Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SHORT-TERM LOAD FORECASTING USING SUPPORT VECTOR REGRESSION WITH AFRICAN-BUFFALO OPTIMIZATION ALGORITHM

INUSA SANI MAIJAMA'A[1], YUHANIS YUSOF[2], MOHAMAD FARHAN MOHSIN[3]

[1]School of Computing, University Utara Malaysia, Malaysia
[1] Department of Computer Science, Hussaini Adamu Federal Polytechnic, Nigeria
[2]School of Computing, University Utara Malaysia, Malaysia
[3]School of Computing, University Utara Malaysia, Malaysia

ABSTRACT

Short-term electric load forecasting is a critical issue due to idiosyncrasies associated with production and consumption of electricity as it cannot be stored in large quantity for future use. Electricity load forecasting remains a challenging task due to non-linearity and uncertainty of associated forecasting variables. This paper presents a hybrid technique based on Support Vector Regression (SVR) optimized using African Buffalo Optimization (ABO) algorithm to predict fourteen (14) days ahead of electricity consumption. Comparison of forecasting performance against PSO and GA as optimizing algorithms for SVR hyper-parameters based on Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Deviation (MAD) has been carried out. Experiment results has shown that ABO algorithm was able to determine optimal parameters of SVR better than state-of-the-art swarm-based optimization algorithms.

Keywords: *African Buffalo Optimization (ABO), Energy Forecasting, Machine Learning, Support Vector Regression (SVR), Swarm Intelligence*

1.0 INTRODUCTION

Electrical energy is one of most vital commodities that drives today's economic activities [1]. However, electric energy cannot be reserved in larger quantities for future usage. Hence, there is a need for accurate forecasting of required consumption quantities at any given time to avoid shortage and wastage. Importance of electric load forecasting ranges from its vital role in daily operational management of a power utility, such as energy transfer scheduling, unit commitment, load dispatch, and so on. Emergence of load management strategies makes it highly desirable to develop accurate, fast, simple, robust and interpretable load forecasting models in order to achieve higher reliability and better operational management.

The importance of forecasting in power sector cannot be overemphasized. This is due to the fact that small margin of increased accuracy can result into significant yield of profit as reported by [2]. The authors argued that small improvement of forecasting accuracy by one percent (1%) could result into cost reduction of about 0.1% - 0.03%. This in turn yielded a profit gain of about $1.5 million per year for a medium-size utility installation with a five (5) GW peak load.

Various techniques have been employed for Electric Load Forecasting (ELF). These techniques can be broadly categorised into statistical and computational intelligence [3]. The most popular among statistical techniques are time series techniques that comprises of Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and its variants [4], and Exponential Smoothing models [5]. However, statistical techniques were found to be poor of performance [6] due to their inability to deal with complex and non-linearity in multivariate data [7], [8]. Hence, they are not suitable to be used for electric load type of data which is complex and non-linear [9]. These highlighted shortcomings of statistical techniques prompt researchers to explore computational intelligence methods. Artificial Neural Network (ANN) and Support Vector Machines (SVM) [10] have been the most popular computational intelligence methods in time series studies. ANN is a family of learning methods that are founded based on the working of biological neural brain, mostly employed to determine estimate value of values of unknown functions [11].

Despite the satisfactory forecasting result recorded by ANN-based techniques in literature [4], [6] as a result of its ability to solve non-linear and complex problems, yet the technique is time consuming and is subject to over training which led to overfitting [12]. This is because ANN adopts Empirical Risk Minimisation (ERM) approach [4] which focus on minimising training errors. Another shortcoming of ANN is the presence of numerous control parameters that have to be optimized [13]. These control parameters include but not limited to number of hidden layers and type of activation function required per layer type. This makes ANN-based models to be complicated and not suitable to be used in power consumption forecasting.

In order to overcome the problems of ANN, SVM introduced by Vapnik [14] were proven to be most effective due to its adoption of Structural Risk Minimisation (SRM) approach [15]. The SRM approach focuses on minimising the generalisation error instead of minimising training errors as done by ERM. This makes SVM able to overcome the problem of overfitting therefore capable to achieve good generalisation.

The Support Vector Regression (SVR), as a version of SVM, that is meant for regression task [15], has proven to be powerful in the field of load forecasting [16]. However, the generalisation performance of SVR relies on two parameters values [17] which are cost error (C) and tube size (ε) [18]. Manual selection of these parameters can be a complex task. This necessitates the need to find best approach in determining the optimal value of these parameters in order to get best generalisation of SVR that would eventually lead to higher forecasting accuracy.

Two major approaches became popular in optimization process viz grid search and metaheuristics techniques. As for the grid search techniques, it has been reported that they are computationally expensive and usually reported high error rate [19]. This makes grid search technique unsuitable to optimize SVR parameters. Hence, opening the alternative of using swarm intelligence (SI) techniques.

The use of SI based techniques as a means of determining optimal values for SVR parameters through hybridisation [3] has being yielding satisfactory results [20] . Recently a new swarm-based optimization algorithm namely African Buffalo Optimization (ABO) has been introduced by Odili [21]. ABO algorithm has been used in different areas of optimization including team formation [22], Travelling Salesman Problem (TSP) [23], selection of biodiversity conservation area with a given constraint [24] and parameter tuning of PID controller [25]. Various studies have reported successes of ABO due to its fast convergence, ability to track the best position, and speed of each buffalo as well as the movement of best buffalo towards better exploration [26].

Due to the above-mentioned reported successes of ABO, this study proposes the use of ABO as an optimizing algorithm for SVR hyper-parameters. Such approach strengthens the methods in time series forecasting, particularly in the area of swarm intelligence for short term forecasting.

2.0 RELATED LITERATURE

Numerous research studies found in literature demonstrated the capability of swarm intelligence techniques for optimization processes in different areas. It is evident in literature that SVR hyper-parameters have been optimized using different swarm intelligence-based algorithms. Siyang et, al., [27] use an ABC to optimize SVR parameters in multi-strategic procedure for predicting annual total electricity consumption in China. Authors in [3] employs cuckoo search algorithm to optimize SVR for household electric as short-term electric load forecasting. Ming-Wei et, al., [28] utilize SVR as prediction algorithm to forecast microgrid electric load as oppose to large power grid. The authors argued that SVR is time consuming for learning when sample is large. As such, they employ fruitfly optimization algorithm (FOA) to optimize SVR parameters. Chou et. al., [29] use SVR for prediction of exchange rate between Canadian dollar and United States dollar (USD). The authors optimize the SVR parameters using enhanced firefly algorithm (FA) based on sliding-window technique. Jui-Sheng Chou & Phma [30] predict the scour depth effect caused by flowing water against bridge. The authors use SVR as prediction algorithm optimized with Firefly algorithm. Authors in [31] use flower pollination algorithm (FPA) as SVR parameters optimization

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

algorithm for prediction of algal colony growth on façade structure. Verma et. al., in [32] used optimized SVR for cement compressive strength based on multivariate parameters. The authors used PSO and Symbiotic Organization Search (SOS) as the optimization algorithms for the SVR parameters.

As presented, optimization of SVR hyper-parameters using swarm intelligence techniques is an established practice in research community. However, ABO algorithm as a recently introduced swarm intelligence algorithm has not been employed for SVR hyper-parameters optimization, hence performance of hybrid algorithm of SVR-ABO is yet to be established in forecasting area.

3.0 THEORETICAL CONCEPT

3.1 Support Vector Regression

This section present the theory behind SVR equations as given by [14]. Considering a data training set, L, represented by:

$$L = \{(x_1, y_1)(x_2, y_2)\ldots(x_n, y_n)\} \qquad (1)$$

where $x \in X \subset \mathbb{R}^n$ are the training inputs and $y \in Y \subset \mathbb{R}$ are the training outputs.

Assume a non-linear function, *f(x)*given by:

$$f(x) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b \qquad (2)$$

where $\boldsymbol{w}$ is the weight vector, b is the bias, and $\boldsymbol{\varphi}(\boldsymbol{x}_i)$ is the high dimensional feature space, which is linearly mapped from the input space $\boldsymbol{x}$. Assume further that the goal is to fit the data $\boldsymbol{L}$ by finding a function *f(x)* that has a largest deviation $\boldsymbol{\varepsilon}$ from the actual targets $\boldsymbol{y_i}$ for all the training data $\boldsymbol{L}$, and at the same time is as small as possible.

Therefore, Eq. (2) is transformed into a constrained convex optimization problem as follows:

$$\begin{aligned} &\textit{minimise} \quad \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} \\ &\textit{subject to:} \begin{cases} y_i - (\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b) \le \varepsilon \\ y_i - (\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b) \ge \varepsilon \end{cases} \end{aligned} \qquad (3)$$

Where $\varepsilon \ge 0$ is a user defined and the maximum acceptable deviation. Hence, Eq (3) can be rewritten as:

$$\textit{minimise} \quad \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} \qquad (4)$$

$$\textit{subject to:} \begin{cases} y_i - \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) - b \le \varepsilon \\ \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b - y_i \le \varepsilon \end{cases}$$

The goal of the objective function in Eq. (4) is to make “**w**” as small as possible while satisfying the constraints. In order to obtain the solution of Eq. (4), slack variables have to be introduced to cope with possible infeasible optimization problems. The objective function in Eq. (2.8) was built based on assumption that there is an existing *f(x)* that provide exact accurate feasible classification. However, such solution is not always feasible in real life therefore necessitated the need for error tolerance. Building on that idea, the primal formulation was put forward as by Vapnik in [14]:

$$\textit{minimise} \quad \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} + \boldsymbol{C}\sum_{i=1}^{m}(\xi_i^+ + \xi_i^-) \qquad (5)$$

$$\textit{subject to:} \begin{cases} y_i - \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) - b \le \varepsilon + \xi_i^+ \\ \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) + b - y_i \le \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \qquad \ge 0 \end{cases}$$

where C > 0 is a pre-specified regularization constant and represents the weight of the loss function.

The first term in the objective function ($\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}$) represents the regularized term and makes the function as "flat" as possible whereas the second term ($C\sum_{i=1}^{m}(\xi^+ + \xi^-)$) called the empirical term, measures the ε-insensitive loss function. According to Eq. (5), all data points whose y-values differ from *f(x)* by more than ε, are penalized.

The slack variables, ξ^+ and ξ^- corresponds to the size of this excess deviation for upper and lower deviations, respectively, as represented graphically in Fig. 1. The ε-tube is the largest deviation and all the data points inside this tube do not contribute to the regression model since their coefficients are equal to zero. For more details about SVR, refer to [33]

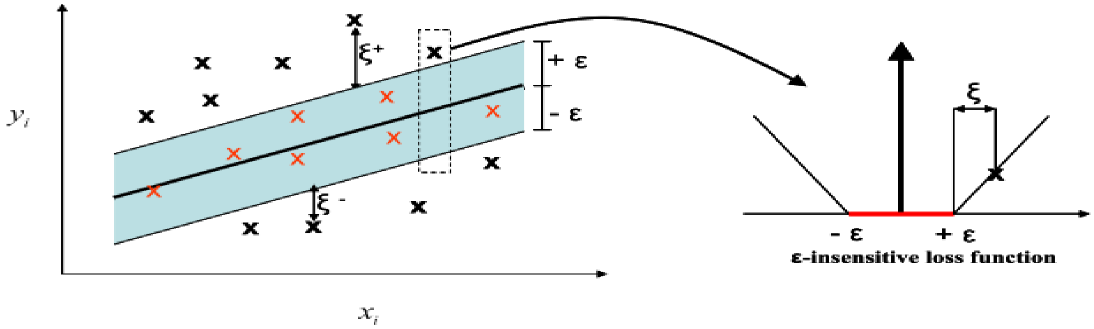

Figure 1: Slack variables and error tolerance of SVR

The left part in figure 1 above depicted the tube of ε, accuracy and points that do not meet this accuracy. The black dots located on or outside the tube are support vectors. While the right part shows the (linear) ε-insensitive loss function in which the slope is determined by C. Data points outside this tube or lying on this tube are used in determining the decision function and they are called support vectors. Eq. (5) assumes Vapnik ε-insensitive loss function as shown in Fig.1 and is defined as follows:

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \quad (6)$$

Lagrangian multipliers $\alpha_i^+, \alpha_i^-, \eta_i^+, \eta_i^-$ are introduced in order to eliminate some of the primal variables in Eq. (5). Therefore, the equivalent Lagrangian of Eq. (5) is as follows:

$$L_p = \frac{1}{2}\boldsymbol{w^T w} + \boldsymbol{C}\sum_{i=1}^{\boldsymbol{m}}(\xi^+ + \xi^-) - \sum_{i=1}^{\boldsymbol{m}}(\eta_i\xi^+ + \eta_i\xi^-) - \sum_{i=1}^{\boldsymbol{m}}\alpha_i^+(\varepsilon + \xi^+ - y_i + \boldsymbol{w^T\varphi(x_i)} + b) - \sum_{i=1}^{\boldsymbol{m}}\alpha_i^-(\varepsilon + \xi^- - y_i + \boldsymbol{w^T\varphi(x_i)} + b) \quad (7)$$

Subject $\alpha_i^+, \alpha_i^-, n_i^+, n_i^- \ge 0$.

The partial derivatives of L_p with respect to primal variables (w, b, ξ^+, ξ^-) resulted into vanishing of the saddle point for optimality. After applying Krush-Kuhn-Tucker condition, the resulting equation becomes

$$g = -\frac{1}{2}\left(w\left(y_i + y_j\right)\right) \quad (8)$$

where y_i, y_j represent the support vectors

Several kernel functions have been proposed in literature, in this work the focus is put on the widely used radial basis function (RBF), which is defined in (6):

$$K\left(x_i, x_j\right) = e^{-\gamma\left(\|x_i, x_j\|^2\right)} \quad (6)$$

Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where γ is the kernel parameter that is always greater than zero (0). The solution of (5) can be written as in eqn 7:

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{m}(\alpha_i^+ - \alpha_i^-)\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) + b \quad (7)$$

The SVR generalization performance strongly depends on the proper setting of C, ε and γ. The hyper-parameter C determines the trade-off between the model complexity and the degree to which deviations larger than ε are tolerated. A poor choice of C could lead to an imbalance between model complexity minimization (MCM) and empirical risk minimization (ERM). The hyper-parameter ε controls the width of the ε-insensitive zone, and its value affects the number of support vectors (SV) used to construct the regression function. If ε is set too large, the insensitive zone will have ample margin to include data points; this would result in too few SVs selected and lead to unacceptable 'flat' regression estimates. The parameter γ represents the width of RBF kernel. If γ is set too small, the SVR will tend to overfit the training data. On the other hand, a too large γ would make SVR not flexible enough for complex function approximation.

3.2 African Buffalo Optimization (ABO) Algorithm

African Buffalo Optimization (ABO) belongs to the class of swarm intelligence, and was proposed by Odili et al. [34]. This metaheuristic algorithm models the foraging and defending behavior of African buffaloes. The unique features of these animals include extensive memory capacity, communal lifestyle, and democratic lifestyle [34]. They utilize '*waaa*' and '*maaa'* sounds to communicate danger and safety, respectively. Thus, their organizational lifestyle could be mapped to these unique characteristics [23], [35].The "*waaa*" sound is denoted by w_i, while the "*maaa*" denoted by m_i, and the learning parameters were denoted by l_1 and l_2. Other parameters are global best (bg_{max}), the personal best ($bp_{max(k)}$) positions. The basic ABO is controlled by two equations, namely democratic Eqn. (2.22) and location update Eqn. (2.23). Algorithm 2.1 shows the basic flow of ABO. The algorithm subtracts the "waaa" value (w_k) asking the animals to explore the search space from the maximum vector (bg_{max} and $bp_{max(k)}$) which is further multiplied by the learning parameters (l1 and l2) [9]. The result is supplied by the "*maaa*" (m_k) value, this indicates that the herds should remain in that location and continue grazing. The equation of the exploitation and exploration stages of ABO algorithm are depicted in Eqn. (8) and Eqn. (9) respectively, while the basic ABO algorithm is presented in algorithm 8 below [34].

$$m_{k+1} = m_k + l_1(bg_{max} - w_k) + l_2\left(bp_{\max(k)} - w_k\right) \quad (8)$$

$$w_{k+1} = \frac{(w_k + m_k)}{\lambda} \quad (9)$$

4.0 METHODOLOGY

In this section, we present the process involved in developing the power consumption forecasting model of household appliances. Figure 2 shows all the steps involved for the construction of the forecasting model. Firstly, we collected the power consumption dataset, the dataset was then pre-processed through resampling and scaling. Four (4) forecasting models were then developed based on SVR, PSO, GA, and ABO algorithms. The developed models were trained using training and validation data. The overall performance of the developed models was then evaluated on test data using several performance metrics.

The detailed methodology is presented in the following subsequent sections.

4.1 Dataset Description

The dataset used in this study is an electric consumption history recorded through smart meters that were strategically placed in a home. The dataset recorded is a time-series and multivariate dataset collected and donated to UCI by Georges Herbrail and Alice Berard of EDF R&D, Clamart, France. The description of the data set is as shown in table 1.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Properties of household appliances dataset

Sno	Attribute	Description
1.	Dataset Name	Household Appliances
2.	Characteristics	Multivariate, Time-series
3.	Number of Attributes	Nine (9)
4.	Attribute Type	Real Values
5.	No of instances	2,075,259

The dataset consists of 2, 075,259 instances of electricity consumption recordings at the interval of ten minutes. However, during preprocessing stage, the data was resampled into days, thereby resulting into 1,112 data samples.

The dataset was partitioned into training, validation and testing sets. The training set consists of nine hundred and ninety-nine (991) data samples that ranges from 16-12-2006 through 30-11-2009. The validation dataset consists of ninety-one (91) data samples that ranges from 01/09/2009 through 30/11/2009. While the testing dataset consists of fourteen (14) data samples for the first two weeks of the month of December 2009.

The dataset was scaled to values between {1 and -1} in order to remove the difference in magnitude of features in order to improve model learning. The training and validation datasets were scaled separately from testing dataset in order to avoid data leakage between training and test data. Sample of the scaled dataset was presented in table 2 below.

Table 2: Sample of scaled dataset

Sno	F1	F2	F3
0	-1.04702	-0.71647	-0.56846
1	-0.38682	-1.19299	1.128183
2	-0.96907	-0.39464	-1.33329
3	-0.08434	-0.34441	-0.65926
4	2.031809	-1.3762	1.22832
5	2.445771	-1.39077	1.287591

where *F1, F2* and *F3* represent input features of the dataset.

4.2 Evaluation Metrics

Four (4) performance metrics were used for ascertaining the level of accuracy of the developed forecasting model. The aim is to obtain a small error value; the smaller the value obtained, the better the forecast is. The proposed model has been evaluated using Mean Absolute Percentage Error (MAPE) [5], [36], [37], Root Mean Squared Error (RMSE) [38]–[40], Mean Absolute Error (MAE) [15], [41] and Mean Absolute Deviation (MAD) [42]. These performance metrics are mathematically represented as in equation 10, 11, 12 and 13 respectively.

$$MAPE = \frac{1}{N}\left[\sum_{n=1}^{N} \left|\frac{y_n - \acute{y}_n}{y_n}\right|\right] * 100\% \qquad (10)$$

Where N represents number of observations, y_n and $\acute{y}_n$ represent the n_{th} observed and forecast values respectively.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_n - \acute{y}_n)^2}{N}} \qquad (11)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_n - \acute{y}_n)| \qquad (12)$$

$$\mathrm{MAD} = \frac{1}{m}\sum_{I=1}^{m} |y_n - \acute{y}_n| \qquad (13)$$

4.3 Model Construction

To develop the proposed hybrid SVR-ABO algorithm, the ABO algorithm was used to optimize the SVR hyper-parameters. To achieve this, the RMSE was used as the cost optimization function during the training. Initial values for the ABO algorithm were set accordingly.

Sequence of the process involved in SVR-ABO includes 1) initializing ABO population in the search space and 2) calculating the fitness of each buffalo based on its position in the search space. All buffaloes were individually evaluated and the buffalo with the smallest error of prediction based on RMSE metric, is considered as the best for that iteration. The overall best buffalo among the best buffaloes determined at each iteration is considered

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as the global best. The training process was conducted with varied population size and different iterations in the to determine the effect of population size and number of iterations on algorithm performance. At second stage, the proposed algorithm was trained based on parameters shown in Table 2.

Upon reaching the termination criteria, the values of the position of the overall best buffalo are considered as the required optimal hyper-parameters for the SVR model. The optimized SVR model was used to forecast the test dataset. Figure 2 illustrate the SVR-ABO forecasting process.

The forecasting model was developed using Python programming language with scikit-learn implementation of LibSVM module coupled with Numpy for matrix manipulation, Pandas for dataframes processing and Matplotlib for visualization

Table 3: PSO, GA and ABO algorithms parameters

PSO		**GA**		**ABO**	
Parameter	Value	Parameter	Value	Parameter	Value
Population size	100	Swarm size	100	Population size	100 1000
Maximum Iterations	1000	Maximum Iterations	1000	Maximum Iterations	
Cognitive acceleration (C_1)	0.5	Elite chromosomes	5	Democratic Parameter (l_1)	0.4
Social acceleration (C_2)	0.5	Selection method	Roulette wheel	Location update Parameter (l_2)	0.6
Inertia weight (w)	0.9	Cross-over function	Partially matched Crossover (PMX)	Lambda	0.4
Mutation function	Uniform	Mutation function	Uniform		
Mutation rate	0.1	Mutation rate	0.1		

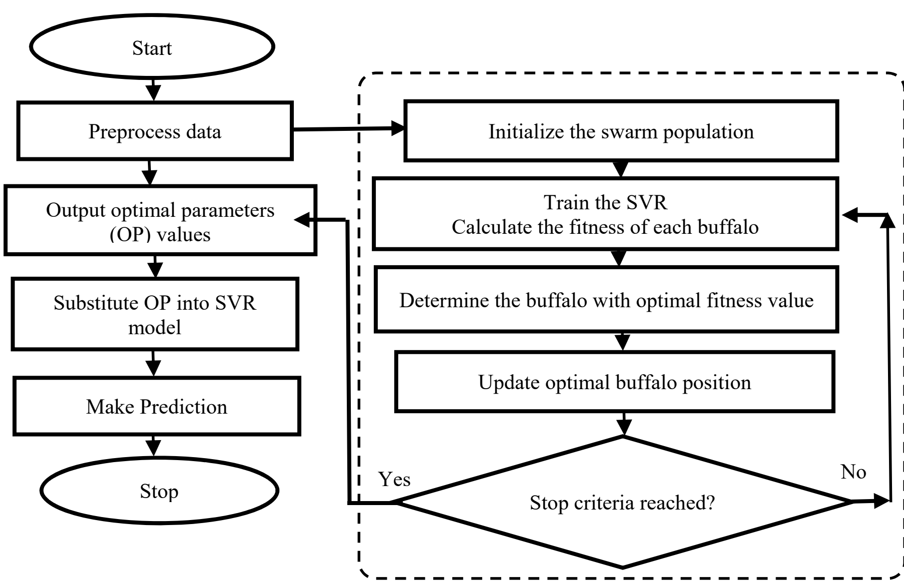

Figure 2: SVR-ABO hybrid optimization process

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.0 RESULT AND DISCUSSION

Several experiments were performed in order to ascertain the performance of the developed model. In order to determine the impact of population size and number of iterations on the accuracy of the algorithm, different population sizes ranging from fifty (50) through five hundred (500) based on varied number of iterations from fifty (50) through three hundred (300) using RMSE as the evaluation metric were used in the experiment. The result obtained is presented as shown in table 4.

Table 4: Impact of population size and varied iterations on model performance

		Number of Iterations					
		5	10	15	20	25	Standard SVR
Population	50	3421.1486	5020.2718	3394.3485	5020.2718	3057.1559	2252.8273
	60	**2517.7853**	5020.2718	5020.2718	3387.7175	5020.2718	2252.8273
	70	2939.6027	5020.2718	2927.7567	5020.2718	3054.4402	2252.8273
	80	3360.4206	2519.2041	3070.8977	5020.2718	3266.3524	2252.8273
	90	3335.1991	2819.0940	2592.8076	2977.3690	2692.5842	2252.8273
	100	3450.9419	2443.0290	3277.7027	3147.0606	5020.2718	2252.8273

However, upon careful analysis, it can be deduced that population size and number of iterations does not add to the performance of the proposed algorithm. Hence, values presented in table 2 were used to perform the second experiment. The performance of the proposed SVR-ABO algorithm was compared with SVR-PSO, SVR-GA and standard SVR algorithms. The result obtained has shown that SVR-ABO algorithm has the highest forecasting accuracy compared to the benchmarked algorithms. The performance of the SVR-ABO algorithm and comparison of algorithms performance is graphically presented in figure 3 and figure 4 respectively.

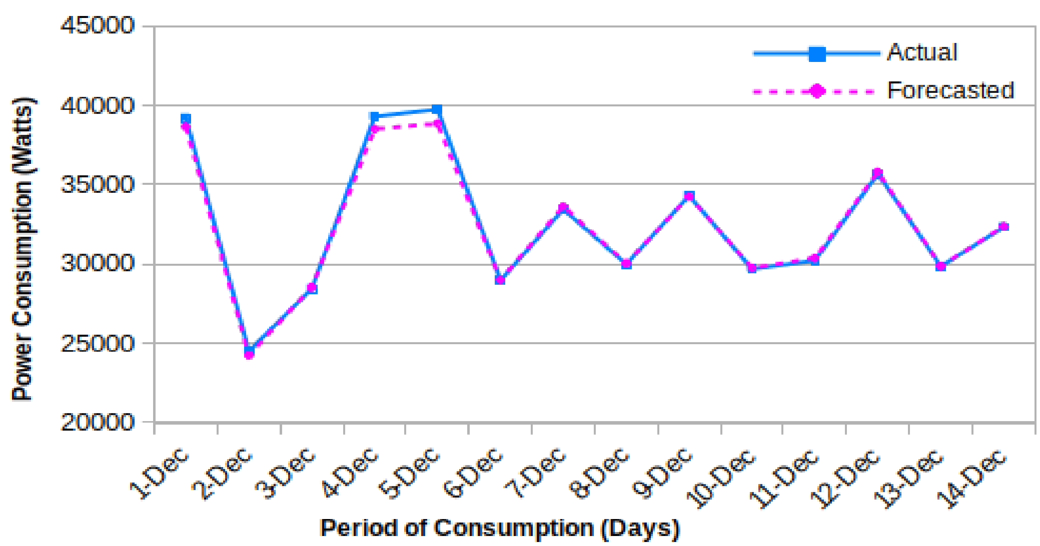

Figure 3: Prediction of power consumption using optimized SVR-ABO

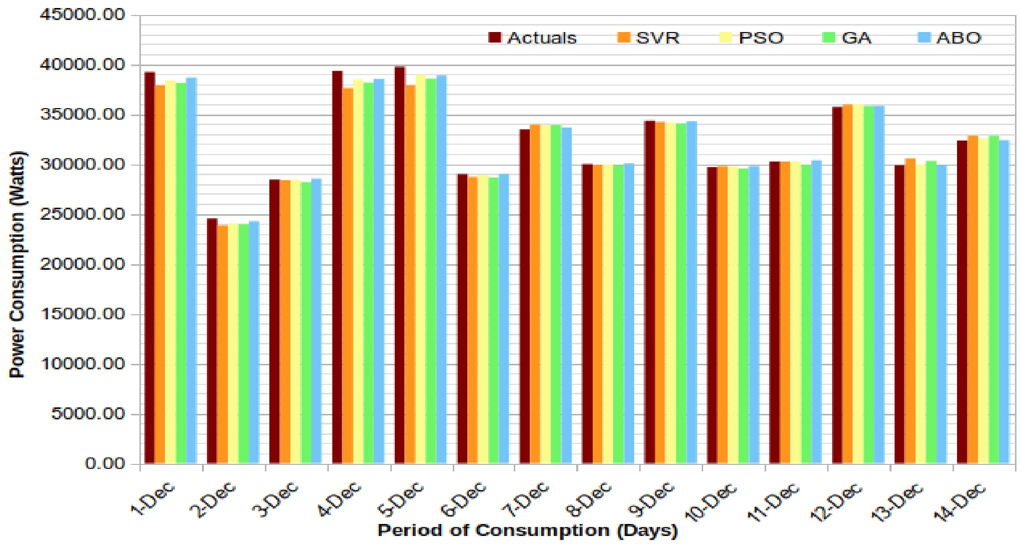

Figure 4: Performance Comparison of the proposed model

In order to determine the degree of accuracy of the proposed model, percentage error has also been computed of each algorithm against the actual values of the forecasting period as presented in table 5. The SVR-ABO algorithm has shown less variation of deviation from the actual values as compared to the rest of the benchmarked algorithms as shown in figure 5.

Table 5: Comparative analysis of daily forecasting errors

		SVR		PSO-SVR		GA-SVR		ABO-SVR	
Days	**Actual Power Consumption (Kwh)**	**Forecasted Value**	**PE (%)**	**Forecasted Value**	**PE (%)**	**Forecasted Value**	**PE (%)**	**Forecasted Value**	**PE (%)**
1-Dec	39201.01	37877.56	-3.38	38375.70	-2.11	38116.90	-2.77	38665.11	-1.37
2-Dec	24521.28	23833.35	-2.81	24055.96	-1.90	23955.77	-2.31	24237.82	-1.16
3-Dec	28447.37	28374.51	-0.26	28395.89	-0.18	28176.02	-0.95	28545.47	0.34
4-Dec	39304.32	37620.50	-4.28	38448.51	-2.18	38140.86	-2.96	38515.99	-2.01
5-Dec	39753.97	37872.47	-4.73	38944.82	-2.04	38549.91	-3.03	38864.02	-2.24
6-Dec	28978.70	28697.32	-0.97	28873.81	-0.36	28647.52	-1.14	28999.42	0.07
7-Dec	33458.09	33942.52	1.45	33878.03	1.26	33895.07	1.31	33644.46	0.56
8-Dec	29991.01	29907.70	-0.28	29966.74	-0.08	29850.61	-0.47	30048.67	0.19
9-Dec	34311.97	34215.15	-0.28	34149.83	-0.47	34056.09	-0.75	34265.64	-0.14
10-Dec	29687.68	29770.18	0.28	29748.45	0.20	29552.73	-0.45	29768.17	0.27
11-Dec	30228.74	30217.75	-0.04	30213.97	-0.05	29895.27	-1.10	30372.90	0.48
12-Dec	35694.51	35956.29	0.73	35982.41	0.81	35804.32	0.31	35820.31	0.35
13-Dec	29863.98	30558.24	2.32	29962.61	0.33	30297.82	1.45	29873.22	0.03
14-Dec	32326.26	32828.27	1.55	32529.64	0.63	32823.81	1.54	32371.34	0.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

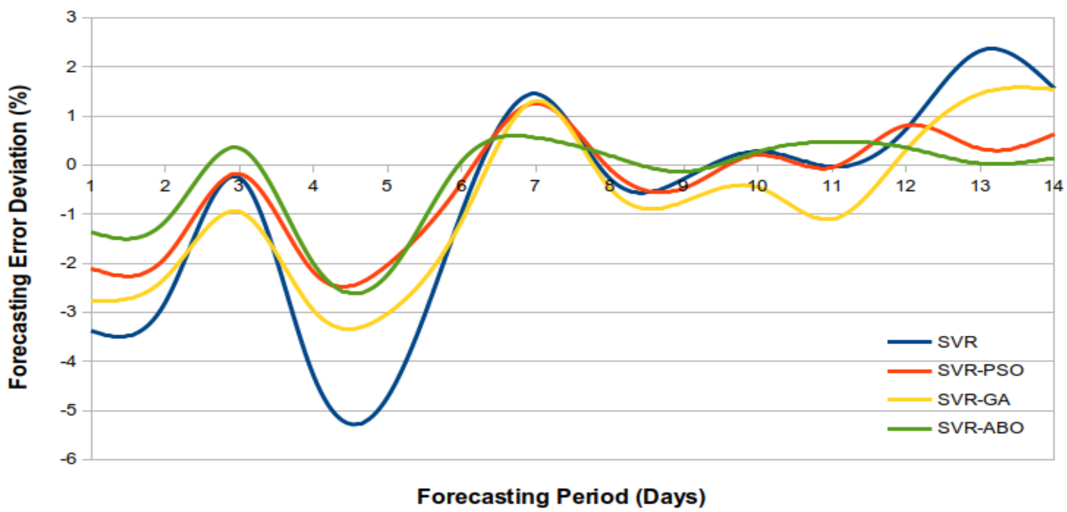

Figure 5: Comparison of Forecasting Error Deviation of Four algorithms

To further analyze the performance of the proposed algorithm, absolute error deviation from the actual values was also computed and compared to benchmarked algorithms. It was observed that the proposed algorithm has fewer absolute percentage errors in comparison to the rest of the algorithms. This ascertains the superiority of the proposed algorithm in terms of forecasting accuracy with less deviation from the actual values. Table 6 and figure 6 respectively present the values of daily Absolute Percentage Error (APE) and comparison of the four algorithms based on the computed APE.

Table 6: Comparative analysis of daily absolute forecasting errors

		SVR		PSO-SVR		GA-SVR		ABO-SVR	
Days	**Actual Power Consumption (Kwh)**	**Forecasted Value**	**APE (%)**	**Forecasted Value**	**APE (%)**	**Forecasted Value**	**APE (%)**	**Forecasted Value**	**APE (%)**
1-Dec	39201.01	37877.56	3.38	38375.70	2.11	38116.90	2.77	38665.11	1.37
2-Dec	24521.28	23833.35	2.81	24055.96	1.90	23955.77	2.31	24237.82	1.16
3-Dec	28447.37	28374.51	0.26	28395.89	0.18	28176.02	0.95	28545.47	0.34
4-Dec	39304.32	37620.50	4.28	38448.51	2.18	38140.86	2.96	38515.99	2.01
5-Dec	39753.97	37872.47	4.73	38944.82	2.04	38549.91	3.03	38864.02	2.24
6-Dec	28978.70	28697.32	0.97	28873.81	0.36	28647.52	1.14	28999.42	0.07
7-Dec	33458.09	33942.52	1.45	33878.03	1.26	33895.07	1.31	33644.46	0.56
8-Dec	29991.01	29907.70	0.28	29966.74	0.08	29850.61	0.47	30048.67	0.19
9-Dec	34311.97	34215.15	0.28	34149.83	0.47	34056.09	0.75	34265.64	0.14
10-Dec	29687.68	29770.18	0.28	29748.45	0.20	29552.73	0.45	29768.17	0.27
11-Dec	30228.74	30217.75	**0.04**	30213.97	**0.05**	29895.27	1.10	30372.90	0.48
12-Dec	35694.51	35956.29	0.73	35982.41	0.81	35804.32	**0.31**	35820.31	0.35
13-Dec	29863.98	30558.24	2.32	29962.61	0.33	30297.82	1.45	29873.22	**0.03**
14-Dec	32326.26	32828.27	1.55	32529.64	0.63	32823.81	1.54	32371.34	0.14

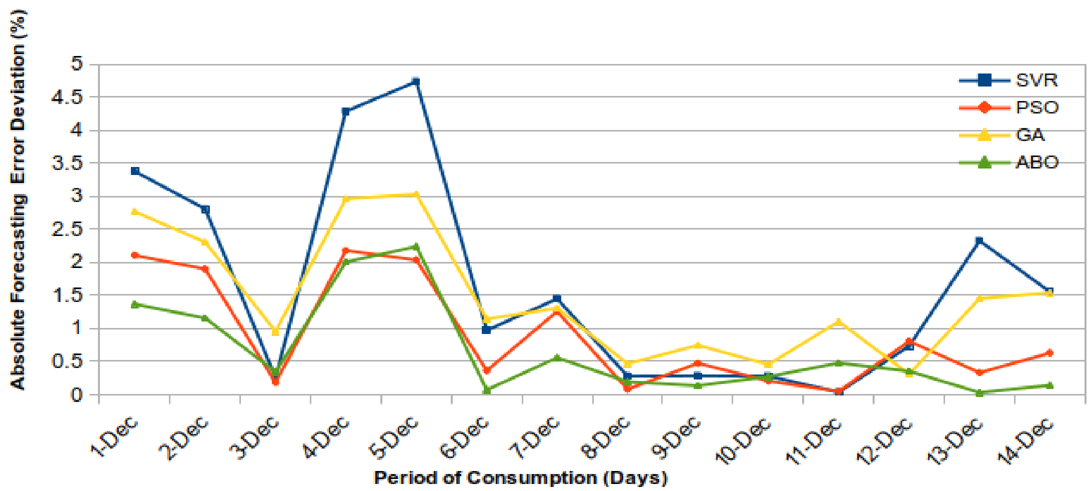

Figure 6: Comparison of Absolute Forecasting Error

Additionally, Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as statistical metrics for determining of forecasting accuracy were also used to determine the degree of performance of the proposed algorithm.

Table 7 presented the comparative performance of SVR, SVR-PSO, SVR-GA and the proposed SVR-ABO algorithms based on MAPE, MAD, RMSE and MAE. Accordingly, the proposed algorithm has demonstrated superiority of accuracy performance (based on MAPE) as presented in table 7.

Table 7: Comparative analysis of algorithms performance

	MAPE	MAD	RMSE	MAE	Accuracy(%)
SVR	2.78E+00	**3.57E+03**	3.31E+02	8.24E+02	97.22
SVR-PSO	1.46E+00	3.85E+03	2.06E+02	4.36E+02	98.54
SVR-GA	2.23E+00	3.25E+04	2.71E+02	6.62E+02	97.77
SVR-ABO	**1.08E+00**	3.84E+03	**1.34E+02**	**3.26E+02**	**98.92**

The following points can be deduced from table 5, 6 and 7 respectively:

i. After careful observations it can be concluded that swarm-based algorithms could determine best values for SVR hyper-parameters better than the default values of standard SVR.
ii. The SVR-ABO algorithm is not significantly affected by the demand fluctuation as shown around third (3rd) and fourth(4th) days as compared to first day or fifth (5th) day consumption. Hence SVR-ABO is an obvious choice for electricity consumption forecast.
iii. The performance of the proposed algorithm shows that it can achieve low prediction error based on three (3) out of four (4) selected metrics (MAPE, RMSE, MAD and MAE). This clearly indicates its superiority against standard SVR, GA-SVR and PSO-SVR based algorithms.

However, despite the remarkable result achieved by SVR-ABO algorithm, there is still room for improvement as the algorithm is also susceptible to fluctuation of consumption like other algorithms.

6.0 CONCLUSION

The excellent performance demonstrated by proposed SVR-ABO algorithm can be attributed to its ability of tracking and memorising position of, as well as the movement of best buffalo at each iteration. This resulted into better exploration and ability to escape local optima entrapment. The proposed algorithm shows that ABO algorithm has the capability of improving the generalisation ability of SVR.

This paper presents an SVR algorithm that has been hybridized with ABO algorithm for electric load forecasting. The results from several experiments executed shows that ABO as a swarm intelligence-based algorithm can achieve superiority in terms accuracy of forecasting against SVR, PSO-SVR, GA-SVR algorithms. The ABO algorithm was used to determine SVR's hyper-parameters in order to improves its forecasting performance. The proposed algorithm achieved a lesser deviation from actual value on individual data point with overall best accuracy achieved as compared to benchmarked state-of-the-art swarm-based algorithm.

REFERENCES:

[1] Wibowo, A. *et al.*, "Combined Multiple Neural Networks and Genetic Algorithm with Missing Data Treatment : Case Study of Water Level Forecasting in Dungun River - Malaysia," no. May, 2018.

[2] Ugurlu, U. *et al.*, "Electricity Price Forecasting Using Recurrent Neural Networks," *Energies*, vol. 2145, no. 5, pp. 75–78, 2018.

[3] Dong, Y. *et al.*, "A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting," *Energies*, vol. 11, no. 4, 2018.

[4] Mat Daut, M. A. *et al.*, "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 1108–1118, 2017.

[5] Hermias, J. P. *et al.*, "Short-term stochastic load forecasting using autoregressive integrated moving average models and Hidden Markov Model," in *2017 International Conference on Information and Communication Technologies, ICICT 2017*, 2018, vol. 2017-Decem, pp. 131–137.

[6] Conejo, A. J. *et al.*, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, 2005.

[7] Mustaffa, Z. *et al.*, "An Application of Grey Wolf Optimizer for Commodity Price Forecasting," *Applied Mechanics and Materials*, vol. 785, pp. 473–478, 2015.

[8] Li, S. Y. *et al.*, "An improved grey wolf optimizer algorithm for the inversion of geoelectrical data," *Acta Geophysica*, vol. 66, no. 4, pp. 607–621, 2018.

[9] Salkuti, S. R., "Short-term electrical load forecasting using radial basis function neural networks considering weather factors," *Electrical Engineering*, vol. 100, no. 3, pp. 1985–1995, 2018.

[10] Huang, B. *et al.*, "Load Forecasting based on Deep Long Short-term Memory with Consideration of Costing Correlated Factor," in *Proceedings - IEEE 16th International Conference on Industrial Informatics, INDIN 2018*, 2018, pp. 496–501.

[11] Jaddi, N. S. *et al.*, "Master-leader-slave cuckoo search with parameter control for ANN optimization and its real-world application to water quality prediction," *PLoS ONE*, vol. 12, no. 1, pp. 1–19, 2017.

[12] Eapen, R. R. and Simon, S. P., "Performance Analysis of Combined Similar Day and Day Ahead Short Term Electrical Load Forecasting using Sequential Hybrid Neural Networks," *IETE Journal of Research*, vol. 2063, no. May, pp. 1–11, 2018.

[13] Yusof, Y. and Mustaffa, Z., "Time Series Forecasting of Energy Commodity," *International Journal of Engineering Research and Applications*, vol. I, 2015.

[14] Corinna, C. and Vladimir, V., "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[15] Al-Musaylh, M. S. *et al.*, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Advanced Engineering Informatics*, vol. 35, no. April 2017, pp. 1–16, 2018.

[16] Velasco, L. C. P. *et al.*, "Day-ahead load forecasting using support vector regression machines," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, pp. 22–27, 2018.

[17] Sarhani, M. *et al.*, "Generalization enhancement of support vector regression in electric load forecasting with model selection," in *Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications - LOPAL '18*, 2018, pp. 1–6.

[18] Peng, L. L. *et al.*, "Hybridizing DEMD and quantum PSO with SVR in electric load forecasting," *Energies*, vol. 9, no. 3, 2016.

[19] Qu, D. *et al.*, "Short-Term Traffic Flow Forecasting Method Based on LSSVM Model Optimized by GA-PSO Hybrid Algorithm," *Discrete Dynamics in Nature and Society*, vol. 2018, 2018.

[20] Zhang, X. *et al.*, "Applying cuckoo search algorithm to solve fractional differential equation based on cubic spline function," *IAENG International Journal of Applied Mathematics*, vol. 50, no. 1, pp. 31–42, 2020.
[21] Odili, J. B. and Noraziah, A., "African buffalo optimization for global optimization," *Current Science*, vol. 114, no. 3, pp. 627–636, 2018.
[22] El-Ashmawi, W. H., "An Improved African Buffalo Optimization Algorithm for Collaborative Team Formation in Social Network," *International Journal of Information Technology and Computer Science*, vol. 10, no. 5, pp. 16–29, 2018.
[23] Odili, J. B. *et al.*, "Performance Analyses of Nature-inspired Algorithms on the Traveling Salesman's Problems for Strategic Management," *Intelligent Automation and Soft Computing*, vol. 8587, no. June, pp. 1–11, 2017.
[24] Almonacid, B. *et al.*, "Selecting a biodiversity conservation area with a limited budget using the binary African buffalo optimization algorithm," *IET Software*, vol. 12, no. 2, pp. 96–111, 2017.
[25] Odili, J. B. *et al.*, "Parameters-Tuning of PID controller for automatic voltage regulators using the African buffalo optimization," *PLoS ONE*, vol. 12, no. 4, pp. 1–17, 2017.
[26] Odili, J. and Kahar, M., "Numerical Function Optimization Solutions Using the African Buffalo Optimization Algorithm (ABO)," *British Journal of Mathematics & Computer Science*, vol. 10, no. 1, pp. 1–12, 2015.
[27] Zhang, S. *et al.*, "Support Vector Regression with Multi-Strategy Artificial Bee Colony Algorithm for Annual Electric Load Forecasting," *Advances in Intelligent Systems and Computing*, vol. 891, pp. 576–585, 2019.
[28] Li, M. *et al.*, "Hybridizing Chaotic and Quantum Mechanisms and Fruit Fly Optimization Algorithm with Least Squares Support Vector Regression Model in Electric Load Forecasting," *Energies*, vol. 11, no. 9, p. 2226, 2018.
[29] Chou, J.-S. and Truong, T. T. H., "Sliding-window metaheuristic optimization-based forecast system for foreign exchange analysis," *Soft Computing*, 2019.
[30] Chou, J. S. and Pham, A. D., "Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information," *Information Sciences*, vol. 399, pp. 64–80, 2017.
[31] Tran, T. H. and Hoang, N. D., "Estimation of algal colonization growth on mortar surface using a hybridization of machine learning and metaheuristic optimization," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 42, no. 6, pp. 929–939, 2017.
[32] Verma, M. *et al.*, "Kernel-based models for prediction of cement compressive strength," *Neural Computing and Applications*, vol. 28, pp. 1083–1100, 2017.
[33] Boser, E. *et al.*, "A Training Algorithm for Optimal Margin Classifiers," in *Proceedings of the fifth annual workshop on computional learning theory*, 1992, pp. 144–152.
[34] Odili, J. B. *et al.*, "African Buffalo Optimization: A Swarm-Intelligence Technique," *Procedia Computer Science*, vol. 76, pp. 443–448, 2015.
[35] Odili, J. B. *et al.*, "Solving Traveling Salesman' s Problem Using African Buffalo Optimization, Honey Bee Mating Optimization & Lin-Kerninghan Algorithm," *World Applied Science Journal*, vol. 34, no. 7, pp. 911–916, 2016.
[36] Wang, M. *et al.*, "Optimized rolling grey model for electricity consumption and power generation prediction of China," *IAENG International Journal of Applied Mathematics*, vol. 49, no. 4, pp. 1–11, 2019.