Aalto University

Eseye, Abinet Tesfaye; Lehtonen, Matti; Tukia, Toni; Uimonen, Semen; Millar, R. John

Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems

Please cite the original version:
Eseye, A. T., Lehtonen, M., Tukia, T., Uimonen, S., & Millar, R. J. (2019). Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems. *IEEE Access*, *7*, 91463-91475. Article 8744520. https://doi.org/10.1109/ACCESS.2019.2924685

Machine Learning Based Integrated Feature Selection Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

ABINET TESFAYE ESEYE, MATTI LEHTONEN, TONI TUKIA, SEMEN UIMONEN, AND R. JOHN MILLAR

Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland

Corresponding author: Abinet Tesfaye Eseye (abinet.eseye@aalto.fi)

ABSTRACT Improved performance electricity demand forecast can provide decentralized energy system operators, aggregators, managers, and other stakeholders with essential information for energy resource scheduling, demand response management, and energy market participation. Most previous methodologies have focused on predicting the aggregate amount of electricity demand at national or regional scale and disregarded the electricity demand for small-scale decentralized energy systems (buildings, energy communities, microgrids, local energy internets, etc.), which are emerging in the smart grid context. Furthermore, few research groups have performed attribute selection before training predictive models. This paper proposes a machine learning (ML)-based integrated feature selection approach to obtain the most relevant and nonredundant predictors for accurate short-term electricity demand forecasting in distributed energy systems. In the proposed approach, one of the ML tools – binary genetic algorithm (BGA) is applied for the feature selection process and Gaussian process regression (GPR) is used for measuring the fitness score of the features. In order to validate the effectiveness of the proposed approach, it is applied to various building energy systems located in the Otaniemi area of Espoo, Finland. The findings are compared with those achieved by other feature selection techniques. The proposed approach enhances the quality and efficiency of the predictor selection, with minimal chosen predictors to achieve improved prediction accuracy. It outperforms the other evaluated feature selection methods. Besides, a feedforward artificial neural network (FFANN) model is implemented to evaluate the forecast performance of the selected predictor subset. The model is trained using two-year hourly dataset and tested with another one-year hourly dataset. The obtained results verify that the FFANN forecast model based on the BGA-GPR FS selected training feature subset has achieved an annual MAPE of 1.96%, which is a very acceptable and promising value for electricity demand forecasting in small-scale decentralized energy systems.

INDEX TERMS Binary genetic algorithm, decentralized energy system, electricity demand forecasting, feature selection, feedforward artificial neural network, fitness evaluation measure, Gaussian process regression, machine learning, smart grid.

I. INTRODUCTION

Decentralized energy system operators, aggregators, suppliers, managers or other stakeholders are challenged by several confronts, varying from inadequate electricity supply to increasing consumption. The electricity demand curves of decentralized energy systems (such as buildings, energy communities, microgrids, virtual power plants, local energy internets, etc.) are different from typical electricity demand curves that represent nation- or region-wide electricity consumptions. This makes the conventional techniques (developed for national or regional electricity demand forecasts) inappropriate for their straightforward application in decentralized energy systems due to two clear reasons.

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

IEEE *Access*

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

In decentralized energy systems, not only the total electricity demand level is many times less than the regional or national demand levels, but also the electricity demand profile manifests more fluctuation and does not generally follow the same profile.

Therefore, the recent deployment of decentralized energy systems calls for appropriate and applicable feature selection (FS) tools and forecasting models for economic and efficient consumption modeling.

Feature selection is a procedure of picking a subset of most important features (attributes, variables, or predictors) for use in predictive model development.

Basic knowledge about statistical models may help to understand an effective feature selection method is important for the final prediction performance of prediction models. However, the effective feature selection approach should be properly designed, implemented and tested for a specific application in question. In the present age of 'Big Data', datasets are full of information with enormous data gathered from millions of Internet-of-Things (IoT) apparatuses and sensors. This makes the data high dimensional and it has become very common to observe datasets with hundreds (even thousands) of variables.

The same is true in the power industry sector, since the concept of smart grids has been developed and being implemented based on the idea of IoT and complex interaction of data among various stakeholders.

When data is presented with very high dimensionality, prediction models generally choke since:

1. Training, validation and testing times increase exponentially with number of variables.
2. Prediction models will have increasing risk of overfitting with increasing number of predictors.
3. Prediction accuracy will face cumulative threat of reducing with increasing number of features.

Therefore, feature selection is a very important element in Data Scientists' workflow. Different research groups have performed various feature selection methods for various applications and scenarios. However, very few of them have coupled and investigated feature selection tools and forecasting models. Moreover, there is no standard and universally agreed feature selection method so far. The R& D for finding the most effective feature selection tools is still ongoing by various independent research groups and institutions. Effective and adaptive feature selection methods shall be designed, implemented and tested for desired applications and scenarios all time as new data sources, policies and algorithms are being emerged all time. These are the main reasons, which have motivated the development and implementation of the FS work in this paper.

Feature selection strategies are important for the following main advantages:

- Reduce computation time
- Decrease data storage requirements
- Simplify models, making them user-friendly
- Improve data understandability and interpretability
- Evade curse of dimensionality
- Enhance generalization, reducing overfitting.

The core argument when applying an FS method is that the original dataset holds some variables which are either duplicated or not important, and can therefore be eliminated without inducing ample damage of information. It has been proved by several researches that, redundant and irrelevant features reduce the accuracy and generalization capability of predictive models.

That is why, nowadays, FS studies has become very popular in AI, Machine Leaning (ML), Deep Learning (DL), and Statistics.

It is demonstrated, by various relevant studies, that the total energy cost (energy production, operation and purchasing costs) can be reduced significantly by applying demand response concept in in small-scale decentralized energy systems. However, techniques for electricity demand prediction in decentralized energy systems have not been properly investigated. Most applicable methods are limited for predicting the amount of electricity demand at large-scale (nation- or region-wide) and disregard the specific electricity demand for smaller entities such as buildings, energy communities, microgrids, virtual power plants, local energy internets, etc., which comprise equivalent significance for energy system optimization, sustainability and efficiency.

Thus, the goal of this paper is to propose and implement a feature selection approach for modeling and forecasting the fluctuating electricity demand in decentralized energy systems in general and buildings in particular. Results will assist distributed energy system stakeholders to efficiently use limited energy resources and regulate dispatchable generation and flexible demand levels. Prediction accuracy has mainly been the indispensable target of forecasting studies. It is soundly revealed in [1] and [2] that the accuracy of prediction models not only relies on the models' configurations and associated learning methods but also on the predictor domain, which is established via the initial predictor space and FS techniques. FS is mostly applied in ML implementations as one of the preprocessing steps where a predictor subset (independent attributes) is found by removing predictors with lower or irrelevant information and highly redundant [3]. However, very few forecasting techniques perform FS before training the prediction models.

Several heuristic optimization approaches have been effectively implemented as searching technique for FS. For instance, these methods contain Particle Swarm Optimization (PSO) [4], Ant Colony Optimization (ACO) [5] and Genetic Algorithm (GA) [6].

GA has gained extensive consideration due to its operability and robust searching ability. GA is one of the artificial intelligent (AI) probabilistic searching algorithms, and has been broadly implemented for several optimization problems [7]. It was inspired by the survival of the fittest principle of the Charles Darwin evolution theory and genetics. The GA search starts from a randomly chosen set of individuals called initial population. It then iterates to find the best individual

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

IEEE *Access*

(solution) through its three main consecutive operators – selection, crossover and mutation. GA uses a performance index called fitness function to calculate the fitness of the individuals over iterations. BGA is a special version of GA which operates by first representing the given feature space (chromosomes or candidate solutions) in binary bit-strings. This makes the BGA well suited for FS problems than the conventional GA.

This paper proposes a machine learning based hybrid feature selection method to obtain the most relevant and nonredundant features for improved short-term forecasting of electricity demand in decentralized energy systems. In the proposed method, the Binary Genetic Algorithm (BGA) is applied for the feature selection process and Gaussian Process Regression (GPR) is used for measuring the fitness score of the features.

To the best of our understanding, there exist very few research works that have performed feature selection work before fitting or training forecasting models. Moreover, as far as we have investigated, the BGA-GPR based hybrid machine learning approach has never been applied for feature selection problem in the domain of electricity demand forecasting.

In this study, the residual of the GPR model is chosen as the fitness evaluation measure. GPR is a powerful algorithm for regression. It is chosen due to its higher capability of fitting nonlinear input-output relationships based on probabilistic distributions over functions. It has few parameters to tune and easily to implement. It can provide a consistent estimation of its uncertainty. GPR can directly apprehend the model (input-output or feature-target relationship) uncertainty. For instance, it directly provides a distribution for the feature selection fitness measure (error) value, rather than just one value as the estimate. This uncertainty is not straight apprehended by most of the other ML or AI tools. Moreover, GPR is able to add a previous knowledge and specification about the behavior of the input-output (feature-target) relationship by using different kernel functions.

Generally, the paper contributions can be regarded as (1) modeling, parameterization and implementation of the BGA and GPR algorithms to suit the feature selection problem in question, and (2) establishment of seamless combination of the two algorithms to work in unison for solving the feature selection problem.

Specifically, this paper has generally the following main contributions:

- Investigate and suggest the relevance of an effective FS approach for improved-performance (accurate) electricity demand forecasting;
- Provide effective and efficient hybrid AI-based FS approach for electricity demand forecasting models;
- Improve electricity demand prediction accuracy through the application of FS before fitting prediction models.

The remaining sections of the paper are outlined below. Section II describes the previous relevant works on FS. Section III describes the dataset and states the feature selection problem. Section IV presents the brief working principle

of the binary GA. Similarly, the theory and mathematical modeling of the GPR model used for the fitness measure in the binary GA is described in Section V. Section VI presents the proposed Binary BGA-GPR based FS strategy. The experimental results and validations are presented in Section VII. Conclusions are finally drawn in Section VIII.

II. RELATED WORKS

FS methods are classified as filter, wrapper and embedded techniques [1]. Filter techniques do not depend on any prediction model and they sort features depending on statistical characteristics. They utilize a correlation score to grade a feature subset. Filter technique based FS methods are generally fast. The Filter FS approach contains correlation-based [8], mutual information-based [9], and principal component analysis-based methods [10]. Filters are generally require less computation time than other FS techniques, but they generate a feature set which is not fitted to a particular forecast model.

Filter techniques are extensively applied in big data analysis due to their computational efficiency. Wrapper techniques evaluate predictor subsets based on their worth to a specific forecaster or classifier. Wrapper techniques assume the FS as a searching problem that prepares various mixes of predictors, assessed, and contrasted with other mixes. Common heuristic AI-based optimization methods stated in Section I are used to monitor the searching procedure. Compared to filter techniques, wrapper techniques reveal improved performance since various predictor sets are assessed by a predictive model or fitting method in every iteration [11]. Embedded techniques merge the feature selection process into the prediction model training. For instance, regularization approaches in model training [1] is one example of embedded type FS method. The LASSO model which normalizes the parameters of linear models with L_1 penalties to reduce uncorrelated coefficients to zero can also be one example of embedded method.

Table 1 presents recently published FS strategies in the demand prediction scope for power system and energy applications.

TABLE 1. Summary of FS strategies.

Application	FS Type	References
Power systems	Filter	[8], [12], [13]
	Wrapper	[14], [15]
Energy systems	Filter	[16]
	Wrapper	[17]

From the wrapper techniques outlined in Table 1, the GA-based feature selection shows best performance in removing duplicated features. Reference [15] devised an integrated strategy to predict electricity consumption; in this work, a conventional GA is used to obtain the best predictor subset hybridized with an adaptive neuro-fuzzy inference system (ANFIS). Reference [18] devised a combination of GA and ACO for FS in the electric load demand prediction;

ANN is employed as a forecaster model to measure the fitness score of the feature subsets. Reference [19] devised a GA-support vector regression (SVR) configuration to predict accommodation-booking demands in hotels, where the GA is employed for both feature selection and SVR parameter optimization. Reference [20] implemented a GA-based feature selection to the demand prediction problems in retail industries of China. Reference [21] proposed a modified GA for FS in the demand prediction problem in the health sector specifically in outpatient department (OPD); deep neural network is employed as a predictive model to evaluate the fitness score of the feature subsets. In addition to demand prediction, GA-based feature selection is often used for knowledge exploration in other important sectors, such as financial market analysis [22], stock price forecasting [23] and financial distress forecasting [24]. Reference [25] proposed a binary GA-based FS for classification problem. It used the k-nearest neighbors (kNN) algorithm for the GA fitness evaluation measure.

Designing a suitable fitness evaluation measure is very important in FS approaches. The fitnesses evaluation measure is used as a performance index for evaluating the suitability of candidate features. Predictors are ranked and selected according to their evaluated values of the fitness function or measure.

The combination of predictors that results the best value of the fitness measure is chosen at the end of the FS algorithm running. This paper uses the residual (error) of the Gaussian Process Regression (GPR) model as the fitness function of the BGA. The GPR model is kernel-based probabilistic distribution over functions. It is chosen in this paper due to its higher capability of fitting nonlinear input-output relationships based on assumptions of probabilistic distribution of given input-output data or function. Besides, it has few parameters to tune and easily to implement.

Following a comprehensive assessment of the above-mentioned genetic algorithm based FS techniques, we discover that a conventional genetic algorithm with the usual framework (conventional GA configuration) is used in most researches [15]. For instance, the initial population (initial chromosome set) is arbitrarily created where the population variety cannot be guaranteed and the occurrence of duplicated predictors may influence the quality of the search procedure. Moreover, the conventional GA works with the continuous features themselves to minimize the desired fitness function (FS evaluation measure). This reduces the efficiency of the algorithm and causes computational complexity and increased total computation time.

Assuming an intelligent heuristic algorithm should be the best option to determine the search target; a research problem exists and shall be addressed by replacing the conventional GA with the BGA and hybridizing it with robust fitness evaluation measure (GPR in this paper). BGA first represents the features as an encoded binary string and works with the binary strings to minimize the GPR-based evaluation measure to obtain the most relevant and nonredundant predictor

subset at the end. BGA is more efficient and stable than the conventional GA. It also reduces computational complexity and execution time compared to the conventional GA.

III. DATASET AND FEATURE SELECTION PROBLEM

The original feature set is constructed through basic assessment of the characteristics of the electricity consumption in decentralized energy systems such as buildings and its association with historical (prior) consumption and external agents. The external agents are electricity market price, seasonality (minute/hour, month and season), weather factors, people's interaction (occupancy). The availability of the data sources for these external agents affecting electricity consumption is also another major factor to construct the original feature space.

The complete feature set for electricity demand predictive model in this FS work consists of lagged electricity demand, seasonal (or calendar) parameters, weather parameters, occupancy, and economic factor (electricity price). Table 2 presents the original feature space or initial dataset for the FS work in this paper. The variables f_i, i = 1, 2, . . . , 24, in Table 2 designate the original predictor dataset (feature space) required for the FS work in this paper. Therefore, the feature domain of the FS is an R^{mxn} matrix, where m = 8760 is the number of samples, which is a one-year (2015) hourly observation of the variables and n = 24 is the size of the feature space (original dataset).

TABLE 2. Feature space of the FS problem.

Feature Index	Feature (f_i)	Unit/Scale	Data Category
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	Period of the day:	1-3	Occupancy (number of people)
6	Holiday/Weekend indicator1	0-1	
7	Holiday/weekend indicator2	0-2	
8	Ambient air temp.	°C	Weather
9	Dew-point temp.	°C	
10	Relative humidity	%	
11	Precipitation	mm/h	
12	Snow depth	cm	
13	Ambient air pressure	hPa	
14	Horizontal visibility	m	
15	Wind direction	deg	
16	Wind speed	m/s	
17	Gust speed	m/s	
18	Cloud cover	0-8	
19	Global solar radiation	Watt/m²	
20	Sunshine duration	s	
21	Electricity price	c£/kWh	Economy
22	Previous 24h average electricity demand	kW	Electricity demand
23	24h lagged electricity demand	kW	
24	168h lagged electricity demand	kW	

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

IEEE *Access*

In this paper, the following optimization problem is solved to obtain the best (relevant and nonredundant) predictor subset from the original dataset given in Table 2.

FS Problem: Given that:

$$f_r \epsilon Z^+, \quad 1 \leq f_r \leq 24, \text{ and } \beta \epsilon R^+ \ni 0 \leq \beta \leq 100 \quad (1)$$

where, f_r is the number of predictors in the lower-dimension (reduced) predictor subset and β is the percentage forecast error. Find a feature subset of f_i from Table 2 such that the objective β and f_r are reduced.

IV. BINARY GENETIC ALGORITHM (BGA)

GA is a population-based heuristic type optimization method that is inspired by the survival of the fittest principle of the Charles Darwin theory of evolution and genetics [26]. The GA operating mechanism involves iterative steps processing a set of chromosomes (candidate solutions) to generate a new population (offsprings) via genetic operators - *selection, crossover and mutation.*

The fitnesses of the nominee solutions (chromosomes) are calculated employing a function generally called objective or fitness function. That means, the objective function provides scores (numeric values) which are employed for grading the existing solutions in the population. BGA is an extended version of the standard GA.

The BGA first represents the candidate solutions as encoded binary strings (binary search space) and works with the binary strings to minimize or maximize the fitness function. BGA is more efficient and stable. It also reduces computational complexity and execution time. Figure 1 shows the flowchart of BGA.
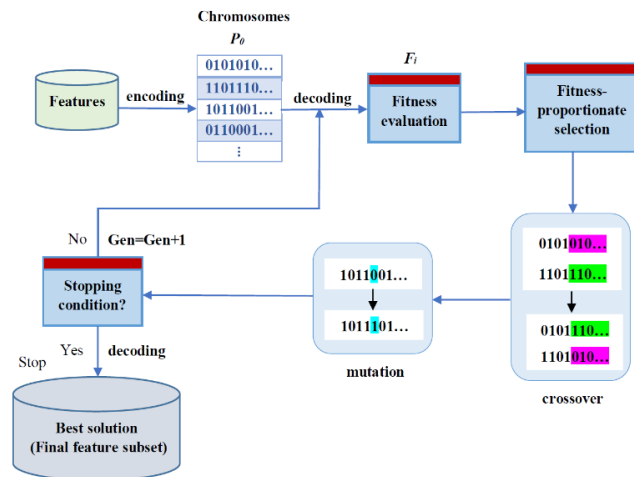

FIGURE 1. Flowchart of BGA for FS.

V. GAUSSIAN PROCESS REGRESSION (GPR)

GPR models nonlinear relationship between input(s) and target(s) based on probabilistic distributions over functions. The Gaussian process (GP) describes the distribution over functions based on the assumption that samples of targets

obtained at any two or more instants in a function trail a joint (multi-variate) Gaussian distribution (GD).

In explicit saying, a GP is described as a group of arbitrary attributes. Any fixed quantity of the attributes can form a joint (multi-variate) GD [27].

In GPR, the target *y* of the function f at input-attribute *x* is defined as:

$$y = f(x) + \varepsilon \quad (2)$$

where, $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$ – normal distribution with zero (0) mean and σ standard deviation. This is the same with the statement stated in linear regression where it is assumed that a sample contains an independent "function" part f(x) and "noise" part ε. In GPR, on the hand, we consider that the function part is also an arbitrary variable that trails a given distribution. The distribution emulates once doubt on the function. The stochasticity on f can be decreased by measuring the function target at various instants. The noise part ε emulates the intrinsic uncertainty in the samples that usually exist regardless of the number of samples made [27].

In GPR, the function f(x) is assumed distributed as a GP:

$$f(x) = GP\left(m(x), k\left(x, x'\right)\right) \quad (3)$$

GP is described by mean and covariance functions. The mean function m(x) emulates the anticipated functional value at input-attribute *x*:

$$m(x) = E[f(x)] \quad (4)$$

That means m(x) is the mean of all the functions in the distribution estimated at x. The initial expected value is usually fixed at zero (i.e., m(x) = 0) to evade costly future calculations and only make decision through the covariance function. Practically, fixing the initial value to zero is generally done by deducting the (initial) mean from each sample. The covariance function k(x, $x\prime$) simulates the relationship between the functional values at various input instants x and $x\prime$:

$$k\left(x, x'\right) = E\left[\left(f(x) - m(x)\right)\left(f\left(x'\right) - m\left(x'\right)\right)\right] \quad (5)$$

The function k is known as the kernel of the GP [28].

A suitable kernel should be selected considering the facts for instance evenness and likely patterns anticipated in the dataset.

A practical consideration is generally to assume the correlation of two instants declines with the distance between the instants. That is, the nearby instants are likely to act more in the same way than instants that are far away from each other. The GPR model in this study employs the squared exponential kernel function defined below:

$$k\left(x, x'|\theta\right) = \sigma_f^2 exp\left(-\frac{1}{2}\sum_{r=1}^{d}\frac{\left(x_r - x'_r\right)^2}{\sigma_r^2}\right) \quad (6)$$

where, σ_r denotes the length scale for forecaster r, r = 1, 2, ..., d and σ_f is the standard deviation of the

data or signal. The unconstrained parametrization θ is defined as:

$$\theta_r = log\sigma_r, \quad for \ r = 1, 2, \ldots, d \tag{7}$$

$$\theta_{d+1} = log\sigma_f \tag{8}$$

The parameters σ_r and σ_f can be regulated to rise or decrease the initial correlation between instants and hence the inconsistency of the function obtained. After the mean and kernel functions are determined, the GP can be used to obtain previous function values and future function values based on the priori samples.

VI. PROPOSED BGA-GPR BASED FEATURE SELECTION

As shown by the flowchart in Figure 2, there are five key sub-operations in BGA, namely - chromosome encoding, objective value calculation, selection methods, genetic operators and stopping condition. The BGA works on binary search domain (chromosome bitstrings). The GA operates the finite binary chromosome set based on the survival of the

FIGURE 2. Flowchart of BGA-GPR based FS.

fittest principle of the evolution theory. A starting population is generated and assessed using an objective function. For binary chromosome employed in this paper, a gene value of '1' indicates that the specific feature pointed by the place of the '1' is chosen. Else, (if '0'), the feature is not chosen for the fitness evaluation.

Employing the place pointer of the variables pointed by the '1s', the individuals are then ordered and according to the orders, the upper k fittest offsprings (Elitism of size k) are chosen to persist with the succeeding generation. Once the selected offsprings are moved directly to the succeeding generation, the other offsprings in the present solution space are permitted to genetically move via the crossover and mutation operators to create crossover and mutation offsprings respectively [26]. The three offsprings namely selection, crossover and mutation then establish the new solution space (new generation). The crossover operator is a fusion of two chromosomes to create crossover offsprings. While the mutation operator is employed for genetic disorder (diversity) of the genes in the chromosomes by tossing the bits based on the mutation likelihood. Following the procedures outlined in Figure 2, the detail operating mechanisms of the proposed GA-GPR FS are described in the following subsections.

A. INITIAL POPULATION

The GA starting solution space, used in this work, is a matrix of size $p \times q$, where p is the number of chromosomes and q is the chromosome length (called Genomelength). p equals the population size and q equals the amount of bits or genes in each individual. It is recommended to let the number of chromosomes equals at least chromosomes length such that the chromosomes in every population encompass the search domain [29].

B. FITNESS EVALUATION

For the BGA to choose the predictor subset, an objective function (BGA driver) should be specified to calculate the discriminative power of each predictor subset. The fitness of each chromosome in the population is assessed employing GPR-based fitness function. In this paper, the fitness of the various subsets of features is evaluated using the MSE (mean squared error) of the GPR model predictive residuals. The GPR model $f(x)$ is fitted for every feature subset. Hence, the MSE of the training target and the GPR model estimate evaluated for each feature subset in the feature search space defined in Table 2 is used as the fitness evaluation measure, and it is defines as follows.

$$fit = \frac{1}{n}\sum_{i=1}^{n}(T_i - f_i)^2 \tag{9}$$

where, T is a vector of training target (electricity demand) and n is amount of training samples or observations.

The aim of the BGA is to minimize the fitness function (MSE) defined in equation (9) by choosing a subset of input features having the best fitness over subsequent iterations. In each chromosome a gene value of '1' shows the specific
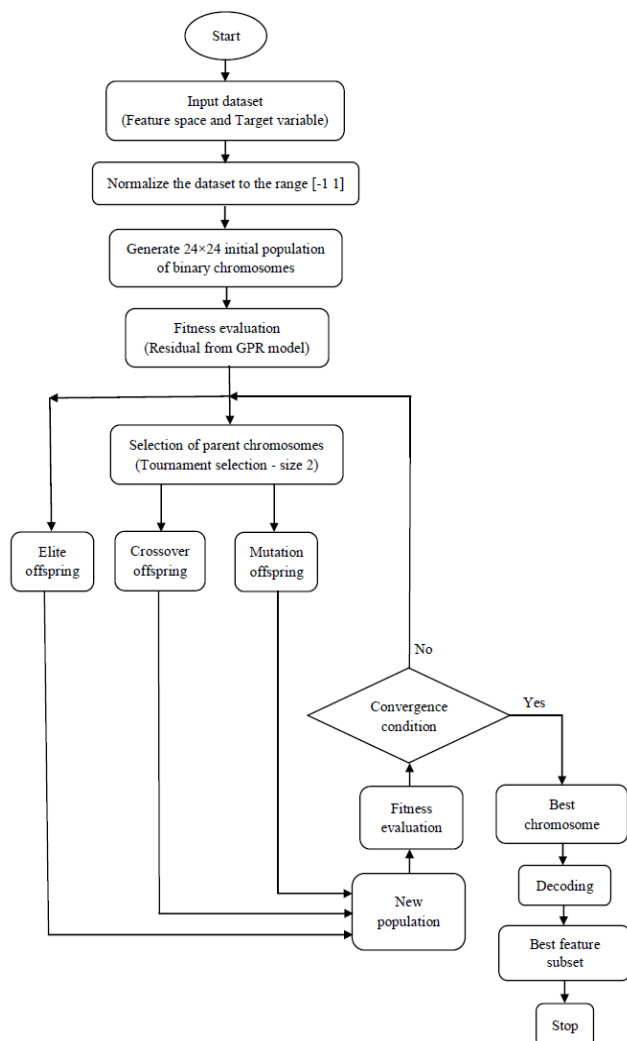
A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

IEEE*Access*

variable pointed by the place of '1' is chosen. If it is '0', the predictor is not chosen for assessment of the chromosome in question. The chromosomes representing the predictors are encoded as bitstrings. As the BGA runs, the individual chromosomes (feature subsets) in the present population are assessed, and their fitnesses are graded based on the GPR model residual or error. Chromosomes with smaller fitness (smaller residual or error) have greater probability of persisting with the next population or mating-pool.

Each iteration of the BGA running guarantees that the BGA decrease the error level and elites the chromosome with the lowest (best) objective function value. The individual chromosome corresponding to the least error level of the fitness evaluation contains the desired most relevant features.

C. REPRODUCTION

Table 3 presents the parameters of the BGA used in this paper. From Table 3, the chromosome length equals 24 as there are an overall number of 24 predictors nominated for the FS works in this paper. The extreme number of iterations (generations) is set to 100 to evade the BGA been stuck by local optimum. Following the fitness evaluation, a new population is produced for the next generation through elitism, crossover and mutation.

TABLE 3. Parameters of BGA.

Parameter	Value
Population size	24
Genomelength	24
Population type	Bitstring
Fitness function	GPR model residual or error
Number of generations	100
Stalled Generation Limit	50
Selection mechanism	Tournament selection
Tournament size	2
Mutation function	Uniform mutation
Mutation rate	0.1
Crossover function	arithmetic crossover (logical XOR)
Crossover fraction	0.8
Elite count	2

In BGA, three kinds of sequential offsprings are formed to create the new population [29]. They are:

1) *Elite offspring*: A selection mechanism should be predetermined in BGA to ensure the population is continuously getting better throughout all the fitness scores or iterations. The selection method assists the BGA to disregard worst individuals and keep only the best chromosomes. There exist several selection methods for BGA, however the Tournament Selection Mechanism (with size 2) is used in this study because of its ease-of-use, swiftness and efficiency [25], [30]. Besides, the tournament selection method imposes better selection burdens on the BGA that results in better convergence rate and ensures the bad candidate solutions are not moved into the succeeding generation. In size 2 tournament selection, two individuals are

chosen out of the solution space following the withdrawal of the elite offsprings and the best of the two individuals, (based on objective function score), is chosen. Tournament selection is carried out repeatedly till the new population is fully populated. Elite offsprings are moved automatically to the following generation, as they are the highest fitted values. However, the number of elite offsprings is limited by the population size. It is generally less than the population size. With size proper setting of the number of elite offsprings (elite count); the BGA chooses the upper elite count best individuals and move them directly into the succeeding generation. The amount of elitism used, as given in Table 3, is two. Hence, the upper 2 offsprings with the best fitness scores are directly taken into the following generation. Therefore, the quantity of the elite offsprings (elite count) = O_1 = 2. That is there are 22 (i.e. 24 - O_1) chromosomes in the population except the elite offsprings. From the rest 22 individuals, crossover and mutation offsprings are then generated.

2) *Crossover offspring*: The BGA crossover operator genetically fuses two chromosomes (parents) to create offsprings for the following generation. Chromosomes from two parents are required to perform the crossover function. These two parents are obtained from tournament selection. The crossover function used in this paper is arithmetic type, which applies a logical XOR operation on the chromosomes of the two parents as they are represented in binary form. The portion of the following generation, excluding the elite offsprings, which are created by the crossover operator is known as crossover offspring. The crossover fraction, which refers the ratio of the number of the crossover offsprings (produced by the crossover operator) to the total number of offsprings other than the elites, used in this paper is 0.8. There is no mutation offspring in the BGA when the crossover ratio sets to one. With the crossover ratio of 0.8, the number of the crossover offsprings is O_2 = round (22*0.8) = 18.

3) *Mutation offspring:* Mutation is a genetic disorder of chromosomes in the population. The BGA implemented in this study used uniform mutation. Using uniform mutation, the BGA creates a set of uniformly distributed random numbers whose size equals the length of the chromosomes. The random number values are related to the index of the bits in the chromosomes. The chromosomes are checked clockwise and for every corresponding bit, the random number is crosschecked with the mutation rate. If the random number at a given index is smaller than the mutation probability, then the gene (bit) at that index is tossed. Else, the gene is remains untossed. This continues from the left most bit to the right most bit of each individual of the mutation offspring. Mutated offsprings are very important for the BGA to have a genetic diversity in the chromosomes, which helps the BGA not to converge to

IEEE *Access*

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

local (suboptimal) solutions due to much similarity of genes in the population.

The quantity of mutation offsprings is $O_3 = 24 - O_1 - O_2 = 24 - 2 - 18 = 4$. This verifies $O_1 + O_2 + O_3 = 24$.

D. CONVERGENCE CONDITION

The BGA terminates when it converges at the desired optimal solution. The optimal solution corresponds to the desired feature subset for the FS problem in question. The termination condition where the BGA ends running is known as convergence or stopping condition. The two convergence conditions used in this paper are the following:

1) Maximum number of generations or iterations
2) Stalled generation limit

The values used for these convergence conditions are given in Table 3.

E. FINAL FEATURE SUBSET

After the BGA attains convergence, the chromosome that resulted in the best fitness score is chosen and decoded as the final feature subset, shown in Figure 3.
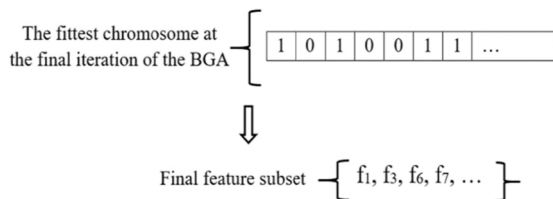

FIGURE 3. Final feature subset decoding.

VII. EXPERIMENTAL RESULTS AND VALIDATION

In this section, the case study for the proposed FS work and the results obtained are discussed. Comparative validation, evaluation of FS results for improved forecasting and quantitative relevance analysis of the FS results are also presented in this section.

A. CASE STUDY

In this paper, the hybrid BGA-GPR based FS approach is developed and applied to four electricity demand datasets obtained from four building types (customer classes) in Otaniemi area of Espoo, Finland. The buildings are Building A (residential building type), Building B (educational building type, contains classrooms and laboratories), Building C (office building type), and Building D (mixed use building, contains computer laboratories and health care center). The buildings have a peak (in 2015) aggregate electricity demand of 221kW, 592kW, 29kW, and 86kW, respectively. The feature space for the FS work in this paper is described in Table 2. The electricity demand datasets of the four buildings are the desired target variables in the proposed FS strategy. A one-year (2015) hourly sample, 8760 values, of both the feature sets and target variables are used for FS work.

B. FS RESULTS

The devised BGA-GPR based FS algorithm is run for each of the four datasets (representing four electricity customer types) separately. The empirical results found for the four datasets are presented in Table 4. The fitness values in Table 4 are calculated using the unnormalized (original dataset format) values of the selected features.

TABLE 4. FS results.

Customer (Dataset) Type	Selected Features	Best Fitness	Without FS
Building A - Residential	1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24	41.21	72.16
Building B - Educational	1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 17, 19, 20, 21, 23, 24	9.48	268.50
Building C - Office	1, 2, 3, 4, 5, 7, 8, 13, 17, 19, 20, 21, 22, 23, 24	0.38	11.44
Building D - Mixed-use	1, 2, 3, 4, 5, 7, 8, 9, 13, 16, 17, 21, 22, 23, 24	0.12	71.87
All (Union)	1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24		

As it is clearly observed from the FS results in Table 4, the number of predictors chosen by the proposed FS strategy is considerably lower than the size of the feature space (number of predictors in the original dataset given in Table 2). This can be due to the availability of irrelevant and redundant information by most of the variables in the original feature space. The BGA finally selects the feature subset which contains the most relevant and nonredundant variables.

Predictors 1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 16, 17, 19, 20, 21, 22, and 24 are selected for Building A.

Similarly, features 1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 17, 19, 20, 21, 23, and 24 are selected for Building B. Variables 1, 2, 3, 4, 5, 7, 8, 13, 17, 19, 20, 21, 22, 23, and 24 are selected for Building C. Attributes 1, 2, 3, 4, 5, 7, 8, 9, 13, 16, 17, 21, 22, 23, and 24 are selected for Building D.

Predictors 1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 16, 17, 19, 20, 21, 22, 23, and 24, which represent hour of the day, day of the week, month of the year, season of the year, period of the day, holiday/weekend indicator2 (with 0-2 values), ambient air temperature, dew-point temperature, ambient relative humidity, ambient air pressure, wind direction, wind speed, gust speed, global solar radiation, sunshine duration, electricity price, previous 24h average electricity demand, 24h lagged electricity demand, and 168h lagged electricity demand respectively, are selected at least for one of the building types.

For the purpose of consistency and making use of similar set of predictors, features selected at least for one of the buildings can be chosen to form the input dataset for short-term forecasting of electricity demands in buildings. Hence, the final selected feature subset consists of 19 features - hour of the day, day of the week, month of the year, season of the year, period of the day, holiday/weekend indicator2, ambient air temperature, dew point, humidity, air pressure,

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

IEEE *Access*

wind direction, wind speed, gust speed, global solar radiation, sunshine duration, electricity price, previous 24h average electricity demand, 24h lagged electricity demand, and 168h lagged electricity demand.

Besides, the average computation time of the devised integrated BGA-GPR based FS algorithm with two-years long hourly sample of 24 initial features is about 30 minutes, using MATLAB simulation environment on a research workstation with Intel Core i7-6820HQ Processor, 2.70 GHz CPU, 16 GB RAM.

C. COMPARISON WITH OTHER FS APPROACHES

To validate the BGA-GPR FS work in this paper, the feature subset results by the proposed GA-GPR FS were compared with feature subset results using other two common FS approaches, namely: Correlation-based feature selection (C FS) and Neighborhood Component Analysis Regression-based feature selection (NCA FS). The Correlation-based FS first calculates the Pearson and Spearman correlations of each feature with the target, and it then takes the maximum of the two correlation coefficients. A feature with correlation value greater than a given threshold value (0.5 in this paper) is selected as relevant feature and included in the final feature subset. The NCA FS operates based on the neighborhood component analysis (NCA) regression model fitted over the feature subsets versus target dataset. The NCA FS obtains the predictor weights (for reduced feature subsets) using a diagonal adaptation of the NCA regression model. The NCA model realizes FS by regularizing the predictor weights.

Table 5 provides the performance comparison of the FS result by the proposed method and other two methods. For the purpose of suitability of comparison, the same fitness function (MSE) modeled as the residual of the GPR model is used. That means each selected feature subset by each method is evaluated for fitness using the GPR model residual. Besides, the fitness values are calculated using the unnormalized values of the selected features.

As shown in Table 5, the proposed BGA-GPR based FS achieved the feature subset with the best fitness value (lowest MSE). Hence, the feature subset selected by the proposed FS strategy contains more relevant and nonredundant features than the other evaluated FS methods. That means, an electricity demand forecasting model whose input dataset constitute the feature subset achieved by the proposed BGA-GPR FS strategy can attain accurate forecast results.

D. EVALUATION OF FS RESULTS FOR IMPROVED FORECASTING

In order to further validate the effectiveness of the obtained FS results, a Feedforward Artificial Neural Network (FFANN) based 24h-ahead electricity demand forecast model was developed for each customer category. The 19 features selected by the devised BGA-GPR FS, presented in Section VII B, form the training input dataset for FFANN forecast model. While the training target variable is the electricity demand of each building.

TABLE 5. Comparison of FS results.

Customer (Dataset) Type	Selected Features			Best Fitness		
	BGA-GPR FS	NCA FS	C FS	BGA-GPR FS	NCA FS	C FS
Building A	1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24	1,2, 3, 4, 6, 7, 8, 9, 10, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24	1, 3, 4, 5, 21, 22, 23, 24	41.21	50.47	54.96
Building B	1, 2, 3, 4, 5, 7, 8, 9, 10, 13, 15, 17, 19, 20, 21, 23, 24	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, 15, 17, 19, 20, 21, 23, 24	2, 3, 4, 5, 6, 7, 10, 13, 19, 20, 21, 22, 23, 24	9.48	14.35	18.67
Building C	1, 2, 3, 4, 5, 7, 8, 13, 17, 19, 20, 21, 22, 23, 24	1, 2, 3, 4, 7, 8, 13, 17, 19, 20, 21, 22, 23, 24	1, 2, 3, 4, 5, 6, 7, 10, 13, 19, 20, 21, 22, 23, 24	0.38	0.50	0.57
Building D	1, 2 , 3, 4, 5, 7, 8, 9, 13,16, 17, 21, 22, 23, 24	1, 2 , 3, 4, 5, 7, 9, 11, 12, 13, 16, 17, 18, 21, 22, 23, 24	1, 2, 3, 4, 5, 6, 7, 13, 19, 21, 22, 23, 24	0.12	0.19	0.43

A two-year (2015 - 2016) time series hourly data of the selected features and target variable were employed to train the FFANN model. The FFANN model parameters was found experimentally. A hidden layer of 10 neurons was used. Moreover, the conventional GA was used to find the optimal parameters of the FFANN weight parameters. The prediction performance of the developed FFANN forecast model was verified with a one-year (2017) window length testing data.

However, to illustrate the forecast results here, the model testing result is given for randomly chosen four weekdays and four weekends/holidays representing the weekdays and weekends of the four seasons of the year: 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). For the

purpose of illustration, the forecast results for one of the buildings, Building B, are shown next. The forecast results are presented for the random eight testing days with one-hour time resolution, and they are depicted in Figures 4 and 5, for the weekdays and weekends/holidays, respectively.
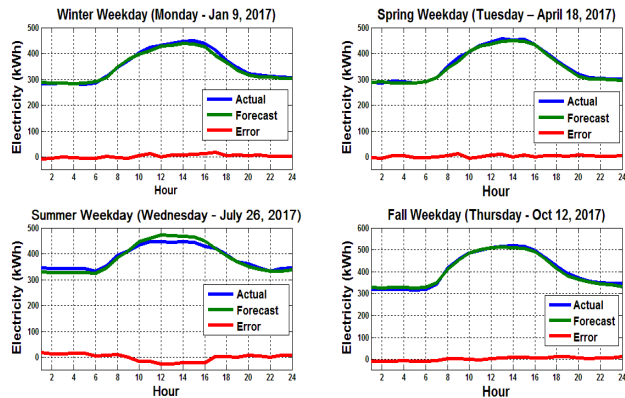

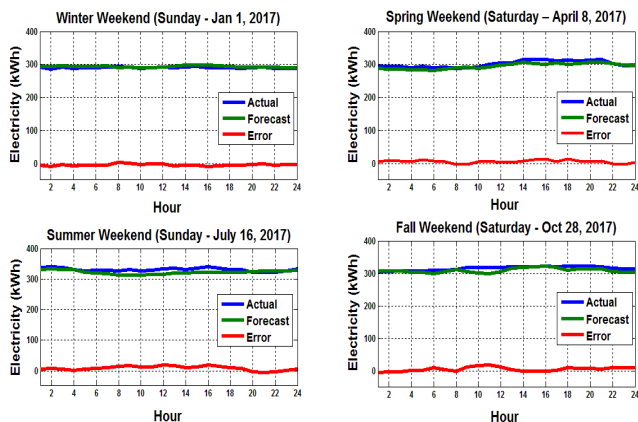**FIGURE 4.** Real vs. forecasted electricity demand for weekdays.

FIGURE 5. Real vs. forecasted electricity demand for weekends/holidays.

As it can be observed in Figures 4 and 5, the forecasts follow the actual electricity demand trends with smaller gaps (errors) between them.

This further verifies the effectiveness of the proposed FS approach in selecting the best features subset that enables the forecast model to achieve improved forecasts that are more accurate.

Moreover, the following criteria were used to evaluate the accuracy of the obtained forecasts:

- Error

$$Error = P_h^a - P_h^f \qquad (10)$$

where, P_h^a and P_h^f are the real and prediction values of the electricity demand at hour h, respectively.

- Mean absolute error (MAE)

$$MAE = \frac{1}{N} \left| P_h^a - P_h^f \right| \qquad (11)$$

where, N is the forecasting horizon and its value is 24 for 24h-ahead forecast.

- Mean absolute percentage error (MAPE)

$$MAPE = \frac{100}{N} \sum_{h=1}^{N} \left| \frac{P_h^a - P_h^f}{P_h^a} \right| \qquad (12)$$

MAE of 7.20kWh, MAPE of 1.96%, and daily peak MAPE of 2.04% are obtained for the one-year (2017) forecast using the proposed BGA-GPR FS results as input dataset for the FFANN based forecast model of the educational building (Building B). Figure 6 shows the schematic illustration of the values of the forecast accuracy evaluation criteria used. Hence, the obtained numerical results further confirm the acceptability of the forecast accuracy achieved and effectiveness of the implemented FS method.
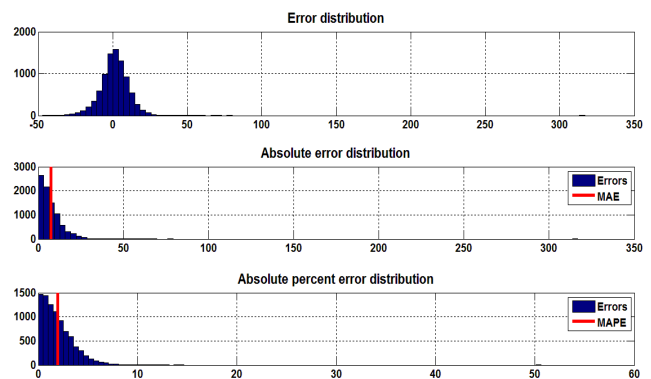

FIGURE 6. Obtained values of forecast accuracy evaluation criteria.

E. QUANTITATIVE RELEVANCE ANALYSIS OF FS RESULTS

In order to quantify the benefits and relevance of the proposed hybrid BGA-GPR based FS method and the selected features, the following metrics are used:

- Computation time reduction:

$$\Delta t_{comp} = \frac{t_{without\_FS} - t_{with\_FS}}{t_{without\_FS}} \qquad (13)$$

where, $t_{without\_FS}$ is the total computation time which includes data preprocessing, forecasting model training, validation, and prediction using the original feature space without FS, t_{with_FS} is the total computation time with the use of the obtained FS results, and Δt_{comp} is the change in total computation time due to FS. Positive value of Δt_{comp} indicates the reduction of computation time requirement of the electricity demand forecasting model due to making use of FS results.

- Dimensionality reduction:

$$\Delta D = \frac{R_{without\_FS}^{m \times n} - R_{with\_FS}^{m \times n_r}}{R_{without\_FS}^{m \times n}} \qquad (14)$$

where, $R_{without\_FS}^{m \times n}$ is a matrix of feature space without FS with m number of samples and n number of features, $R_{with\_FS}^{m \times n_r}$ is a matrix of the reduced feature space with FS

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

IEEE *Access*

with m number of samples and n_r number of reduced features, and Δ D is the change in data dimension due to FS. Positive value of Δ D (n-n_r) indicates the reduction of input data dimension for the electricity demand forecasting model.

- FS fitness value improvement:

$$\Delta fit = \frac{fit_{without\_FS} - fit_{with\_FS}}{fit_{without\_FS}} \qquad (15)$$

where, $fit_{without\_FS}$ is the fitness value of the features without FS with respect to a predefined fitness function (MSE of GPR output and actual target formulated in equation (9)), fit_{with_FS} is the fitness value of the selected features with FS, Δ fit is the change in fitness value due to FS. Positive value of Δ fit indicates the improvement in predictive model fitness value (reduction in MSE value) due to FS.

- Forecasting accuracy improvement:

$$\Delta acc = \frac{acc_{with\_FS} - acc_{without\_FS}}{acc_{without\_FS}} \qquad (16)$$

Here, $acc_{without\_FS}$ is the accuracy of the forecasts without making use of FS results (using the original feature space as training input) and acc_{with_FS} is the accuracy of the forecasts with FS results (using the reduced feature space as training input). $acc_{without\_FS}$ and acc_{with_FS} are defined as follows:

$$acc_{without\_FS} = 100 - MAPE_{without\_FS} \qquad (17)$$
$$acc_{with\_FS} = 100 - MAPE_{with\_FS} \qquad (18)$$

where, $MAPE_{without\_FS}$ is the mean absolute percentage error of the forecasts without FS and $MAPE_{with\_FS}$ is the mean absolute percentage error of the forecasts with FS. Δ acc is the change in forecast accuracy due to FS.

Positive value of Δ acc indicates the improvement of forecast accuracy due to making use of FS results in the forecasting process.

Table 6 presents the values of the metrics defined in equations (13) to (16) to determine the benefits achieved due to the implementation of the devised FS method for short-term electricity demand forecasting.

As clearly shown in Table 6, the implementation of the FS and its integration to the forecasting model has resulted in much improvements compared to the forecasting performance using the original dataset without FS. For example, for the residential building (Building A), the enhancement in fitness value (MSE) using the BGA FS algorithm selected variables to fit the electricity demand by the GPR model is 42.9% over the original dataset (without FS). Similarly, the selected feature subset of the educational (Building B), office (Building C) and mixed-use building (Building D) types have given enhancements of 96.5%, 96.7% and 99%,

TABLE 6. Quantitative relevance analysis of FS results.

Metric	Value (%)			
	Building A	Building B	Building C	Building D
Computation time reduction, $\Delta tcomp$	17.7	26.1	34.6	34.6
Dimensionality reduction, ΔD	20.8	29.2	37.5	37.5
Fitness value improvement, Δfit	42.9	96.5	96.7	99
Forecasting accuracy improvement, Δacc	38.7	81.2	81.9	83.0

respectively, compared to the original dataset, regarding the FS GPR-modeled MSE fitness function.

The reduction in data dimensionality over the original feature space is 20.8%, 29.2%, 37.5% and 37.5% for Building A, B, C and D, respectively. Likewise, the decrease in total computation time is 17.7%, 26.1%, 34.6% and 34.6%, respectively for Building A, B, C and D.

Above all, the improvement of the forecasting accuracy is the most important and major objective of this study. The improvement in prediction accuracy using the BGA-GPR FS selected features to constitute the forecasting model training inputs is 38.7%, 81.2%, 81.9% and 83.0%, respectively.

Thus, the above quantifications and experimental results further demonstrate the relevance of the effective FS for the improvement of the electricity demand forecasting.

VIII. CONCLUSION

This paper proposed and implemented a BGA based feature selection approach for improved short-term electricity demand forecasting models. The approach includes the use of a GPR fitness function to choice a combination of predictors from a given original predictor space. The proposed BGA-GPR FS has given a feature subset that resulted in a better fitness (lower MSE value) than the original dataset with all the initial features. For comparison and validation, features selected by other two feature selection methods were presented. The BGA-GPR features outperformed the other features with respect to the MSE fitness function defined in the GPR framework. Moreover, the proposed BGA-GPR FS was applied to four different electricity demand datasets representing four different customer types (residential, educational, office and mixed-use building energy systems). It achieved the best feature subsets, which can constitute the input datasets for accurate forecasting of electricity demands for all the electricity customer groups. Moreover, a FFANN based 24h-ahead electricity demand forecast model was developed to evaluate the effectiveness of the FS results. The electricity demand forecasting model developed using the obtained FS results has achieved an annual accuracy improvement of 38.7%, 81.2%, 81.9% and 83.0%, respectively for residential, educational, office and mixed-used building types compared to forecasting based on the original feature space

without FS. Therefore, the paper findings verify that the combination of effective feature selection method and forecasting models owns robust forecasting power, compared to forecasting with arbitrary features without predictor selection methods. The work is both novel and effective from application, algorithms hybridization and performance improvement perspectives. The study contributes a new and robust feature selection tool by combining BGA and GPR for improved performance (more accurate) electricity demand prediction problem.

REFERENCES

[1] A. H. Shahana and V. Preeja, "Survey on feature subset selection for high dimensional data," in *Proc. IEEE ICCPCT*, Nagercoil, India, Mar. 2016, pp. 1–4.

[2] H. Li, A. T. Eseye, J. Zhang, and D. Zheng, "A double-stage hierarchical hybrid PSO-ANFIS model for short-term wind power forecasting," in *Proc. IEEE GreenTech*, Denver, CO, USA, Mar. 2017, pp. 342–349.

[3] S. Harde and V. Sahare, "Design and implementation of ACO feature selection algorithm for data stream mining," in *Proc. IEEE ICACDOT*, Pune, India, Sep. 2016, pp. 1047–1051.

[4] B. Tran, B. Xue, and M. Zhang, "A new representation in PSO for discretization-based feature selection," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1733–1746, Jun. 2018.

[5] H. Peng, C. Ying, S. Tan, B. Hu, and Z. Sun, "An improved feature selection algorithm based on ant colony optimization," *IEEE Access*, vol. 6, pp. 69203–69209, 2018.

[6] A. S. Ghareb, A. A. Bakar, and A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Syst. Appl.*, vol. 49, pp. 31–47, May 2016.

[7] M. Gen and R. Cheng *Genetic Algorithms and Engineering Optimization*, vol. 7. Hoboken, NJ, USA: Wiley, 2000.

[8] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowl.-Based Sys.*, vol. 82, pp. 29–40, Jul. 2015.

[9] G. Doquire and M. Verleysen, "Mutual information-based feature selection for multilabel classification," *Neurocomputing*, vol. 122, pp. 148–155, Dec. 2013.

[10] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowl.-Based Sys.*, vol. 24, no. 7, pp. 1024–1032, Oct. 2011.

[11] R. Sikora and S. Piramuthu, "Framework for efficient feature selection in genetic algorithm based data mining," *Eur. J. Oper. Res.*, vol. 180, no. 2, pp. 723–737, Jul. 2007.

[12] S. Voronin and J. Partanen, "Forecasting electricity price and demand using a hybrid approach based on wavelet transform, ARIMA and neural networks," *Int. J. Energy Res.*, vol. 38, no. 5, pp. 626–637, Apr. 2014.

[13] S. Jurado, A. Nebot, F. Mugica, and N. Avellana, "Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques," *Energy*, vol. 86, pp. 276–291, Jun. 2015.

[14] Z. Hu, Y. Bao, and T. Xiong, "Comprehensive learning particle swarm optimization based memetic algorithm for model selection in short-term load forecasting using support vector regression," *Appl. Soft Comput.*, vol. 25, pp. 15–25, Dec. 2014.

[15] S. M. R. Kazemi, M. M. S. Hoseini, S. Abbasian-Naghneh, and S. H. A. Rahmati, "An evolutionary-based adaptive neuro-fuzzy inference system for intelligent short-term load forecasting," *Int. Trans. Oper. Res.*, vol. 21, no. 2, pp. 311–326, Mar. 2014.

[16] Z. Tonkovic, M. Zekic-Susac, and M. Somolanji, "Predicting natural gas consumption by neural networks," *TehnickiVjesnik*, vol. 16, no. 3, pp. 51–61, Sep. 2009.

[17] S. Salcedo-Sanz, J. Munoz-Bulnes, J. A. Portilla-Figueras, and J. D. Ser, "One-year-ahead energy demand estimation from macroeconomic variables using computational intelligence algorithms," *Energy Convers. Manage.*, vol. 99, pp. 62–71, Jul. 2015.

[18] M. Sheikhan and N. Mohammadi, "Neural-based electricity load forecasting using hybrid of GA and ACO for feature selection," *Neural Comp. Apps.*, vol. 21, no. 8, pp. 1961–1970, Nov. 2012.

[19] R. Urraca, A. Sanz-Garcia, J. Fernandez-Ceniceros, E. Sodupe-Ortega, and F. J. Martinez-de-Pison, "Improving hotel room demand forecasting with a hybrid GA-SVR methodology based on skewed data transformation, feature selection and parsimony tuning," in *Proc. 10th Int. Conf., Hybrid Artif. Intell. Syst. (HAIS)*, Bilbao, Spain, vol. 9121, E. Onieva, I. Santos, E. Osaba, H. Quintian, and E. Corchado, Eds. Springer, Jun. 2015, pp. 632–643.

[20] Y. Liu, Y. Yin, J. Gao, and C. Tan, "Wrapper feature selection optimized SVM model for demand forecasting," in *Proc. 9th Int. Conf. Young Comput. Scientists*, G. J. Wang, J. Chen, M. R. Fellows, and H. D. Ma, Eds., vols. 1–5, Nov. 2008, pp. 953–958.

[21] S. Jiang, K.-S. Chin, L. Wang, G. Qu, and K. L. Tsui, "Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department," *Expert Syst. Appl.*, vol. 82, pp. 216–230, Oct. 2017.

[22] R. C. Cavalcante, R. C. Brazileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Syst. Appl.*, vol. 55, pp. 194–211, Aug. 2016.

[23] C.-F. Tsai and Y.-C. Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches," *Decision Support Syst.*, vol. 50, no. 1, pp. 258–269, Dec. 2010.

[24] F. Lin and D. Liang, C.-C. Yeh, and J.-C. Huang, "Novel feature selection methods to financial distress prediction," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2472–2483, Apr. 2014.

[25] B. Oluleye, A. Leisa, J. Leng and D. Dean, "A Genetic Algorithm-Based Feature Selection," *Int. J. Elec. Computer Eng.*, Vol. 5, no. 4, pp. 889–905, Jul. 2014.

[26] J. Holland, *Adaptation in Natural and Artificial Systems*. Cambridge, MA, USA: MIT Press. 1992.

[27] 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.

[28] 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.

[29] T. Mathworks, *Statistics Toolbox User's Guide*. Natick, MA, USA: MathWorks—Apple Hill Campus, 2013.

[30] P. Civicioglu, "Transforming geocentric Cartesian coordinates to geodetic coordinates by using differential search algorithm," *Comput. Geosci.*, vol. 46, pp. 229–247, Sep. 2012.

ABINET TESFAYE ESEYE received the B.Sc. degree in electrical engineering from Hawassa University, Ethiopia, in 2010, the M.Sc. degree in electrical power systems engineering from Bahir Dar University, Ethiopia, in 2012, and the Ph.D. degree in electric power system and its automation from North China Electric Power University, China, in 2018. He was an Assistant Lecturer with the School of Computing and Electrical Engineering, Bahir Dar University, Ethiopia, from 2011 to 2012. He was also a Lecturer and the Head of Electrical and Computer Engineering Department, Mettu University, Ethiopia, from 2012 to 2014. He was a Renewable Energy and Smart Microgrid R&D Engineer with Goldwind Science and Technology Company, Ltd., China, from 2015 to 2018. He is currently with the Department of Electrical Engineering and Automation, Aalto University, Finland, as a Postdoctoral Researcher. His research 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, the IoT, optimization, computational intelligence, forecasting, state estimation, artificial intelligence, machine learning, and deep learning in energy systems.

A. T. Eseye *et al.*: ML Based Integrated FS Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems

IEEE *Access*

MATTI LEHTONEN received the M.S. and Licentiate degrees in electrical engineering from 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 research activities include earth fault problems, and harmonic-related issues and applications of information technology in distribution automation and distribution energy management.

SEMEN UIMONEN was born in Petrozavodsk, Russian Federation, in 1990. He received the M.Sc. degree in electrical engineering from the Aalto University School of Electrical Engineering, Espoo, Finland, in 2015. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Automation, Aalto University School of Electrical Engineering. His research interests include energy and power balance of buildings, the effect of people in the power consumption, and the IoT in power systems targeted at energy efficiency improvements.

TONI TUKIA was born in Kirkkonummi, Finland, in 1988. He received the M.Sc. degree in electrical engineering from the Aalto University School of Electrical Engineering, Espoo, Finland, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Automation, Aalto University School of Electrical Engineering. His research interests include measuring and modeling the energy and power balance of buildings, distributed generation, and the effect of people on the building power consumption.

R. JOHN MILLAR received the bachelor's degree in engineering (mechanical) from the University of Auckland, New Zealand, in 1984, and the master's degree in electrical engineering and D.Sc. (Tech.) from the Helsinki University of Technology, Finland, in 2002 and 2006, respectively. This university is now part of Aalto University, where he continues to work as a Researcher and Lecturer. His main research interests include distribution network planning, underground cable rating, and pedagogical development.

● ● ●