hidden layers, hidden nodes, and solvers.

2.3.2. Bayesian neural network

The BNN model combines neural network operational techniques with Bayesian inference. High accuracy forecasts, low computational complexity, short training time, reduced sensitivity to small datasets, and ease of managing hyperparameters are some of the main benefits of the BNN model [41]. Tuning of hidden layers and nodes was performed to develop an optimized BNN model.

2.3.3. Extreme gradient boosting

XGBoost is an ensemble algorithm that makes use of GBM. XGBoost can control overfitting better than GBM, resulting in improved performance [42]. The XGBoost model was tuned by evaluating different column samples, loss reductions, learning rates, depths, weights, estimators, threads, and subsample sizes.

2.3.4. Gradient boosting machine

The GBM combines iteratively several simple models (i.e., weak learners) to obtain a strong learner with improved forecasting accuracy [43]. Tuning of the learning rate, depth, samples split, and estimators was performed.

2.3.5. k-nearest neighbor

The KNN model assumes that similar instances are close to each other and therefore, predicts the numerical target based on a similarity measure. The KNN algorithm computes the distances between the test data and all the training points and then chooses the k points that are closest to the test data. When applied in regression problems, the KNN model returns the average value of the target [44]. The optimal hyperparameters for the KNN model were determined by tuning the number of neighbors k and the weighted similarity measure.

2.3.6. Multi-layer perceptron

MLP is a class of neural network and is formed from multiple perceptron layers. The nodes of MLP are arranged in input, hidden, and output layers [45]. The activation function, learning rates, hidden layers, hidden nodes, and optimization solvers were tuned to obtain an optimized MLP model.

2.3.7. Multiple linear regression

MLR is an extension of LR and is based on the relationship between a dependent variable and multiple independent variables [24]. MLR relies on a subset of the dataset to obtain the coefficients of the multi-variate fitting. In this case, the train dataset was used to estimate the coefficients of the MLR function. The function was fitted using ordinary least squares (OLS) regression, and it was assumed of first-degree polynomial order, where each input feature had an associated coefficient c_n, with an additional constant. The resulting function after the coefficients' fitting was:

$$c_1 M_{year} + c_2 D_{week} + c_3 T_{day} + c_4 HNL_{day} + c_5 HNL_{week} + c_6 RH + c_7 WS + c_8 GHI + c_9 T_{amb} + c_{10} HI + c_{11} \quad (5)$$

where the coefficients are $c_1 = -6.181$, $c_2 = -45.116$, $c_3 = -1.907$, $c_4 = 0.446$, $c_5 = 0.419$, $c_6 = -0.206$, $c_7 = -14.449$, $c_8 = -0.088$, $c_9 = 2.049$, $c_{10} = 8.101$, and $c_{11} = 191.413$. Since the forecasting horizon was day-ahead, rounded-up historical weather parameters for the next day were used as input (i.e., approximation considering quasi-perfect forecast) to compute the STNLF MLR model.

2.3.8. Random forest

RF is an ensemble method that creates multitude of DTs during the training stage. In regression problems, their output corresponds to the average value of the predictions [46]. Hyperparameter tuning was achieved by determining the optimal number of trees.

2.3.9. Recurrent neural network

RNNs are neural networks that allow prior outputs to be used as inputs, enabling the "memory" functionality [47]. RNNs are commonly employed to tackle problems involving time-series or other sequential input data. The activation function, the fraction of the units to drop for the linear transformation of the inputs, the number of epochs, the number of hidden layers and hidden nodes, and the dimensionality of the output space (i.e., units) were all assessed to determine the optimal RNN hyperparameters.

2.3.10. Support vector regression

SVR models are modified support vector machine (SVM) models that were originally employed for two-class classification problems. The SVR model is utilized for regression problems with non-linear relationships. This model creates a hyperplane (a straight line required to fit the data) and ensures maximum distance between the nearest sample and the hyperplane [48]. The tuning phase comprised of evaluating regularization, error sensitivity, curvature weight, and functions.

2.3.11. Naïve persistence model

The NPM forecasts the next time-step values based on prior observations [49]. Two NPMs were used in this work: the NPM_{day} and the NPM_{week}, which use time-lagged historical net load data of the same timestamp of the immediately preceding day and week, respectively.

2.4. Training and tuning of machine learning models, and testing of the short-term net load forecasting models

The ML models were developed using a 70:30% train and test dataset split approach. The given three-year dataset was divided into a two-year train and a one-year test subsets. Specifically, the first two years of the entire dataset (consisting of 16831 hourly data points) were used for the training process. The last year of the dataset (consisting of 7529 hourly data points) was used for testing the models' accuracy. The same test subset was employed to evaluate the performance of the NPMs.

The train set was also used for tuning the models' hyperparameters. To this end, the hyperparameters of the ANN, GBM, KNN, MLP, RF, RNN, and SVR models were tuned using the grid search. The number of hidden nodes in ANN, BNN, and RNN models was determined using a rule of thumb method. In particular, the number of hidden nodes should be equal to 2/3 of the size of the input layer plus the size of the output layer [50]. On the contrary, XGBoost tuning was performed using random search. Fig. 5 summarizes the optimal hyperparameters for all ML models.

All ML-based models had a low to moderate computational power requirements, except the XGBoost model which had a high computational burden. The computational power requirements depended on the number of hyperparameter combinations evaluated. A higher number of combinations led to a longer computational time for training the models. In the case of XGBoost, there were over 150000 possible combinations of hyperparameters. Thus, the hyperparameters of the XGBoost were optimized using random search (instead of the grid search) to achieve lower computational time. Finally, the NPMs that were implemented based on the assumption that nothing changes between current and forecasting time, exhibited low computational usage time because no training was needed.

The NPMs and the tuned ML models were then used to forecast the net load and to compare their performance on the test set.

2.5. Performance evaluation

The common performance metrics of *RMSE* and *nRMSE* calculated using Eqs. (6) and (7), respectively, were employed to evaluate the forecasting performance of the STNLF models using the test set.

Model	Optimal hyperparameters	
ANN	• Activation function: rectified linear • Learning rate: constant • Number of hidden nodes: 8	• Alpha (learning rate): 10 • Number of hidden layers: 1 • Optimization solver: adaptive moment estimation
BNN	• Number of hidden layers: 1	• Number of hidden nodes: 8
XGBoost	• Column sample by tree: 0.5 • Learning rate: 0.03 • Minimum child weight: 7 • Number of threads: 4	• Gamma (minimum loss reduction): 0.1 • Maximum depth: 3 • Number of estimators: 500 • Subsample: 0.7
GBM	• Learning rate: 0.3 • Minimum samples split: 2	• Maximum depth: 7 • Number of estimators: 750
KNN	• Number of neighbors: 10	• Weights: distance
MLP	• Activation function: rectified linear • Learning rate: constant • Number of hidden nodes (per layer): 11	• Alpha (learning rate): 1 • Number of hidden layers: 3 • Optimization solver: adaptive moment estimation
RF	• Number of trees: 200	
RNN	• Activation function: hyperbolic tangent • Number of epochs: 100 • Number of hidden nodes: 8	• Dropout rate: 0.15 • Number of hidden layers: 1 • Number of units: 50
SVR	• C (regularization): 100 • Gamma (curvature weight): 0.001	• Epsilon (error sensitivity): 0.3 • Kernel (set of mathematical functions): radial basis function

Fig. 5. Optimal hyperparameters for the ML models.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(y_{\text{actual,i}} - y_{\text{forecasted,i}}\right)^2} \quad (6)$$

$$nRMSE = \frac{100}{P_{\max}} \cdot \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(y_{\text{actual,i}} - y_{\text{forecasted,i}}\right)^2} \quad (7)$$

where $y_{\text{actual,i}}$ and $y_{\text{forecasted,i}}$ denote the actual and forecasted net load, respectively, n is the total number of observations, and $P_{\max}$ is the maximum measured net load power of the microgrid.

3. Results

3.1. Benchmarking of short-term net load forecasting models based on hourly averaged data

The scatterplots of the forecasted against the actual net load for each STNLF model are depicted in Fig. 6. The results demonstrated that the BNN, MLR, and RF models provided high accuracies since the forecasts were relatively aligned to the reference line ($y = x$) of the actual net load. In addition, the majority of the ML models exhibited forecasts with low data dispersion within the entire net load range. Both NPMs presented high forecasting variations (high data dispersion) suggesting that the independent variable (i.e., actual net load) does not by itself entirely describe and capture the variance in the dependent variable (i.e., forecasted net load).

The forecasting accuracy results given by the performance evaluation metrics of the STNLF models are presented in Table 3. The BNN model presented the highest forecasting accuracy, with a *nRMSE* of 4.26% (equivalent to 170.28 kW given by the *RMSE*) over the test set period, due to the ability of BNN to capture non-linear and complex

Table 3
Performance evaluation metrics obtained from the developed STNLF models applied to UCY microgrid over the test set period.

Ranking	Model	*RMSE* (kW)	*nRMSE* (%)
1	BNN	170.28	4.26
2	RF	172.56	4.32
3	MLR	202.74	5.07
4	RNN	260.88	6.53
5	NPM_{week}	281.26	7.04
6	NPM_{day}	297.04	7.43
7	XGBoost	299.51	7.49
8	KNN	311.71	7.80
9	GBM	316.62	7.92
10	MLP	325.63	8.15
11	SVR	326.50	8.17
12	ANN	397.65	9.95

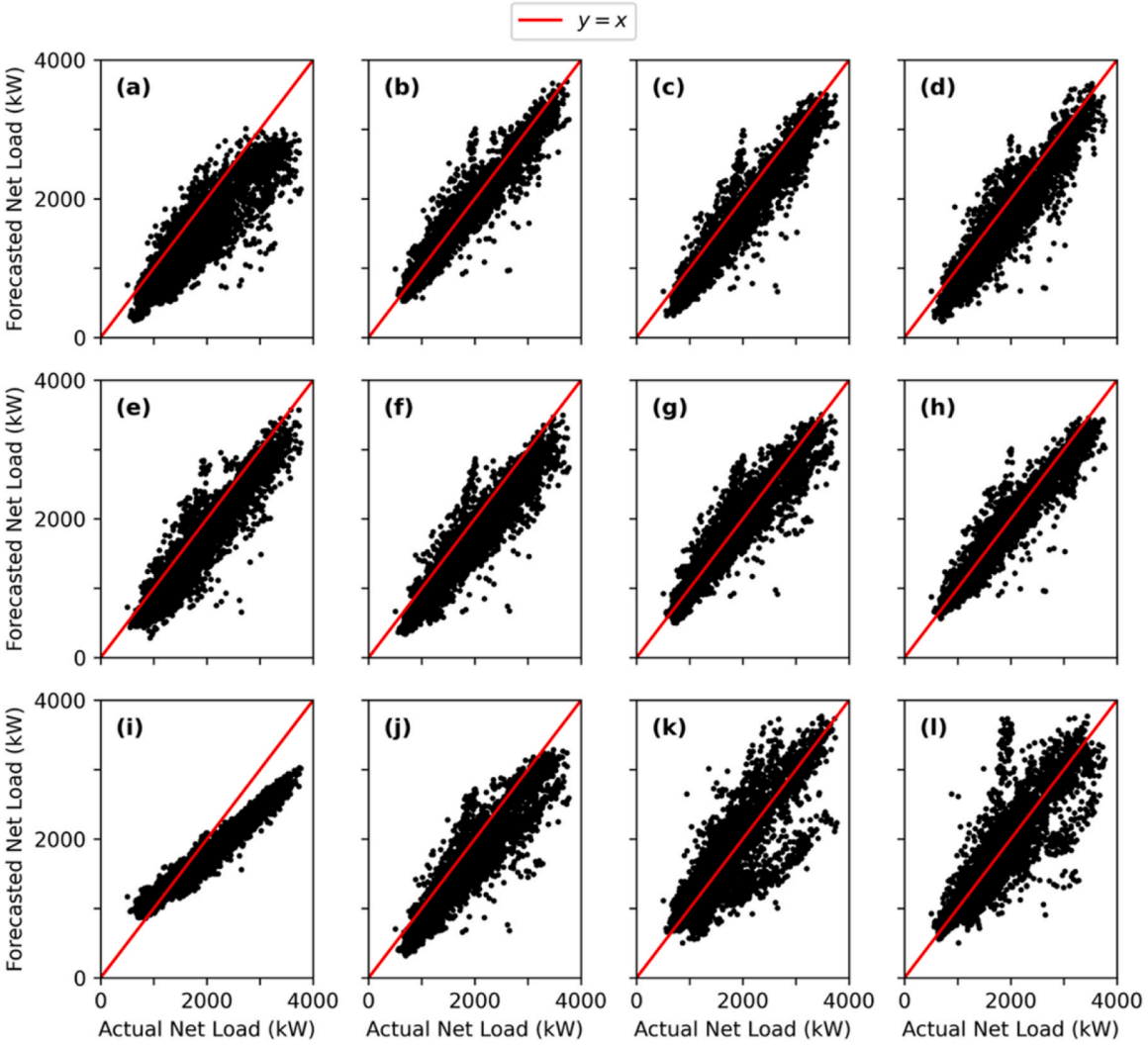


Fig. 6. Scatterplots of the forecasted against the actual net load for the (a) ANN, (b) BNN, (c) XGBoost, (d) GBM, (e) KNN, (f) MLP, (g) MLR, (h) RF, (i) RNN, (j) SVR, (k) NPM_{day}, and (l) NPM_{week} applied to UCY microgrid over the test set period.

behaviors of the varying PV generation and load demand. The BNN model slightly outperformed the second-best RF model, which yielded a *nRMSE* of 4.32%. Furthermore, the MLR and RNN models achieved *nRMSE* values below 7%. Moreover, compared to the NPM_{week} and NPM_{day}, the BNN model reported *nRMSE* improvements of 2.78% and 3.17%, and *RMSE* improvements of 110.98 kW and 126.76 kW, respectively. The BNN model outperformed the simplistic NPMs due to the fact that the net load profiles are not repetitive for the same time-stamp of the previous day or week. Finally, the improvements obtained by the BNN model given by *nRMSE* compared to the XGBoost, KNN, GBM, MLP, SVR, and ANN models ranged from 3.23% (i.e., *RMSE* of 129.23 kW) to 5.69% (i.e., *RMSE* of 227.37 kW).

3.2. Benchmarking of short-term net load forecasting models based on daily averaged data

The daily average error of the developed STNLF models over the test set period is presented in Fig. 7. The observed gaps in the plots are attributed to missing data (due to outages). The lowest daily mean *nRMSE* (3.58%) among the STNLF models was achieved by the BNN model, proving its superiority amongst the investigated models. The RF and MLR models also showed high accuracies, achieving daily mean *nRMSE* of 3.63% and 4.20%, respectively. The remaining STNLF models yielded daily mean *nRMSE* values greater than 5% (between 5.26% and 9.17%). The increased errors (>5%) obtained by the majority of models are attributed to their inability to fully capture the behavior of the net load profiles for all seasons and weather conditions.

A summary of the daily average performance given by the mean and standard deviation (*SD*) is presented in Table 4. The BNN model yielded the lowest daily mean *nRMSE* (3.58%) and *SD* of 2.28%, over a period of a year. The GBM model achieved the lowest daily data dispersion relative to its mean (i.e., a daily *SD* of 1.89%) among the STNLF models. The XGBoost and KNN models provided lower daily *SD* values (1.97% and 2.25%, respectively) compared to the BNN. However, the daily mean *nRMSE* of the three models (GBM, XGBoost, and KNN) ranged between 7.22% and 7.68%, indicating lower reliability of their mean values compared to the BNN model. The low daily mean *nRMSE* and *SD* values obtained from the BNN model demonstrated its ability to provide highly accurate forecasts with the data clustered closely around its mean.

3.3. Performance evaluation of the best performing short-term net load forecasting model in different seasons and weather conditions

The best performing ML model (i.e., the BNN) was further evaluated to assess its reliability during the different seasons (i.e., winter, spring, summer, and autumn). Table 5 outlines the seasonal mean *nRMSE*

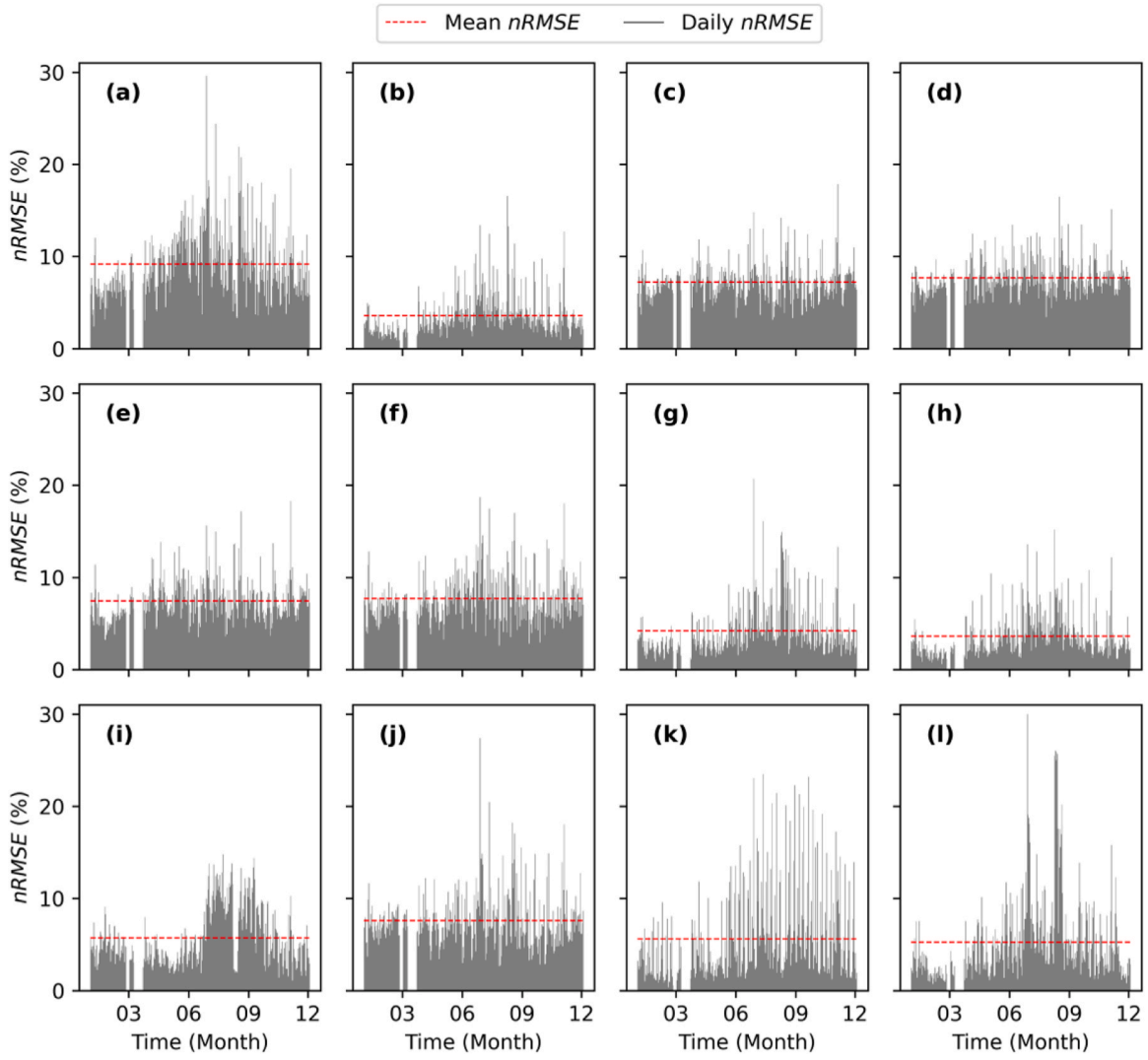


Fig. 7. Daily *nRMSE* for the (a) ANN, (b) BNN, (c) XGBoost, (d) GBM, (e) KNN, (f) MLP, (g) MLR, (h) RF, (i) RNN, (j) SVR, (k) NPM_{day}, and (l) NPM_{week} applied to UCY microgrid over the test set period.

Table 4
Daily mean *nRMSE* and *SD* obtained from the developed STNLF models applied to UCY microgrid over the test set period.

Plot in Fig. 7	Model	Daily *nRMSE*		
		Mean		*SD* (%)
		Ranking	Value (%)	
(a)	ANN	12	9.17	3.77
(b)	BNN	1	3.58	2.28
(c)	XGBoost	7	7.22	1.97
(d)	GBM	10	7.68	1.89
(e)	KNN	8	7.46	2.25
(f)	MLP	11	7.71	2.58
(g)	MLR	3	4.20	2.80
(h)	RF	2	3.63	2.30
(i)	RNN	6	5.72	3.09
(j)	SVR	9	7.60	2.97
(k)	NPM_{day}	5	5.61	4.81
(l)	NPM_{week}	4	5.26	4.63

obtained from the BNN model applied to the UCY microgrid. The lowest seasonal mean *nRMSE* (2.40%) was obtained during the winter season as it was the period with the lowest mean actual net load and measured *GHI*. During the spring and autumn seasons the obtained mean *nRMSE* values were 3.22% and 3.36%, respectively. Conversely, the highest mean *nRMSE* (5%) was obtained during the summer season due to the increased uncertainty arising from the high mean actual net load demand and higher irradiation values. Although there were variations in the seasonal mean *nRMSE* values, the results ranged from 2.40% to 5%, indicating robust forecasting behavior by the BNN model regardless of the season.

The obtained daily BNN forecasts at different weather conditions are depcited in Fig. 8. The almost identical profiles of the actual and forecasted net load proved the reliability of the BNN model for diverse meteorological conditions throughout the year. Specifically, the obtained daily *nRMSE* values ranged from 0.93% to 1.32% for the different selected days.

Table 6 is provided to gain further insights about the BNN performance for days with different weather conditions. Specifically, among the two days with the lowest mean and maximum *GHI*, the lowest error (daily *nRMSE* of 0.93%) was achieved for the partly cloudy day due to the less cloudy weather that favors the BNN model. A slightly higher daily *nRMSE* of 1.32% was observed for the mostly sunny day compared to the other days due to the increased daily mean actual net load demand and unstable radiation profile during critical hours. For the sunny day, the BNN model yielded an improved daily *nRMSE* (1.20%) compared to the mostly sunny day due to the stable sunny weather.

The seasonal mean *nRMSE* and daily *nRMSE* values obtained from the BNN model, proved that the model is robust and adaptable to the seasonality of solar PV energy production, meteorology, and changes in the electric load consumption profiles of the microgrid throughout the year.

3.4. Discussion

In this study a three-year dataset was used to train, tune, and test the models. The utilization of a three-year dataset enables the collection of data under different microgrid operational scenarios, and the acquisition of measurements under varying weather conditions and seasons. This facilitates the capture of the intermittent and uncertain behaviors of RES generation and load demand that significantly influence the performance of the model. In addition, a supervised dataset split approach was followed in this study whereby the three-year dataset was divided into a 70:30% train and test set. This is a common split approach regime which allows sufficient information for training the ML models, while also maintaining adequate data for testing the performance of STNLF models. The two-year duration of the train set enables the development of accurate and robust ML models (due to large amount of data used for the training process), while the yearly test set allows the performance validation of the models throughout the year (under different weather conditions and seasons). The obtained results were derived from performance evaluations and comparisons made using the third year of the dataset (i.e., the test set). This approach was adopted to ensure a rigorous evaluation of the models' performance on unseen data, which is crucial for validating the robustness and reliability of the models. The performance of the models in this paper verified the effectiveness of the methodology employed and aligned with expectations based on theoretical underpinnings and empirical evidence from related works [28], [29], [32]–[34]. Specifically, by employing various metrics and analyzing the results, the BNN outperformed all other models for the investigated microgrid. The comparison highlighted the effectiveness of the BNN model over other models in handling the complexities and uncertainties in solar-integrated microgrids.

Table 5
Seasonal mean *nRMSE* obtained from the BNN model applied to UCY microgrid.

Season	Seasonal irradiance (W/m^2)		Seasonal mean actual net load (kW)	Seasonal mean *nRMSE* (%)
	Mean *GHI*	Maximum *GHI*		
Winter	116	704	990	2.40
Spring	285	1063	1203	3.22
Summer	325	1080	1982	5.00
Autumn	196	943	1616	3.36

A simplified version of the best performing model has been successfully deployed to provide NLF for all the substations of the electricity grid in Cyprus. This real-world application demonstrates the model's potential for practical use and its ability to contribute towards the efficient management and operation of electricity grids. The effectiveness of the model at actual environment suggests that it can be adapted for similar practical applications in renewable microgrids. It is important to note that the application of the model to a real application requires modifications and adaptations based on the specific characteristics and requirements of the microgrid system under study. Along this context, the model needs to be refined to account for the increased variability and uncertainty associated with high-RES penetration, as well as to adapt to the specific operational and regulatory frameworks of the investigated microgrid.

Finally, it must be noted that the developed STNLF models were validated on a specific microgrid with a particular set of characteristics (e.g., moderate level of PV penetration). Therefore, the transferability and replicability of the models to other microgrid settings and climates remain an open question and a limiting factor. Another important limitation is the availability of high-quality data for training data-driven models. Specifically, low availability of data, inefficient quality of datasets, and low-performing hyperparameter optimization are some of the factors that negatively affect the performance of the ML models, leading to forecasts with reduced accuracies. Moreover, external factors such as sudden or extreme changes in weather conditions, changes in operational policies, or unexpected equipment failures, reduce the models' ability to adapt to such unforeseen circumstances, thus affecting the accuracy and reliability of the forecasts. Lastly, the transferability and scalability of the ML models can be limited due to compatibility issues between forecasting platforms and cyber-security regulatory requirements.

4. Conclusions

The rising penetration of RES in microgrids necessitates the transition from STLF to STNLF in order to enable their efficient operation and management, and to reduce uncertainty. Given that the accuracy of STNLF is significantly affected by the model used, this paper proposed a novel direct STNLF methodology that develops and compares ten ML

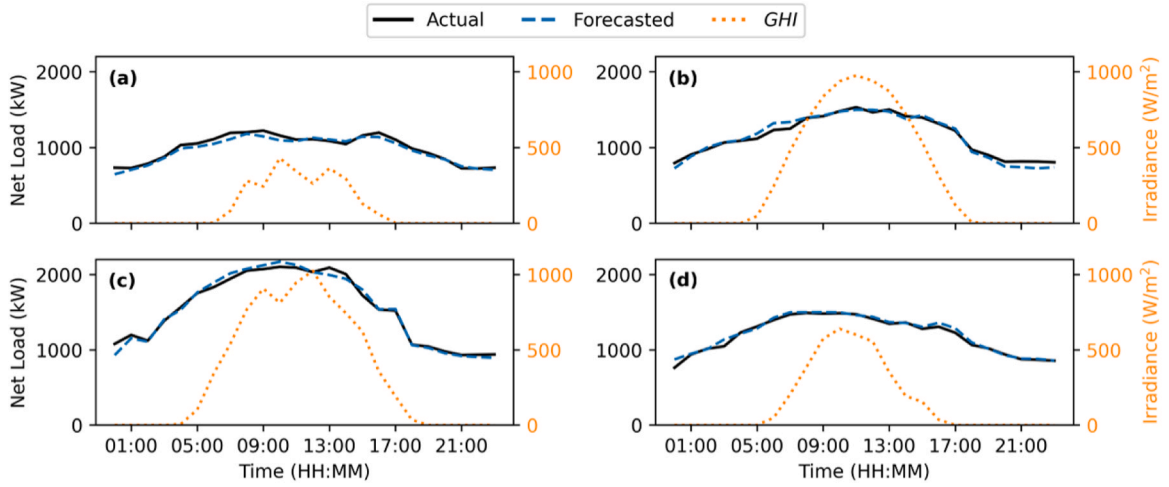


Fig. 8. Actual and forecasted daily net load profiles given by the BNN model for (a) mostly cloudy, (b) sunny, (c) mostly sunny, and (d) partly cloudy weather applied to UCY microgrid.

Table 6
Daily *nRMSE* obtained from the BNN model applied to UCY microgrid for specific days with different weather conditions.

Plot in Fig. 8	Month-Day	Weather	Daily irradiance (W/m^2)		Daily mean actual net load (kW)	Daily *nRMSE* (%)
			Mean *GHI*	Maximum *GHI*		
(a)	02-01	Mostly cloudy	103	430	993	1.08
(b)	04-28	Sunny	321	975	1152	1.20
(c)	06-16	Mostly sunny	343	1024	1541	1.32
(d)	10-30	Partly cloudy	155	639	1193	0.93

models (i.e., ANN, BNN, XGBoost, GBM, KNN, MLP, MLR, RF, RNN, and SVR) and two NPMs (i.e., NPM_{day} and NPM_{week}) applicable to a low-voltage microgrid with moderate PV penetration.

The supervised ML models were developed using historical, categorical, and computed data. Data pre-processing was conducted prior to model development to ensure high quality data. The ML models were then constructed using a 70:30% train and test set approach.

Based on the hourly and daily averaged performance results obtained for the investigated STNLF models, the BNN model outperformed all other models. Specifically, the results revealed that the BNN model yielded the lowest error with daily mean *nRMSE* of 3.58%. The RF and MLR models achieved slightly higher forecasting errors. In addition, daily mean *nRMSE* values $\geq$ 5.26% were obtained from the remaining models. Furthermore, the reliability of the optimally constructed BNN model was verified in different meteorological seasons and weather conditions. Overall, the findings of this study demonstrated that the BNN model achieved robust, adaptable, and highly accurate STNLF for a solar-integrated microgrid throughout the year.

The methodology proposed in this study provides information for the development of accurate STNLF models applicable to solar-integrated microgrids. To this end, an optimally BNN model was constructed for STNLF to facilitate the efficient management and control of a microgrid with RES penetration. While the model has shown a practical application, further modifications and adaptations are needed to enable its broader use in microgrids with high levels of RES penetration. This will be addressed in future studies along with the validation of models' performance in different microgrid characteristics and environments/climates.

CRediT authorship contribution statement

Georgios Tziolis: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Javier Lopez-Lorente:** Conceptualization, Formal analysis, Software, Validation, Visualization, Writing - review & editing. **Maria-Iro Baka:** Data curation, Writing - review & editing. **Anastasios Koumis:** Software, Writing - review & editing. **Andreas Livera:** Writing - review & editing. **Spyros Theocharides:** Writing - review & editing. **George Makrides:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Visualization, Writing - review & editing. **George E. Georghiou:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that has been used is confidential.

Acknowledgment

This work was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Cyprus Research and Innovation Foundation (RIF) in the framework of the project "ELECTRA" with protocol number: INTEGRATED/0918/0071.

References

[1] International Energy Agency (IEA), "Renewable electricity," 2022. ⟨https://www.iea.org/reports/renewables-2022/renewable-electricity⟩.
[2] G. Aburiyana, H. Aly, T. Little, Direct net load forecasting using adaptive neuro fuzzy inference system, IEEE Electr. Power Energy Conf. (2021) 131–136, https://doi.org/10.1109/EPEC52095.2021.9621457.
[3] M. Beichter, K. Phipps, M.M. Frysztacki, R. Mikut, V. Hagenmeyer, N. Ludwig, Net load forecasting using different aggregation levels, Energy Inform. 5 (1) (2022) 1–21, https://doi.org/10.1186/S42162-022-00213-8.
[4] H. Shaker, H. Chitsaz, H. Zareipour, D. Wood, On comparison of two strategies in net demand forecasting using Wavelet Neural Network, North Am. Power Symp. NAPS 2014 (2014), https://doi.org/10.1109/NAPS.2014.6965360.
[5] D.W. van der Meer, M. Shepero, A. Svensson, J. Widén, J. Munkhammar, Probabilistic forecasting of electricity consumption, photovoltaic power generation

and net demand of an individual building using Gaussian processes, Appl. Energy 213 (2018) 195–207, https://doi.org/10.1016/J.APENERGY.2017.12.104.
[6] Y. Wang, N. Zhang, Q. Chen, D.S. Kirschen, P. Li, Q. Xia, Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV, IEEE Trans. Power Syst. 33 (3) (2018) 3255–3264, https://doi.org/10.1109/TPWRS.2017.2762599.
[7] J. Lopez Lorente, X. Liu, D.J. Morrow, Spatial aggregation of small-scale photovoltaic generation using Voronoi decomposition, IEEE Trans. Sustain. Energy 11 (4) (2020) 2677–2686, https://doi.org/10.1109/TSTE.2020.2970217.
[8] A.Al Mamun, M. Sohel, N. Mohammad, M.S. Haque Sunny, D.R. Dipta, E. Hossain, A comprehensive review of the load forecasting techniques using single and hybrid predictive models, IEEE Access 8 (2020) 134911–134939, https://doi.org/10.1109/ACCESS.2020.3010702.
[9] A. Langevin, M. Cheriet, G. Gagnon, Efficient deep generative model for short-term household load forecasting using non-intrusive load monitoring, Sustain. Energy Grids Netw. 34 (2023), 101006, https://doi.org/10.1016/J.SEGAN.2023.101006.
[10] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, IEEE Trans. Smart Grid 10 (1) (2019) 841–851, https://doi.org/10.1109/TSG.2017.2753802.
[11] V.H. Nguyen, Y. Besanger, Q.T. Tran, Self-updating machine learning system for building load forecasting - method, implementation and case-study on COVID-19 impact, Sustain. Energy Grids Netw. 32 (2022), 100873, https://doi.org/10.1016/J.SEGAN.2022.100873.
[12] L. Xu, M. Hu, C. Fan, Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks, J. Build. Eng. 46 (2022), 103853, https://doi.org/10.1016/J.JOBE.2021.103853.
[13] L. Hernández, C. Baladrón, J.M. Aguiar, B. Carro, A. Sánchez-Esguevillas, J. Lloret, Artificial neural networks for short-term load forecasting in microgrids environment, Energy 75 (2014) 252–264, https://doi.org/10.1016/J.ENERGY.2014.07.065.
[14] A. Moradzadeh, S. Zakeri, M. Shoaran, B. Mohammadi-Ivatloo, F. Mohammadi, Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms, Sustain 12 (17) (2020) 7076, https://doi.org/10.3390/SU12177076.
[15] J. Forcan, M. Forcan, Optimal placement of remote-controlled switches in distribution networks considering load forecasting, Sustain. Energy Grids Netw. 30 (2022), 100600, https://doi.org/10.1016/J.SEGAN.2021.100600.
[16] S. Sepasi, E. Reihani, A.M. Howlader, L.R. Roose, M.M. Matsuura, Very short term load forecasting of a distribution system with high PV penetration, Renew. Energy 106 (2017) 142–148, https://doi.org/10.1016/j.renene.2017.01.019.
[17] C. Gilbert, J. Browell, B. Stephen, Probabilistic load forecasting for the low voltage network: forecast fusion and daily peaks, Sustain. Energy Grids Netw. 34 (2023), 100998, https://doi.org/10.1016/J.SEGAN.2023.100998.
[18] L. Burg, G. Gürses-Tran, R. Madlener, A. Monti, Comparative analysis of load forecasting models for varying time horizons and load aggregation levels, Energies 14 (21) (2021) 1–16, https://doi.org/10.3390/en14217128.
[19] D. Solyali, A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus, Sustainability 12 (9) (2020) 3612, https://doi.org/10.3390/SU12093612.
[20] A. Parrado-Duque, S. Kelouwani, K. Agbossou, S. Hosseini, N. Henao, and F. Amara, "A comparative analysis of machine learning methods for short-term load forecasting systems," 2021 IEEE Int. Conf. Commun. Control. Comput. Technol. Smart Grids, SmartGridComm 2021, pp. 270–275, 2021, doi: 10.1109/SMARTGRIDCOMM51999.2021.9632002.
[21] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, M. Elaroussi, A comparative study of predictive approaches for load forecasting in smart buildings, Procedia Comput. Sci. 160 (2019) 173–180, https://doi.org/10.1016/J.PROCS.2019.09.458.
[22] A. Groß, A. Lenders, F. Schwenker, D.A. Braun, D. Fischer, Comparison of short-term electrical load forecasting methods for different building types, Energy Inform. 4 (3) (2021) 1–16, https://doi.org/10.1186/S42162-021-00172-6.
[23] N. Amral, C.S. Özveren, D. King, Short term load forecasting using multiple linear regression, Proc. Univ. Power Eng. Conf. (2007) 1192–1198, https://doi.org/10.1109/UPEC.2007.4469121.
[24] A.Y. Saber and A.K.M.R. Alam, "Short term load forecasting using multiple linear regression for big data," 2017 IEEE Symp. Ser. Comput. Intell. *SSCI 2017 - Proc.*, vol. 2018-Janua, pp. 1–6, Feb. 2018, doi: 10.1109/SSCI.2017.8285261.
[25] V. Mayrink and H.S. Hippert, "A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting," *2016 IEEE Lat. Am. Conf. Comput. Intell. LA-CCI 2016 - Proc.*, Mar. 2017, doi: 10.1109/LA-CCI.2016.7885697.
[26] D. van der Meer, J. Widén, and J. Munkhammar, "Comparison of strategies for net demand forecasting in case of photovoltaic power production and electricity consumption," in 33rd European Photovoltaic Solar Energy Conference and Exhibition, 2017, pp. 2723–2728, doi: 10.4229/EUPVSEC20172017–6BV.3.86.
[27] X. Chen, F. Xu, G. He, Z. Li, F. Wang, K. Li, et al., Decoupling based monthly net electricity consumption prediction model considering high penetration of distributed solar PV systems, Sustain. Energy Grids Netw. 32 (2022), 100870, https://doi.org/10.1016/J.SEGAN.2022.100870.
[28] P. Kobylinski, M. Wierzbowski, K. Piotrowski, High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources, Int. J. Electr. Power Energy Syst. 117 (2020), 105635, https://doi.org/10.1016/J.IJEPES.2019.105635.
[29] E. Garcia-Garrido, M. Mendoza-Villena, P.M. Lara-Santillan, E. Zorzano-Alba, and A. Falces, "Net demand short-term forecasting in a distribution substation with PV power generation," in 2019 International Conference on Power, Energy and Electrical Engineering (PEEE 2019), 2019, doi: 10.1051/e3sconf/202015201001.
[30] A. Stratman, T. Hong, M. Yi, and D. Zhao, "Net load forecasting with disaggregated behind-the-meter PV generation," Conf. Rec. - IAS Annu. Meet. (IEEE Ind. Appl. Soc., vol. 2022-Octob, 2022, doi: 10.1109/IAS54023.2022.9940025.
[31] S.E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D.B. Smith, M. Minakshi, From load to net energy forecasting: short-term residential forecasting for the blend of load and PV behind the meter, IEEE Access 8 (2020) 224343–224353, https://doi.org/10.1109/ACCESS.2020.3044307.
[32] A. Kaur, L. Nonnenmacher, C.F.M. Coimbra, Net load forecasting for high renewable energy penetration grids, Energy 114 (2016) 1073–1084, https://doi.org/10.1016/j.energy.2016.08.067.
[33] G. Tziolis, A. Koumis, S. Theocharides, A. Livera, J. Lopez-Lorente, G. Makrides, et al., "Advanced short-term net load forecasting for renewable-based microgrids," 2022 *IEEE Int. Smart Cities Conf.*, pp. 1–6, Sep. 2022, doi: 10.1109/ISC255366.2022.9922157.
[34] G. Tziolis, A. Livera, J. Montes-Romero, S. Theocharides, G. Makrides, G. E. Georghiou, Direct short-term net load forecasting based on machine learning principles for solar-integrated microgrids, IEEE Access (2023), https://doi.org/10.1109/ACCESS.2023.3315841.
[35] G. Tziolis, C. Spanias, M. Theodoride, S. Theocharides, J. Lopez-Lorente, A. Livera, et al., Short-term electric net load forecasting for solar-integrated distribution systems based on Bayesian neural networks and statistical post-processing, Energy 271 (2023), 127018, https://doi.org/10.1016/J.ENERGY.2023.127018.
[36] G. Tziolis, J. Lopez-Lorente, M.-I. Baka, A. Koumis, A. Livera, S. Theocharides, et al., "Comparative analysis of machine learning models for short-term net load forecasting in renewable integrated microgrids," 2022 2nd Int. Conf. Energy Transit. Mediterr. Area (SyNERGY MED), pp. 1–5, Oct. 2022, doi: 10.1109/SYNERGYMED55767.2022.9941378.
[37] C. Schoen, A new empirical model of the temperature–humidity index, J. Appl. Meteorol. Clim. 44 (9) (2005) 1413–1420, https://doi.org/10.1175/JAM2285.1.
[38] D. Bolton, The computation of equivalent potential temperature, Mon. Weather Rev. 108 (7) (1980) 1046–1053, https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2.
[39] A. Livera, M. Theristis, E. Koumpli, S. Theocharides, G. Makrides, J. Sutterlueti, et al., Data processing and quality verification for improved photovoltaic performance and reliability analytics, Prog. Photovolt. Res. Appl. 29 (2) (2021) 143–158, https://doi.org/10.1002/PIP.3349.
[40] M.Q. Raza, A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, Renew. Sustain. Energy Rev. 50 (2015) 1352–1372, https://doi.org/10.1016/J.RSER.2015.04.065.
[41] S. Theocharides, C. Spanias, I. Papageorgiou, G. Makrides, S. Stavrinos, V. Efthymiou, et al., A hybrid methodology for distribution level photovoltaic power production forecasting verified at the distribution system of Cyprus, IET Renew. Power Gener. 16 (1) (2022) 19–32, https://doi.org/10.1049/RPG2.12296.
[42] J. Sauer, V.C. Mariani, L. dos Santos Coelho, M.H.D.M. Ribeiro, M. Rampazzo, Extreme gradient boosting model based on improved Jaya optimizer applied to forecasting energy consumption in residential buildings, Evol. Syst. 1 (2021) 1–12, https://doi.org/10.1007/S12530-021-09404-2.
[43] S. Touzani, J. Granderson, S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings, Energy Build. 158 (2018) 1533–1543, https://doi.org/10.1016/J.ENBUILD.2017.11.039.
[44] R. Zhang, Y. Xu, Z.Y. Dong, W. Kong, and K.P. Wong, "A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts," IEEE Power Energy Soc. Gen. Meet., vol. 2016-Novem, Nov. 2016, doi: 10.1109/PESGM.2016.7741097.
[45] Y. Xie, Y. Ueda, M. Sugiyama, A two-stage short-term load forecasting method using long short-term memory and multilayer perceptron, Energies 14 (18) (2021) 5873, https://doi.org/10.3390/EN14185873.
[46] G.F. Fan, L.Z. Zhang, M. Yu, W.C. Hong, S.Q. Dong, Applications of random forest in multivariable response surface for short-term load forecasting, Int. J. Electr. Power Energy Syst. 139 (2022), 108073, https://doi.org/10.1016/J.IJEPES.2022.108073.
[47] H. Eskandari, M. Imani, M.P. Moghaddam, Convolutional and recurrent neural network based model for short-term load forecasting, Electr. Power Syst. Res. 195 (2021), 107173, https://doi.org/10.1016/j.epsr.2021.107173.
[48] H. Liu, Y. Tang, Y. Pu, F. Mei, D. Sidorov, Short-term load forecasting of multi-energy in integrated energy system based on multivariate phase space reconstruction and support vector regression mode, Electr. Power Syst. Res. 210 (2022), 108066, https://doi.org/10.1016/J.EPSR.2022.108066.
[49] H.S. Hippert, D.W. Bunn, R.C. Souza, Large neural networks for electricity load forecasting: are they overfitted? Int. J. Forecast. 21 (3) (2005) 425–434, https://doi.org/10.1016/J.IJFORECAST.2004.12.004.
[50] S. Karsoliya, Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture, Int. J. Eng. Trends Technol. 3 (6) (2012).