Nanomaterials Characterization Using Hybrid Genetic Algorithm Based Support Vector Machines

Odedele Timothy Oladele
Raw Materials Research and Development Council, Abuja Nigeria
E-mail: odedelermrdc@yahoo.com

***Abstract*—Nanotechnology is the process that develops novel materials at size of 100 nm or less and has become one of the most promising areas of human endeavor. Because of their intrinsic properties, nano-particles are commonly employed in electronics, photovoltaic, catalysis, environmental and space engineering, cosmetic industry and even in medicine and pharmacy. However, recent toxicological studies have shown evident toxicity of some nano-particles to living organisms (toxicity), and their potentially negative impact on environmental ecosystems (ecotoxicity). Characterization is the connection between an abstract material model and its real world behavior. Until recently, relatively simple testing procedures are available for the characterization of engineering materials. However, the large number of nanoparticles and the variety of their characteristics including sizes and coatings show that it is only rational to develop an approach that avoids testing every single nanoparticle produced. The modeling of the material is becoming increasingly difficult and complex such that it requires the use of complex numerical models. A trend is being established where characterization is accomplished through a combination of numerical modeling and experimental testing. Several researchers have carried out analytical and numerical studies on modeling of materials but failed to give a simple model to predict the physico-chemical properties of nano-materials. Computational intelligent techniques such as artificial neural network (ANN), fuzzy logic, genetic algorithm and support vector machine (SVM) are successfully used to solve complex problems. In this paper, a hybrid genetic algorithm tuned support vector machine classifier (GA-SVMC) model is developed to predict the toxicity of nano-materials.**

***Index Terms*—nanotecnology, nanomaterials, support vector machines, genetic algorithm, toxicity, characterization**

I. Introduction

The term "nanotechnology" covers processes associated with the creation and utilization of structures in the 1 nanometer (nm) to 100nm range. Nanofabrication involves engineering at the atomic length scale. Engineering at this scale makes it feasible to create, atom by atom, fibers which are very small in diameter but extremely strong. In the health care domain, nanofabrication can be utilized to fabricate extremely minute probes that can detect disease by examining individual strands of DNA. Nanofabrication makes it possible to manufacture capillary systems for providing nutrients to man-made replacement organs. The unique properties of these [nanotechnology] materials are a double edged sword because they can be tailored for beneficial properties and at the time may also have unknown consequences, such as new toxicological and environmental effects. The following examples illustrate how the same nanotechnology material may be both potentially beneficial and potentially harmful to human health and environment.

Nanoscale silver is highly effective as an antibacterial agent in wound dressings, clothing, and washing machines, but there are general concerns that widespread dispersion of nanoscale silver in the environment could kill microbes that are vital to waste water treatment plants and to ecosystems. Some beneficial bacteria, for example, break down organic matter, remove nitrogen from water, aid in animal digestion, protect against fungal infestations, and even aid some animals in defense against predators [1]-[3].

Due to size of nanoscale particles they may have the potential to penetrate the blood-brain barrier, a structure that protects the brain from harmful substances in the blood but also are such that they may affect the delivery of therapeutic agents. The characteristics of certain nanoscale materials may assist the development of pharmaceuticals to purposefully and beneficially cross this barrier and deliver medicine directly to the brain to treat, for example, a brain tumor. The harmful aspect, however, is the possibility of nanoscale particles to unintentionally pass through the blood-brain barrier causing harm to humans and animals [1]-[3].

It is also generally believed that certain nanoscale materials are highly chemically reactive due to their high surface-to-volume ratio. This is a property which could be positively applied in catalysis, treatment of groundwater contamination, and site remediation. This property which is also being explored for use in protective masks and clothing as defense against chemical and biological agents can potentially result in cell damage in animals.

Carbon nanotubes (CNTs) despite potential uses in a wide range of applications (e.g., materials, batteries, memory devices, electronic displays, transparent conductors, sensors, medical imaging), can exhibit

Manuscript received November 1, 2013; revised February 25, 2014.

doi: 10.12720/ijmse.2.2.107-114

properties similar to asbestos fibers and might become lodged in organs (e.g., lungs, kidneys, livers), harming humans and animals [1]-[3].

Considering the Environmental, Health, Safety (EHS) issues, nanomaterials play decisive roles in their distribution through environment, ecosystem and human body. Due to their biological activities/or unique properties they may gain access into human body through the main ports of entry such as the skin, lungs, gastrointestinal tract. Several toxicological studies have reported that nanomaterials can be cytotoxic, neurotoxic, genotoxic and ecotoxic [4].

These apprehensions of the potential EHS effects of nanomaterials constitute serious barrier to nanotechnology transfer towards business perspectives. It is quite obvious that the current body of knowledge of how nanoscale materials might affect humans and the environment is insufficient to assess, address, and manage the potential risks. While there is agreement on the need for more EHS research, there are differing views on the level of funding required, how it should be managed, and related issues. To accomplish this, there is need to develop sensitive analytical methodologies, tools and an acceptable protocol for screening, characterization and monitoring of nanomaterials in the work place, laboratory, homes and environment. Therefore, considering the EHS issues there is serious need to develop and design predictive models for nanomaterials toxicity using computational intelligent systems (Artificial neural network, neuro-fuzzy systems, hybrid support vector machines and fuzzy inference systems). The objective is to develop computational/predictive model used to establish knowledge domains, risk modeling and nano-informatics capabilities to reliably assist decision making.

Therefore, in order to accomplish this, the following are necessary:

- Development of computational intelligent predictive models for nanomaterials toxicity.
- Development of standardized methods, risk evaluation, risk assessment and management protocol.
- Information sharing, common database for research that uses standard protocols to generate knowledge

Recent advances in machine learning methods have provided attractive alternatives for constructing interpretation models of complex nanomaterials. Here, Support Vector Machines (SVMs), a class of a learning machine that is formulated to output regression models and classifiers of competitive generalization capability, has been explored to determine its capabilities for determining the relationship, both in regression and classification, between physicochemical properties and human health effect which is the main focus. This paper will therefore focus on the capability of SVMs to model physicochemical properties and toxic effect of nanomaterials [4].

Specifically, the capabilities of GA tuned SVM regression and classification will be examined for appropriate prediction purposes on the beneficial and toxic effects of specific nano-particles.

Section I gives a brief introduction. Section II highlights the existing computational model approach for nanomaterial characterization. Section III describes the proposed SVM technique and approach for estimating the optimal SVM parameter settings. In Section IV, detailed numerical data for training and testing the model. SectionV discusses the results of the study. Section VI highlights the conclusion of the study and Section VII presents the software developed for this study.

II. Existing Models

The existing models for the characterization of nanomaterials that could address this problem and find possible solution include Quantitative Structure-Activity Relationship (QSAR), Numerical techniques (finite Element, Classical Laminated theory) and Atomistic, Molecular Methods.

A. QSAR Methodology

QSAR methodology is a quantitative or qualitative relationship between the chemical structure and the biological activity being modeled. The property being modeled is called end point while the form of relationship is called the algorithm. QSAR has been mainly developed for small organic compounds with diverse structural types [5].

The basic requirements to develop a QSAR are (i) large dataset that provides experimental values of a biological activity/property for a group of chemicals and (ii) Molecular structure and/or property data.

B. Finite Element Methods

Finite element methods (FEM) have been used for a wide variety of applications including problems in mechanical, biological, and geological systems. The FEM objective is to provide a numerical, approximate solution to initial-value and boundary-value problems including time-dependent processes. The method uses a variational technique for solving the differential equations wherein the continuous problem described by the differential equation is transformed into the equivalent simultaneous equations and the solution is found solving the systems of linear equations [6]. In the FEM, the physical structure of the domain of interest is broken into simple sub-domains (elements) that are interconnected and stiffness is derived for each element. Global stiffness matrix is obtained and solved subject to initial and boundary values.

C. Atomistic, Molecular Methods

At the atomistic or molecular level, the applicable methods are molecular mechanics, molecular dynamics, and coarse-grained, Monte-Carlo simulation. Molecular models encompassing thousands and perhaps millions of atoms can be solved by these methods and used to predict fundamental, molecular level material behavior [7]. The methods exhibit both static and dynamic attributes. For example, molecular mechanics can establish the minimum-energy structure statically and molecular dynamics can resolve the nanosecond-scale evolution of a

molecule or molecular assembly. These approaches can model both bonded and non-bonded forces (e.g., Van der Waals and electrostatic) but cannot explicitly account for bond cleavage.

III. Genetic Algorithm Tuned Support Vector Machine Classification

A. Overview of Support Vector Machines

Vapnik [8] proposed the support vector machines (SVMs) which was based on statistical learning theory. The basic principles of support vector machines is to map the original data x into a feature space with high dimensionality through a non-linear mapping function and construct hyper plane in new space. The problem of classification can be represented as follows. Given a set of input-output pairs $Z = \{(x1, y1), (x2, y2), ..., (x\ell, y\ell)\}$, construct a classifier function f that maps the input vectors $x \in X$ onto labels $y \in Y$. In binary classification the set of labels is simply $Y = \{-1, 1\}$. The goal is to find a classifier $f \in F$ which will correctly classify new examples (x, y), i.e. $f(x) = y$ for examples (x, y), which were generated under the same probability distribution as the data [9]-[11] Binary classification is frequently performed by finding a hyper-plane that separates the data, e.g. Linear Discriminant Analysis (LDA. There are two main issues to consider when we use a separating hyper-plane:

- The problem of learning this hyperplane is an ill-posed one because there is no unique solution and many solutions may not generalize well to the unseen examples.
- The data might not be linearly separable.

SVMs tackle the first problem by finding the hyper-plane that realizes the maximum margin of separation between the classes. [12] A representation of the hyper-plane solution used to classify a new sample *xi* is:

$$Y = f(x) = wi\phi(x) + b \quad (1)$$

where $wi, \phi(x)$ is the dot-product of the weight vector w and the input sample, and b is a bias value. The value of each element of w can be viewed as a measure of the relative importance of each of the sample attributes for the classification of a sample. It has been shown that the optimal hyperplane can be uniquely constructed by solving the following constrained quadratic optimization problem [2]

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_I \quad (2a)$$

subject to $yi(\|w\| + b) \geq 1 - \xi i, i = 1, ..., \ell$

$$\xi i \geq 0, i=1, ..., \ell \quad (2b)$$

In linearly separable problem, the solution minimizes the norm of the vector w which increases the flatness (or reduces the complexity) of the resulting model and thereby improves its generalization ability. With non-linearly separable hard-margin optimization, the goal is simply to find the minimum $\|w\|$ such that the hyper-plane $f(x)$ successfully separates all ℓ samples of the training data. The slack variables ξi are introduced to allow for finding a hyper-plane that misclassifies some of the samples (soft-margin optimization) because many datasets are not linearly separable. The complexity constant $C>0$ determines the trade-off between the flatness and the amount by which misclassified samples are tolerated. A higher value of C means that more importance is attached to minimizing the slack variables than to minimizing $\|w\|$. Instead of solving this problem in its primal form of (2a) and (2b), it can be more easily solved in its dual formulation by introducing Langrangian multiplier α [12]:

$$\text{Maximize } W(\alpha) = \sum_{i=1}^{l} \alpha i + \tfrac{1}{2} \sum_{i,j=1}^{l} \alpha i \alpha j y i y j \langle xi, xj \rangle \quad (3a)$$

Subject to $C \geq \alpha i \geq 0$,

$$\sum_{i=1}^{l} \alpha i y i = 0 \quad (3b)$$

In this solution, instead of finding w and b the goal now is find the vector α and bias value b, where each αi represents the relative importance of a training sample I in the classification of a new sample. To classify a new sample, the quantity

$f(x)$ is calculated as:

$$f(x) = \sum_{i=1}^{sv} \alpha i y i \langle xi, xj \rangle + b \quad (4)$$

where b is chosen so that $yif(x) = 1$ for any i with $C>\alpha i>0$. Then, a new sample xs is classed as negative if $f(xs)$ is less than zero and positive if $f(xs)$ is greater than or equal to zero. Samples xi for which the corresponding αi are non-zero are known as *support vectors* since they lie closest to the separating hyper-plane. Samples that are not support vectors have no influence on the decision function. Training an SVM entails solving the quadratic programming problem of (3a) and (3b). There are many standard techniques that could be applied to SVMs, including the Newton method, conjugate gradient and primal-dual interior-point methods. For the experiments reported here the SVM implementation uses the Sequential Minimal [13]. One key aspect of the SVM model is that the data enters the above expressions (3a and 4) only in the form of the dot product of pairs. This leads to the resolution of the second problem mentioned above, namely that of non-linearly separable data. The basic idea with SVMs is to map the training data into a higher dimensional feature space via some mapping $\varphi(x)$ and construct a separating hyperplane with maximum margin. This yields a non-linear decision boundary in the original input space. By use of a kernel function, $K(x, z) = \langle \varphi(x), \varphi(z) \rangle$it is possible to compute the separating hyperplane without explicitly carrying out the mapping into feature space. [12] Typical types of kernels are:

- Linear Kernel: $K(x, z) = \langle x, z \rangle$
- Polynomial Kernel: $K(x, z) = (\langle x, z \rangle)^d$
- RBF Kernel: $K(x, z) = \exp(-\|x-z\|^2/2\sigma 2)$
- Sigmoid Kernel: $K(x, z) = \tanh(\gamma * \langle x, z \rangle - \theta)$

This condition ensures that the solution of (3a) and (3b) produces a global optimum. The functions that satisfy Mercer's conditions can be as kernel functions.

B. Genetic Programming for Selecting SVM Parameters

Support vector machine (SVM), which is a statistical learning theory based machine learning method, is gaining popularity due to its many attractive features and promising generalization performance. Some prominent features of SVM are: (i) the ability to model non-linear relationships, (ii) SVM generalization performance does not depend on the dimensionality of the input space, (iii) the regression function is related to a quadratic programming problem whose solution is global and in general unique. Apart from these features, SVM also has a drawback that limits the use of SVM on academic and industrial platforms: there are free parameters (SVM hyper-parameters and SVM kernel parameters) that need to be defined by the user. Since the quality of SVM regression models depends on a proper setting of these parameters, the main issue when we use SVM is how to set these parameter values (to ensure good generalization performance) for a given training data set. The existing approaches to setting SVM parameters and a practical method for selecting the values of C (the regularization parameter) and ε (the radius of the insensitive tube) are summarized [14], [15]. However, all these approaches (including the one proposed in [14], [15] are formed on the basis of prior knowledge, user expertise, or experimental trial, and hence there is no assurance that the parameter values obtained are truly optimal. On the other hand, further complication on the optimal parameter selection arises through the fact that the SVM generalization performance depends on all of these parameters (both hyper-parameters and kernel parameters) together. This means that the interaction of SVM parameters has to be considered jointly, and that a separate optimization of each parameter is not adequate enough to find the optimal regression model [16]. In view of the above, in the practical application of SVM regression, usually a time-consuming grid search method is invoked to estimate the optimal SVM parameter settings [17]. When applying grid search method, one might need to increase the parameter range and / or decrease the step size to increase the accuracy of the optimal solution. However, this will result in a cumbersome time consuming search process. The GA typically starts off with a random population of individuals, each encoding a function or expression. This population is evolved by selecting better individuals for recombination and using their offspring to create a new population (generation). Mutation is employed to encourage discovery of new individuals. This process is continued until some stopping criteria is met, e.g. homogeneity of the population.

The approach presented here combines the two techniques of SVMs and GP, using the GP to select parameters for a SVM. The algorithm is based on the principle of survival of the fittest which tends to retain information from generation to generation [18]. In this paper genetic algorithm is used to search for better combination of C, ε and kernel (d and γ) parameters to maximize the generalization performance of SVM models. An overview of the proposed GK SVM is shown in Fig. 1.

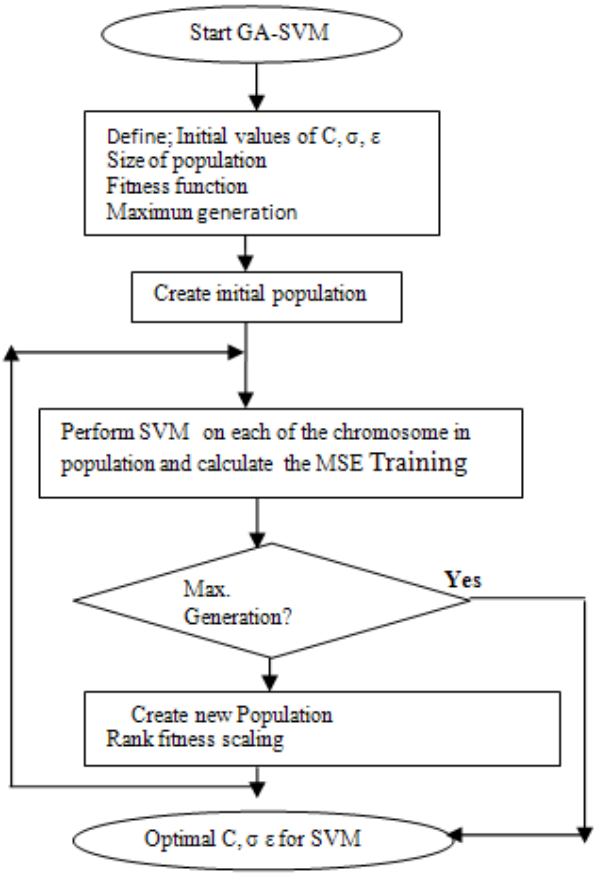

Figure1. GA-SVM Procedure

The main steps in the building of a GA are:

- Create a random population of chromosomes generated randomly.
- Evaluate the fitness of each individual by building an SVM and test it on the training data

The fitness function=$1/N\sum_{i=0}^{n}(d-yi)^2$, where N is the total nuber of data, d is the actual value and y_i is the predicted values.

- Select the two parent chromosomes from population according to fitness function. The roulette principle is used for selection.
- Crossover: Here using crossover probability, crossover of parents is done to form new offspring (children). In crossover, chromosomes are paired randomly.
- Perform random mutation on the newly created offspring
- Replace the old population with the offspring
- Repeat Steps 2 to 5 until the population has converged
- Build final SVM using the fittest chromosome found

IV. Numerical Experiments

Prediction performance of the resulting models depends on the size and quality of the training data. Each

data record consists of input and output data. Input data are derived from physicochemical properties of the materials as shown in Table I.

TABLE I. NANOMATERIALS PHYSICOCHEMICAL PROPERTIES

Measured Attribute	Toxicity potential index	Score
Size characteristics	X_1 average particle size X_2 aggregate size	15 10
Surface area and charge characteristics	X_3 surface area X_4 specific surface area X_5 surface charge X_6 hydrophobicity X_7 point of zero change X_8 zeta potential	3 2 8 5 2 5
Chemical composition structure	X_9 aspect ratio X_{10} bulk density	15 10
Reactivity characteristics	X_{11} degradability X_{12} hydrolysis rate X_{13} biodegradation rate X_{14} photolysis rate X_{15} Redox reaction rate	9 6 3 4 3
Partitioning characteristics	X_{16} solubility X_{17} volatility X_{18} partition coefficient	12 6 7

A. Purpose of Study

The objective of this study is to classify the nanomaterials toxicity based on size, surface area and charge, chemical composition structure, reactivity and partitioning characteristics. The flow chart for nanomaterial characterization/classification is as shown in Fig. 2.

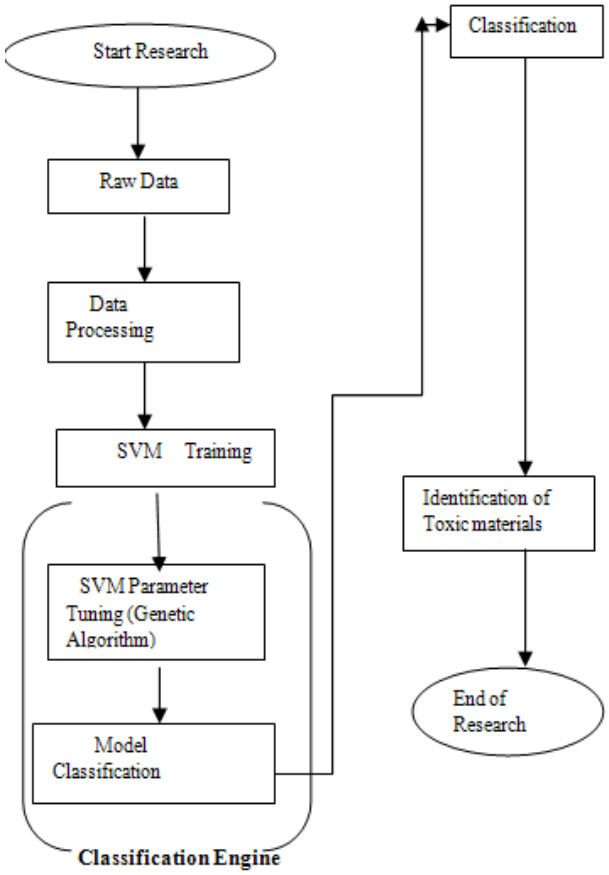

Figure 2. Flow chart of the nanomaterial classification/characterization

B. Data Sets

The data of this study were adopted to represent different species of a nanomaterial. 20 experimental samples, 25 samples for testing data random select from database.

C. Nanomaterials Toxicity Potential Indicators

In this study, using size, surface area and charge, chemical composition structure, reactivity and partitioning characteristics are measured attributes. There are 15 measure indexes (as in Table I). Table I shows nanomaterials toxicity potential indicators.

D. Process

Step1: Data preprocess and variable selection

In this study, the measured attribute are size, surface area and charge, chemical composition structure, reactivity and partitioning characteristics. Table I is showed as toxicity potential index. From the Table I, the measured attributes values of each listed nano-material are $(X_1*15+X_2*10)/25$ in size characteristics, $(X_3*3+X_4*2+X_5*8+X_6*5+X_7*2+X_8*5)/25$ in surface area and charge characteristics, $(X_9*15+X_{10}*10)/25$ in chemical composition structure $(X_{11}*9+X_{12}*6+X_{13}*3+X_{14}*4+X_{15}*3)/25$ in reactivity characteristics and $(X_{16}*12+X_{17}*6+X_{18}*7)/25$ in partitioning characteristics. Table II is denoted as the grade of measured attributes (pretreatment training data).

Step 2: Sample data processing

In this study, 15 toxicity potential indexes were used, and select 20 training sample data, 25 testing sample data from listed species of a nanomaterial. The training experiments were conducted on a small data set. According to Table I, we have the grade of measured attributes. The measured attributes are size, surface area and charge, chemical composition structure, reactivity and partitioning characteristics, y is sample data decision attribute. If listed nanomaterial is toxic (T) then *y* is 1, and otherwise *y* is -1

Step 3: Solve the nanomaterial toxicity evaluation problem

In Computational Intelligent Nanomaterials Toxicity (CINT) software (developed by the author), genetic algorithm was used to find the best parameter C and kernel parameters.

The result confirmed that the classification precision of the SVM with radial function (RBF) kernel function was high as 100% when γ and C where 0.55 and 1. Then the best parameter of C and γ was selected to train the whole training set, we have 11 support vector index sets.
The outputs from NCIS software are
Accuracy=100%
MSE=0.0
Squared correlation coefficient=1

V. RESULTS AND DISCUSSION

The sample data used for testing are as shown in Table III. There are two types of errors namely Type I and Type

II errors. Type I refers to a situation when toxic material is classified as non- toxic material. Type II refers to non-toxic material being classified as toxic material. The predicted result is as listed in Table III. The results of testing (external validation check were summarized in Table IV. We observe form these results that the hybrid genetic algorithm-support vector machines modeling scheme performs satisfactorily for predictive correlations. The proposed model shows a high accuracy in predicting toxicity class with a stable performance, and achieved the lowest absolute percent relative error typeI and typeII errors, lowest root mean square error, and the highest correlation coefficient among other correlations for the used two distinct data sets. The interface for data input for SVM implementation is as shown in Fig. 3. A plot of the experimental and predicted data versus the input data is as shown in Fig. 4.

TABLE II. NANOMATERIALS TRAINING DATASET

Listed species of a nanomaterial	Size characteristics	Surface area and charge characteristics	Chemical composition structure	Reactivity characteristics	Partitioning characteristics	Attribute of y
1	0.23	0.20	0.09	0.20	0.22	-1
2	0.18	0.18	0.10	0.21	0.23	-1
3	0.16	0.18	0.08	0.17	0.18	-1
4	0.19	0.11	0.12	0.18	0.16	-1
5	0.20	0.22	0.11	0.19	0.20	-1
6	0.24	0.2	0.09	0.20	0.23	-1
7	0.23	0.14	0.06	0.20	0.21	-1
8	0.20	0.08	0.07	0.10	0.12	1
9	0.18	0.09	0.05	0.18	0.15	1
10	0.19	0.12	0.03	0.12	0.13	1
11	0.22	0.13	0.04	0.15	0.15	-1
12	0.16	0.10	0.07	0.14	0.13	1
13	0.19	0.09	0.11	0.12	0.15	1
14	0.15	0.18	0.16	0.10	0.13	-1
15	0.18	0.20	0.20	0.08	0.07	-1
16	0.12	0.17	0.18	0.13	0.15	-1
17	0.21	0.18	0.10	0.12	0.10	-1
18	0.19	0.18	0.12	0.09	0.08	-1
19	0.22	0.19	0.09	0.14	0.15	-1
20	0.20	0.15	0.15	0.09	0.08	-1

TABLE III. TEST SAMPLE DATA

Listed species of a nanomaterial	Size characteristics	Surface area and charge characteristics	Chemical composition structure	Reactivity characteristics	Partitioning characteristics	Attribute of y
1	0.22	0.22	0.20	0.10	0.20	-1
2	0.21	0.20	0.18	0.10	0.21	-1
3	0.17	0.16	0.20	0.08	0.17	-1
4	0.18	0.19	0.15	0.12	0.18	-1
5	0.22	0.20	0.22	0.20	0.19	-1
6	0.21	0.22	0.20	0.09	0.20	-1
7	0.22	0.23	0.18	0.06	0.20	-1
8	0.24	0.20	0.08	0.09	0.10	1
9	0.19	0.18	0.12	0.05	0.11	1
10	0.18	0.19	0.15	0.03	0.12	1
11	0.20	0.22	0.13	0.08	0.15	-1
12	0.19	0.18	0.10	0.07	0.14	1
13	0.20	0.19	0.09	0.12	0.12	1
14	0.17	0.16	0.18	0.16	0.10	-1
15	0.17	0.18	0.22	0.20	0.08	-1
16	0.13	0.12	0.17	0.18	0.15	-1
17	0.27	0.21	0.18	0.12	0.12	1
18	0.18	0.19	0.20	0.12	0.10	-1
19	0.20	0.22	0.15	0.09	0.14	-1
20	0.23	0.21	0.15	0.18	0.10	-1
21	0.22	0.20	0.13	0.09	0.15	-1
22	0.19	0.18	0.15	0.07	0.18	1
23	0.17	0.16	0.10	0.12	0.14	1
24	0.18	0.19	0.18	0.16	0.18	-1
25	0.17	0.16	0.20	0.20	0.09	-1

TABLE IV. TOXICITY PREDICTION RESULTS

Method	Number of samples	Type I error	TypeII error	Error	Accuracy
GA-SVM	11	0%	0%	0	100%

7.0 Software Implementation

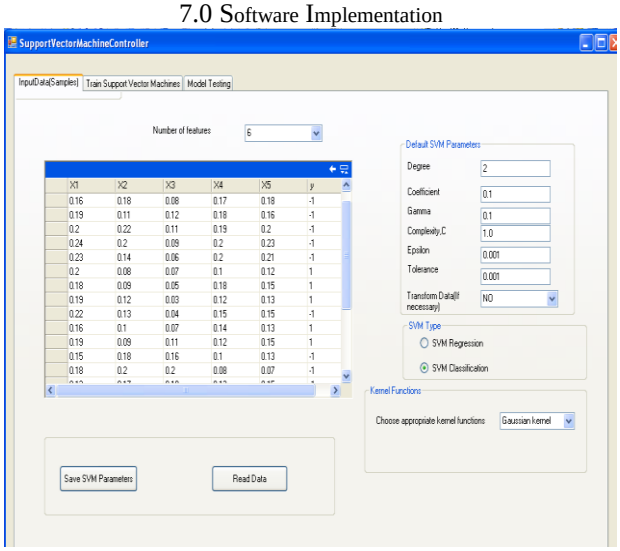

Figure 3. Data input for SVM implementation

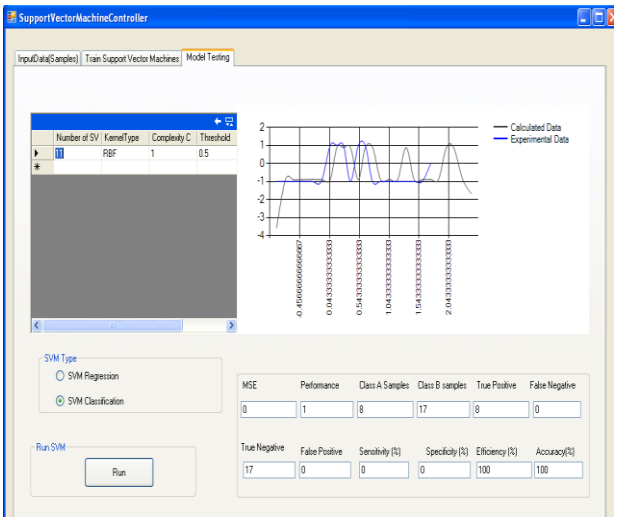

Figure 4. Plot of Calculated and experimental data versus input data

VI. CONCLUSION

This study develops a novel model to search the optimal values of SVM parameters, to increase the accuracy of prediction and ability of generalization of SVM and the proposed model (GA-SVM) were applied to a dataset on nanomaterial toxicity. First, this study found that the GA yields different optimal values of the parameters of SVM given various datasets. The classification of nanomaterial (toxic or non toxic) using GA-Support Vector Machines is a work that is aimed to start with an in-depth study and understanding of the various aspects of Support Vector Machines (SVM) and Genetic Algorithm, which is used widely for classification and regression purposes. The study and understanding of the SVM technique and its role in classification tasks are done. The classification process follows supervised learning model, in which the available or known data is used to train the model based on similarity measures. With the trained model, the unseen or new data is classified into either of the classes. The non-linear input space is handled using Gaussian kernel function, which is chosen for this application. This technique is then implemented in the Microsoft *C#* programming language to perform data classification task for the nanomaterial toxicity data set. This approach is rather less complex than the traditional computational modeling used for material classification and characterization.

REFERENCES

[1] John F. Sargent Jr., *Nanotechnology and Environmental, Health and Safety: Issues for Consideration*, DIANE Publishing, 2008.

[2] K. L. Dreher, "Health and environmental impact of nanotechnology: Toxicological assessment of manufactured nanoparticles," *Toxicol. Sci.*, vol. 77, pp. 3-5, 2004.

[3] R. D. McNeil, J. Lowe, T. Mastroianni, J. Cronin, and D. Ferk, "Barriers to nanotechnology commercialization," *Report Prepared for U.S. Department of Commerce, Technology Administration*, Sep. 2007.

[4] A. Gajewicz, B. Rasulev, *et al.*, "Advancing risk assessment of engineered nanomaterials: Application of computational approaches," *Advanced Drug Delivery Reviews*, vol. 64, no. 15, pp. 1663-1693, 2012.

[5] A. P. Worth, A. Bassam, A. Gallegos, *et al.*, *The Characterization of (Quantitative Structure-Activity Relationship: Preliminary Guidance*, 2005.

[6] T. O. Odedele, *Introductory Numerical Analysis and Computer Methods for Scientists and Engineers*, 2008.

[7] M. W. Schmidith, K. K. Baldridge, *et al.*, "General atomic and Molecular electronic structure System," *J Comput. Chem.*, vol. 14, no. 11, pp. 1347-1363, 1993.

[8] V. Vapnik, *Statistical Learning Theory*, New York: John Wiley and Sons, 1998.

[9] Sch ölkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, 2001.

[10] Sch ölkopf, P. Simard, A. Smola, and V. Vapnik, "Prior knowledge in support vector kernels," *Advances in Neural Information Processing Systems*, pp. 640-646, 1998.

[11] Sch ölkopf, A. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.

[12] Sch ölkopf, S. Mika, C. Burges, *et al.*, "Input space vs. feature space in kernel-based methods," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1000-1017, 1999.

[13] Sch ölkopf, A. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, pp. 1207-1245, 2000.

[14] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks*, vol. 17, pp. 113-126, 2004.

[15] X. F. Yuan and Y. N. Wang, "Parameter selection of SVM for function approximation based on chaos optimization," *Journal of Systems Engineering and Electronics*, vol. 19, pp. 191-197, 2008.

[16] Z. Y. Luo, P. Wang, Y. G. Li, W. F. Zhang, W. Tang, and M. Xiang, "Quantum-inspired evolutionary tuning of SVM parameters," *Progress in Natural Science*, vol. 18, pp. 475-480, 2008.

[17] K. Ito and R. Nakano, "Optimizing support vector regression hyper-parameters based on cross-validation," in *Proc. International Joint Conference on Neural Networks*, 2003, pp. 2077-2082.

[18] B. Üst ün, W. J. Melssen, M. Oudenhuijzen, and L. M. C. Buydens, "Determination of optimal support vector regression

parameters by genetic algorithms and simplex optimization," *Analytica Chimica Acta*, vol. 54, pp. 292-305, 2005.

Engr Odedele Timothy Oladele who hails from Ipetumodu, Osun state of Nigeria was born on 29th April 1958. The educational background is as follows: B.Sc. Petroleum Engineering (Second class Upper Division1984) University of Ibadan, Ibadan Post Graduate Diploma Computer Science (Upper Credit.2008) Federal University of Technology, Minna, M.Sc Computer Science -University of Ibadan, Ibadan (In view). He works with Raw Materials Research & Development Council, Abuja Nigeria. He is currently an Assistant Director (New and advanced Materials Division). He has published several books which include:
Introductory Numerical Analysis and Computer Methods for Scientists and Engineers.
Computer Aided Project Appraisal: Kagara Talc Processing Plant as a case study
Using Computer software Packages-MS Word, Excel, Access, Powerpoint Microsoft Project, Visual Basic.Net and SQL Database
Computational Algorithm Using Visual Basic .Net Programming language: Application in Numerical Methods and Statistics
Technical Brief on New and Advanced Materials
Computer-Aided Advanced Composite Design
His seminar/conference papers include:
Computer-Aided Project Evaluation: Kagara Talc Processing Plant as a case study
Information Technology for an effective raw material development and utilization: RMRDC˙S Perspective
Potentials of New and Advanced Materials Development in Nigeria
Promoting the Utilization of Local Materials in the Construction Industry
Computer analysis of the laminated composite plate under combined thermal/hygral loads
Process equipment design using case-based reasoning
SOFTWARES DEVELOPED include:
Project Appraisal Software
Human Resources Management system
Numerical and statistical solutions software
Petroleum Well log Interpretation software
Drilling and Well Design software
Computer Aided Advanced Composite Design
PRIZE WON includes 1983/84 Department of Petroleum Engineering Prize.
PROFESSIONAL BODIES include:
Nigerian Society of Engineers (NSE)
Council for Regulation of Engineering in Nigeria (COREN)
Nigeria Computer Society(NCS)

