A New Hybrid Model for Energy Consumption Prediction Based on Grey Wolf Optimization

Asmaa Wahba, Reda El-khoribi, and Shereen Taie

***Abstract*—The prediction of building energy consumption (BEC) facilitates an effective energy management system based on the comprehensive understanding of the energy reduction potential, contributing to the reduction in climate variations. Several factors influence the energy efficiency of buildings. Therefore, a suitable technique that considers these factors must be implemented to predict BEC. Herein, a hybrid prediction model that combines a metaheuristic technique, namely, the gray wolf optimization (GWO) algorithm, with a machine learning algorithm, namely, support vector machine (SVM), (hereinafter referred to as GWO–SVM) is proposed based on 10-fold cross validation. Several machine learning and statistical techniques are employed to predict the energy consumption and show the robustness of the proposed model, including SVM, artificial neural networks, a hybrid genetic algorithm–SVM model, and the multiple linear regression. The energy consumption prediction models are evaluated on five real datasets (1) to predict the monthly energy consumption of four governmental sectors in the US (residential, industrial, commercial, and transportation) using two environmental parameters from January 1973 to May 2021 and (2) to predict the BEC, particularly hourly consumption in 2010, using eight environmental parameters employed for short- and long-term predictions. Results show that for the annual prediction, the GWO–SVM model outperforms all the other models with a prediction accuracy of 98.012% and an execution time of 10 min. These findings indicate that the proposed GWO–SVM model achieves a better accuracy and prediction time in short- and long-term predictions than the other models.**

***Index Terms*—building energy prediction, machine learning, Gray Wolf optimizer, support vector machine, Hybrid model, Genetic Algorithm, artificial neural network.**

I. Introduction

The higher the growing number of people the more business organizations construct extra houses and buildings, which are the principal cause of greenhouse gasses.

To reduce the energy consumption and quantity of carbon dioxide (CO2) emission of these new blocks, we consider an efficient implementation of energy at the initial design phase. Moreover, existing buildings could make energy effective with a good organization of energy and smart restorations to enhance its performance.

Energy consumption is increasing with a nonstop growth of the building structures. This growth makes it difficult to achieve the purpose of energy-saving and emission reduction, which demands more efficient techniques to recognize and predict the building energy consumption. Reference [1] shows that countries with an increasing population like Chinese lack power sources, which is needed to effectively predict the consumption of electricity using optimization techniques by applying an evolution algorithm with grey models.

According to [2], building usage increased to reach 40% of the overall consumption of universal resources and rose to reach the defined energy red line. Therefore, millions of dollars could be saved annually if energy consumption is predicted because of the improvement in buildings or consumption reduction. Hence, the need arises for a prediction technique for the energy consumption of buildings. Forecasting building energy is useful and important to make progress in energy performance. Many companies and management systems have conducted studies on energy reduction, especially in buildings. Based on their findings, the problem is complicated because of the nonlinearity of variable relationships and many influencing factors.

The challenge to create a prediction technique for building energy consumption relies on many factors, which must be considered properly employed as various parameters affect the energy consumption. These factors can be divided into four different groups: building sectors, types of energy, time measurement, and energy input-parameters. Building sectors: Hundred research papers have classified buildings into categories as commercial buildings, research and education buildings, and residential sector [3]. Types of energy: Almost 74% of energy prediction studies are based on the prediction of total buildings energy consumption, 15% of studies focused on the prediction of buildings cooling and heating load, while 11% of prediction cases used other energy parameters like physical environmental features. The relationship between these factors could be explored via machine learning algorithms and regression models [4]–[6]. Time measurement: According to the surveyed paper [3] load prediction is the length of time to be predicted. Seventy percent of the predictions were short-period load prediction, which predicts (hour, sub-hour, and day-ahead), and 30% for long-period load prediction, which are (weekly, monthly, and annual). Energy input-parameters: weather-related data were used mostly, particularly outdoor temperature, wind speed, solar radiation

Manuscript received November 23, 2020; revised March 10, 2022.

Asmaa Mohamed Wahba is a postgraduate student of Computer Science Department, Faculty of Computers and Information - Fayoum University, Fayoum City, Egypt, (corresponding author, e-mail: am2053@fayoum.edu.eg).

Reda A. El-khoribi is a professor of Information Technology Department, and the dean of Faculty of Computers and Information-Cairo University, Giza 12511, Egypt (e-mail: ralkhoribi@staff.cu.edu.eg).

Shereen A. Taie is an associate professor of Computer Science Department, Faculty of Computers and Information- Fayoum University, Fayoum, Egypt (e-mail: sat00@fayoum.edu.eg).

humidity, surface pressure, and heating and cooling degree per day. Among all these factors, the first three are mainly used in the prediction process as they have a major effect on the gain and loss of the building's heating [7]-[9].

Machine learning black-box modeling type will be useful, when the whole input parameters are known and measured, while no physical parameters or details are available. The model helps with buildings when they are already built. The principle of these techniques was implemented to solve regression, prediction, and classification problems, which include forecasting building energy usage [10]. The major working places for societies throughout the daytime are commercial buildings. Moreover, those working places that consume a large amount of energy as the working period in the tunneling process can be optimized by controlling the usage of machines to reduce energy consumption [11]. Sometimes there is need for an efficient power saving strategy that could implement approaches to analyze the design of inner electrical services [12], or the power control of the electronic intensity for indoor electrical lights [13], especially in developing countries, when loads of energy are used up. Thus, in this research, we consider energy consumption in a general and commercial building. Outdoor temperature and power consumption are taken as quantitative original factors to discover the potential of buildings power reduction. This paper aims to predict energy consumption using a big dataset with several parameters and test the performance with multiple linear regression (MLR), support vector machine (SVM), a genetic algorithm with SVM (GA-SVM), neural network, and the proposed model to compare their results to come up with a suitable technique.

The SVM is widely used in the prediction, classification [14], and fault diagnosis as the dependent relationship between accuracy and kernel parameters leads to using various optimization techniques for its parameters. The grid is the standard technique used for SVM optimization, though it is insufficient in performance and time-consumer [15]. Then, several algorithms have been proposed to cover these limitations, such as using particle swarm optimization (PSO) and a genetic algorithm to optimize the SVM parameters, which give good accuracy but also consume more time.

Grey wolf optimization (GWO) gives better performance and execution speed. In cases of optimization, it has been proved in [16] that GWO optimization results are superior to that of PSO and GA. Several types of research have been made for predicting building power consumption with machine learning mechanisms, especially building consumption for a short time. Furthermore, many studies were conducted over a long time with several parameters to give better results.

The rest of this paper is prepared as follows. Section II demonstrates the related work. Section III presents the preliminaries of the applied techniques with a comprehensive comparison. Section IV illustrates the proposed hybrid model GWO-SVM, while Section V shows the experimental results and discussion. Finally, Section VI presents the conclusion.

II. RELATED WORKS

G. Ciulla and A. D'Amico [17] proposed an MLR method for solving building energy consumption (BEC) performance of its different situations and determine the heating and cooling a load of energy with a small number of variables. The proposed model gave better performance for three main results for heating-degree, cooling-degree, and energy-demand were measured with R2 as 0.90, 0.96, and 0.95, with MAE as 3.66, 5.54, 6.88, and with RMSE as 4.54, 6.75, and 8.00, respectively.

Zhitong Ma et al. in [18] proposed a forecasting approach for building power usage in seven zones of China based on a SVM. The proposed model used various factors, such as climate data and economic parameters. The proposed model CV < 3%, and it gives similar results in MSE and r2, which indicate accurate predictions.

Hafiz Muhammad et al. in [19] implemented a machine learning system using a complicated neural network, with an extreme learning machine to predict electricity consumption using a dataset from 2015 to 2017. The proposed model gives better performance than the current machine learning as (ANN- SVR-MLR) according to MAE, MAPE, MSE with high error-value, and RMSE with low error-value.

In [20], Duangjai and Eakasit developed GWO with extra packs to extend the search area for predicting a building's heating and cooling load using 768 different residential structures. The dataset consists of eight variables describing the physical characteristics of each building. The proposed model compared with six ML techniques, such as ANN, GSGP, MLP, SVR, EMARS, and random forecasts measure the statistical error by computing MAE, RMSE, and MRE with values 0.323, 1.394, and 1.546, respectively.

Adel Alshibani [21] also proposed ANN model For the BEC prediction in Saudi Arabia applied the model on 352 datasets for really different types of school buildings. They distinguished the main factors that affect BEC, especially in school, and analyzed the relationship of their effectiveness on the total EC. The model recognized 11 factors that could affect EC; therefore, they were all used to identify the most influential factor. The presented model accuracy is 87.5%.

In [22], Shuyu Dai et al. proposed an enhanced hybrid strategy of the GWO. The model applied PSO to give the first initial population's positions to use the GWO for the position updating of the population, and then give the ideal features to be applied in the SVM for the categorization purpose. There was a comparison between it and other techniques based on SVM, such as GA and PSO. The proposed model's accuracy is 89.33%, which outperformed the other two compared methods that have an accuracy of 87.36% and 85.37%.

In [23], Shuyu Dai et al. proposed a developed hybrid model to optimize SVM with GWO and DE (differential evolution) to forecast the financial power grid for five years since 2018 in China. The dataset used was collected since 2010. Meanwhile, the proposed model was applied in different ways to prove the efficiency as DE-GWO-SVM, GWO-SVM, BP, and SVM for measuring the performance MAPE, MAE, and RMSE were calculated that gave 1.70, 6.31%, and 9.36, respectively.

In [24], Tian et al. presented a hybrid prediction procedure for electricity consumption of buildings using FCM-GWO-BP. They showed that the historical building's dataset clusters through FCM and BP for the reduction of data related noise in multiple categories and the hybrid GWO-BP applied for error optimization of BPNN in every category. The proposed procedure achieved 0.225 reduction in RMSPE, whereas that of FCM was 0.135, which means the accuracy of the prediction model was enhanced by almost 75%.

The contribution of this paper is to propose a new automatic model GWO-SVM that predicts the energy consumption for buildings in two periods; six months and one year. The proposed model overcomes problems of other techniques by providing a hybrid meta-heuristic prediction model GWO- SVM. Furthermore, a comprehensive comparison was applied on the most common techniques in this problem SVM, GWO-SVM, MLR model, GA, and Artificial Neural Network using the same dataset, which proves the effectiveness of the proposed model in BEC prediction. The proposed model has a robust performance in accuracy and prediction time.

III. PRELIM IN AREAS

This section presents a brief idea concerning the core concepts of SVM, GWO- SVM, MLR Model, GA, Artificial Neural Network, and GWO.

A. Support vector machine:

The SVM is a supervised learning group of processes (which means SVM perform training process on a group of a labeled dataset so that SVM learns from it to categorize the new input dataset) considered as regression, categorization, and prediction tool, which applies a theory of machine learning for the optimization of forecasting accuracy and avoiding data over-fit [25].

The SVM is accessible with many barriers in R, Matlab, and other programming languages [26], such as libsvm, which is used in this study. Moreover, SVM is a frame for the classification of datasets by finding the optimal separating edge "hyperplane" between two nonlinear classes with kernel rules having the margin at the maximum distance between two classes [27]. It also estimates the linear association between input dataset and output targets as shown in equation (1).

$$f(x) = Wtx + \Theta \tag{1}$$

Where w represents weight and θ represents bias. The used dataset represented through $\{(xxi,\ yyi) |\ i{=}n1\}$, where $xxi \in \mathbb{R}m$ with m features and $yyi \in \mathbb{R}$, and n shows the samples number [28].

TABLE I
MAJOR STEPS OF SVM

Steps	Process
Step 1:	Preprocessing: eliminate empty data
Step 2:	Segmentation: Split data to x (have all parameters) and y (only the target class to be predicted)
Step 3:	Modeling: Make SVM Model
Step 4:	Prediction: apply the prediction on X to predict y and compute Accuracy.
Step5:	Optimization: apply tuning function to optimize SVM parameters then apply Step 4

B. Multiple Linear Regression:

Regression is a method that analyzes and presents a changeable effect on one or several variables of another variable [29].

MLR shows the dependency of variables on various independent variables or on one variable change. Those variables that affect other variables are named predictors (indicated with X), and the affected variables are named responses (indicated with Y). It allows mean function to have more shapes than linear lines, even though such allowance prevents random shapes [30]. In this research, the MLR will be applied because our data include various predictors and one response (Energy Consumption). The MLR is estimated using equation (2) as:

$$Y = \beta\_0 + X\_1\beta\_1 + X\_2\beta\_2 + \ldots + Xk\beta\_k + \varepsilon \tag{2}$$

where Y is the response variable or dependent, β_0 is the intercept, X1, X2,…,Xk are predictor variables or independent, β1, β2,…, βk are the regression coefficients of X1, X2,…,Xk, respectively, which show the change of Y associated with each unit change in X. ε is the random error, which represents the difference in detected and fitted values of the linear model.

TABLE II
MAJOR STEPS OF MLR

Steps	Process
Step 1:	Preprocessing: check for empty data
Step 2:	Fitting regression model: Fit data using Least Square Method of estimation In R language, function (lm) will be proper.
Step 3:	Prediction: Use predict function for the approximation of energy consumption based on regression model.
Step 4:	Evaluation: evaluate the model using Root Mean Square error (RMSE) method and other metrics.

C. Artificial Neural Network

Artificial neural network enables us to defeat the limitations of old strategies in resolving difficult issues by gaining knowledge from given models to make approximations by making mapping between input and output layers [31].

The ANN simulates constructions of the human mind that consists of many cells or nodes known as artificial neurons basically organized in three layers called input, output, and one or more hidden layers [32]. Each layer is a group of interconnected nodes connected by edges that interact via signals. So, the received neuron that receives a signal could operate, then transfer it to another connected neuron [33]. These signals have weight wg (j=1,2,3,….z). Therefore, every node makes summation for the weights from input neurons to it and computes activation function, which shows the nonlinearity of the neuron's output [34].

Recently, ANN is used in predicting a building's energy consumption because of its ability of solving nonlinear problems and complicated problems [35].

TABLE III
MAJOR STEPS OF ANN

Steps	Process
Step 1:	Preprocessing: eliminate empty data
Step 2:	Segmentation: Split data to into train and test set
Step 3:	normalizing the data set using min-max method and scale it in (0;1) interval for accurate result
Step 4:	Fitting the model via gml () function.
Step 5:	building ANN model
Step 6:	Prediction: predict energy consumption for the test set and compute accuracy.

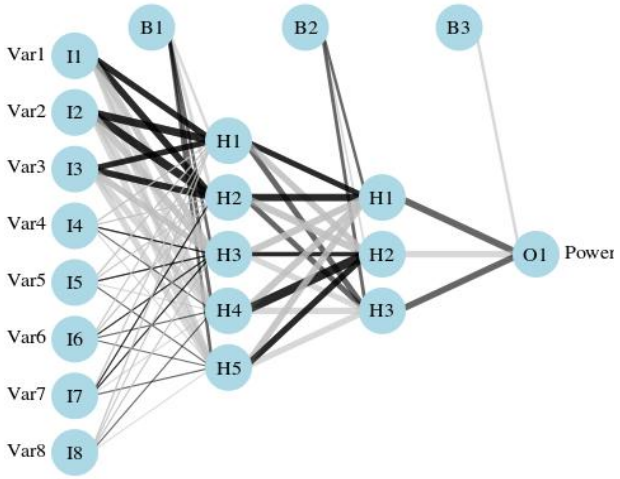

Fig.1. ANN graph

Fig. 1. The applied network of two hidden layers with the network structure given as 8:5:3:1. The eight input parameters will be illustrated in the Dataset section and five neurons for the first hidden layer, three for the other hidden layer, and one neuron for the output layer. Using the potent function, we represent the model with the following labels (I: input, H: hidden, O: output, B: bias).

D. Genetic Algorithm:

The GA has the greatest advantage because of its capability to use accumulated data of premier minor space to search for useful spaces and shift next searches [36]. This algorithm is different from the nonlinear standard optimization methodologies by keeping a population of candidate solutions; it actually looks for the best candidate. The basic characteristic of GA owns a chromosome that could be coded for a certain length as a sequence of characters. This is because a population is a set of chromosomes.

The fitness function (F) links the GA and the current problem. It is the only criterion to measure the quality of the independent solutions to enhance the forecasting accuracy [37].

$$\min\ f = \frac{1}{n}\sum_{j}^{d}\frac{(D_{actual} + D_{estimated})}{D_{actual}} \quad (3)$$

where D_actual and D_estimated are real and predictable power consumption, respectively, and m is the observed number. Meanwhile, the advantage GA is that it concentrates on a point in the chromosome search area, while focusing on a population of chromosomes. The GA operation and the number of chromosomes is directly related. In other words, when the number of chromosomes is small, the GA motion would be low and only a minor search space part would be searched [38].

Furthermore, GA chooses the top genes of the first generation and proceeds them toward the following generation. Then, with crossover and mutation, a developing process would be modeled where only genes adaptable to the environment will exist. First, a solution population will be initialized randomly. Then, top genes are chosen through the probability of their suitability via the selection operator. After that, the picked genes will be moved to the following generation without ruining their characteristics via a crossover operator. The crossover operation takes probability values from 0.2 to 0.8. Finally, the mutation process searches for great solutions by evolving or keeping original solutions, and the user's input can control the mutation probability [38].

TABLE IV
MAJOR STEPS OF GA-SVM

Steps	Process
Step 1:	Preprocessing: eliminate empty data
Step 2:	Segmentation: Use 5-fold cross validation to Split Data to into train set and test set.
Step 3:	Define the upper bound and lower bound of svm Parameters where:
Step 4:	Lower<-{cost=le-4,gamma=le-2,epsilon=le-3}
Step 5:	Upper <- {cost=200 , gamma=2 ,epsilon=1.7}
Step 6:	Initialize a population
Step 7:	Model SVM using train-data as its data using "e1071" library
Step 8:	Define fitness function
Step 9:	Run genetic algorithm using "GA " library with Population size=500 and invoke step 3,6 then execute Selection, crossover operators and mutation process

E. Grey Wolf Optimization:

GWO [39] algorithm is a meta-heuristic optimization procedure proposed by Mirjalili et al. (2014). It gives perfect accuracy results in optimization and global search problems because of its minimal complexity, a smaller number of control constraints, and greater search efficiency [40].

The GWO simulates the grey wolves' hierarchical leadership structure and hunting technique in nature, the troop of wolves' is organized hierarchically into four layers where the top wolf alpha (α) is the decision-maker for the most important problems in the whole wolf troop, and it is considered as the best solution. The second wolf beta (β) assists the leader. The third wolf delta (δ) follows the orders of α and β performs tasks. The lowest wolf omega (ω) is referred to as the rest of the subordinates that obey the main three layers.

Performing the hunting procedure of grey wolves mathematically, we use the following equations (4) and (5).

$$\vec{D} = |\vec{C}.\vec{X}_p(t) - \vec{X}(t)| \tag{4}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}.\vec{D} \tag{5}$$

Where t represents the immediate iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}_p$ is the location vector of the prey position, and $\vec{X}$ shows the location vector of a grey wolf. The vectors $\vec{A}$ and $\vec{C}$ are estimated as in (6) and (7), respectively.

$$\vec{A} = 2\vec{a}.\vec{r}_1 - \vec{a} \tag{6}$$

$$\vec{C} = 2.\vec{r}_2 \tag{7}$$

Where components of $\vec{a}$ are reduced linearly from two to zero in the loop of iterations, and r1, r2 are random vectors in the interval [0, 1].

In simulating the hunting method of grey wolves, assume the best solution is α, all data needed for the position of the prey are β and δ. So, the top three solutions are stored, and the other search agent's ω updates their locations according to the new position of the three best search agents. The updating process of the grey wolves' positions is applied as in (8), (9), (10) [41]- [42]:

$$\vec{D}_\alpha = |\vec{C}_1.\vec{X}_\alpha - \vec{X}|, \vec{D}_\beta = |\vec{C}_2.\vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_2.\vec{X}_\delta - \vec{X}|$$

(8)

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1.(\vec{D}_\alpha)|, \vec{X}_2 = \vec{X}_\beta - \vec{A}_2.(\vec{D}_\beta), \vec{X}_2 = \vec{X}_\delta - \vec{A}_2.(\vec{D}_\delta)$$

(9)

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{2} \tag{10}$$

A. *The Proposed Model*

The proposed hybrid model consists of three phases; preprocessing phase, optimization phase of SVM based-GWO optimization algorithm, and prediction phase.

Fig. 2 shows the comprehensive framework used in applying GWO-SVM Algorithm.

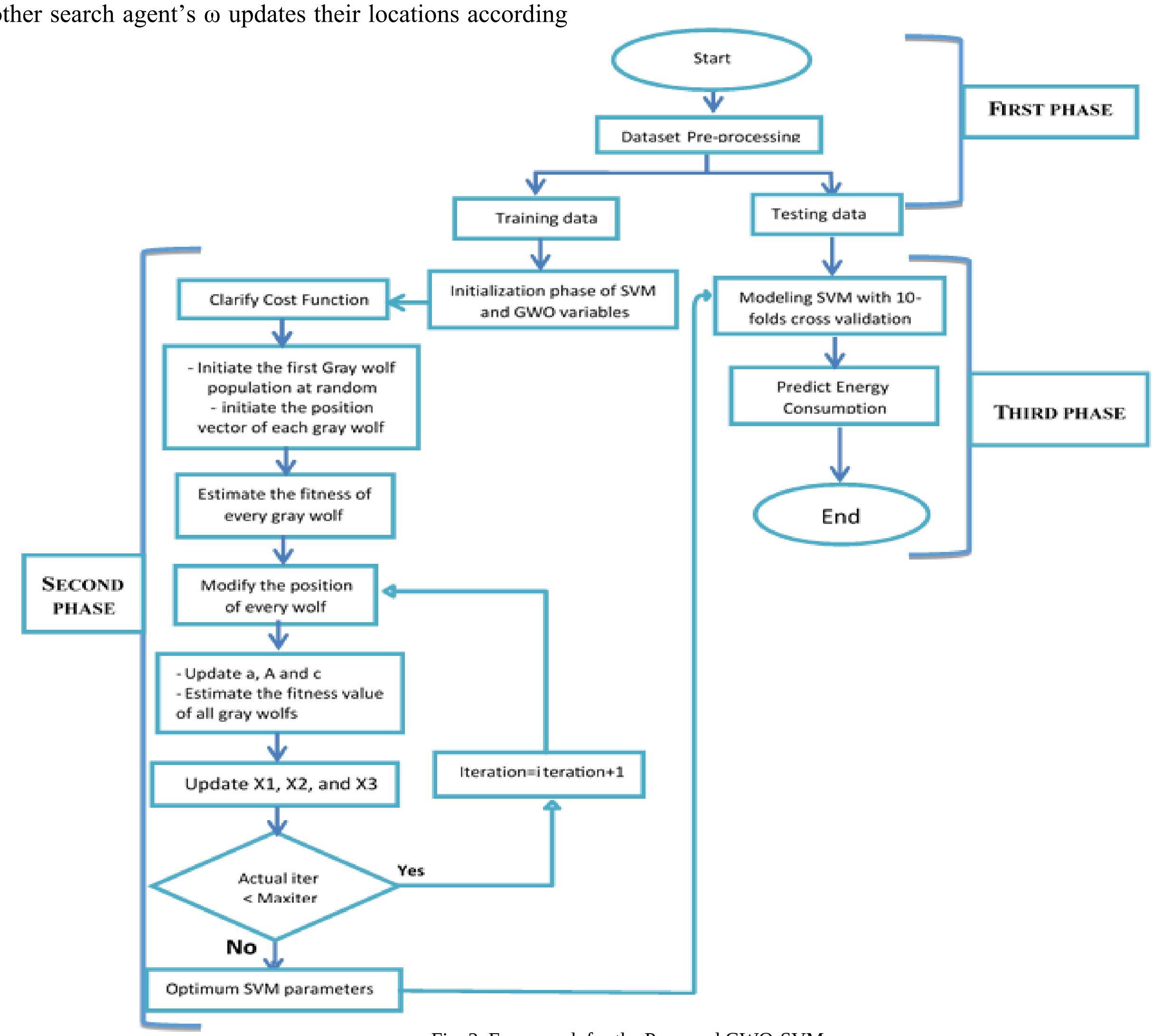

Fig .2. Framework for the Proposed GWO-SVM

IV. THE PROPOSED HYBRID MODEL GWO-SVM

The proposed hybrid model is to predict BEC and energy in general for two periods, short and long period. Moreover, the model is applied using five different datasets. For the robustness of the proposed model, different models were applied to demonstrate that the proposed model has higher accuracy and least time consumption. The details will be illustrated in the following subsections.

1. Pre Processing Phase

This phase prepares the data to make a prediction model and enhance its robustness depending on the three major stages.

In the first stage, it eliminates missing and redundant data that does not give an advantage to that individual object when training the model.

The second stage split data into training and testing since the model needs much training. Therefore, a big chunk of data with 80% is for training and 20% is for testing.

The third stage normalizes the dataset since the input dataset has multiple groups with various dimensions, which affect the prediction efficiency and the reliability of the accuracy. Moreover, this stage helps in the creation of a standard scale of data parameters without changing variations in the range of features. This means that the training and testing input data were normalized independently using the min-max method for normalization according to the following:

$$X_i^{normalized} = \frac{xi - x_{\min}}{x_{\max} - x_{\min}} \qquad (11)$$

where $X_i^{normalized}$ is the data normalized value, X_i is a data input, X_{min} is the smallest data value in the whole dataset, and X_{max} is the largest data value in the whole dataset.

2. Optimization of SVM Based-GWO Phase

In this phase, the optimization of SVM parameters which are kernel variable σ and penalty coefficient c is done using the proposed GWO algorithm according to the following steps.

TABLE V
MAJOR STEPS OF GWO-SVM

Steps	Process
Step 1:	Initiate number of variables of GWO, size of Population, range of variables which indicate the lower and upper bound of each variable, Parameters associated with SVM, the upper limit number of iterations and the penalty of coefficient vectors.
Step 2:	Initiate the first Gray wolf population at random and the position vector of each gray wolf.
Step 3:	Clarify Cost Function and estimates the fitness value of every grey wolf and selects alpha, beta and delta using SVM.
Step 4:	Modify the position of every wolf (X1, X2, X3) with best position for each one using equation (12).
Step 5:	Estimate the fitness value of all gray wolfs with new values and update their positions.
Step 6:	If the fitness value of the new position is better, then Check termination formula, if it meets the goal, make alpha the optimum solution for given problem. Otherwise, incriminate iteration by 1 and go to step 4.

3. Prediction phase:

This phase handles predicting BEC using the optimized SVM features that involve C cost-parameter and σ kernel-parameter of radial-basis kernel. Subsequently, the GWO produces optimal features and SVM parameters, which have a significant effect on the model prediction accuracy.

The selected parameters are used as input for SVM modeling to perform the prediction process using optimal variables obtained from the GWO. The process is performed according to the following steps:

Step 1: If the termination criteria is achieved, take SVM parameters to the model SVM using parameters from GWO algorithm on testing data.

Step 2: Predict BEC.

Step 3: Measure accuracy by computing the four metrics.

V.EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Set

The energy consumption datasets utilized in this paper are all real from energy departments in the USA. We selected five datasets of energy consumption in different regions of USA.

The dataset parameters could be split into three types: the meteorological dataset (outdoor air temperature, relative humidity, wind speed, air pressure, heating degree days, cooling degree days), time parameter (hour of the day, date), and historic dataset (energy consumption).

Dataset A:

The publicly available "Monthly and Annual Energy Consumption by Sector" dataset used in this research, which consists of four datasets is about monthly historic and recent data of energy consumption from January 1973 to May 2021. In USA, the total power consumed by the end-user of four different sectors: commercial, industrial, residential, and transportation sector. Seven zones include New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, and West South Central, with two meteorological parameters of heating and cooling degrees representing the outdoor air temperature, which is available at the energy department of the "Data.gov" website.

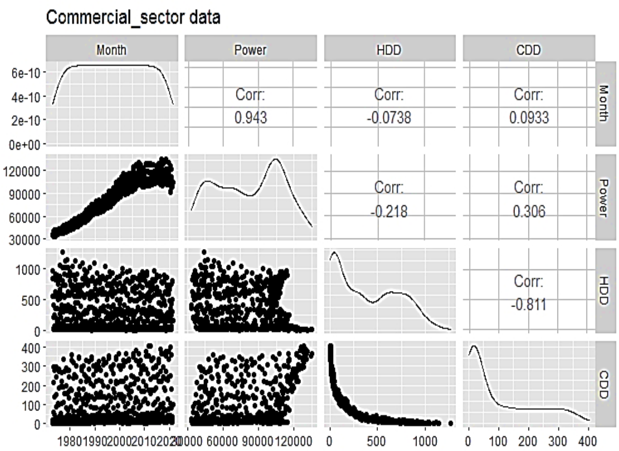

Fig. 3. Visualization graph of Dataset A parameters

Fig. 3. The applied parameters density and scatter plots as well as the illustration of the correlations between parameters for commercial sector data according to the applied dataset; the commercial sector is the most consumers for energy.

Data set B:
The fifth dataset is about commercial buildings include weather parameters that have the most effect on BEC [43]. The dataset comprises eight variables (outdoor air temperature, energy consumption, cooling degree days above zero, heating degree days, surface pressure, relative humidity at ten meters, wind speed at ten meters and precipitation) with the size of 34941 records to predict short-term energy consumption periods six months, and long-term for one year. The dataset used is from the U.S Department of Energy and a dataset collected from the past and current estimations from centers of Industrial Assessment all over the countries “https://iac.university/” [44].

The applied dataset is a combination of two categories: one is the energy usage from the U.S. and the other from the Department of Energy, which is connected with the other meteorological dataset from the National Center for Environmental Information to provide climate and historical weather data obtained from UCI Machine Learning Repository SML 2010 Dataset. The repository is collected from a monitor system mounted in a commercial building over one year. The data range from January 2010 to December 2010 every 15 minutes is 34941 records used for the aim of this study, which provides data about energy consumption and outdoor temperature for buildings in Fremont, California, and United States.

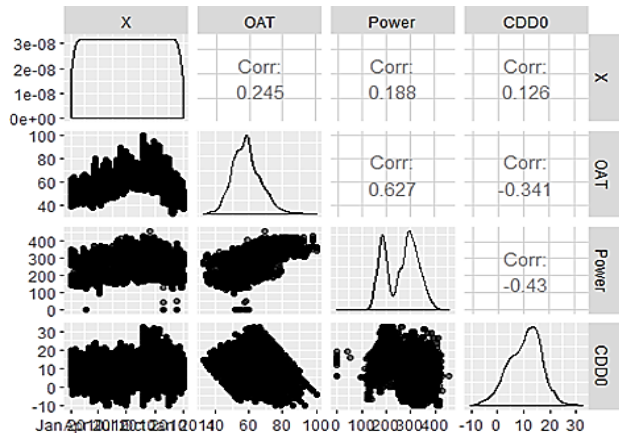

Fig. 4. Visualization graph of the energy consumption parameters

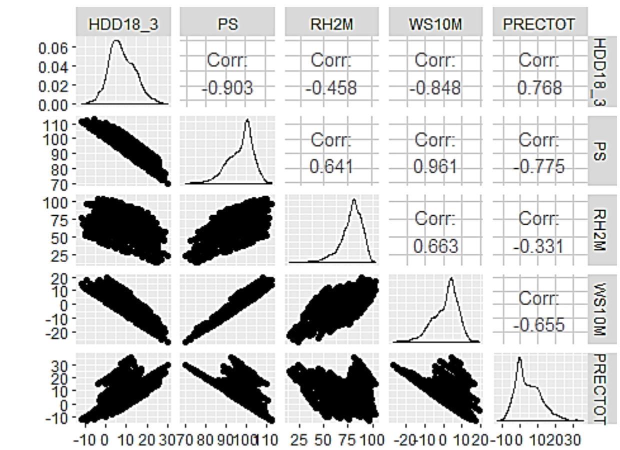

Fig. 5. Visualization graph for the environmental parameters

Fig.4 and Fig. 5 are histograms for dataset B parameters, which give the nearest look to the parameters and decide beforehand that a worthy model would be produced.

Table VI represents the applied dataset parameters, showing their summarized names and their corresponding units of measurement

Table VII shows the descriptive statistics: min, median, mean, and max values of the dataset parameters, which is important understanding and analyzing the model. It helps to determine the most affected parameters on energy consumption (like outdoor air temperature (OAT), PS, and RH2M). This shows that there is a linear association between power consumption and OAT. Specifically, when the OAT degree rises, the power consumption also raises.

TABLE VI
SAMPLE PARAMETERS

parameter	Name	Measurement unit
Outdoor Air Temperature	OAT	degrees Fahrenheit or Celsius
Energy consumption	Power	kWh
Cooling Degree Days Above 0	CDD0	°F Or ° Celsius
Heating Degree Days	HDD18_3	° Celsius
Surface Pressure	PS	Pascal (Pa)
Relative Humidity at 10 Meters	RH2M	Grams per cubic meter
Wind Speed at 10 Meters	WS10M	knot (nautical mile per hour = 0.51 m sec-1 = 1.15 mph). or miles per hour
Precipitation	PRECTOT	Millimeters

TABLE VII
SAMPLE PARAMETERS

	Energy Factors			
parameters	min	median	mean	max
OAT	33.00	58.00	58.58	100.00
Power	0.0	276.7	261.6	457.9
	Environmental Factors			
parameters	min	median	mean	max
CDD0	-10.320	11.18	10.58	32.62
HDD18_3	-12.000	7.160	7.829	30.000
PS	69.98	98.87	97.44	112.63
RH2M	14.28	72.49	77.98	103.07
WS10M	-62.920	2.450	1.217	19.170
PRECTOT	-12.880	3.180	4.625	35.620

In this study, the implemented experiments have been conducted using the R environment along with e107, GA, metaheuristicOpt, hydroGOF, Metrics and MLmetrics open-source libraries, and caret packages.

A. Prediction results and comparative analysis

To evaluate the prediction performance error of the proposed algorithms and the other applied techniques, the four metrics criteria, namely, RMSE, MAE, MAPE, and MSE are applied as follow:

The MAPE estimates the difference in the forecast and actual values in the absolute percentage (In. (12)) [45]:

$$MAPE = \frac{1}{n}\sum\frac{(yi' - yi)}{yi} * 100 \quad (12)$$

The MSE combines the estimation of the variance between the main model target and what is predicted with bias of the prediction (In. (13)) [51]:

$$MSE = \sum_{i=1}^{n}\frac{(yi' - yi)2}{n} \quad (13)$$

The MAE demonstrates the linear variation of actual and forecast values, which give small weight values to Outliers (data points that are out of the linear model) (In. (14)) [46]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(yi' - yi)| \quad (14)$$

Several models have been applied in this research; therefore, we are looking for the best model for predicting power consumption based on various environmental factors.

The RMSE shows the quadratic variance of forecast and actual values under the square root. It gives perfect results because using the squared error allows that positive and negative errors do not drop each other (In. (15)).

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(yi' - yi)2}{n}} \quad (15)$$

where $\hat{y}_i$ is the predicted value, y_i is the experimental value, and n is the observation number. We measure the model accuracy through the RMSE value as stated in the NDEP (National Digital Elevation Guidelines) and FEMA guidelines, where, Accuracy = 1.96 × RMSE.
To prove the high performance of the proposed model in the prediction of energy consumption, five various prediction models, namely MLR, ANN, SVM, GA-SVM, and GWO-SVM with the proposed hybrid model are constructed and estimated. Four metrics are estimated to show the prediction accuracy of the proposed hybrid model. The fifth models' evaluation was compared using various datasets as showed above in Fig. 6–9. We verified the efficiency of GWO-SVM over the other models.

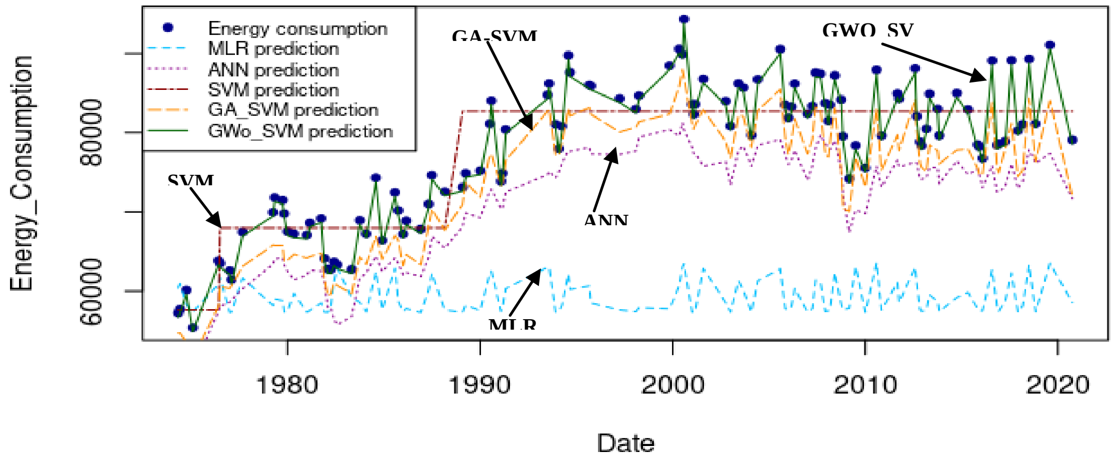

Fig. 6. Industrial Sector Prediction Results

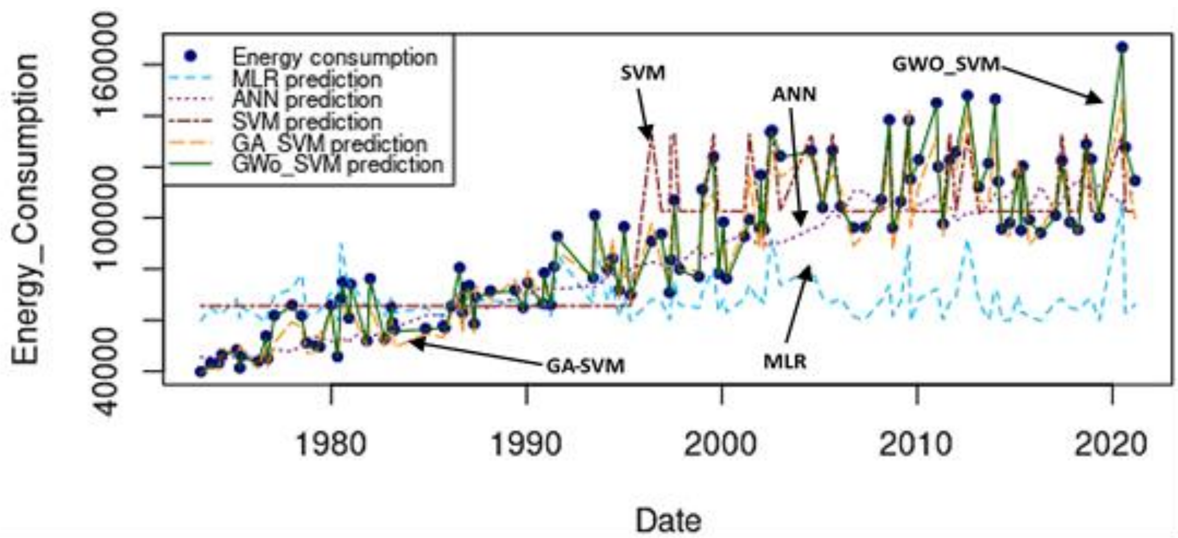

Fig. 7. Residential Sector Prediction Results

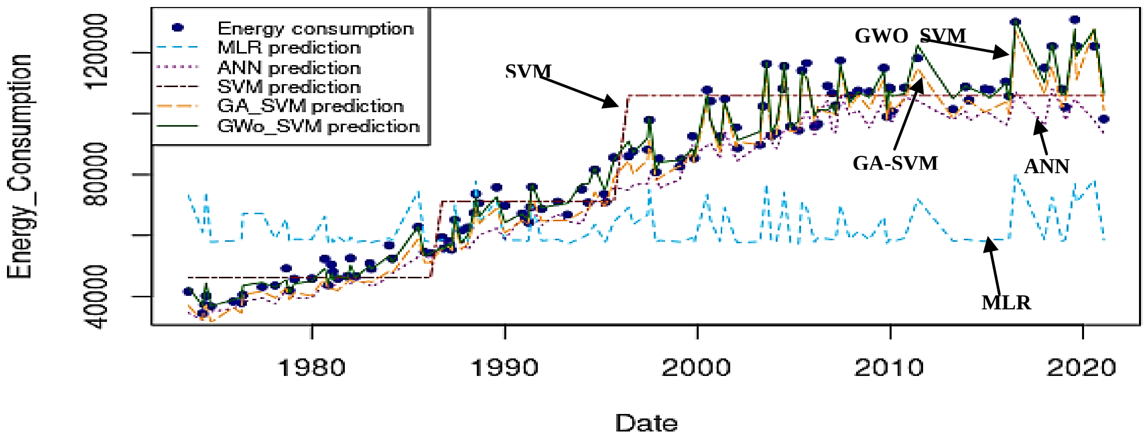

Fig. 8. Commercial Sector Prediction Results

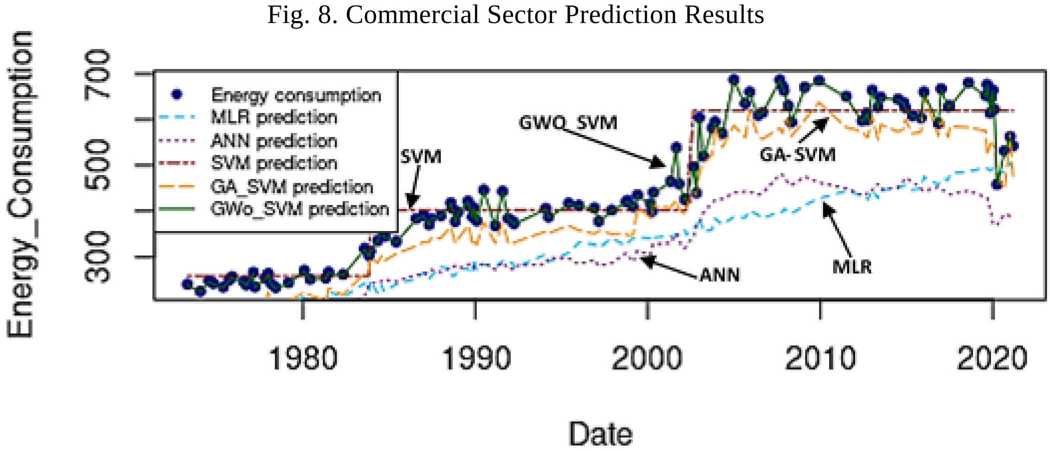

Fig. 9. Transportation Sector Prediction Results

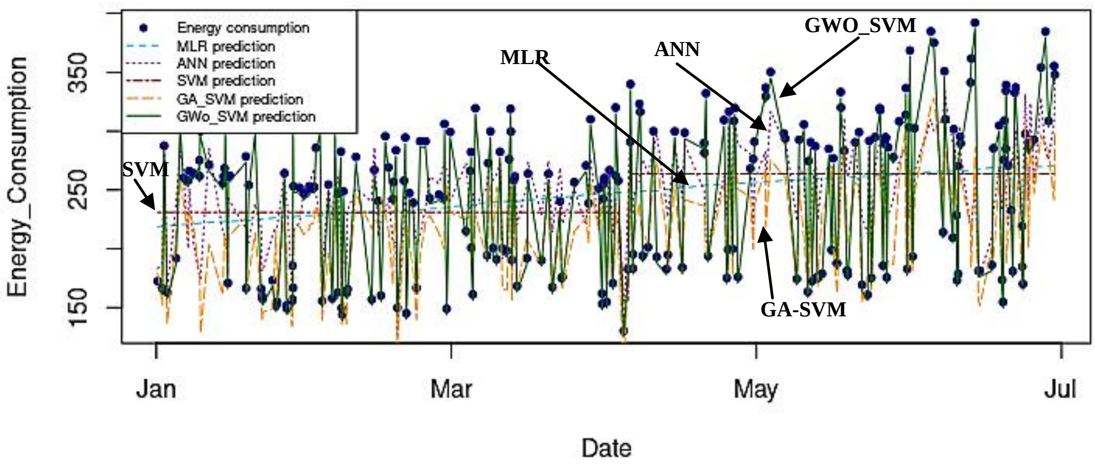

Fig. 10-short-period prediction of dataset B

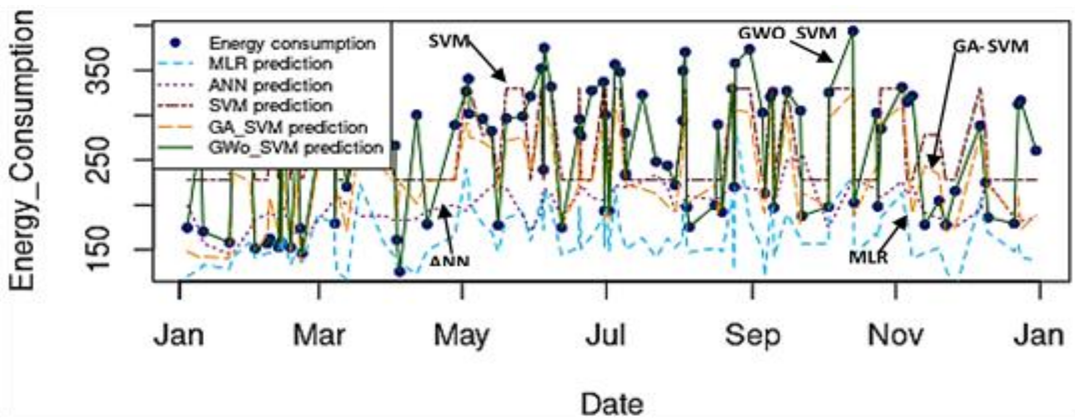

Fig. 11. Long-period Prediction of dataset B

Fig. 6–9 show the prediction rate of every model for the dataset A, the five various models performance indicators of their prediction with the real energy consumption points on the four sectors datasets.

Fig. 10–Fig 11 shows the prediction rate of every model for dataset B, the five various models' performance indicators of their prediction with the real energy consumption points in a short and long period of 6-months and year prediction respectively, where: The 5 model's prediction results.

Our result shows that the proposed model GWO-SVM reaches every point with a slight variance, whereas the GA-SVM and optimized SVM are very far from the actual energy consumption, which is clearly shown in the accuracy of graphs and prediction figures.

The proposed GWO-SVM model has a high accuracy result than other techniques as shown in Fig. 12–Fig 13. It shows the accuracy value for predicting BEC for one year, and the execution time also is the best among other applied techniques as shown in Fig. 14.

Fig. 12–Fig. 14 confirm that the proposed GWO-SVM algorithm is more efficient than the GA-SVM, SVM, ANN, and MLR, thus showing the superiority of GWO-SVM algorithm in the prediction process of BEC.

Fig. 12–Fig. 14 and Table VIII–Table X show that, the accuracy of SVMs without applying GWO is so low, and its value is 81.03%. However, it reached 98.012% using GWO-SVMs, which indicates that the accuracy increased by≈ 17.09%, but the execution time increased by 9.04 minutes comparing to the running of the original SVM not the Tuned SVM which decreased with 50 minutes. Where SVM consumed only 6 seconds but with small accuracy, so its parameters were optimized using SVM Tune Function, which consumed one hour.

The GWO-SVM achieved a progress also in the accuracy and execution time compared to the accuracy of GA-SVM, which is 91.33% less than the accuracy of GWO-SVM that increased by ≈ 7. 02% and decreased the execution time.

Furthermore, ANN achieved an accuracy of 75.07% in one day, which is very low and not comparable with the accuracy or time of GWO-SVM that surpasses it with ≈ 23.722%. Moreover, MLR has the minimal number of accuracy among different algorithms, which equal to 37.28% although its execution time is better than that of GA-SVM, but still, the GWO-SVM outperformed them in accuracy and execution time.

Fig 14 shows the execution time, which proves the effectiveness of the proposed GWO-SVM algorithm than other applied algorithms. Where SVM took 6 seconds in run but with small accuracy, so the optimization function "Tune" applied to enhance it but with more time than the original SVM", it consumed one hour. It indicates the superiority of GWO-SVM algorithm in accuracy and execution time.

Table VIII–Table X represents performance metrics (RMSE, MSE, MAPE, and MAE errors) of the five applied models and their values for the used dataset A and dataset B.

Table X shows prediction metrics of Dataset B for two different time periods. The forecasting of short-period estimated for six-months prediction and the long-period estimated for one year prediction.

The metric results confirm that the hybrid GWO-SVM surpasses the other techniques as it gives better accuracy than GA-SVM with 32%, surpasses ANN with 45%, MLR with 24%, and SVM with 6%.

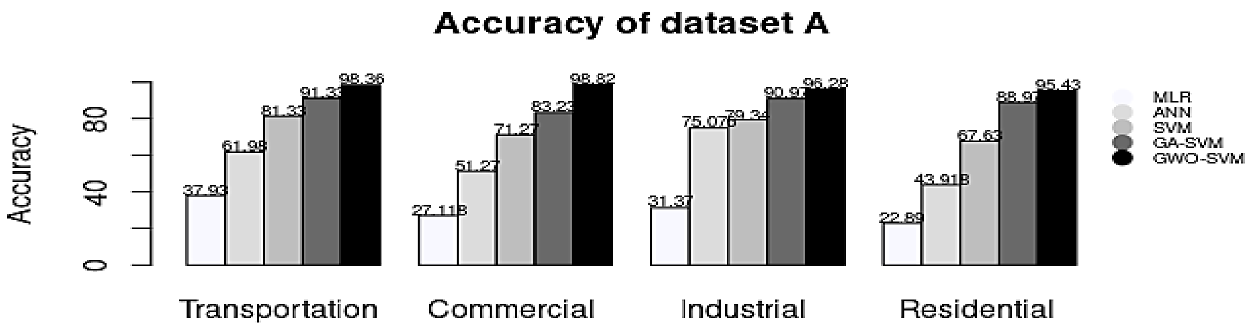

Fig. 12. Accuracy comparison graph for Dataset A

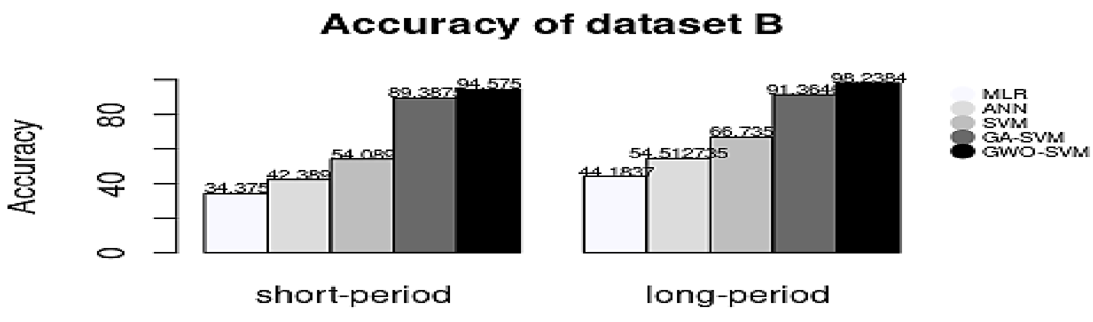

Fig. 13. Accuracy comparison graph for short and long period for Dataset B

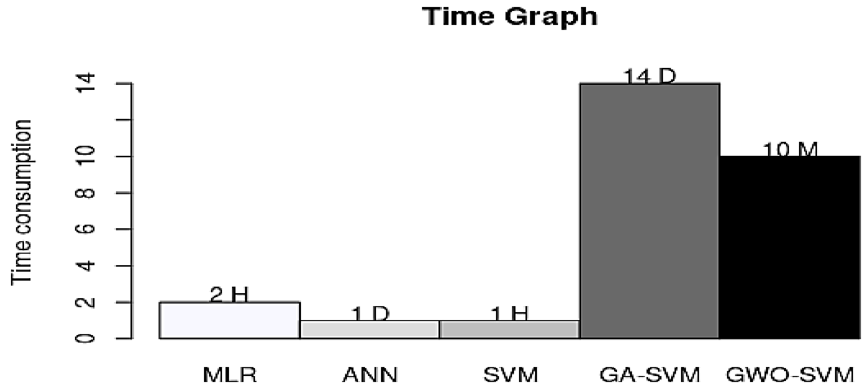

Fig. 14. Time consumption graph of 5 different methods
Where: H for number of Hours, M For minutes and D for days.

TABLE VIII
RESULTS PERFORMANCE OF THE FIVE APPLIED TECHNIQUES FOR DATASET A

Models	Industrial				Commercial			
	MAE	MSE	MAPE	RMSE	MAE	MSE	MAPE	RMSE
MLR	26262.3	347.8148	0.5125995	27575.96	31463.91	110092139	0.599702	33180.14
ANN	12692.63	170189975	0.1956782	13045.69	81033.07	73031	81033.07	85458.44
SVM	3467.451	18768589	0.04398204	4332.273	7866.715	108334167	0.09352581	10408.37
GA-SVM	2574.123	10660344	0.03487474	3265.018	4996.839	38556971	0.06114464	6209.426
GWO-SVM	1275.537	3061200	0.01630367	1749.628	2185.377	8767616	0.02722656	2961.016

TABLE IX
RESULTS PERFORMANCE OF THE FIVE APPLIED TECHNIQUES FOR DATASET A

Models	Transportation				Residential			
	MAE	MSE	RMSE	MAPE	MAE	MSE	MAPE	RMSE
MLR	126.2016	522.3557	152.5387	0.3042642	32340	3061200	0.5532712	39554.42
ANN	74.14095	7092.503	84.217	0.194082	15311.9	3930688	0.2038274	19825.97
SVM	43.53867	2797.652	52.89283	0.106117	13276.82	267501796	0.1568927	16355.48
GA-SVM	26.73905	1474.059	38.39348	0.0601985	4927.009	39171841	0.05567542	6258.741
GWO-SVM	13.55468	347.8148	18.64979	0.02945243	4726.172	37607163	0.05362217	6132.468

TABLE X
RESULTS PERFORMANCE OF THE FIVE APPLIED TECHNIQUES FOR DATASET B

Models	Short-period				Long-period			
	MAE	MSE	RMSE	MAPE	MAE	MSE	MAPE	RMSE
MLR	48.6445	3219.143	56.7375	0.272869	40.24384	2640.719	0.164318	51.38793
ANN	40.506	2365.56	48.6370	0.162126	37.7877	2518.11	0.1526563	50.1807
SVM	35.55744	2155.434	46.42665	0.154294	29.03963	1671.403	0.1273365	40.8828
GA-SVM	15.39259	608.9675	24.67727	0.061849	9.12102	194.573	0.0353219	13.9489
GWO-SVM	8.456711	188.8061	13.74067	0.035355	1.17999	13.9427	0.0111436	3.7340

VI. CONCLUSION

This study proposed a hybrid GWO-–SVM algorithm based on 10-fold cross- validation, which optimizes the energy consumption prediction, and employs other machine learning techniques such as ANN and SVM, and statistical regression MLR, and metaheuristic GA-–SVM. To solve the complex problem of energy consumption, nonlinearity of environmental parameters, and prediction time, we employed real- datasets containing the data of monthly energy consumption of four government sectors in the US (commercial, residential, transportation, and industrial sectors) from January 1973 to May 2021. Using two additional environmental parameters as heating and cooling degree days for the training and testing of the proposed model, the fifth dataset containing the hourly energy consumption data for a commercial building was considered. This dataset contained 34,941 records of eight variables (OAT, energy consumption, cooling degree days above zero, heating degree days, surface pressure, relative humidity at 10 m, wind speed at 10 m, and precipitation) for the validation and prediction of two periods (six months and one year). We compared the performance of each technique with the proposed algorithm using several performance metrics (RMSE, MSE, MAPE, and MAE).

The compared techniques achieved prediction accuracies in the range of 22.28%–91.41%. The proposed model achieves an accuracy of 98.012% and a prediction time of 10 min for the one-year prediction. Proving the high efficiency and robustness of the proposed model is over the other compared models.

REFERENCES

[1] Miaomiao Wang, Qingwen Luo, Lulu Kuang, and Xiaoxi Zhu, "Optimized Rolling Grey Model for Electricity Consumption and Power Generation Prediction of China," *IAENG* International Journal of Applied Mathematics, Vol.49, no.4, pp577-587, 2019.

[2] Soares, N., Costa, J.J., Gaspar, A.R. and Santos, P., 2013. Review of passive PCM latent heat thermal energy storage systems towards buildings' energy efficiency. Energy and buildings, 59, pp.82-103.

[3] Li, K., Xue, W., Tan, G. and Denzer, A.S., 2020. A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms. Building Services Engineering Research and Technology, 41(1), pp.108-127.

[4] T. Hong, S.C. Taylor-Lange, S. D'Oca, D. Yan, S.P. Corgnati, Advances in research and applications of energy- related occupant behavior in buildings, Energy Build. 116 (2016) 694–702.

[5] A. Kavousian, R. Rajagopal, M. Fischer, Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings, Energy Build. 99 (2015) 220–230.

[6] Z. Guo, Z.J. Wang, A. Kashani, Home appliance load modeling from aggregatedsmart meter data, IEEE Trans. Power Syst. 30 (1) (2015)254–262.

[7] Li K, et al. Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energy Build 2015; 108: 106–113.

[8] Jovanovic´ RZˇ , Sretenovic´ AA and Zˇ ivkovic´ BD.Ensemble of various neural networks for prediction of heating energy consumption. Energy Build 2015; 94:189– 199.

[9] Virote J and Neves-Silva R. Stochastic models for building energy prediction based on occupant behavior assessment. Energy Build 2012; 53: 183–193.

[10] Aleksandra, A.S., JOVANOVI, R.Ž., NOVAKOVI, V.M., NORD, N.M. and Branislav, D.Ž., 2018. Support vector machine for the prediction of heating energy use. Thermal Science, 22(4), pp.S1171-S1181.

[11] Xuanyu Liu, Yao Zhang, and Kaijuu Zhang, "Optimization Control of Energy Consumption in Tunneling System of Earth Pressure Balance Shield Tunneling Machine". Engineering Letters, vol.28, no.2, pp551-558, 2020.

[12] M.J. Mbunwe, B.O. Anyaka, and U.C. Ogbuefi, "An Effective Energy Saving Design to Maximize The Use of Electricity in Developing Country". Lecture Notes in Engineering and Computer Science; Proceeding of The World Congress on Engineering and Computer Science 2019, 22-24 October, 2019,San Francisco, USA, PP188-197.

[13] Joy Dassgupta, and Peter Pudney, "Real-time Control of Indoor Lighting to Minimise Energy Use," Lecture Notes in Engineering and Computer Science; Proceeding of The World Congress on Engineering 2019, 3-5 July, 2019, London, U.K., PP181-186..

[14] Lan Zhang, Lina Luo, Lei Hu, and Maohua Sun, M., "An SVM-Based Classification Model for Migration Prediction of Beijing," Engineering Letters, vol.28, no.4, pp1023-1030, 2020.

[15] Zhou, D.Y., Xu, Q., Xue, X.Y., Zhang, F.F. and Liang, X.M., 2009. Characterization of polymethoxylated flavones in Fructus aurantii by off-line two-dimensional liquid chromatography/electrospray ionization-ion trap mass spectrometry. Journal of pharmaceutical and biomedical analysis, 49(2), pp.207-213.

[16] Mirjalili, S., Mirjalili, S.M. and Lewis, A., 2014. Grey wolf optimizer. Advances in engineering software, 69, pp.46-61.

[17] Ciulla, G. and D'Amico, A., 2019. Building energy performance forecasting: A multiple linear regression approach. Applied Energy, 253, p.113500.

[18] Dong, B., Cao, C. and Lee, S.E., 2005. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings, 37(5), pp.545-553.

[19] Faisal, H.M., Javaid, N., Sarfraz, B., Baqi, A.,Bilal, M., Haider, I. and Shuja, S.M., 2019, March.Prediction of Building Energy Consumption Using Enhance Convolutional Neural Network. In Workshops of the International Conference on Advanced Information Networking and Applications (pp. 1157-1168). Springer, Cham.

[20] Jitkongchuen, D. and Pacharawongsakda, E., 2019, January. Prediction Heating and cooling loads of building using evolutionary grey wolf algorithms. In 2019 joint international conference on digital arts, media and technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON) (pp. 93-97). IEEE.

[21] Alshibani, A., 2020. Prediction of the Energy Consumption of School Buildings. Applied Sciences, 10(17), p.5885.

[22] Wei, Y., Ni, N., Liu, D., Chen, H., Wang, M., Li, Q., Cui, X. and Ye, H., 2017. An improved grey wolf optimization strategy enhanced SVM and its application in predicting the second major. Mathematical Problems in Engineering, 2017.

[23] Dai, S., Niu, D. and Han, Y., 2018. Forecasting of power grid investment in china based on support vector machine optimized by differential evolution algorithm and grey wolf optimization algorithm. Applied Sciences, 8(4), p.636.

[24] Tian, Y., Yu, J. and Zhao, A., 2020. Predictive model of energy consumption for office building by using improved GWO-BP. Energy Reports, 6, pp.620-627.

[25] Dong, Bing, Cheng Cao, and Siew Eang Lee. 2005. "Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region." Energy and Buildings 37 (5): 545–53.

[26] Dai, S., Niu, D. and Han, Y., 2018. Forecasting of power grid investment in china based on support vector machine optimized by differential evolution algorithm and grey wolf optimization algorithm. Applied Sciences, 8(4), p.636.

[27] Paudel, S., Nguyen, P.H., Kling, W.L., Elmitri, M., Lacarrière, B. and Corre, O.L., 2015. Support vector machine in prediction of building energy demand using pseudo dynamic approach. arXiv preprint arXiv:1507.05019.

[28] Cortes, C. & Vapnik, V. 1995. Support-Vector Networks. Machine Learning, Vol. 2, pp 273-297.0

[29] Tso, Geoffrey K. F., and Kelvin K. W. Yau. 2007. "Predicting Electricity Energy Consumption: A Comparison of Regression Analysis, Decision Tree and Neural Networks." Energy 32 (9): 1761–68.

[30] Iain Pardoe. 2012. "Applied Regression Modeling – A Business Approach." John Wiley & sons.

[31] Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M. and Zhao, X., 2018. A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews, 82, pp.1027-1047.

[32] Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Wasserman,W. Applied Linear Statistical Models; Irwin: Chicago, IL,USA, 1996.

[33] Zhao, H.X. and Magoulès, F., 2012. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6), pp.3586-3592.
[34] Cortes, C. & Vapnik, V. 1995. Support-Vector Networks. Machine Learning, Vol. 2, pp 273-297.0
[35] Lee, S., Jung, S. and Lee, J., 2019. Prediction model based on an artificial neural network for user-based building energy consumption in South Korea. Energies, 12(4), p.608.
[36] Du, L., Peeta, S. and Kim, Y.H., 2012. An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks. Transportation Research Part B: Methodological, 46(1), pp.235-252.
[37] H. Ceylan, H. Ozturk, "Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach", Energy Conversion and Management (45), pp. 2525–2537,2004.
[38] Azadeh, A., Tavakkoli-Moghaddam, R. and Tarverdian, S., Electrical Energy Consumption Estimation by Genetic Algorithm and Analysis of Variance. Working Paper, nd.
[39] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw. 2014;69:46–61.
[40] Kamel, S.R., YaghoubZadeh, R. and Kheirabadi, M., 2019. Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. Journal of Big Data, 6(1), p.90.
[41] Mirjalili, Seyedali and Mirjalili, Seyed Mohammad and Lewis, Andrew, “Grey wolf optimizer”, In Journal of Advances in Engineering Software, vol. 69 , pp. 46-61, 2014,Elsevier.
[42] El-Gaafary, Ahmed AM and Mohamed, Yahia S and Hemeida, Ashraf Mohamed and Mohamed, Al-Attar A,“Grey wolf optimization for multi input multi output system”,In Universal Journal of Communications and Network , vol.1, no.1, pp. 1-6, 2015.
[43] Chae, Y.T., Horesh, R., Hwang, Y. and Lee, Y.M., 2016. Artificial neural network model for forecasting sub- hourly electricity usage in commercial buildings. Energy and Buildings, 111, pp.184-194.
[44] Centers of Industrial Assessment open repository Dataset Science 25-01 May, 1973, USA, available at the website: https://iac.university/
[45] Xuan Z, Xuehui Z, Liequan L, Zubing F, Junwei Y, Dongmei P. Forecasting performance comparison of two hybrid machine learning models for cooling load of a large- scale commercial building. J Build Eng 2019;21:64–73.
[46] Ahmad T, Chen H. Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. Energy Build 2018;166:460–76.
[47] Yezioro A, Dong B, Leite F. An applied artificial intelligence approach towards assessing building performance simulation tools. Energy Build 2008;40:612– 20.