Extreme Learning Machines and Support Vector Machines Models for Email spam detection

Sunday Olusanya Olatunji
Computer Science Department, College of Computer Sciences and Information Technology, University of Dammam, Dammam, Kingdom of Saudi Arabia
oluolatunji.aadam@gmail.com; osunday@uod.edu.sa

Abstract— Extreme Learning machines (ELM) and Support Vector Machines have become two of the most widely used machine learning techniques for both classification and regression problems of recent. However the comparison of both ELM and SVM for classification and regression problems has often caught the attention of several researchers. In this work, an attempt has been made at investigating how SVM and ELM compared on the unique and important problem of Email spam detection, which is a classification problem. The importance of email in this present age cannot be overemphasized. Hence the need to promptly and accurately detect and isolate unsolicited mails through spam detection system cannot be over emphasized. Empirical results from experiments carried out using very popular dataset indicated that both techniques outperformed the best earlier published techniques on the same popular dataset employed in this study. However, SVM performed better than ELM on comparison scale based on accuracy. But in term of speed of operation, ELM outperformed SVM significantly.

Keywords— Extreme Learning Machines; Support vector machines; Email; Spam; Non-Spam; Spam detector; computational intelligenc)

I. INTRODUCTION

Extreme Learning machines (ELM) was recently introduced as a learning algorithm to be used in training single hidden-layer feed forward neural network and within a very short time it has distinguished itself as one of the widely used and successful techniques for both prediction and classifications problems [1]–[4]. Meanwhile, Support vector machines that has its origin in statistical learning theory has distinguished itself as one of the best if not the best machine learning based techniques in recent times with several successful applications in several fields, for both classification and regression, often with excellent results [5]–[9]. However the comparison of both ELM and SVM for classification and regression problems has often caught the attention of several researchers in recent time [10]–[12]. In this paper, the comparison between ELM and SVM is investigated for the unique and important problem of Email spam detection, which happens to be a classification problem, in order to achieve better accuracy in the detection process.

It is an established fact that electronic mail (Email) has become extremely popular among people nowadays. In fact hardly can people do without sending or receiving email messages on daily basis several times within few minutes or hours. In essence, the importance of email in our present day life is very clear and self convincing. Despite the huge benefits of emails, unfortunately its usage has been bedeviled with the huge presence of unsolicited and sometimes fraudulent emails that have often caused several and huge damages to individuals and corporate establishments both psychologically and financially. To mitigate the effect of such email abuses, there has always been the need to promptly detect and isolate such unwanted emails through what is popularly referred to as spam detection system. Spam detection facilitates separating spam email from non-spam emails thereby making it possible to prevent spam email from getting into the inbox of users. Thus it could be stated that spam detection is the first step and the most important stage in the email filtering process to ensure spam mails are prevented from entering the user's inbox, particularly in this age of huge spam mails due to bulk mailing tools that has pushed up the amount of spam emails in a skyrocketed manner.

Several spam detection models have been proposed and tested in literatures but still the reported accuracy still begs for more work in this direction in order to achieve better accuracy. Authors in [13] made use of artificial neural network based model for spam detection but only succeeded in achieving 86% accuracy which is still considered far from the ideal. In [14], the authors applied naves Bayes approach while incorporating cost-sensitive multi-objective Genetic Programming for feature extraction and used it for Spam detection but achieved an accuracy of 79.3% correctly detected email types. Moreover in [15], the authors presented a spam detection system based on interval type-2 fuzzy sets, exploring the capabilities of fuzzy logic of type-2 category, but could only succeeded in obtaining spam detection accuracy of 86.9% for the testing set. The authors in [16] made use of genetic algorithm based hybrid on the same dataset but were only able to push the accuracy to 90% accuracy for the testing set. Furthermore and of recent is the work of [17], where an hybrid model consisting of smart hybridization of negative selection algorithm (NSA) with particle swarm optimization (PSO) was presented. They were able to push the accuracy to 91.22% for the testing set. Considering the work done so far and the performance accuracy obtained till date, it is clear that there is still the need to further explore the possibility of achieving better results using the same popular datasets. This work is thus set to come up with

978-1-5090-5538-8/17/$31.00 ©2017 IEEE

alternative models that could hopefully push the accuracy level higher than previous models.

Empirical studies' outcomes showed that both techniques performed better than the best earlier published techniques on the same popular dataset employed in this study, with SVM and ELM achieving 94.06% and 93.04% testing accuracy respectively as compared to the 91.22% testing accuracy obtained in [17] which was the best in literature so far. Therefore, this indicated that SVM and ELM respectively achieved 3.11% and 2.0% improvement over the best reported model in literature [17]. However, SVM performed better than ELM on comparison scale based on accuracy. Specifically SVM achieved 94.06% testing accuracy while ELM achieved 93.04% accuracy, which indicated just 1.1% performance improvement that SVM achieved over ELM. But regarding the comparison based on speed of operation, ELM performed better than SVM. With the very little difference in the accuracy of the two models, it could be comfortably stated that when real time system is needed then ELM could be a favorable option otherwise SVM standout in this area of application.

The remaining part of this work is organized as follow. Section II contains the proposed models. Section III contains empirical studies that include dataset description, and experimental setup or methodology. Section IV presents results and discussion with detailed comparison while section V contains the conclusion and recommendation emanating from this work.

II. Proposed Machine Learning Models

A. Extreme Learning Machines (ELM)

Extreme learning machine (ELM) is a learning algorithm for single-hidden layer feed-forward neural networks (SLFNs), which chooses its hidden nodes randomly and then determines the output weights of SLFNs analytically [18][19]. This new learning algorithm for SLFNs was proposed as a means to overcoming the perennial problems of the classical feed-forward neural networks (FFNN), particularly its slow gradient based learning algorithms and the iterative tuning of its parameters. The proposed ELM uniquely work in such a way that it is tuning-free and avoids using the gradient based learning algorithms that consumes lots of time. Further details of this interesting technique could be found in [18], [19].

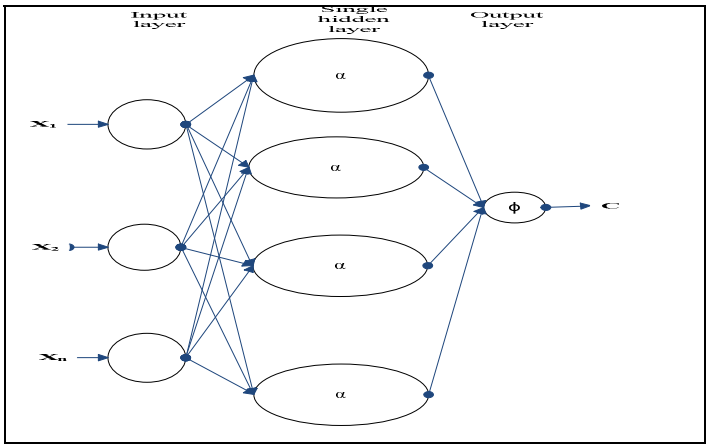

Fig. 1. Schematic Representation of ELM ($x_1 \ldots x_n$ are the input values, ϕ is the activation function, and C represent the classification output) [Adapted with modification from [4]].

Firstly, we need to define the standard SLFN. Assuming that there are N samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$, then the standard SLFN with $\tilde{N}$ hidden neurons and activation function g(x) is defined as [18], [19] :

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i . x_j + b_i) = o_j, j = 1,...,N \qquad (1)$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector that connects the i^{th} hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector that connects the i^{th} neuron and the output neurons, and b_i represents the threshold of the i^{th} hidden neuron. The "." in $w_i . x_j$ stands for the inner product of w_i and x_j.

It must be noted that in the case of email spam, x_i will be those 57 independent variables. That is, there are inputs $x_{ij}, i = 1,.....,N; j = 1,.....,M$ and target t_i, which is the class label indicating whether the current sample i is a spam email or not, where N=number of data points and M=57, which is the number of independent variables. Details about the dataset are presented later in section III.

The aim of the SLFN is to minimize the difference between network output (o_j) and the target (t_j). This can be expressed mathematically as:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i . x_j + b_i) = t_j, j = 1,...,N \qquad (2)$$

Or, it could be written compactly as:

$$H\beta = T \qquad (3)$$

where

$$\mathbf{H}(w_1,\ldots,w_{\tilde{N}},b_1,\ldots,b_{\tilde{N}},x_1,\ldots,x_N)=\begin{bmatrix} g(w_1 x_1+b_1) & \ldots & g(w_{\tilde{N}} x_{\tilde{N}}+b_{\tilde{N}}) \\ . & & . \\ . & \ldots & . \\ . & & . \\ g(w_1 x_N+b_1) & \ldots & g(w_{\tilde{N}} x_N+b_{\tilde{N}}) \end{bmatrix}_{N\times\tilde{N}} \quad (4a)$$

$$\beta=\begin{bmatrix} \beta_1^T \\ . \\ . \\ . \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}\times m} \quad \text{and} \quad \mathrm{T}=\begin{bmatrix} T_1^{\,T} \\ . \\ . \\ . \\ T_{\tilde{N}}^{\,T} \end{bmatrix}_{N\times m} \quad (4b)$$

As proposed by Huang and Babri in [20], [21] H is referred to as the output matrix of the neural network. Based on the aforementioned, the training procedures for the ELM based classifier can be summarized in the following algorithmic steps. See [18], [19] for further details on the workings of ELM algorithm.

B. Support Vector Machines (SVM)

Support vector machines (SVM) is a statistical based machine learning techniques with unique ability to model complex relationships among variables [22]. It uniquely addresses the curse of dimensionality through the use of generalization control technique. Curse of dimensionality often limit the performance of machine learning techniques in the face of few data simples but for support vector machines it has distinguish itself as a unique technique with ability to perform excellently even in the face of few data samples [22]–[24]. In support vector machines, the formulation leads to a global quadratic optimization problem with box constraints, which is readily solved by interior point methods [22], [25]. Support vector machines is uniquely empowered through its kernel functions to easily map non-separable problems to higher dimensions where they become easily separable.

Generally, in prediction and classification problems, the purpose is to determine the relationship among the set of input and output variables of a given dataset $D=\{Y,X\}$ where $X\in R^p$ represents the n-by-p matrix of p input variables also know as predictors or independent variables. It may be noted that $Y\in R$ for forecasting or regression problems and $Y\subseteq R$ for classification problems. Now for the case of classification as we have in this work: Suppose $D=\{y_i,x_{i1},\ldots,x_{ip}\}$ is a training set for all $i=1,\ldots,n$ of input variables X_j where $[\mathrm{X_j}=(\mathrm{x_{1j}},\ldots.\,\mathrm{x_{nj}})^{\mathrm{T}}]$ for $j=1,\ldots,p$, and the output variables, $Y=\begin{pmatrix} y_1 & \ldots & y_n \end{pmatrix}^T$. The lower case letters $x_{i1},x_{i2},\ldots,x_{ip}$ for all $i=1,\ldots,n$ refer to the values of each observation of the input variables, and $y=k$ to the response variable Y to refer to class A_k for all $k=1,2,\ldots,c$, where $c\geq 2$, but in this case of spam detection, $c=2$ representing spam or not spam two classes labels.

In what follows, the basic ideas behind SVM for pattern recognition, especially for the two-classes classification problem are briefly described and reader are referred to [22], [25] for a full description of the technique.

According to [22], [25] the goal of two-classes SVM is to construct a binary classifier or derive a decision function from the available samples which has a small probability of misclassifying a future sample. The proposed SVM implements the following idea: it maps the input vectors $\vec{x}\in R^d$ into a high dimensional feature space $\Phi(\vec{x})\in \mathrm{H}$ (see figure 1) and constructs an Optimal Separating Hyperplane (OSH), which maximizes the margin, which is the distance between the hyper plane and the nearest data points of each class in the space H (see figure 2). Different mappings construct different SVMs. The mapping $\Phi(\cdot)$ is performed by a kernel function $K(\vec{x_i},\vec{x_j})$ which defines an inner product in the space H. The decision function implemented by SVM can be written as [22], [25]:

$$f(\vec{x})=\mathrm{sgn}\left(\sum_{i=1}^{N} y_i\alpha_i\cdot K(\vec{x},\vec{x_i})+b\right) \quad (5)$$

Where the coefficients α_i are obtained by solving the following convex Quadratic Programming (QP) problem:

Maximize

$$\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\cdot y_i y_j\cdot K(\vec{x_i},\vec{x_j})$$

Subject to $0\leq\alpha_i\leq C$ (6)

$$\sum_{i=1}^{N}\alpha_i y_i=0 \qquad i=1,2,\ldots..N.$$

In the equation (6), C is a regularization parameter which controls the tradeoff between margin and misclassification error. These x_j are called Support Vectors only if the corresponding $\alpha_i>0$. Further details on the algorithmic implementation of SVM utilized in this work could be found in [26].

III. EMPIRICAL STUDIES

For the empirical work, the popular and earlier used dataset [27] by several researchers was acquired. Computational intelligence methodologies and procedures based on SVM and ELM were then followed to arrive at the final outcome of the empirical works.

A. Dataset Description

The dataset utilized in this work is the popular and often used corpus bench-mark spam dataset that are acquired from email spam messages [27]. The dataset is made of 57 features (predictors attributes) and 1 target attributes, which is the class label in the corpus indicating the status of each dataset sample or instance as either being a spam or non-spam, represented by 1 or 0 respectively. Further details regarding the dataset could be found in [27].

B. Performance measures Criteria for Model Evaluation

The need to have acceptable performance measure criteria for any proposed model is well established since it is the basis for determining how promising or otherwise a proposed model is. In this work, the accuracy, which is the overall percentage of correctly classified samples is majorly used. Also the time taken particularly during the training and testing periods for each model are considered as quality measure. The description of accuracy, as a popular performance measure, is provided below.

Accuracy (Acc): is the overall percentage of samples correctly classified. It is represented by the equation:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \qquad (7)$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

C. Experimental Settings and Implementation

The experimental procedures implemented here followed strictly the computational intelligence approach. The acquired dataset was first divided into training and testing set in the ratio 7:3 respectively using the stratify sampling approach. Then the training dataset was first introduced to the models for training and validation. Thereafter the reserved 30% of the dataset was then used to test the system in order to ascertain the performance accuracy of the proposed model.

It must be noted at this point that the implementation of the two compared models were carried out using written codes in MATLAB 2012b environment, although some SVM and ELM implementation functions from SVM toolbox in [26] and ELM codes in [28] were made use of where necessary respectively. As for the SVM, the implementation proceeded by providing the training set to the system to train and generate necessary support vectors that will be used during the testing face to achieve the needed identification of the mail messages as either spam or not. After successfully training the SVM model and validation, the testing set that have been kept away from the system are then presented to be used for the actual testing that will determine the performance accuracy of the proposed system. Similar procedures were also followed for the ELM model.

IV. RESULTS AND DISCUSSION

This section contains the results of the empirical works and the comparison between the two proposed models and then their comparison with earlier used models on the same dataset employed in this work.

A. Results of ELM and SVM compared

Experimental results from simulations indicated that the two proposed, developed and compared ELM and SVM models showed an improved accuracy when compared with earlier used models on the same dataset. Performance measures outcomes for the proposed SVM based spam-detection system and ELM based spam-detection system are first presented below and thereafter, it is followed by the comparison of the accuracy of these two models with those of earlier used models on the same dataset.

Finally, the overall result for both training and testing set are summarized in the table below:

TABLE I. SUMMARY OF RESULTS FOR BOTH TRAINING AND TESTING SET

	Accuracy(%)		**Time (s)**	
	SVM	**ELM**	**SVM**	**ELM**
Training	95.87	94.94	114.40	0.94
Testing	94.06	93.04	1.47	0.92

From the above results, it is clear that the two proposed and compared spam detector models achieved excellent results requisite of any preferred formidable model. Although SVM model perform better than the ELM in this case (SVM achieved 1.1% improvement over ELM) thereby indicating the superior performance of SVM when it comes to spam detection in term of accuracy. But regarding the speed of operation, ELM performed better than SVM. In fact ELM with a training time of 0.94 second achieved 99.18% training speed improvement over that of SVM that had 114.4 (i.e. ELM was 121.7 times faster than SVM during the training process). This is a great fit considering that the training speed is the most important in model building as major part of the time spent are usually during the training phase of model development. As for the time taken

during the testing phase, ELM also achieved shorted possible time compared to SVM.

B. results of ELM & SVM compared with earlier works

Furthermore, in order to appreciate and make the improvement provided by these two proposed models clearer, their accuracies comparison to other earlier published schemes implemented on the same dataset are presented below to support result discussion.

TABLE II. COMPARISON OF TESTING ACCURACY FOR THE PROPOSED SVM AND ELM BASED SPAM DETECTORS WITH OTHER EARLIER PUBLISHED CLASSIFIERS ON THE SAME DATASET IN TERM OF ACCURACY

Classifiers	Accuracy (%)
Proposed SVM-spam detector	94.06
Proposed ELM-spam detector	93.04
NSA-PSO [17]	91.22
PSO [17]	81.32
NSA [17]	68.86
BART [29]	79.3
IT2 Fuzzy Set [15]	86.9

From the results in table II above, it could be clearly seen that the two proposed SVM and ELM based spam detector in this work clearly outperformed all the earlier used classifiers on the same dataset as earlier cited in [17]. Specifically, it could be noticed that the proposed SVM and ELM spam detectors respectively achieved improvement of 3.11% and 2.0% over the best among the other earlier schemes, which was the hybridized negative selection algorithm-particle swarm optimization (NSA–PSO) proposed in [17]. NSA-PSO was reported to be the hybrid of NSA and PSO in order to achieve better accuracy [17], yet both ELM and SVM based classifiers in this work outperformed all the three including the hybrid scheme, perhaps due to the systematic parameter search procedures implemented coupled with the excellent reputation of both ELM and SVM in various previous research findings.

Thus it is very clear from the obtained accuracy as presented in table II that both ELM and SVM based spam detectors implemented in this work clearly outperformed earlier schemes implemented on the same dataset. This work has further corroborated the often reported superior performance of SVM and ELM models in various fields of applications [7], [9], [24], [30], [31].

V. CONCLUSION

In this work, the comparison of ELM and SVM as two of the very popular and recent successful computational intelligence techniques have been carried out on the problem of email spam detection. The two models have been proposed, trained and tested using popular and often used standard database. Empirical results from simulation indicated that the proposed SVM based scheme outperformed the ELM in term of accuracy while ELM outperformed SVM in term of speed of operation. Specifically SVM achieved 94.06% testing accuracy while ELM achieved 93.04% accuracy, which indicated just 1.1% performance improvement that SVM achieved over that of ELM. Since the improvement offered by SVM over ELM accuracy is minimal, it therefore suggest that in situation where time of detection is very important like in real time systems, then ELM spam detector should be given preference over SVM spam detector. This is in tandem with the outcomes of previous researches [11], [32], [33], which compared SVM with ELM on other problem areas. In all these cases, ELM has always demonstrated speed advantage over SVM while SVM often achieve better performance over ELM even though with just minimal difference in performance accuracy or even at par.

Furthermore, testing accuracy for these two compared system (ELM & SVM) were also compared with those of other recently published spam detector schemes tested on the same popular database used in this study. The need for a better and more accurate email spam detector scheme is definitely germane, hence the two spam detector models implemented in this research work came appropriately and timely as two improved schemes over the best among the previous methods used on the same dataset. In fact, the best accuracy reported in literature was 91.22% testing accuracy [17], which indicated that SVM and ELM respectively achieved 3.11% and 2.0% improvement over the best reported model in literature on the same dataset. Thus, the encouraging outcomes recorded in this work has further corroborated the unique reputation of SVM and ELM as two of the very viable and reliable prediction or classification tools with excellent performance in different field of applications. As a result of the promising results achieved in this work, efforts shall be made next, to investigate any possible means to improve upon the performance while also exploring the unique capability of ELM and SVM classifiers in other germane areas where accurate prediction or classification outcomes are highly desirable, for instance in biomedical predictions to save lives and facilitate preemptive diagnosis of diseases. Finally, the comparison of ELM and SVM models in different field have caught the attention of researchers in recent time and this work present the first attempt at comparing ELM and SVM on the problem of email spam detection. Hopefully this work should spur further works in this direction.

Acknowledgment

The author will like to acknowledge the University of Dammam, Kingdom of Saudi Arabia for some of the facilities utilized during the course of this research.

References

[1] T. Mantoro, A. Olowolayemo, and S. O. Olatunji, *Mobile user location determination using extreme learning machine*. IEEE, 2010, pp. D25–D30.

[2] S. O. Olatunji, I. A. Adeleke, and A. Akingbesote, “Data Mining Based on

Extreme Learning Machines for the Classification of Premium and Regular Gasoline in Arson and Fuel Spill Investigation," *J. Comput.*, vol. 3, no. 3, pp. 130–136, 2011.

[3] S. O. Olatunji, Z. Rasheed, K. A. Sattar, A. M. Al-Mana, M. Alshayeb, and E. A. El-Sebakhy, "Extreme Learning Machine as Maintainability Prediction model for Object-Oriented Software Systems," *J. Comput. Vol. 2, Issue 8, August 2010*, vol. 2, no. 8, pp. 42–56, 2010.

[4] S. O. Olatunji, A. Selamat, A. Abdulraheem, and A. A. Abdul Raheem, "A hybrid model through the fusion of type-2 fuzzy logic systems, and extreme learning machines for modelling permeability prediction," *Inf. Fusion*, vol. 16, no. 2014, pp. 29–45, Mar. 2014.

[5] K. O. K. O. Akande, T. O. Owolabi, and S. O. S. O. Olatunji, "Investigating the effect of correlation-based feature selection on the performance of support vector machines in reservoir characterization," *J. Nat. Gas Sci. Eng.*, vol. 22, pp. 515–522, Jan. 2015.

[6] A. E. El-Sebakhy, "Forecasting PVT properties of crude oil systems based on support vector machines modeling scheme," *J. Pet. Sci. Eng.*, vol. 64, no. 1–4, pp. 25–34, 2009.

[7] T. O. Owolabi, K. O. K. O. Akande, and S. O. S. O. Olatunji, "Application of computational intelligence technique for estimating superconducting transition temperature of YBCO superconductors," vol. 43, pp. 143–149, 2016.

[8] T. O. Owolabi, K. O. Akande, and S. O. Olatunji, "Estimation of Surface Energies of Transition Metal Carbides Using Machine Learning Approach," *Int. J. Mater. Sci. Eng.*, no. June, pp. 104–119, 2015.

[9] M. O. Ibitoye, N. A. Hamzaid, A. K. Abdul Wahab, N. Hasnan, S. O. Olatunji, and G. M. Davis, "Estimation of electrically-evoked knee torque from mechanomyography using support vector regression," *Sensors (Switzerland)*, vol. 16, no. 7, 2016.

[10] Q. Xu, Q. Xu, Xu, and Qingsong, "A Comparison Study of Extreme Learning Machine and Least Squares Support Vector Machine for Structural Impact Localization," *Math. Probl. Eng.*, vol. 2014, pp. 1–8, 2014.

[11] G.-J. Cheng, L. Cai, and H.-X. Pan, "Comparison of Extreme Learning Machine with Support Vector Regression for Reservoir Permeability Prediction," in *2009 International Conference on Computational Intelligence and Security*, 2009, pp. 173–176.

[12] S. O. Olatunji, "Comparison of Extreme Learning Machines and Support Vector Machines on Premium and Regular Gasoline Classification for Arson and Oil Spill Investigation," *ASIAN J. Eng. Sci. Technol.*, vol. 1, no. 1, pp. 1–7, 2011.

[13] F. G. Levent Özgür, Tunga Güngör and F. Gürgen, "Spam Mail Detection Using Artificial Neural Network and Bayesian Filter," pp. 505–510, 2004.

[14] Y. Zhang, H. Li, M. Niranjan, and P. Rockett, "Applying Cost-Sensitive Multiobjective Genetic Programming to Feature Extraction for Spam E-mail Filtering," Springer Berlin Heidelberg, 2008, pp. 325–336.

[15] R. Ariaeinejad and A. Sadeghian, "Spam detection system: A new approach based on interval type-2 fuzzy sets," in *2011 24th Canadian Conference on Electrical and Computer Engineering(CCECE)*, 2011, pp. 000379–000384.

[16] F. Temitayo, O. Stephen, and A. Abimbola, "Hybrid GA-SVM for Efficient Feature Selection in E-mail Classification," *ISSN*, vol. 3, no. 3, pp. 2222–1719, 2012.

[17] I. Idris and A. Selamat, "Improved email spam detection model with negative selection algorithm and particle swarm optimization," *Appl. Soft Comput.*, vol. 22, no. September 2014, pp. 11–27, Sep. 2014.

[18] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," *International Joint Conference on Neural Networks (IJCNN2004)*, vol. 2. Budapest, Hungary, pp. 985–990, 2004.

[19] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing, Elsevier*, vol. 70, no. 1–3, pp. 489–501, 2006.

[20] H. Guang-Bin and A. B. Haroon, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Trans. Neural Networks*, vol. 9, no. 1, pp. 224–229, 1998.

[21] G. B. Huang and H. A. Babri, "Feedforward neural networks with arbitrary bounded nonlinear activation functions. 9(1):224–229," *IEEE Trans Neural Netw.*, vol. 9, no. 1, pp. 224–229, 1998.

[22] Cortes and V. Vapnik, "Support vector networks.," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[23] T. O. Owolabi, K. O. Akande, and S. O. Olatunji, "Development and validation of surface energies estimator (SEE) using computational intelligence technique," *Comput. Mater. Sci.*, vol. 101, pp. 143–151, Apr. 2015.

[24] A. A. A. A. Adewumi, T. O. Owolabi, I. O. I. O. Alade, and S. O. S. O. Olatunji, "Estimation of physical, mechanical and hydrological properties of permeable concrete using computational intelligence approach," *Appl. Soft Comput.*, vol. 42, pp. 342–350, Feb. 2016.

[25] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer , N.Y. , 1995.

[26] S. Canu and Y. Grandvalet and V. Guigue and A. Rakotomamonjy, "SVM and Kernel Methods Matlab Toolbox," *Perception Systemes et Information, INSA de Rouen, Rouen, France*, 2008. .

[27] J. S. Mark Hopkins, Erik Reeber, George Forman, "SpamBase Dataset. Hewlett-Packard Labs; 1501 Page Mill Rd.; Palo Alto; CA 94304," 1999.

[28] G.-B. Huang, "MATLAB Codes of ELM Algorithm," *http://www.ntu.edu.sg/home/egbhuang/ELM_Codes.htm*. 2006.

[29] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "Bayesian Additive Regression Trees-Based Spam Detection for Enhanced Email Privacy," in *2008 Third International Conference on Availability, Reliability and Security*, 2008, pp. 1044–1051.

[30] T. O. Owolabi, K. O. Akande, and S. O. Olatunji, "Computational intelligence method of estimating solid- liquid interfacial energy of materials at their melting tem- peratures," *J. Intell. fuzzy Syst.*, 2016.

[31] T. O. Owolabi, M. Faiz, S. O. S. O. Olatunji, and Idris.K.Popoola, "Computational intelligence method of determining the energy band gap of doped ZnO semiconductor," vol. 101, pp. 277–284, 2016.

[32] S. A. Mahmoud and S. O. Olatunji, "Automatic recognition of off-line handwritten Arabic (Indian) numerals using support vector and extreme learning machines," *Int. J. Imaging*, vol. 2, no. 9 A, pp. 34–53, 2009.

[33] S. O. Olatunji and H. Arif, "Identification Of Erythemato-Squamous Skin Diseases Using Extreme Learning Machine And Artificial Neural Network," *Ictact J. Soft Comput.*, vol. 6956, no. October, 2013.