Short-Term Forecasting of Heat Demand of Buildings for Efficient and Optimal Energy Management Based on Integrated Machine Learning Models

Abinet Tesfaye Eseye and Matti Lehtonen

***Abstract*—The increasing growth in the energy demand calls for robust actions to design and optimize energy-related assets for efficient and economic energy supply and demand within a smart grid setup. This article proposes a novel integrated machine learning (ML) technique to forecast the heat demand of buildings in a district heating system. The proposed short-term (24h-ahead) heat demand forecasting model is based on the integration of empirical mode decomposition (EMD), imperialistic competitive algorithm (ICA), and support vector machine (SVM). The proposed model also embeds an ML-based feature selection (FS) technique combining binary genetic algorithm and Gaussian process regression to obtain the most important and nonredundant variables that can constitute the input predictor subset to the forecasting model. The model is developed using a two-year (2015–2016) hourly dataset of actual district heat demand obtained from various buildings in the Otaniemi area of Espoo, Finland. Several variables from different domains such as seasonality (calendar), weather, occupancy, and heat demand are used to construct the initial feature space for FS process. Short-term forecasting models are also implemented using the Persistence approach as a reference and other eight ML approaches: artificial neural network (ANN), genetic algorithm combined with ANN (GA-ANN), ICA-ANN, SVM, GA-SVM, ICA-SVM, EMD-GA-ANN, and EMD-ICA-ANN. The performance of the proposed EMD-ICA-SVM-based forecasting model is tested using an out-of-sample one-year (2017) hourly dataset of district heat consumption of various building types. Comparative analysis of the forecasting performance of the models was performed. The obtained results demonstrate that the devised model forecasts the heat demand with improved performance evaluated using various accuracy metrics. Moreover, the devised model achieves outperformed forecasting accuracy enhancement, compared to the other nine evaluated models.**

Manuscript received September 9, 2019; revised December 21, 2019 and January 1, 2020; accepted January 15, 2020. Date of publication February 10, 2020; date of current version September 18, 2020. This work was supported by the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Espoo, Finland. Paper no. TII-19-4151. *(Corresponding author: Abinet Tesfaye Eseye.)*

The authors are with the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland (e-mail: abinet.eseye@aalto.fi; matti.lehtonen@aalto.fi).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2970165

***Index Terms*—Building, data-driven model, district heating, energy management, forecasting, machine learning (ML), smart grid.**

I. Introduction

THE RECENT growth of the energy demand has been challenging the energy generation and its delivery. Energy reliability and efficiency can be attained by implementing optimal strategies at the generation and demand sides. In the current smart grid context, smart grids encompass both electricity and heating systems. A prediction platform is important for both electricity and heating networks at the demand side.

There are different types of heating systems to deliver heat energy to thermal loads. District heating system (DHS) is a system for dispensing heat energy produced at a central place via a network of highly insulated pipes for residential, commercial, and industrial heating needs such as space heating/cooling and water heating/cooling.

It has been found that buildings consume a large quantity of energy. According to the Center for Clean Air Policy, buildings share almost 40% of the global energy demand [1] and as Eurostat, buildings share 38.1% of the energy demand in the European Union (EU), much greater than any other area, comprising transport (33.3%), and industry/factory (25.9%) [2]. Specifically, heating demands share about 55% of the energy demands of buildings globally [3]. The implementation of efficient and optimal building energy management system (BEMS) is anticipated to produce a peak saving of 8% of the consumption in the EU [4]. For reducing the power consumption and enhance compliance with the EU policies on buildings energy efficiency [5], it is necessary to regulate effectively the available Heating, Ventilation, and Air Conditioning (HVAC) systems. Therefore, heat energy demand (consumption of HVAC systems) forecasting is vital for optimal, efficient, and smart energy management. This largely assists the control and management of BEMSs, in the contemporary smart grid context.

In the DHS, several developments have been made to achieve effective operational control from the economic and ecological aspects. Nevertheless, most previous works are bounded to the generation side [6]–[8]. However, for improved and flexible planning and operation of smart grids, it is important to consider

the energy demand sector where decentralized planning and operation monitoring are possible. It assists smart grids to further decrease energy demands, emissions, fuel consumptions (in combined heat and power generation) and rush hour demands. This can be realized via optimal management and decentralized energy saving approaches, and through increasing of local generations close to the consumption centers. That is why the focus of this article is to assist smart building energy management from the demand control side using integrated machine learning (ML) model to predict the heat energy consumption of buildings integrated in DHS.

The heat demand curves of small-scale energy systems such as microgrids, residential entities, commercial areas, industrial units, buildings, and homes (offices or single rooms) have quite different characteristics from the typical heat consumption curves that generally represent national or regional demands. In small-scale energy units, the amount of heat demand is many times less than national- or regional-level demands, the heat demand profile manifests higher fluctuation, and it is much more stochastic. This makes the conventional modeling and forecasting methodologies established for national- or regional-level demands unsuitable for straightforward use in small-scale energy entities. That is why modeling and forecasting of energy (electricity, heat, gas, and water) demands of small-scale energy systems have become a hot research and development issue in both academia and industry since the last few years.

The prior works in the area of heat demand prediction can generally be classified as classical (statistical) and data-driven techniques [9], [10]. The classical methods use equations that define the physical characteristics of systems to estimate the outcome (value of the target variable). The data-driven techniques relate to artificial intelligence (AI), ML, and deep learning (DL) tools where observations of system inputs and outputs are gathered to develop the forecasting model. The observed data are, then, employed to describe model of the system [9].

Numerous research groups have recently taken into account the use of AI, ML, and DL to build data-driven systems for improved modeling and forecasting energy generations and demands. This is due to the fast development and deployment of the Internet-of-Things, smart meters, sensor networks, big data, and cloud computing technology to collect and process large amount of system data.

Various data-driven AI/DL/ML approaches have been employed for heat demand prediction in different application scenarios, for instance, support vector regression [11], multiple regression [12], and artificial neural network (ANN) [6], [13]. Idowu *et al.* [10] proposed a data-driven method to forecast the heating energy demand of apartment buildings in a DHS. The method takes into account the external and internal factors affecting the heat consumption of the buildings and used them as input for the forecasting model. Several ML-based forecasting models were also implemented and compared for varying prediction horizons up to 24 h. Several ML tools for day-ahead forecasting of heat consumption of thermal loads in DHS network in the city of Riga, Latvia are implemented in [14]. The forecast results are used as inputs for decisions in day-ahead operation planning and whole electricity market participation. The forecasting tools were implemented using ANN, polynomial regression, and hybrid of the former two techniques. Performance comparison among these tools was carried out and it is shown that the hybrid tool performs better.

The data-driven models have shown considerable accuracy improvements over the classical methods. Data-driven methods have the advantages of establishing models from big data, easy adaptability, and fast model parameter updating capability [15]. The fast parameter updating capability is especially very important for energy demand forecasting problems because of the nonstationary behavior of energy demand profiles.

The literature survey in this article, as reported above, indicates that most of the prediction models have aimed to predict large-scale heat energy usage at the national or regional levels. Nevertheless, there exists quite limited number of research works on prediction of heat energy consumptions in small-scale energy systems. In addition, there have been quite few previous works on data-driven-based heat demand forecasting for small-scale energy systems. Therefore, this article can contribute to the field of heat demand modeling and forecasting for small-scale energy systems in general and buildings in particular by making use of data-driven integrated ML models.

Besides, most of the heat demand forecasting models by the prior researchers employed randomly or subjectively selected predictors (features or variables) to form the input dataset for the prediction model. This has become a major challenge to maintain the performance consistency of forecasting models over different operation scenarios and changing conditions. To overcome this problem, we use an ML [hybrid of binary genetic algorithm (BGA) and Gaussian process regression (GPR)] feature selection (FS) methodology to find the most relevant and nonredundant variables to establish the training input data for the proposed heat demand forecasting model.

Moreover, most of the previous data-driven ML energy demand prediction models have employed a supervised back propagation (BP) training technique to find their parameters. However, BP training methods are often trapped by suboptimal (local) solutions and unable to attain global (system-level) optimum and this, in turn, reduces the accuracy of the forecasting models. To alleviate this problem, we use the imperialistic competitive algorithm (ICA) optimization algorithm to search for the global optimum values of the ML [support vector machine (SVM)] model parameters. The ICA, unlike the BP algorithms, is capable to find global optimum solution. Hence, optimal parameter set of the SVM model and, thus, accurate heat demand forecasts are obtained, in this article, through the parameter optimization process using the ICA algorithm. In addition, important advantages of SVMs over ANNs are that they have a simple geometric interpretation, few parameters to adjust, and give a sparse solution. Unlike ANNs, the computational complexity of SVMs does not rely on the dimensionality of the input space. ANNs use empirical risk minimization, while SVMs use structural risk minimization. Besides, SVMs are less prone to overfitting problems than ANNs. The ICA has few parameters to adjust, easy to implement, and fast convergent compared to the other evolutionary algorithms.

Furthermore, unlike the prior works on data-driven ML energy demand forecasting, this article employs a feature extraction technique using the empirical mode decomposition (EMD) technique to extract the most relevant features of the target variable (heat demand) for use in the prediction model training.

Therefore, the proposed data-driven ML-based heat demand forecasting approach in this article consists of integrated EMD-ICA-SVM model augmented with BGA-GPR-based FS process. This approach is novel and proposed mainly because of its improved training mechanism, higher accuracy, and short learning time. As far as we have investigated, this is the first ML-based approach applying the integrated EMD-ICA-SVM model augmented with BGA-GPR-based FS method for energy demand modeling and forecasting.

Moreover, the proposed integrated EMD-ICA-SVM heat demand prediction model is compared with persistence, ANN (BP trained ANN), genetic algorithm (GA)-ANN (GA trained ANN), ICA-ANN (ICA trained ANN), SVM (BP trained SVM), GA-SVM (GA trained SVM), ICA-SVM (ICA trained SVM), EMD-GA-ANN, and EMD-ICA-ANN models to demonstrate its effectiveness with respect to different forecasting performance evaluation metrics. This article is the extension of [16].

The key contributions of the article are as follows.

1) This article presents a novel and effective integrated ML model for 24h-ahead district heat demand forecasting in small-scale energy systems (buildings).
2) This article improves forecasting accuracy by the augmenting robust ML-based FS technique to the forecasting model.
3) The article evaluates the performance the proposed forecasting model over different building types (customer classes) such as residential, educational, office and mixed-use.
4) This article enhances prediction accuracy, taking into account accuracy values achieved by other nine models.

The rest of this article is structured as follows. Section II describes the proposed forecasting approach. Section III presents the data collection and preparation methods. The features selection strategy is provided in Section IV. The forecasting model configuration and development is presented in Section V. The forecasting performance evaluation metrics are defined in Section VI. The experimental findings and discussions are presented in Sections VII. Section VIII concludes this article.

II. Proposed Forecasting Approach

The day-ahead hourly forecast results of the proposed model are aimed to be employed as input information for efficient and optimal energy management decisions in smart grid operation planning. Specifically, the forecasts are employed as input information for smart BEMS and/or other operational control systems, as shown in Fig. 1.

This article devises a novel heat demand forecasting approach using the hybridization of the EMD, ICA, and SVM (integrated EMD-ICA-SVM). The approach also includes an FS methodology using the combination of the BGA and GPR (integrated BGA-GPR). First, the most important and nonredundant variables that can constitute the input dataset for the integrated

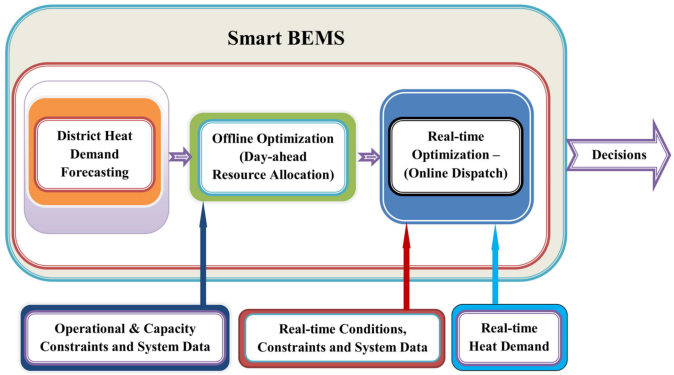

Fig. 1. Heat demand forecast for BEMS.

EMD-ICA-SVM forecasting model are chosen by the integrated BGA-GPR FS method. Several candidate variables are collected from different domains to form the initial feature space for the FS process, and, then, the final selected features are used as predictor (input) subset for the forecast model training. The BGA in the FS is the main feature selector tool while the GPR is the performance measure tool used to score or rank the variables' importance. The EMD in the hybrid EMD-ICA-SVM forecasting model is used to extract the important characteristics of the heat demand profile. It converts the original heat demand time series data into a set of improved subseries data before the model training commences. The ICA is used to find the optimal values of the SVM parameters. The SVM is the basic regression tool or prediction engine of the forecast model.

The historical values of the selected predictors (input features) and the EMD extracted subseries of the heat demand data are used to the train SVM model using the ICA parameter optimization process. Once the trained model is obtained, the future values of the predictors are fed to the trained model to obtain the forecasted subseries of the heat demand. Finally, the forecasted heat demand time series data is obtained by applying the inverse (reconstruction) EMD on the forecasted subseries.

Fig. 2 illustrates the proposed heat demand forecasting approach. As shown in Fig. 2, the proposed forecasting approach consists of three stages. The first stage (Stage #1) obtains the predictors. The second stage (Stage #2) fits the forecast model. The third stage (Stage #3) forecasts the future values of the heat energy demand based on the fitted (trained) forecast model.

The detail working principle and mathematical formulations of the FS methodology and the forecaster model will be presented in Sections IV and V, respectively. The sequential operations and calibrations of the proposed forecasting approach are illustrated with the flowchart shown in Fig. 3. A two-year (2015–2016) 1-h resolution dataset of the initial feature space and the target variable (heat demand) are used for the FS process.

In addition, the two-year (2015–2016) 1-h resolution dataset of the selected predictors and subseries of the target variables are used to learn and validate the ICA-SVM forecast model. The proposed forecast model can be retrained every time (with 24 h rolling period) when new daily training dataset is available.

The efficacy of the devised prediction model is tested with a one-year 1-h resolution out-of-sample (2017) dataset.

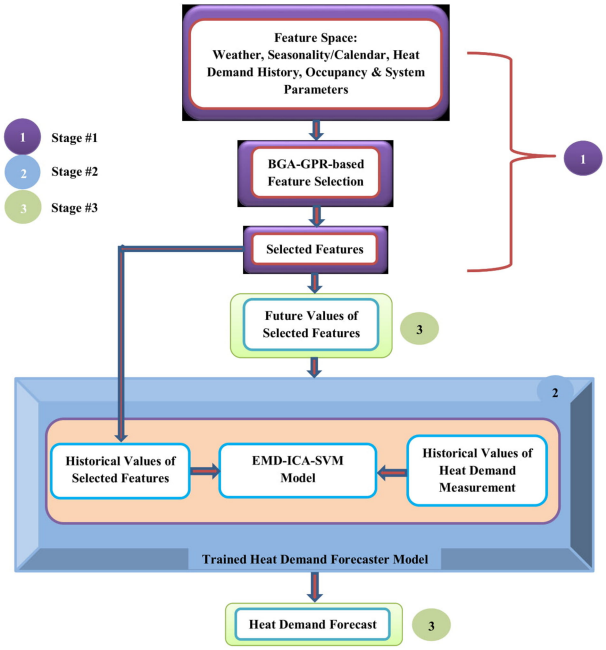

Fig. 2. Proposed heat demand forecasting approach.

The forecast results and performance evaluation analysis will be discussed in Section VII.

III. Data Collection and Preprocessing

The dataset used to construct the proposed forecasting model consists of variables (f_i) collected from different domains or sources, namely seasonality (calendar), weather, heat demand, and demography (occupancy of people). The initial feature space is established through basic assessment of the characteristics of the heat demand data and its association with the previous consumption (time-lag-similarity) and exogenous factors (i.e., the domain variables, f_i). The accessibility of the data sources for the domain variables is also another key issue to build the initial predictor space. In total, 20 variables are used to establish the initial feature space for the FS process. Among those variables, 11 of them are weather variables, four of them are seasonal variables, three of them are heat demand-derivative variables, and two of them are occupancy-representative variables. The target variable is of course the heat demand data. Table I shows the list of the variables and their associated domain (i.e., type or source).

A three-year (2015–2017) window length of historical values of the variables are collected with 1-h resolution. The two-year (2015–2016) historical data are used for the FS task and forecast model training and validation. While the one-year (2017) data are employed for the model testing (forecasting).

The district heat demand data are actual data collected from various building types (residential, educational, office, and mixed-use) in the Otaniemi area of Espoo, Finland.

The weather data are collected from the nearby (to the buildings) weather stations (open source) of the Finnish Meteorological Institute [17]. The seasonality data are taken from the official

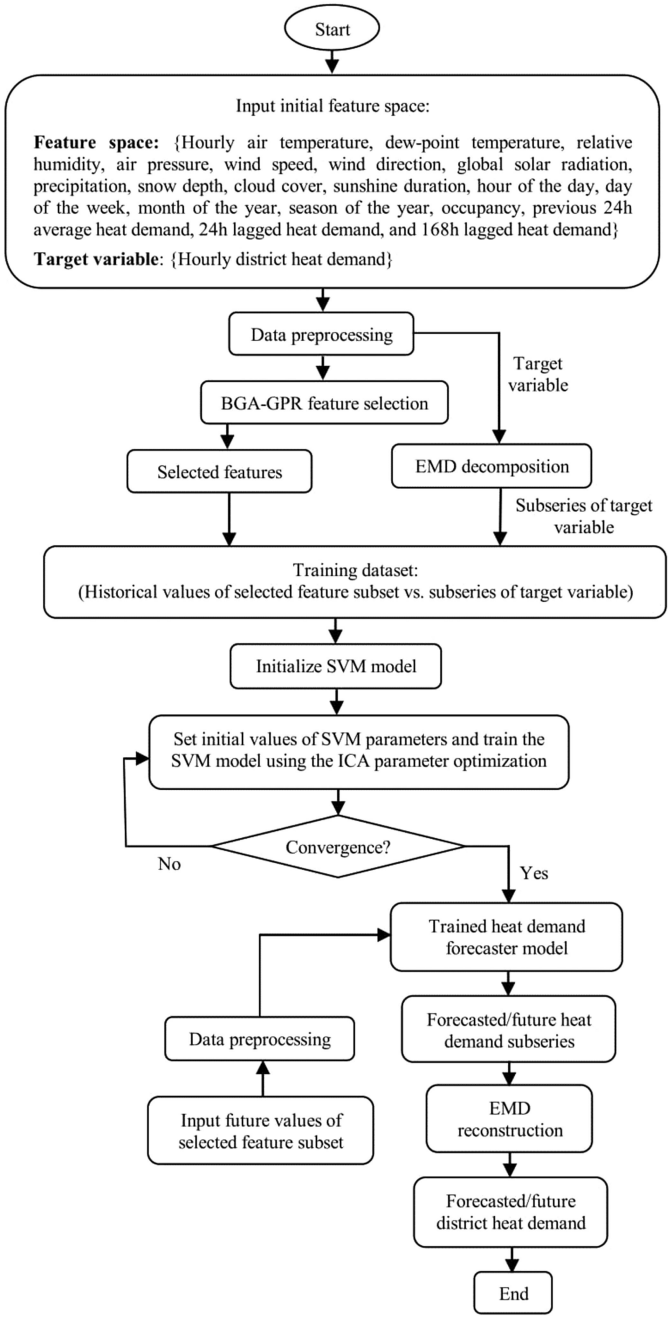

Fig. 3. Proposed district heat demand forecasting algorithm.

annual calendar of the country where the buildings are located [18]. The people occupancy is also the domain variable in the feature space.

Occupancy represents the number of heat energy users (people) in the building at each time interval. It has been found that building energy consumptions have a high correlation with their user (people) occupancies [19].

However, due to limitation of time and resources for direct measurement of occupancy at the buildings, we have used indirect representation of buildings occupancy using two additional variables, the holiday indicator and period of the day. The holiday indicator variable has binary values (0 or 1), where

TABLE I
FEATURE SPACE VARIABLES AND THEIR DOMAIN

Feature Index (f_i)	Feature/Variable	Unit/Scale	Domain
1	Hour of the day	1-24	Seasonality /Calendar
2	Day of the week	1-7	
3	Month of the year	1-12	
4	Season of the year	1-4	
5	Holiday/weekend indicator	0-1	Occupancy
6	Period of the day	1-3	
7	Air temperature	°C	Weather
8	Dew-point	°C	
9	Relative humidity	%	
10	Precipitation	mm/h	
11	Snow depth	cm	
12	Air pressure	hPa	
13	Wind direction	deg	
14	Wind speed	m/s	
15	Cloud cover	0-8	
16	Global solar radiation	Watt/m2	
17	Sunshine duration	s	
18	Previous 24h average heat demand	kW	Heat demand
19	24h lagged heat demand	kW	
20	168h lagged heat demand	kW	

0 indicates the particular hour is in a holiday (or weekend) and 1 specifies the hour is in a working day (in weekday). The list of holidays and calendar information in the years 2015, 2016, and 2017 are taken from [18]. The period of the day is another variable used to represent the occupancy. It has three numeric values (1, 2, or 3), where a value of 1 indicates an idle mode (nonworking hours) in the time period [9PM, 7AM), 2 indicates working mode in the time period [7AM, 5PM), and 3 indicates cool-down mode in the time period [5PM, 9PM). Hence, these two predictor variables, the holiday indicator and the period of the day, are used as representative variables of the people occupancy in the buildings. The last three variables (f_{18}, f_{19}, and f_{20}) in Table I are derived from the original heat demand data based on the time-lag similarity behavior of the heat consumption profile. They are important variables to demonstrate whether the heat demand at the current time depends on the demand at the previous times.

Regarding the data preprocessing task, some of the predictors must be treated to further suitable forms and synchronized timestamps ahead of the BGA-GPR FS, EMD extraction, and ICA-SVM learning tasks. All the dataset values are converted into hourly mean values. The predictors and target datasets are timestamped in one-hour resolution to synchronize the timestamp dissimilarities of the various data sources. The weather stations timestamps are in UTC and, thus, they are changed to the local time to fit them with the time zone of the city (Espoo, Finland) where the buildings are located (i.e., heat demand data timestamps).

IV. FEATURE SELECTION

FS is the process of finding a cluster of the most significant features (attributes, variables, or predictors) for use in forecasting model development. It enhances prediction accuracy, reduces model training time, and overfitting [19]. The major premise when using an FS is that the initial dataset contains some attributes that are either repetitive or superfluous, and can, thus, be eradicated without persuading loss of relevant information. Several research works have demonstrated that repetitive or superfluous variables decrease the performance, flexibility, and generalization potential of forecasting models.

Quite few studies have done FS ahead of learning energy demand forecasting models. This article devises an ML-based integrated predictor selection methodology to find the most significant and nonrepetitive features for enhanced short-term building heat demand prediction. The devised FS methodology employs the BGA for the attribute selection task and GPR for measuring the fitness score or importance of the variables. The variables are sorted and chosen based on their calculated value of the GPR residual. The BGA runs to minimize the GPR residual (i.e., performance index or evaluation measure) for obtaining the most significant and nonrepetitive variables. The subset of variables (feature subset) that gives the best value of the performance index is selected at the end.

The feature space for the heat demand forecasting contains several candidate variables from various domains as given in Table I. The variables f_i, i = 1, 2, …, 20, in Table I form the predictor space required for the FS task. Hence, the predictor space is m by n matrix, where m = 17 544 is the number of observations, which is a two-year (2015–2016) hourly sample of the predictors and n = 20 is the size of the predictor space (i.e., number of candidate variables in the original dataset).

The BGA is a type of GA that works by first designating (encoding) the given predictor space (candidate solutions) with binary bitstrings. This brands the BGA very suitable for FS applications than the traditional real-valued GA [20]. The BGA operates on the set of candidate solutions (chromosomes) to generate new chromosomes (offsprings) using three genetic operators known as selection, crossover, and mutation. The fitness of the chromosomes is evaluated by defining a fitness function. The fitness function gives numerical values that are used to sort the chromosomes.

The GPR is a powerful algorithm for regression. It has a robust competence of approximating a nonlinear relationship between variables using probabilistic distributions [21]. GPR has few adjustable parameters and it is simple for implementation. The GPR output y of the function f at feature x is defined as

$$y = f(x) + \varepsilon \tag{1}$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ is a normal distribution with zero mean and σ standard deviation. The function f(x) is considered as a Gaussian Process (GP) distribution and its GP formulation is given below

$$f(x) = GP(m(x), k(x, x')). \tag{2}$$

The GP is expressed by the mean and covariance functions. The mean m(x) designates the expected value of the function at x

$$m(x) = E[f(x)]. \tag{3}$$

The covariance k(x, x') describes the relation between the values of the function at different points x and x'

$$k(x, x') = E\left[(f(x) - m(x))(f(x') - m(x'))\right]. \quad (4)$$

The covariance k is generally called the GP kernel [22]. The GPR in this article uses the squared exponential kernel expressed beneath

$$k(x, x'|\theta) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{r=1}^{d}\frac{(x_r - x'_r)^2}{\sigma_r^2}\right) \quad (5)$$

where σ_r symbolizes the scale-length of the GPR predictor r, r $= 1, 2, \ldots,$ d and σ_f is the standard deviation of the GPR training dataset. The parameter θ is expressed as follows:

$$\theta_r = \log\sigma_r \text{ , for } r = 1, 2, \ldots, d \quad (6)$$

$$\theta_{d+1} = \log\sigma_f. \quad (7)$$

σ_r and σ_f are controlled to increase or diminish the correlation between the points and, therefore, manage the distribution of the function. Following the determination of the mean and kernel, the GP is, then, applied to find the prior and future values of the function using the historical observations (dataset).

The flowchart of the BGA-GPR-based FS methodology for the proposed heat energy demand prediction approach is shown in Fig. 4.

The fitness (suitability) of the variable subsets is calculated employing the mean squared error (MSE) of the GPR model residual. The GPR model is established for all the variable subsets using the features whose position in the chromosomes is indexed by "1." Thus, the MSE of the actual target and GPR output is employed as the fitness function (performance index), and it is defined by

$$fit = \frac{1}{n}\sum_{i=1}^{n}(T_i - y_i)^2 \quad (8)$$

where T is a vector of the actual target (heat demand), y is a vector of the GPR model output (estimate), and n is length of the model training dataset that equals 17 544 samples (a two-year (2015–2016) hourly record).

The objective of the BGA is to minimize the fitness function formulated in (8) by selecting a feature subset from the original variable space that can give the best fitness value (lowest MSE) through iterations.

A gene value of "1" in the chromosomes indicates that the particular variable in that position is selected, whereas a value of "0" means that the variable indexed in that position is not selected. While the BGA is running, the chromosomes (candidate variables) in the population are evaluated, and their fitness values are ranked. The chromosome with a better fitness value (lower MSE) has a higher probability of persisting with the following generation.

At the end of the BGA running (i.e., convergence condition is met), the chromosome associated with the best fitness value holds the desired important input variables for the heat demand prediction. This chromosome is, therefore, selected and decoded to form the desired variable subset as illustrated in Fig. 5.

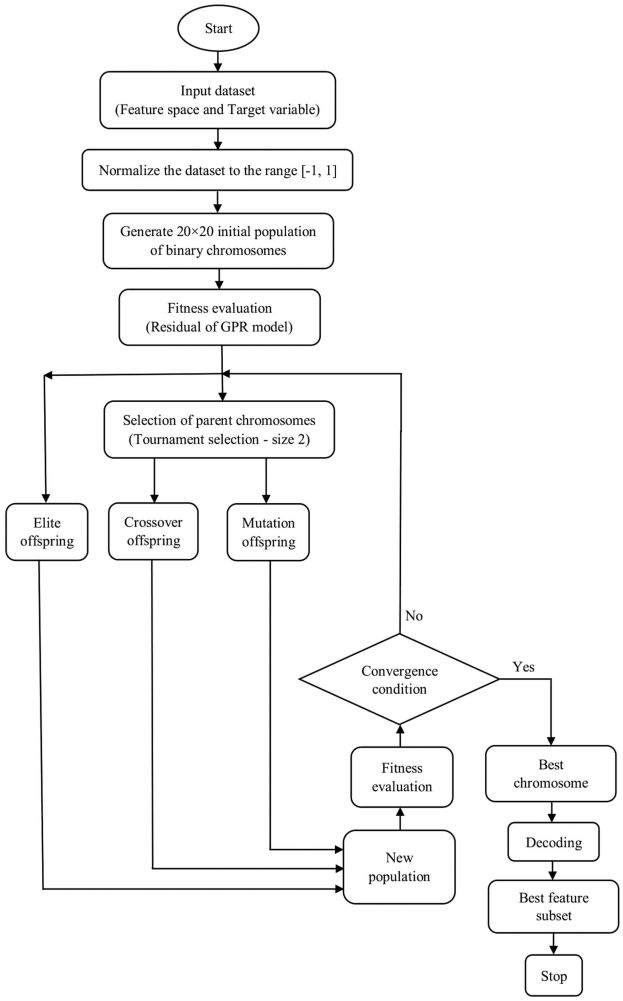

Fig. 4. Flowchart of BGA-GPR-based FS for short-term heat demand forecasting.

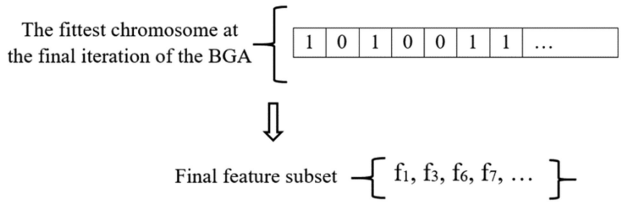

Fig. 5. Final feature subset decoding.

The selected (final) feature subset is used as the training input for the proposed ML-based forecasting model, which will be discussed in the following section.

V. Proposed Forecasting Model

As discussed in the above section sections, the proposed heat demand forecasting model is developed by combining the EMD, ICA, and SVM. The article major contributions in this regard can be considered as 1) modeling (formulation), parameterization and implementation of the EMD, ICA, and SVM algorithms to suit the heat demand forecasting problem in question, and 2)

establishing seamless integration of the three algorithms to work in unison for solving the forecasting problem. The operating mechanism, mathematical formulation, and implementation of each of these algorithms (or models) in the integrated EMD-ICA-SVM forecasting model will be discussed in the following sections.

A. Empirical Mode Decomposition

The magnitudes of the heat demand curves of the buildings vary in each time instant. Observation of the nonstationary and nonlinear behavior of the heat demand curves of the buildings calls for the decomposition of the heat demand raw data (time series) into various subseries elements, to extract the most useful characteristic-feature of the data. In this article, this is performed by the EMD technique. The EMD decomposes the heat demand time series data into a set of subseries that contain more important information than the original raw data. Thus, the decomposed data can be used to forecast the heat energy consumption more accurately. The rational for the effectiveness of the decomposed data in the prediction process is due to the enhanced data decomposition capability of the EMD technique.

By making use of the EMD, the analogical time representation of the frequency presence in the initial district heat demand data is performed to obtain the real-time frequency of the data.

The real-time frequency is obtained by using the Hilbert transform (HT) [23]. The HT for a mono-component signal $g(t)$ is described below

$$h(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau} d\tau. \tag{9}$$

In (9), $g(t)$ and $h(t)$ designate the conjugate pair that describes time series $f(t)$ given below

$$f(t) = g(t) + jh(t). \tag{10}$$

The polar coordinate representation of (10) is expressed below

$$f(t) = c(t) e^{j\emptyset(t)} \tag{11}$$

where

$$\begin{aligned} c(t) &= \sqrt{g(t)^2 + h(t)^2} \\ \emptyset(t) &= \arctan\left(\frac{h(t)}{g(t)}\right) \end{aligned} \tag{12}$$

where $c(t)$ and $\varphi(t)$ designate the real-time amplitude and phase of $f(t)$. $c(t)$ and $\varphi(t)$ represent the suboptimal (neighborhood) representation of an amplitude- and phase-angle-oscillating trigonometrical fit to $g(t)$. The real-time frequency $\omega(t)$ is obtained from the real-time phase expressed below

$$\omega(t) = \frac{d\emptyset(t)}{dt} = \frac{\dot{h}(t) g(t) - h(t) \dot{g}(t)}{g^2(t) + h^2(t)}. \tag{13}$$

$\omega(t)$ is essentially feasible if $\varphi(t)$ is a mono-component signal. Since $\varphi(t)$ is derived from $g(t)$, $g(t)$ must be a mono-component signal as well. However, almost all real-world signals, especially weather variables and building heat demand, are not mono-component. The work by Huang *et al.* [23] described the EMD technique for decomposing multicomponent signals into mono-component subsignals. By using the EMD, the heat demand data series is expressed by an aggregation subseries data called intrinsic mode functions (IMFs). Each generated IMF is mono-component quantity that must satisfy the requirements below.

1) The quantity of zero touches and the quantity of extreme points should be same or vary, at extreme, by one.
2) The mean of the wraps found by local maximum points and minimum points must be zero at every instant.

The IMFs are derived from the initial data by employing a sifting technique. In the sifting technique, bottom and top wraps are formed by introducing an interpolating curve via the neighborhood minimum and maximum points. The wraps average q_1 is deducted from $g(t)$ to find the initial element d_1. To create the IMFs, the sifting process is successively carried out j times to d_i, till the IMFs are found

$$d_{1k} = d_{1_{(j-1)}} - q_{1_j}. \tag{14}$$

The sifting procedure ends if the standard-deviation of two serial outcomes is less than a specified threshold value. The initial IMF contains information about the peak frequency is expressed as

$$a_1 = d_{1_j}. \tag{15}$$

Then, a_1 is deducted from the first signal, and the residue p_1, which holds information about the lesser frequency elements, is expressed as

$$p_1 = g(t) - a_1. \tag{16}$$

In the process of the EMD decomposition, p_1 is considered like the beginning signal. The procedure iterates, a_2, is calculated, etc., till either of the requirements below are satisfied

1) p_n or a_n has lower energy;
2) p_n is monotonic.

Utilizing the aforementioned process, the decomposition of the initial multicomponent signal $g(t)$ is expressed as follows:

$$g(t) = \sum_{i=1}^{n} a_i + p_n. \tag{17}$$

Each IMF (a_i) is constrained to HT by (13) to determine the real-time frequencies. Since $g(t)$ is a multicomponent signal, it contains greater than one frequencies. Therefore, the HT-based decomposition of $g(t)$ is given by following expression [23]:

$$g(t) = Re\left\{\sum_{i=1}^{n} c_i(t) \exp\left(j \int \omega_i(t)\, dt\right)\right\}. \tag{18}$$

Here, $c_i(t)$ and $\omega_i(t)$ are real-time quantities whose values change instantaneously. The HT assisted Huang EMD preprocessor, expressed by (18), is called the Hilbert–Huang transform (HHT).

The EMD provides a comprehensive, adaptively flexible, and almost orthogonal designation of the initial function compared to Fourier and Wavelet transforms [24]. Therefore, in this article, the EMD is selected for the devised day-ahead district heat demand forecasting model development because of its better performance for extracting important detail behaviors of the

TABLE II
PSEUDO-CODE OF THE EMD DECOMPOSITION PROCESS

Pseudo code of improved HHT EMD decomposition process
(1) Start
(2) Preset: $p_0 = g(t)$, and $i = 1$
(3) Obtain IMF i
(a) Preset: $d_{i(j-1)} = p_i$, $j = 1$
(b) Obtain the neighborhood minimum and maximum points of $d_{i(j-1)}$
(c) Cubic spline the neighborhood minimum and maximum points to create bottom and top wraps to $d_{i(j-1)}$
(d) Calculate the average $q_{i(j-1)}$ of the bottom and top wraps of $d_{i(j-1)}$
(e) Set $d_{ik} = d_{i(j-1)} - q_{i(j-1)}$
(f) Verify whether d_{ij} meets the IMF criteria
(g) If d_{ij} meets the IMF criteria then let $\text{IMF}_i = d_{ij}$, otherwise proceed to procedure (b) by increasing j by one (i.e., $j = j + 1$)
(4) Verify the cross-correlation μ_i between IMF i and $g(t)$
• Save IMF i when $\mu_i \geq \lambda$, otherwise remove IMF i and combine it with residue p_i, and proceed towards procedure (2)
(5) Set: $p_{i+1} = p_i - \text{IMF}_i$
(6) Check convergence/completion
• If p_{i+1} till contains at minimum two extreme points then proceed to procedure (2), otherwise the feature extraction procedure is over and proceed to procedure (7)
(7) Collect the IMFs and residue
• Obtain the extracted IMFs, IMF_i and residual, p_{i+1}
(8) End

target variable. That is, the amplitudes of the IMFs generated from the EMD decomposition (of the heat demand) are used as target variables for the proposed ICA-SVM-based heat demand forecasting model.

Table II presents the pseudo-code of the EMD decomposition process [25], [26].

B. Imperialist Competitive Algorithm

The ICA optimization technique is employed in order to optimally adjust the SVM model parameters for the sake of achieving a higher forecasting accuracy. The ICA is a newly emerging optimization method in the class of evolutionary algorithms. It was inspired by the imperialistic competition in the socio-political evolution of humans for developing a powerful imperialistic authority [27].

The ICA optimal solution search begins with an initial population of N_{country}, which are classified into two distinct categories called imperialists and colonies based on the functional values of the solution that are indicated with (N_{imp}) and (N_{col}), respectively. In the ICA terminology, the combination of imperialists and colonies is said to be empires. Each colony in the first population is proportionally distributed to the empires according to the powers of the imperialists. The power of empires is inversely proportional to their cost (functional value). The normalized cost of imperialists is given below

$$C_n = c_n - \max_i (c_i) \tag{19}$$

where c_n is the nth imperialist cost and C_n is the associated normalized cost of the imperialist. Assuming this normalized cost is computable for all the imperialists, the normalized power

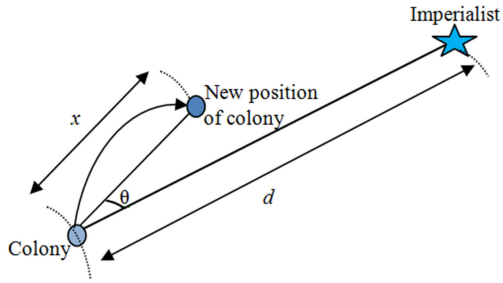

Fig. 6. Movement of colonies toward their pertinent imperialist.

of each imperialist is given by

$$P_n = \left| \frac{C_n}{\sum_{i=1}^{N_{\text{imp}}} C_i} \right|. \tag{20}$$

At the assimilation stage, every empire's colonies start moving toward their target imperialist, right away the establishment of the initial empires. Fig. 6 shows the movement of the colonies toward their relevant imperialist by x units and θ deviations (θ is also called assimilation angle coefficient), where x and θ are random parameters that enjoy uniform distributions

$$\begin{aligned} x &\in U\ (0,\ \beta \times d) \\ \theta &\in U\ (-\gamma,\ \gamma) \end{aligned} \tag{21}$$

where β is called assimilation coefficient, and it is a number defined to be slightly bigger than 1; d is the straight distance from the imperialist to the colony; and γ, called revolution rate, is a parameter that controls the angular rotation from the initial position.

The total cost of each empire is composed of the authority of the imperialist and the colonies, and defined by

$$\begin{aligned} T.C._n = &\ \text{Cost}\left(\text{imperialist}_n\right) \\ &+ \xi.\text{mean}\left\{\text{Cost}\left(\text{colonies of empire}_n\right)\right\} \end{aligned} \tag{22}$$

where T.C._n is the nth empire absolute cost; and ξ, called power coefficient, is positive number slightly less than 1.

According to the philosophy of imperialism, in the process of social and political struggles of human beings, imperialists achieve to arrive at a wider territory than earlier by improving their power to catch additional colonies. Consequently, the power of the stronger imperialist will get an increase in the course of the struggle, while the weaker imperialists will get lower power rank. The weaker empires gradually collapse when they drop their possessions (colonies). This procedure will go on until, at the end, the strongest imperialist rules and all the countries are controlled by that imperialist and become its colonies.

Similarly, in the ICA optimization process, all the empires but the strongest one will finally downfall. This sole empire will control all the colonies. Mathematically speaking, this sole empire is the desired optimal solution of the optimization problem in question.

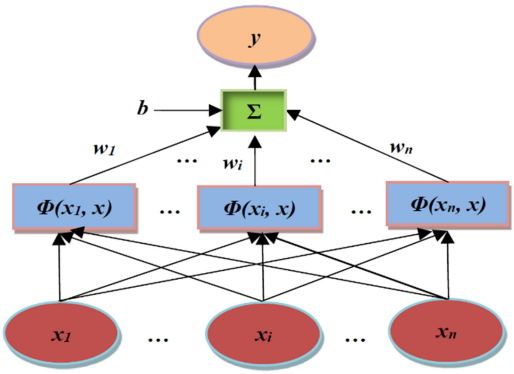

Fig. 7. Structure of SVM.

C. Support Vector Machine

SVM is a type of nonparametric model that fundamentally operates based on kernels. Vapnik [28] created the fundamentals of SVMs in 1995. SVMs are getting substantial credits nowadays because of plenty of evident features and practical advantages. SVMs have been widely applied to forecasting, classification, and clustering tasks. They are formulated based on the structural-risk-minimization theory that has been demonstrated to have better performance than the standard empirical-risk-minimization theory, which is employed for ANN formulation [29]. The SVM basic operational principle is mapping datasets to higher dimension representative hyperplanes using nonlinear mappings. Linear regression in hyperplanes is related to nonlinear regression in reduced-dimension planes. This is formulated as follows [30]:

$$y\left(x\right) = w.\Phi\left(x\right) + b\,;\;\; \Phi:\; R^n \rightarrow\; R^N \tag{23}$$

where $y \in R^N$ is the learning target; $x \in R^n$ is the input (selected feature subset); b is bias parameter; $w \in R^N$ is weight parameter; $\Phi(x)$ is the nonlinear mapping function; and Φ: $R^n \rightarrow R^N$ is the mapping that transforms the original learning space (input variables x) to the hyperplane.

Fig. 7 depicts the structure of SVM, where the inputs x_i is converted to output y by the mapping function $\Phi(\cdot)$. The SVM output y is the sum of the weighted $\Phi(x_i)$ and the bias.

A specific SVM model called linear-epsilon-insensitive SVM (ε-SVM) is employed in this article. The ε-SVM objective function is formulated using the ε-insensitive loss function. The SVM parameters w and b are optimally found by solving the objective function described below

$$\begin{aligned} \min\quad & \left\{\frac{1}{2}w^T w + \gamma\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)\right\} \\ \text{subject to}:\quad & y_i - w.\Phi\left(x_i\right) - b \le \varepsilon + \xi_i \\ & w.\Phi\left(x_i\right) + b - y_i \le \varepsilon + \xi_i^* \\ & \xi_i,\ \xi_i^* \ge 0 \end{aligned} \tag{24}$$

where ξ_i and ξ_i^* are supplementary parameters; γ is the normalization parameter; N is length of the learning dataset; and ε is the loss parameter.

The objective function given in (24) is a constrained quadratic program and is conveniently solved by solving its corresponding dual problem expressed as

$$\begin{aligned} \min\quad & \left\{\begin{array}{l} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_i - \alpha_i^*\right).\Phi\left(x_i, x_j\right).\left(\alpha_j - \alpha_j^*\right) \\ -\sum_{i=1}^{N}\left(\alpha_i + \alpha_i^*\right).\varepsilon + \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right).y_i \end{array}\right\} \\ \text{subject}\quad \text{to}:\quad & \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right) = 0\,;\ \alpha_i,\ \alpha_i^* \ge 0. \end{aligned} \tag{25}$$

Solving for the positive Lagrange multipliers (α_i–α_i*), the simplified expression of the SVM regression output y is given as

$$\hat{y}\left(x\right) = \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right).K\left(x,\ x_i\right) + b \tag{26}$$

where $K(x_i,\ x_j) = \Phi(x_i).\Phi(x_j)$ is called the SVM kernel. The RBF kernel expressed in (27) is used in this article

$$K\left(x_i,\ x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\sigma^2}\right) \tag{27}$$

where σ is a Gauss parameter (width of the kernel) and it describes the influence zone of the support vectors in the search space (learning domain).

As discussed in the above sections, this article proposes the ICA optimization technique (training algorithm) to globally search the *ε-SVM* model parameters for improved prediction accuracy. The SVM model parameters are formulated as parameters of the ICA optimization process. The fitness function formulated by (24) is employed as the cost function in the ICA optimization process. The target of the devised hybrid forecasting model is to achieve a lowest value for the cost function that corresponds to a lowest forecasting error. The ICA successively iterates by solving the cost function until it reaches the optimal SVM parameters that give the desired (lowest) forecasting error. The ICA has the benefit of computational simplicity and few design parameters.

VI. Forecasting Accuracy Evaluation

To evaluate the accuracy of the devised integrated EMD-ICA-SVM district heat demand forecasting model, the mean absolute percentage error (MAPE), root MSE (RMSE), normalized mean absolute error (NMAE), and forecast skill (FCS) metrics are employed. These evaluation metrics are calculated in terms of the actual heat demand data, and their formulations are given next.

The MAPE is given as

$$\text{MAPE} = \frac{100}{N}\sum_{h=1}^{N}\left|\frac{H_h^a - H_h^f}{H_h^a}\right| \tag{28}$$

where H_h^a and H_h^f are the actual and forecasted values of the heat demand at hour h, respectively, and N is the prediction horizon and its value is 24 for day-ahead (24h-ahead) forecast.

TABLE III
FS RESULTS

Customer (Dataset) Type	Selected Features	Best Fitness ($\times 10^{-6}$)	Without FS ($\times 10^{-6}$)
Building A - Residential	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 16, 18, 19, 20	24.78	52.35
Building B - Educational	1, 2, 3, 4, 5, 7, 8, 12, 16, 18, 19, 20	21.58	42.97
Building C - Office	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 14, 16, 18, 19, 20	17.24	33.01
Building D - Mixed-use	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 14, 16, 19, 20	11.24	19.56
All (Union)	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 14, 16, 18, 19, 20		

The RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum\nolimits_{h=1}^{N}\left(H_h^a - H_h^f\right)^2}. \tag{29}$$

The NMAE is expressed as

$$\mathrm{NMAE} = \frac{1}{N}\sum_{h=1}^{N}\frac{\left|H_h^a - H_h^f\right|}{H_{\mathrm{peak}}} \tag{30}$$

where H_{peak} is the peak aggregate heat demand. Annual peak heat demand in the test year is used for H_{peak}, in this article.

The FCS evaluates the performance of forecasting models by referring the prediction accuracy (or error) obtained by the models to the accuracy (or error) obtained by the persistence model (reference model). For day-ahead hourly forecast, the Persistence forecast is given by

$$H_h^f(t) = H_h^a(t-24). \tag{31}$$

The FCS is, then, calculated by relating the RMSE of the desired model forecast with the persistence forecast, as formulated below [25], [31]

$$\mathrm{FCS} = 1 - \frac{\mathrm{RMSE}_{\mathrm{Model}}}{\mathrm{RMSE}_{\mathrm{Persistence}}}. \tag{32}$$

A FCS value of 1 indicates a perfect model, and 0 indicates the model's RMSE is the same as the reference's RMSE (no improvement from Persistence). A negative FCS tells lower effectiveness of the model than the reference.

VII. CASE STUDY AND EXPERIMENTAL RESULTS

District heat demand data of four building types in the Otaniemi area of Espoo, Finland have been used to train, validate, and test the proposed forecasting model. The FS and forecasting model is implemented for each building. The buildings are Building A (residential building), Building B (educational building, which contains classrooms and laboratories), Building C (office building), and Building D (mixed-use building, which contains computer laboratories, classrooms, offices, and health center). The buildings have a peak (in the three-year period: 2015–2017) aggregate district heat demand of 720, 2390, 60, and 280 kW, respectively. A two-year (2015–2016) hourly data are used for the FS, forecast model training, and validation tasks. While a one-year (2017) hourly data are employed for the forecast model testing.

The BGA-GPR-based FS is executed for each of the four buildings. The obtained FS results are given in Table III.

As shown in Table III, the number of features selected by the devised FS method is noticeably smaller than the size of the predictor space (number of candidate variables in Table I). That means there have been insignificant and repeated information by most of the features in the original predictor space. The BGA finally chooses the predictor subset, which holds the most significant and nonrepetitive features for each building as given in Table III. In addition, the FS results show that the values of the fitness measure (MSE) with the FS are considerably better (lower) than without FS. This confirms the importance of FS before fitting (training) forecasting models.

The fact that the FS results do not look much different across the buildings is due to the proximity of the buildings (the building are located in the same district of a city) where most of the external variables are the same and have similar impact on each building.

For the sake of consistency, the predictors chosen at least for one of the buildings are selected to constitute the input dataset for the 24h-ahead prediction of the heat demand of the buildings. Thus, the final chosen feature subset contains 15 variables: f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{12}, f_{14}, f_{16}, f_{18}, f_{19}, and f_{20}, which respectively represent the hour of the day, day of the week, month of the year, season of the year, holiday/weekend indicator, period of the day, air temperature, dew-point, relative humidity, air pressure, wind speed, solar radiation, previous 24 h average heat load, 24 h lagged heat load, and 168 h lagged heat load.

The forecasting model training, validation, and testing is done based on the hourly dataset of the selected feature subset.

In order to illustrate the forecast (test) results in this section, randomly chosen four weekdays and four weekends/holidays designating the weekdays and weekends in the four seasons of the test year (2017) are used. These days are summer weekday (Wednesday-July 26, 2017), summer weekend (Sunday-July 16, 2017), fall weekday (Thursday-Oct 12, 2017), fall weekend (Saturday-Oct 28, 2017), winter weekday (Monday-January 9, 2017), winter holiday (Sunday-January 1, 2017), spring weekday (Tuesday-April 18, 2017), and spring weekend (Saturday-April 8, 2017). Hence, specific days with better heat demand curves are not purposely taken. This will help to obtain an uneven forecasting accuracy over the test year that demonstrates the real heat demand profile in the buildings.

The prediction results are illustrated for the randomly selected test days with 1-h time resolution. The heat demand predictions by the devised integrated EMD-ICA-SVM model are shown in Figs. 8 to 9, for the weekdays and weekends/holidays, respectively. The forecast plots are shown for some of the pilot buildings, due to limitation of space. As shown in Figs. 8 and 9, the forecasts by the devised hybrid EMD-ICA-SVM model track the actual heat consumption trends with small gaps (errors).

Table IV provides the values of the evaluation criteria that are employed to calculate the error (or accuracy) of the proposed integrated EMD-ICA-SVM model for 24h-ahead forecasting of building district heat demand. The values are given for one of the pilot buildings.

As given in Table IV, the proposed forecasting model has achieved very accurate values (MAPE $< 10\%$ and FCS $> 50\%$). It has given accurate accuracy values for the other pilot buildings

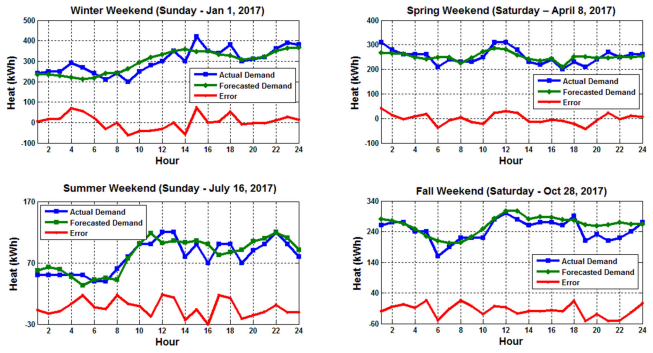

Fig. 8. Real versus forecasted district heat demand in weekends for Building A—residential building.

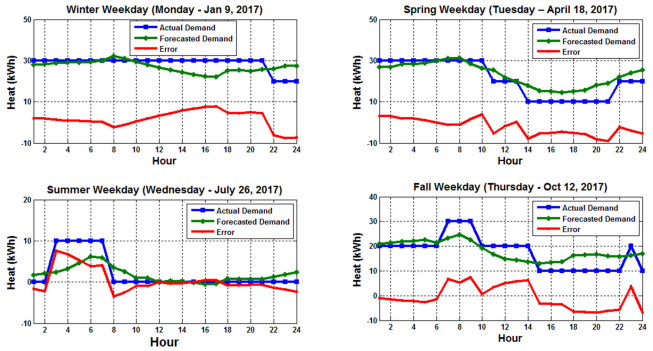

Fig. 9. Real versus forecasted district heat demand in weekdays for Building C—office building.

TABLE IV
FORECASTING ERROR ANALYSIS FOR BUILDING A—RESIDENTIAL BUILDING

Day Type		MAPE (%)	RMSE (kW)	NMAE (%)	Forecast Skill (%)
Winter	Weekday	3.14	18.03	1.91	62.41
	Holiday	4.00	19.82	2.11	86.74
Spring	Weekday	3.96	18.21	2.04	54.16
	Weekend	2.91	11.97	1.39	84.42
Summer	Weekday	6.42	7.22	0.85	12.94
	Weekend	7.78	8.27	0.99	39.55
Fall	Weekday	5.15	13.89	1.53	32.73
	Weekend	3.86	14.91	1.67	82.62
Average	Weekday	4.67	14.34	1.58	40.56
	Weekend	4.64	13.74	1.54	73.33
Total Average		4.65	14.04	1.56	56.94

as well. Specifically, the model has achieved an excellent performance (MAPE of 4.65%) for the residential building type and relatively lowest accuracy (MAPE of 5.88%) for the educational building type. The MAPEs obtained for the office and mixed-use buildings are 5.42% and 4.81%, respectively.

Figs. 10–13 present comprehensive performance comparisons between the proposed hybrid EMD-ICA-SVM-based forecasting model and the other nine models (Persistence, ANN, GA-ANN, ICA-ANN, SVM, GA-SVM, ICA-SVM, EMD-GA-ANN, and EMD-ICA-ANN), with respect to the MAPE metric. All the evaluated forecasting models have used the same input variables that are selected by the proposed BGA-GPR FS algorithm. That means, all the models are with FS. The comparisons are based on hourly values of the actual and forecasted heat demands for the complete testing year (2017). The monthly mean MAPE values are shown by the bar plots for the purpose of saving the plotting space.

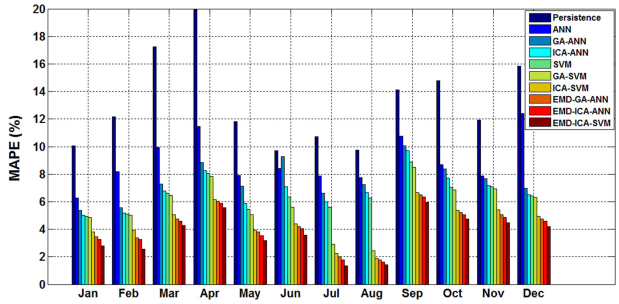

Fig. 10. Comparison of forecasting models w.r.t MAPE, based on the full one-year (2017) testing dataset for Building A—residential building.

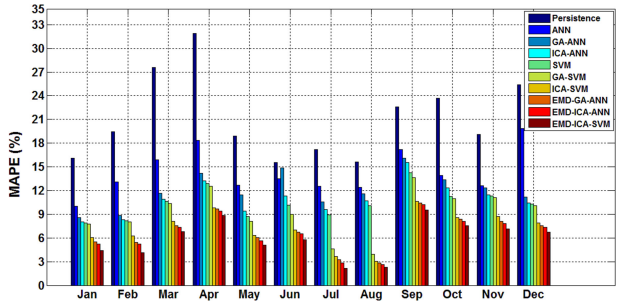

Fig. 11. Comparison of forecasting models w.r.t MAPE, based on the full one-year (2017) testing dataset for Building B—educational building.

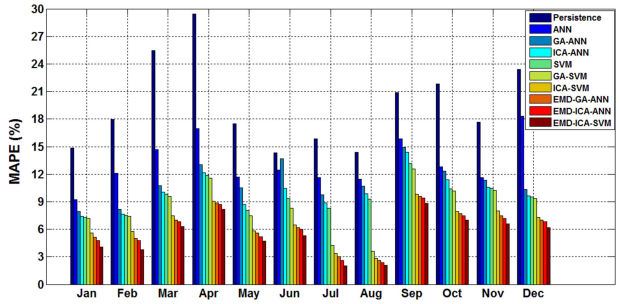

Fig. 12. Comparison of forecasting models w.r.t MAPE, based on the full one-year (2017) testing dataset for Building C—office building.

As seen in Figs. 10–13, the proposed model provides the lowest forecast error (MAPE). It outperforms all the other evaluated models. Quantitatively, the proposed model MAPE improvement over the other nine models is respectively 64.18%, 33.57%, 24.95%, 21.42%, 19.93%, 17.63%, 15.12%, 13.23%, and 11.25% for the residential building, 55.51%, 33.03%, 27.78%, 23.09%, 20.79%, 16.17%, 14.02%, 12.61%, and 11.10% for the

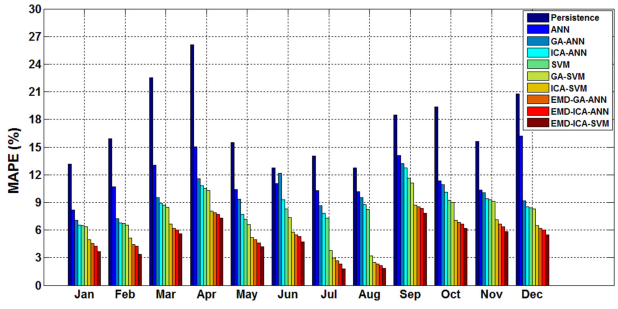

Fig. 13. Comparison of forecasting models w.r.t MAPE, based on the full one-year (2017) testing dataset for Building D—mixed-use building.

educational building, 63.59%, 42.13%, 27.31%, 12.06%, 6.35%, 5.14%, 4.32%, 3.24%, and 2.07% for the office building, and 74.33%, 56.08%, 43.37%, 22.89%, 18.64%, 17.23%, 14.41%, 11.56%, and 7.75% for the mixed-use building. The same learning data were employed for all models and each model was applied with its best values of parameters and configuration.

Moreover, the annual MAPEs obtained are almost the same as the MAPEs achieved for the randomly chosen individual testing days (shown in Table IV for Building A) in the respective seasons. This verifies the consistency of the proposed model performance throughout the year. In addition, the devised integrated EMD-ICA-SVM model still outperforms all the other evaluated models.

To summarize the discussion on the experimental results obtained in this article, the devised integrated EMD-ICA-SVM model achieves very accurate forecasts tested over one-year data of different building types. It has given very low forecast error and higher FCS, outperforming other nine evaluated models. Besides, the average computation time to generate the day-ahead hourly forecasts (excluding FS and training times) is about 5 s using MATLAB R2019a on a stand-alone research workstation with Intel Xeon W-2133 CPU @ 3.60GHz 3.60 GHz Processor and 32 GB RAM. Even the training time of the proposed forecast model is in the order of few minutes (less than 10). Thus, this may be interesting to run it on smaller building energy management hardware without the need to go for high performance computing or cloud platforms. Therefore, the devised model is both novel and considerably effective for a short-term (24h-ahead) building heat energy demand forecasting.

VIII. Conclusion

In this article, a novel and effective ML model (integrated EMD-ICA-SVM) was proposed and implemented for 24h-ahead forecasting of district heat demand of buildings using a predictor subset selected by a BGA-GPR-based FS methodology. The FS process selected a feature subset containing 15 variables to constitute the training input of the forecasting model. A two-year (2015–2016) hourly data were used for the FS, forecast model training, and validation processes. The performance of the proposed model was tested with a one-year (2017) hourly data. The model was implemented and tested for four building types. It had the ability to learn any time when new training dataset is available. The application of the proposed model for 24h-ahead building heat demand forecasting was novel and effectively successful. It achieved very accurate values (MAPE$<$10% and FCS $>$50%). It gave improved forecasting accuracy values for all pilot buildings. Specifically, it achieved an excellent performance (MAPE of 4.65%) for the residential building type and relatively lowest accuracy (MAPE of 5.88%) for the educational building type. In addition, the MAPEs obtained for the office and mixed-use buildings are 5.42% and 4.81%, respectively. The performance of the devised model has also been compared with the persistence model and other eight ML models (ANN, GA-ANN, ICA-ANN, SVM, GA-SVM, ICA-SVM, EMD-GA-ANN, and EMD-ICA-ANN). The devised model outperforms all the evaluated models. The MAPE improvement by proposed model over the other nine models (persistence, ANN, GA-ANN, ICA-ANN, SVM, GA-SVM, ICA-SVM, EMD-GA-ANN, and EMD-ICA-ANN) is, respectively, 64.18%, 33.57%, 24.95%, 21.42%, 19.93%, 17.63%, 15.12%, 13.23%, and 11.25% for the residential building, 55.51%, 33.03%, 27.78%, 23.09%, 20.79%, 16.17%, 14.02%, 12.61%, and 11.10% for the educational building, 63.59%, 42.13%, 27.31%, 12.06%, 6.35%, 5.14%, 4.32%, 3.24%, and 2.07% for the office building, and 74.33%, 56.08%, 43.37%, 22.89%, 18.64%, 17.23%, 14.41%, 11.56%, and 7.75% for the mixed-use building. Therefore, the presented illustrative experimental findings, numerical results, and performance comparisons verify the capability and suitability of the proposed integrated ML model for the short-term (24h-ahead) heat demand forecasting. The obtained forecast results can be used as input information for smart (adaptive), efficient, and optimal decisions in the energy management, market participation, and flexibility management of smart grids.

References

[1] Center for Clean Air Policy (CCAP), "Success stories in building energy efficiency," [Online]. Available: https://ccap.org/resource/success-stories-in-building-energy-efficiency/. (accessed on: Jan. 5, 2019).
[2] European Parliamentary Research Service, "Energy efficiency in buildings," [Online]. Available: https://epthinktank.eu/2016/07/08/energy-efficiency-in-buildings/. (accessed on: Jan. 7, 2019).
[3] Int. Energy Agencies, *Transition to Sustainable Buildings Paris*, IEA, 2013.
[4] P. Ferreira, A. Ruano, S. Silva, and E. Conceicao, "Neural networks based predictive control for thermal comfort and energy savings in public buildings," *Energy Build*, vol. 55, pp. 238–251, Aug. 2012.
[5] European Commission. "Directive 2010/31/EU of the european parliament and council on energy performance of buildings," Brussels, Belgium, 2010.
[6] K. Kato, M. Sakawa, and S. Ushiro, "Heat load prediction through recurrent neural network in district heating and cooling systems," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2008, pp. 1401–1406.
[7] M. Wang and Q. Tian, "Application of wavelet neural network on thermal load forecasting," *Int. J. Wireless Mobile Comput.*, vol. 6, no. 6, pp. 608–614, 2013.
[8] H. Gadd and S. Werner, "Daily heat load variations in Swedish district heating systems," *Appl. Energy*, vol. 106, pp. 47–55, Jun. 2013.
[9] N. Fumo, "A review on the basics of building energy estimation," *Renew. Sustain. Energy Rev.*, vol. 31, pp. 53–60, Mar. 2014.
[10] S. Idowu, S. Saguna, C. Åhlund, and O. Schelén, "Forecasting heat load for smart district heating systems: A machine learning approach," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Venice, Italy, Nov. 2014, pp. 554–559.
[11] L. Wu, G. Kaiser, D. Solomon, R. Winter, A. Boulanger, and R. Anderson, "Improving efficiency and reliability of building systems using machine learning and automated online evaluation," in *Proc. IEEE Long Island Syst., Appl. Technol. Conf.*, 2012, pp. 1–6.

[12] T. Catalina, V. Iordache, and B. Caracaleanu, "Multiple regression model for fast prediction of the heating energy demand," *Energy Build.*, vol. 57, pp. 302–312, Feb. 2013.
[13] M. Sakawa and S. Ushiro, "Cooling load prediction in a district heating and cooling system through simplified robust filter and multi-layered neural network," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 1999, pp. 995–1000.
[14] R. Petrichenko, K. Baltputnis. A. Sauhats, and D. Sobolevsky, "District heating demand short-term forecasting," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur.*, Milan, Italy, 2017, doi: 10.1109/EEEIC.2017.7977633.
[15] A. Kusiak, M. Li, and Z. Zhang, "A data-driven approach for steam load prediction in buildings," *Appl. Energy*, vol. 87, no. 3, pp. 925–933, Mar. 2010.
[16] A.T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Day-ahead prediction of building district heat demand for smart energy management and automation in decentralized energy systems," in *Proc. IEEE 17th Int. Conf. Ind. Informat.*, Helsinki-Espoo, Finland, Jul. 2019, pp. 1694–1699.
[17] Finnish Meteorological Institute (FMI), [Online]. Available: https://FMI.fi/. (accessed on: Nov. 3, 2018).
[18] Timeanddate.com. Calendar for Year 2015, 2016 and 2017. [Online]. Available: https://www.timeanddate.com/calendar/. (accessed on: Nov. 6, 2018).
[19] A.T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems," *IEEE Access*, vol. 7, pp. 91463–91475, 2019.
[20] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Adaptive predictor subset selection strategy for enhanced forecasting of distributed PV power generation," *IEEE Access*, vol. 7, pp. 90652–90665, 2019.
[21] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, Aug. 2018.
[22] F. Jäkel, B. Schölkopf, and F. A. Wichmann, "A tutorial on kernel methods for categorization," *J. Math. Psychol.*, vol. 51, no. 6, pp. 343–358, Dec. 2007.
[23] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London*, vol. 454, pp. 903–995, 1998.
[24] Z. K. Peng, P. W. Tse, and F. L. Chu, "A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mech. Syst. Signal Process.*, vol. 19, pp. 974–988, 2005.
[25] D. Zheng, M. Shi, Y. Wang, A. T. Eseye, and J. Zhang, "Day-Ahead wind power forecasting using a two-stage hybrid modeling approach based on SCADA and meteorological information, and evaluating the impact of input-data dependency on forecasting accuracy," *Energies*, vol. 10, 2017, Art. no. 1988.
[26] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. J. Millar, "Short-term forecasting of electricity consumption in buildings for efficient and optimal distributed energy management," in *Proc. IEEE Int. Conf. Ind. Informat.*, Helsinki-Espoo, Finland, Jul. 2019, pp. 1103–1110.
[27] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," in *Proc. IEEE Congr. Evol. Comput.*, 2007, pp. 4661–4667.
[28] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
[29] S. R. Gunn, "Support vector machines for classification and regression," Faculty Eng., Sci. Math., Univ. Southampton, Southampton, U.K., Tech. Rep. 256459, 1998.
[30] K. R. Miller and V. Vapnik, *Using Support Vector Machine for Time Series Prediction*, Cambridge, MA, USA: MIT, pp. 243–253, 1999.
[31] C. Coimbra, J. Kleissl, and R. Marquez, "Overview of solar forecasting methods and a metric for accuracy evaluation," in *Solar Resource Assessment and Forecasting*, J. Kleissl, , Ed. Waltham, MA, USA: Elsevier, 2013.

Abinet Tesfaye Eseye received the B.Sc. degree in electrical engineering from Hawassa University, Awasa, Ethiopia, in 2010, the M.Sc. degree in electrical power systems engineering from Bahir Dar University, Bahir Dar, Ethiopia, in 2012, and the Ph.D. degree in electric power system and its automation from North China Electric Power University, Beijing, China, in 2018.

He was an Assistant Lecturer with the School of Computing and Electrical Engineering, Bahir Dar University, Bahir Dar, Ethiopia from September 2011 to July 2012. He was also a Lecturer and the Head of the Department of Electrical and Computer Engineering, Mettu University, Metu, Ethiopia, from September 2012 to August 2014. He worked in the position of Renewable Energy and Smart Microgrid R&D Engineer with the Goldwind Science and Technology Company, Ltd., Beijing, China, from January 2015 to July 2018. He is currently working with the Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland, as a Postdoctoral Researcher. His research experiences and interests include electric power system, distributed generation, hybrid energy system, microgrid, renewable energy, energy internet, demand response, energy economics and market, techno-economic and feasibility analysis, IoT, optimization, computational intelligence, forecasting, state estimation, artificial intelligence, and machine/deep learning for power/energy systems.

Matti Lehtonen received the M.S. and Licentiate degrees in electrical engineering from the Aalto University School of Science and Technology (formerly Helsinki University of Technology), Espoo, Finland, in 1984 and 1989, respectively, and the D.Sc. degree from the Tampere University of Technology, Tampere, Finland, in 1992.

Since 1987, he has been with VTT Energy, Espoo, and since 1999, he has been with the School of Electrical Engineering, Aalto University, where he is a Professor of IT applications in power systems. His main activities include earth fault problems, harmonic related issues, and applications of information technology in distribution automation and distribution energy management.