

## Imbalanced data classification using support vector machine based on simulated annealing for enhancing penalty parameter

Hussein Ibrahim Hussein<sup>1</sup>, Said Amirul Anwar<sup>2</sup>

<sup>1,2</sup>Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Kampus Alam UniMAP, Pauh Putra, 02600 Arau, Perlis, Malaysia

### ABSTRACT

For pattern cataloguing and regression issues, the support vector machine (SVM) is an eminent and computationally prevailing machine learning method. It's been effectively addressing several concrete issues across an extensive gamut of domains. SVM possesses a key aspect called penalty factor C. The choice of these aspects has a substantial impact on the classification precision of SVM as unsuitable parameter settings might drive substandard classification outcomes. Penalty factor C is required to achieve an adequate trade-off between classification errors and generalisation performance. Hence, formulating an SVM model having appropriate performance requires parameter optimisation. The simulated annealing (SA) algorithm is employed to formulate a hybrid method for evaluating SVM parameters. Additionally, the intent is to enhance system efficacy to obtain the optimal penalty parameter and balance classification performance at the same time. Our experiments with many UCI datasets indicate that the recommended technique could attain enhanced classification precision.

**Keywords:** Imbalanced data, support vector machine, penalty parameter, simulated Annealing.

### Corresponding Author:

Hussein Ibrahim Hussein  
Faculty of Electronic Engineering Technology  
Universiti Malaysia  
Perlis, Malaysia  
[Husseinsarhan45@yahoo.com](mailto:Husseinsarhan45@yahoo.com)

### 1. Introduction

Structural risk minimization (SRM) algorithm based, a new classification approach was presented by Cortes and Vapnik [1], generally known as Support vector machine (SVM). The algorithm was quickly deployed for many cataloguing jobs because of its efficiency in identifying handwritten characters, wherein it outclassed accurately trained neural networks. SVMs carried out effective classification in other uses like time series estimation, bioinformatics and pattern classification. Burges (1998) issued an inclusive tutorial about the SVM classifier algorithm. SVM can process massive feature spaces since SVM training is performed so that classified vector dimensions do not exert undue influence on SVM performance compared to the influence of a typical classifier. Therefore, it is said to be particularly effective in bigger classification problems. Furthermore, SVM-based classifier is said to have sound generalization attributes as against traditional classifiers, because SVM classifier training requires a systematic reduction in misclassification risk. On the other hand, conventional classifiers are typically trained to minimise empirical risk [2]. Numerous techniques have been proposed to address this SVM issue. Huang and Wang [3] recommended a genetic algorithm (GA) method for constraint refinement[4]. Ren and Bai [5] too offered twin methodologies for constraint refinement in SVM: particle swarm optimization (PSO) SVM and GA-SVM. A classifier using the hybrid ant colony optimisation (ACO) technique concurrently identifies the best possible feature subset and works on SVM parameter optimisation; this technique was formulated by Huang [6]. Also, Lin et al. [7] worked on parameter value computation and SVM feature selection using the Simulated Annealing concept. The simulated annealing algorithm optimises an SVM by addressing the issue of the system being stuck at local optima. It works by facilitating non-optimal steps to be selected based on probability values. The technique was outlined separately by Kirkpatrick et al. [8]. Simulated annealing chooses an explanation in each repetition via examining whether neighbour explanation is preferable to present scenario. If yes, the novel explanation would be acknowledged unreservedly. Nevertheless,

if the proximal solution is not better, acceptance relies on probability values considering its difference from the proximal solution; the present solution value is also considered. This paper assesses the combination of SVM and Simulated Annealing to determine optimal parameters for the betterment of SVM accuracy[9]. Our experimental outcomes have proved that the recommended technique has greater precision in comparison to SVM. The remainder of the paper is organised as follows. Sect. 2 contains a brief overview of the literature. Sect. 3 presents the proposed SA-SVM algorithm. The experimental results are presented in Section 4, and the findings are presented in Section 5.

## 2. Related work

This field has witnessed over two decades of continuous research; however, imbalanced data-based learning is still researched extensively. Initially, binary problems with skewed distributions were the focus of this domain. Over time, the concept has grown extensively. Advances in data mining, machine learning, and big data capabilities have provided insightful information concerning imbalanced learning. Such advances have also led to numerous challenges being discovered. Hybrid techniques are now getting increased attention, and algorithms are being continuously made better. Table 1 lists the associated research conducted for SVM parameter enhancement.

Table 1. Optimization of SVM penalty parameter

Authors	Context of Study	Method	Contribution
Kuo-PingWu [10]	Classification	Grid search and SVM	Suggested the use of the distance between feature space and clusters to identify kernel parameters.
Iwan Syarif [11]	Classification	Grid search, GA, and SVM	Proposed the use of a combination of the Genetic Algorithm (GA) and the Grid Search Algorithm for SVM parameter optimisation.
Yao Chang [12]	Traffic sign recognition	HOG-SVM and GS	Concurrent use of grid search (GS) and HOG-SVM is suggested. Subsequently, grid search is used for SVM parameter optimisation.
Wu et al. [13]	Multidimensional time series	CPSO-g-SVRM	The e-insensitive loss function is replaced by the Gaussian loss function for SVR in order to reduce the influence of noise on regression values. The suggested g-SVRM technique is used, and its parameters are optimised using the chaotic PSO technique.
Xueying Zhang [14]	Speech recognition system	PSO and SVM	This research proposes using particle swarm optimisation to ascertain optimal parameters used for a speech recognition system based on SVM.
Yaxiong Zhang [15][16]	Classification	GA and SVM	SVM was used to augment a predictive modelling method that relied on identifying the right combination of training and test data by employing the sphere exclusion technique and SVM parameter optimisation using the GA.
M.R. Gauthama Raman [17]	Classification	(HG - GA) and SVM	A robust intrusion detection mechanism has been suggested using hypergraph-based Genetic Algorithm (HG - GA) for SVM feature identification and parameter regulation.

## 3. Methodology

This section outlined the recommended technique to categorise the imbalanced dataset by using Simulated Annealing and SVM. The dataset as well as the performance metrics utilised to corroborate the recommended technique are also deliberated.

### 3.1. Data set

The datasets originate from the UCI Machine Learning Repository which is accessible at the link: <http://www.ics.uci.edu/~mllearn/MLRepository.html>. Thorough information comprising the number of attributes, data size, data attributes and class distribution can be observed in Table 2. Initially, just the ten data sets that comprise binary imbalanced data (Australian, heart C, heart-statlog, ionosphere, liver, Pima, hepatitis, breast cancer, kidney and German credit data) are utilised. However, the experimental outcomes indicate that our recommended approach has sound ability of dealing with imbalanced data.

Table 2. Details of imbalanced dataset

Dataset name	Number of instances	Attributes number	Missing value	Attribute characteristics
Australian	690	14	Yes	Categorical, Integer, Real
Heart-statlog	270	13	No	Categorical, real
Heart C	303	14	Yes	Categorical, Integer, Real
Liver	345	7	No	Categorical, Integer, Real
Ionosphere	351	34	No	Integer, Real
Hepatitis	155	19	Yes	Categorical, Integer, Real
Pima	N/A	20	N/A	Categorical, Integer
Breast cancer	286	9	Yes	Categorical
kidney	400	25	Yes	Real
German credit data	1000	20	N/A	Categorical, Integer

### 3.2. Simulated annealing

Simulated annealing (SA) is a universal search algorithm that first recommended by Metropolis et al. [18] and then made popular by Kirkpatrick et al. [8]. SA borrows its elementary understanding from metallurgy. Metal molecules slowly attain a low-energy state and witness gradual crystallisation with a reduction in temperature. Every grain is expected to ultimately have lowest energy if the metal is initially heated to an adequately high temperature and cooling is kept gradual. Metropolis proposed a technique that enhances search results and also prevents the local optima problem. Furthermore, the “cooling” process, similar to metal cooling, facilitates SA to ultimately converge to results that indicate global optima. Parameter optimisation is typically performed using techniques like analytical gradient, genetic algorithms, Monte Carlo, and numerical gradient. Global optima for parameters are identified using Simulated Annealing (SA). Though SA requires relatively more time, it determines more accurate solutions compared to other techniques. The present study uses SA for identifying optimal SVM parameter values. The present study suggests an SA-augmented SVM to assess the effectiveness association with every C value that corresponds to iterations conducted using parameter T. A novel technique has been proposed for minimising T so that a gradual convergence towards global optimality can be attained. Parameter T is multiplied by value V, which lies in the [0,1] range. The SA algorithm relies on the initial and subsequently probability state. The critical decision concerns retaining the present state or moving to the subsequent state ( $C_j$ ) based on value comparisons. If the new state ( $C_j$ ) has a fitness value greater than the present state, it is accepted. On the other hand, if the new state has relatively less fitness value, its acceptance is associated with a probability value based on the problem objective. Such probabilistic transitions gradually move the system to a close-to-optimal state. Individual iterations comprise random neighbour selection. If the chosen neighbour has accuracy higher than the present state, the neighbour is selected as the present state, and the parameters (C) are updated accordingly. SA accepts solutions ( $C_j$ ) considering probability value P that is influenced by T. Figure 1. depicts the working of the SA technique.

$$P(c_i, c_j, T) = \begin{cases} c_j & \text{if } f(c_j) \geq f(c_i) \\ e^{-\frac{f(c_i) - f(c_j)}{T}} & , \text{otherwise} \end{cases} \quad (1)$$

where T is the temperature and  $f(c_j)$  and  $f(c_i)$  are the fitness scores. The worst  $f(c_i)$  and new solution  $f(c_j)$  are depicted in Fig1.

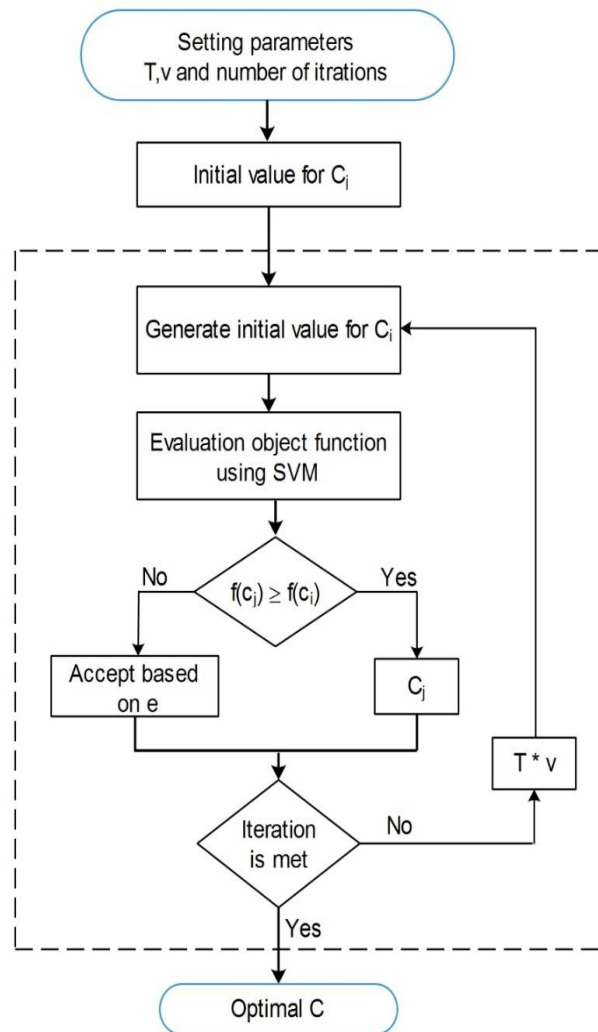


Figure 1. Tuning parameter selection

The primary objective is to provide the algorithm additional time to evaluate the search space using the “temperature” parameter. In high-temperature cases, the algorithm works like a random search where all transitions are accepted, notwithstanding their effect, facilitating better search.

The proposed technique relies on SA to find the C factor value for the SVM classifier as well as for its training set. The search space is relatively better because of the use of a single parameter, C.

### 3.3. Support Vector Machine

Support vector machines (SVM) were initially proposed by [1] and [19] using the Vapnik-Chervonenkis (VC) framework augmented using structural risk minimisation (SRM) (Vapnik, 1995, 1998). The framework works by identifying the ideal balance between training set error reduction and margin maximisation. The model has optimal generalisation capability while also remaining free from overfitting. Another benefit of using SVM is convex quadratic programming, where the local minima trap is avoided because only global minima are reported. Consider a training set comprising input and output, where the input contains feature extraction aspects  $(x_1, x_2, x_3, \dots, x_n)$  along with the output outcomes classes  $\{(y_1, y_2, y_3, \dots, y_N), (x_N, y_N)\}$ , where  $x_i \in \text{input features}$  and  $y_i \in \{\text{class} - 1, \text{class} + 1\}$ . The precise value of (y) can be computed using a set of weights  $w_i$ . The present study uses margin maximisation to ascertain the hyperplane.

$$\begin{aligned}
 f(x) &= w^T x_i + b \\
 f(x) &= \sum_i \lambda_i y_i (x_i^T x + b) \\
 f(x) &\geq 1, \quad \forall x \in \text{class 1} \\
 f(x) &\leq -1, \quad \forall x \in \text{class 2}
 \end{aligned} \tag{2}$$

$$H = \frac{|g(x)|}{\|w\|} = \frac{1}{\|w\|}$$

$w^T$  contains vector weights, while  $f(x)$  denotes the features sets corresponding to the classes,  $\lambda_i$  denotes the dual function when the training condition  $0 \leq \lambda \leq C$  is satisfied,  $x$  denotes the training set,  $y$  denotes the output classes,  $H$  denotes SVM hyperplanes,  $g(x)$  denotes feature count,  $b$  denotes bias that belongs to omega 0.

Originally, support vector machines (SVMs) were formulated for binary classification. The question of how to successfully extend it for multiclass grouping is still being researched. Multiclass SVM uses two different approaches. The first involves building and integrating multiple binary classifiers, while the second involves explicitly considering all data in a single optimisation formulation. Multiclass problems can be split into several binary problems, each of which can be addressed independently. **Pseudocode for proposed method**

---

**Algorithm 1: SA-SVM**

---

```

1 Input:
2  $C = c_i \in N, i \in \{1,2,\dots,N\}$  – the start state of N samples of solutions;
3  $k =$  number of iterations.
4  $T =$  number of temperatures.
5  $f =$  fitness function
6  $v =$  value for reduce the temperature
7  $S = \{(x_i, y_i) | x_i \in R^n, y_i \in m, i \in \{1,2,\dots,N\}\}$  – the set of N training samples and class;
8  $Z = \{z_i | z_i \in R^m, i \in \{1,2,\dots,t\}\}$  – the set of t test samples;
9 Begin:
10 for  $k \leftarrow 1$  to  $n - n$  the max number iterations
9      $C' \leftarrow$  neighbor(C): pick a random neighbor;
10     If  $f(C') \geq f(C)$ :
11          $C \leftarrow C'$ ;
12          $f =$ SVM
13     else
14          $C \leftarrow e^{\frac{f(C)-f(C')}{T}}$ ;
15          $T \leftarrow T * V$ ;
16 Stack function  $\leftarrow C$ 
17 Output:
18 select best solution C
19 Initialization
20  $Y \leftarrow \emptyset$ 
21 Computation:
22 for  $z_i \in Z$  do
23  $N \leftarrow$  the nearest k neighbours to  $z_i$  from S according to;
24  $f \leftarrow$  the discriminant function of C trained on N find by SA;
25  $y \leftarrow$  the label predicted by applying f on  $z_i$ ;
26  $Y \leftarrow Y \cup \{y\}$ ;
27 Output:
28  $Y = \{y_i | y_i \in N i \in \{1,2,\dots,1\}\}$  - the set of predicted labels for the test samples in Z.

```

---

**3.4. Evaluation metric**

Considering the unique characteristics of class-imbalanced problems, the traditional classifier specifically classifies minority class samples as majority and achieves a higher global precision score. However, the accuracy rating for minority-class samples is poor. The proposed strategy entails equating the class and deleting duplicates to work for class equalisation and eliminating duplicates that interfere with the classification process. For substantiating the efficacy and usefulness of the recommended approach, precision, recall, and F-score are utilised for evaluating the class-imbalanced classification which utilises the confusion matrix, as indicated by formulas (3) and (4).

$$Precision = \frac{TP}{TP+FP} \text{ and } Recal = \frac{TP}{TP+FN} \tag{3}$$

Precision and recall fail to offer significant information individually; and hence they should be utilised together. Essentially, the precision-recall breakeven point (BEP) is an assimilated efficacy indicator which is utilised coupled with the F1 measure. The below-mentioned expression for F1 is the harmonic mean of precision (P) and recall (R):

$$F_1 = 2PR/(P + R) \quad (4)$$

Precision-recall BEP is that point where recall and precision are in equation. It is typically calculated by contrasting precision, and the arithmetic mean of recall. It is essential that the BEP performance indicator be computed separately for each class. The mutual efficacy of an approach, as gauged by taking into account all classes, could be ascertained by utilising the micro-average or macro-average BEP. In case of macro average, all groups are assigned equal weight. Conversely, the micro-average method renders similar weight to each document.

#### 4. Experiment results and assessment

In order to assess the efficacy of our proposed procedure, we employed our methodology in a python, with a platform comprising Intel® Core™ i7 CPU @ 7GHz and 8.00 GB RAM along with Windows 10. For measuring the suggested approach's performance, several benchmarks from the UCI machine learning repository were chosen, and the stated three sets of experiments were carried out. It is to be noted that we have divided the data as 70% for training and 30% for testing. This has been deployed for all original datasets for assessing the precision of every class, by utilising methods like precision, F-measure, overall accuracy, and recall for these original data sets. Table 3 shows the precision of the original data set for each class. Through imbalanced data sets, the evaluation metrics of the classifier's performance are separately obtained with regards to the data distribution, ensuring that the apt metric is chosen for this type of issues. Table 3 depicts the average values for precision, wherein SVM is utilised for every original dataset for carrying out the average test of precision so as to warrant a pertinent statistical behaviour. This table shows the experiment outcomes for the original SVM minus execution of the recommended approach, wherein underperformance is seen in the outcomes for few of the datasets which pertain to the regular imbalanced assessment metrics.

Table 3. Recall, precision, F-score and accuracy rate for original dataset

Dataset	Precision	Recall	F-score	Accuracy
Australian	0.85	0.83	0.83	83.09
Heart-statlog	0.77	0.77	0.76	76.54
Heart C	0.87	0.86	0.86	86.23
Liver	0.54	0.58	0.52	57.69
Ionosphere	0.79	0.79	0.79	79.24
Hepatitis	0.84	0.81	0.82	81.39
Pima	0.48	0.69	0.57	69.26
Breast cancer	0.68	0.71	0.68	70.93
kidney	0.97	0.97	0.97	96.66
German credit	0.70	0.73	0.70	73.0

The experimental outcomes presented above indicate that learning without the suggested technique leads to lower performance specific to the assessment indicators for regular imbalanced data. Also, the SVM technique achieved poor recall and precision for all datasets tested during the study.

To assess the efficacy of the suggested technique, the selected imbalanced dataset was experimented upon. Imbalanced data classification accuracy required computing the optimal value of the regulating parameter. The present study uses SA for parameter optimisation. Table 4 shows the particulars of the SA parameter.

Table 4. The best parameter (C)

Dataset	T	V	iterations	parameter C
Australian	3	0.9	300	0.5
Heart-statlog	3	0.8	100	1
Heart C	4	0.7	120	1

Liver	2.5	0.9	10	0.013
Ionosphere	2	0.8	5	0.01
Hepatitis	3.5	0.8	300	0.1
Pima	3	0.7	300	3
Breast cancer	4	0.7	250	0.5
kidney	1.5	0.8	20	0.1564994
German credit	2.5	0.9	5	1

Table 5. Classification accuracies for our approach

Dataset	Precision	Recall	F-score	Accuracy
Australian	0.86	0.86	0.86	86.091
Heart-statlog	0.80	0.80	0.81	81.220
Heart C	0.87	0.88	0.88	87.231
Liver	0.65	0.61	0.65	65.194
Ionosphere	0.80	0.80	0.80	80.00
Hepatitis	0.77	0.90	0.88	85.744
Pima	0.71	0.70	0.71	71.80
Breast cancer	0.80	0.81	0.81	80.56
kidney	0.99	0.98	0.97	97.52
German credit	0.77	0.76	0.77	75.43

In terms of imbalanced datasets, and to warrant the legitimacy of our recommended model, we have used it to determine the precision of every class, where a noteworthy enhancement with regards to estimation for every class and high precision was seen for every class as against when SVM was applied on the same data set. SVM exhibits a variable average precision for every class on the majority of the data sets. We observed an enhancement in classification precision, thus making the classifier competent enough to classify better compared to SVM. as can be seen in Table 5. The parameters cast a pronounced effect on efficacy and competence of SVM. Parameter adjustment not just pertains to the under fitting and over fitting of the training data but also has an impact on the outcomes of the validation sets. The best parameter pair (C) is then utilised for generating the model for training. Perform the prediction test on every test set; subsequently, the corresponding training set is expected to produce another optimal parameter value.

## 5. Conclusion

Viewed as amid the most prevalent machine learning algorithms, support vector machines (SVM) are extensively deployed for object recognition and dataset classification. Nevertheless, this technique has a problem-specific parameter and ascertaining its optimal value is a difficult task; this impacts the precision of the algorithm. In the present study, we suggested using an SA-augmented SVM algorithm. Hybridisation facilitates better results on SVM parameter optimisation. Experimental outcomes on UCI datasets with diverse sizes have shown that the algorithm has superior precisions as against the SVM algorithm whereas the time consumed for computation is judicious as well.

## References

- [1] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [2] W. Aoudi and A. M. Barbar, "Support vector machines: A distance-based approach to multi-class classification," in *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, 2016, pp. 75–80.
- [3] C.-F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Appl. Soft Comput.*, vol. 12, no. 2, pp. 807–818, 2012.
- [4] F. Abayaje, S. A. Hashem, H. S. Obaid, Y. S. Mezaal, S. K. Khaleel, "A miniaturization of the UWB monopole antenna for wireless baseband transmission," *Periodicals of Engineering Natural Science*, vol.

- 8, no. 1, pp. 256-262, 2020.
- [5] Y. Ren and G. Bai, "Determination of optimal SVM parameters by using GA/PSO.," *JCP*, vol. 5, no. 8, pp. 1160–1168, 2010.
- [6] C.-L. Huang, "ACO-based hybrid classification system with feature subset selection and model parameters optimization," *Neurocomputing*, vol. 73, no. 1–3, pp. 438–448, 2009.
- [7] S.-W. Lin, Z.-J. Lee, S.-C. Chen, and T.-Y. Tseng, "Parameter determination of support vector machine and feature selection using simulated annealing approach," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1505–1512, 2008.
- [8] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *J. Stat. Phys.*, vol. 34, no. 5–6, pp. 975–986, 1984.
- [9] A. S. Hussein, R. S. Khairy, S. M. M. Najeeb, and H. ALRikabi, "Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 5, 2021.
- [10] K.-P. Wu and S.-D. Wang, "Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space," *Pattern Recognit.*, vol. 42, no. 5, pp. 710–717, 2009.
- [11] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM parameter optimization using grid search and genetic algorithm to improve classification performance," *Telkomnika*, vol. 14, no. 4, p. 1502, 2016.
- [12] C. Yao, F. Wu, H. Chen, X. Hao, and Y. Shen, "Traffic sign recognition using HOG-SVM and grid search," in *2014 12th International Conference on Signal Processing (ICSP)*, 2014, pp. 962–965.
- [13] R.-J. Kuo, C. M. Chen, T. W. Liao, and F. C. Tien, "Hybrid of artificial immune system and particle swarm optimization-based support vector machine for Radio Frequency Identification-based positioning system," *Comput. Ind. Eng.*, vol. 64, no. 1, pp. 333–341, 2013.
- [14] X. Zhang and Y. Guo, "Optimization of SVM parameters based on PSO algorithm," in *2009 Fifth International Conference on Natural Computation*, 2009, vol. 1, pp. 536–539.
- [15] Y. Zhang, "An improved QSPR method based on support vector machine applying rational sample data selection and genetic algorithm-controlled training parameters optimization," *Chemom. Intell. Lab. Syst.*, vol. 134, pp. 34–46, 2014.
- [16] B. K. Mohammed, M. B. Mortatha, A. S. Abdalrada, H. T. ALRikabi, and N. Sciences, "A comprehensive system for detection of flammable and toxic gases using IoT," *Periodicals of Engineering Natural Sciences*, vol. 9, no. 2, pp. 702-711, 2021.
- [17] M. R. G. Raman, N. Somu, K. Kirthivasan, R. Liscano, and V. S. S. Sriram, "An efficient intrusion detection system based on hypergraph-Genetic algorithm for parameter optimization and feature selection in support vector machine," *Knowledge-Based Syst.*, vol. 134, pp. 1–12, 2017.
- [18] I. Wegener, "Simulated annealing beats metropolis in combinatorial optimization," in *International Colloquium on Automata, Languages, and Programming*, 2005, pp. 589–601.
- [19] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.