Traffic Congestion Detection using Whale Optimization Algorithm and Multi-Support Vector Machine

Bodanapu Sony, Ankit Chakravarti, M. Madhusudhan Reddy

Abstract: ***Currently, urban traffic is an evolving issue that is closely related to economic factors and population growth. Several developing and developed countries are seeing increases in vehicle growth per each passing year. So, traffic flow congestion forecasting is one of the important aspects in intelligent transport system that aims to improve mobility, influence travel behavior, and save energy. In this research paper, an effective system; Whale Optimization Algorithm (WOA) and Multi-Support Vector Machine (MSVM) classifier was proposed in order to improve the performance of road traffic congestion detection. Here, the proposed system was experimented by using a simulated dataset, which was collected for beibei tunnel (4.2 km long road in yuwu free-way). The resulting section confirms that the proposed system enhanced the classification accuracy in road traffic congestion detection up-to 9.3% related to other existing systems.***

Index Terms: ***Data normalization; Multi support vector machine; Traffic congestion detection; Whale optimization algorithm.***

I. INTRODUCTION

In recent decades, road traffic congestion is one of the emerging traffic problems that negatively affects other road user's, which results in high pollution, unnecessary delay, stress, high fuel consumption, etc. As vehicle population endures to increase, the traffic management associated to congestion is an unavoidable consequence [1-3]. In order to enhance the human life quality, it is important that the problem of traffic congestion is needed to be addressed [4]. In intelligent transport system, an effective system to reduce road traffic congestion should comprise of electronic information system to manage, control, and improve the variable speed limits measured by computer systems, real-time transport information, dynamic signal timings and transportation through real time route prediction systems [5-6]. By applying intelligent transport system to road traffic congestion creates a new traffic management opportunity that makes the best use of existing transport networks [7]. Generally, the intelligent transport systems operate in both wired and wireless networks. Though, the modern applications show preference to wireless communication, due to its versatility and ease of implementation [8].

In the last a few decades, several techniques are applied for regulating the traffic flow on road networks like deep neural networks [9], improved particle swarm optimization [10], etc. These conventional approaches are not effective in managing the large datasets, so a new system is developed in this research study to manage and control the road traffic congestion for minimizing the intensity and duration of exposure to vehicular pollution, suffered by the commuters at traffic signals. Here, the input data was collected for beibei tunnel (4.2 km long road in yuwu free-way). The collected data was pre-processed by applying data normalization for scaling the heterogeneous set of numerical data. After data pre-processing, WOA was used to select the optimal attributes. The output of WOA specifies, which attributes were vital in classifying the road traffic congestion stages. These optimized attributes were given as the input for MSVM classifier for classifying the road traffic congestion stages like smooth traffic, re-current congestion, non-recurrent congestion of downstream, and non-recurrent congestion of upstream.

This research paper is organized as follows. Section II survives several recent papers on traffic congestion forecasting. In section III, detailed explanation about the proposed system is presented. In section IV, execution of the proposed system is assessed by using (MATLAB 2018b) simulation. The conclusion is done in section V.

II. Literature review

Several new systems are developed by the researchers in traffic congestion forecasting. In this section, a brief evaluation of a few essential contributions to the existing literature papers are presented.

M. Ma, and S. Liang, [11] developed a new co-ordinated optimization algorithm for forecasting the traffic congestion in a free-way network. Here, bi-level programming approach was utilized to address the issues between the traffic control path and network path. In the path level, an optimal control method was developed for relieving congestion in each and every path of the test network. Further, in the network level, the optimal control methodology concentrates on cost optimization and traffic equilibrium. The effectiveness and efficiency of the developed system was evaluated by testing on a traffic network calibrate. The experimental outcome showed that the developed system effectively balances the traffic load and also highly enhances the free-way networks service level. However, the developed system lags with the concern towards handling the missing attributes in the collected dataset.

Revised Manuscript Received on April 15, 2019.

Bodanapu Sony, Civil engineering department, KL University, Guntur district-522502, A.P, India. And assistant prof. in GNIT, IBP, RR Dist.,
Dr. Ankit Chakravarti, Civil engineering department, KL University, Guntur district-522502, A.P, India.
M. Madhusudhan Reddy, Research scholar, Civil engineering department, KL University, Guntur district-522502, A.P, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

X. Zhang, *et al*, [12] presented a robust and accurate traffic congestion detection system (Hierarchical Fuzzy Rule-Based System (HFRBS)-Genetic Algorithms (GA)) for an enormous number of input data. The developed system effectively reduces the input data size without the loss of information. To perform this operation, a hierarchical structure of FRBS was optimized by applying GA, which combines the membership functions of lateral tuning, variable ranking and selection, and rule base optimization. In the experimental outcome, the developed system was evaluated by using a benchmark dataset in order to analysis the efficiency of the developed system. The major problem with this method was the creation of complex operations that highly increases the system complexity.

E. Walraven, *et al*, [13] developed a new methodology for optimizing the traffic flow on the basis of reinforcement learning. In this research paper, Q-learning and Markov decision process were utilized for finding the policies to assign speed limits. In order to enhance the scalability and performance of the developed algorithm, artificial neural network was used for efficiently learning the policies. The simulation results show that the developed algorithm performs well in small road networks, which was very essential in the reinforcement of larger road networks and real traffic engineering applications. The developed methodology was only suitable for the datasets with limited attributes and features, or else, it leads to the inconsistency with realistic applications.

X. Yu, *et al*, [14] developed a new descriptor in order to calculate the campus traffic congestion level. In this literature paper, back propagation neural network and Markov model were additionally used for detecting the campus traffic congestion. In the experimental phase, the developed system achieved better classification accuracy related to other state of the art systems. The major drawback in the developed system was insufficient improvement in future extraction and parameter optimization, which were very crucial in the case of traffic congestion forecasting.

To address the above mentioned concerns and also for improving the traffic congestion forecasting, a new supervised system (WOA-MSVM) was developed in this research paper.

III. PROPOSED SYSTEM

For traffic congestion forecasting, most of the existing systems requires lengthy time data, prior knowledge, and need to recognize the presence of congestion, which is very difficult in most of the circumstances because the high-ways extend across thousands of kilo-metres. To over-come these concerns, a new system is developed in this research paper. In this scenario, the proposed system consists of four steps; data collection, pre-processing, data optimization and classification. The Fig. 1 represents the work flow of proposed system and the detailed description about the proposed system is described below.

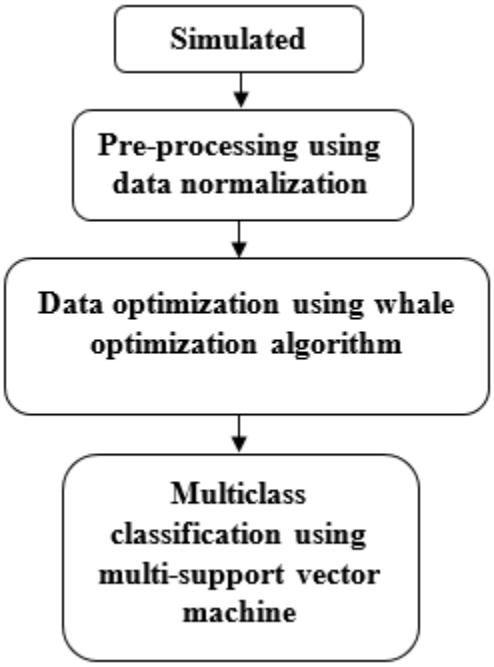

Fig. 1. Work Flow of Proposed System

A. Data collection

In this research study, the beibei tunnel (4.2km long road in yuwu free-way) is under-taken as a test road. There is a detector at the up-stream of beibei tunnel and the detector collects information about the vehicle speed, occupancy, and flow rate, averagely for five minutes. The collected dataset is an unstructured or imbalanced data that comprises of 68 pieces of recurrent congestion, 53 pieces of non-recurrent congestion, and 793 pieces of smooth traffic flow data.

B. Pre-processing of collected data

After collecting the numerical data, pre-processing is carried-out to enhance the quality of input data or to convert the unstructured data into structured data. Generally, the collected data are high dimensional in nature, so it is essential to convert high dimensional data into low dimensional data that helps to reduce the "curse of dimensionality" issue. In recent scenario, there are numerous approaches available for data pre-processing. Among available approaches, a suitable approach is undertaken in this research study for pre-processing the collected data. Hence, the pre-processed numerical data are utilized for further operations like optimization and classification. Here, a pre-processing approach (data normalization) is used for converting the unstructured data into structured data. The data normalization is utilized for scaling the heterogeneous set of numerical data, which is accomplished to set the maximum value of one and the baseline to zero.

3. Optimizing the pre-processed data using WOA

After pre-processing the data, optimal features or attributes are selected by using WOA. It is a new meta-heuristic approach that mimics the Hump-back whales. In this algorithm, the optimization starts by creating a random

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

population of whales. Initially, the created whales search for the optimum prey's location and then attach by any one of these approaches: bubble-net or encircling [15]. In the encircling approach, the hump-back whales enhance their location on the basis of best location, which is determined by the equations (1) and (2).

$$D = |B \odot P^*(t) - P(t)| \quad (1)$$

$$P(t+1) = |P^*(t) - A \odot D| \quad (2)$$

Where, A and B are represented as coefficient values, D is denoted as the distance between the position vector of a whale $P(t)$ and a prey $P^*(t)$, and t is stated as the iteration number. The coefficient values are evaluated by using the equations (3) and (4).

$$A = 2l \odot r - l \quad (3)$$

$$B = 2r \quad (4)$$

Where, l is represented as the linearity value that ranges from 0 to 2 and r is denoted as the random vector $\in [0,1]$.

In addition, the bubble net approach is performed on the basis of two methods. The first method is named as shrinking encircling that helps to reduce the values l and A and the second method is the spiral updating position. It is used to mimic the helix shaped movement of hum-back whales around prey, which is determined by the equation (5).

$$P\ (t+1) = \acute{D} \odot e^{ba} \odot \cos(2\pi a) + P^*(t) \quad (5)$$

Where, $\acute{D} = |P^*(t) - P(t)|$ is represented as the distance between the prey and whale, a is denoted as the random value [-1, 1], b is indicated as the constant, which is used to find the logarithmic spiral shape and $\odot$ is specified as the element-by-element multiplication.

The whale swims around the victim using a shrinking circle and spiral shaped path con-currently, which is mathematically denoted in the equation (6).

$$P(t+1) = \begin{cases} P^*(t) - A \odot D & if\ p \geq 0.5 \\ \acute{D} \odot e^{ba} \odot \cos(2\pi a) + P^*(t) & if\ p < 0.5 \end{cases} \quad (6)$$

Where, $p \in [0,1]$ is denoted as the random value that determines the probability of selecting either the spiral approach or shrinking encircling approach for adjusting the whales position.

In exploration section, the hump-back whales search for a prey. The whale position is updated by calculating a random search agent rather than the best search agent that is described in the equations (7) and (8).

$$D = |B \odot P_{rand} - P(t)| \quad (7)$$

$$P(t+1) = |P_{rand} - A \odot D| \quad (8)$$

Where, P_{rand} is stated as the random position that is calculated from the current population. Fig. 2 illustrates the flow chart of WOA.

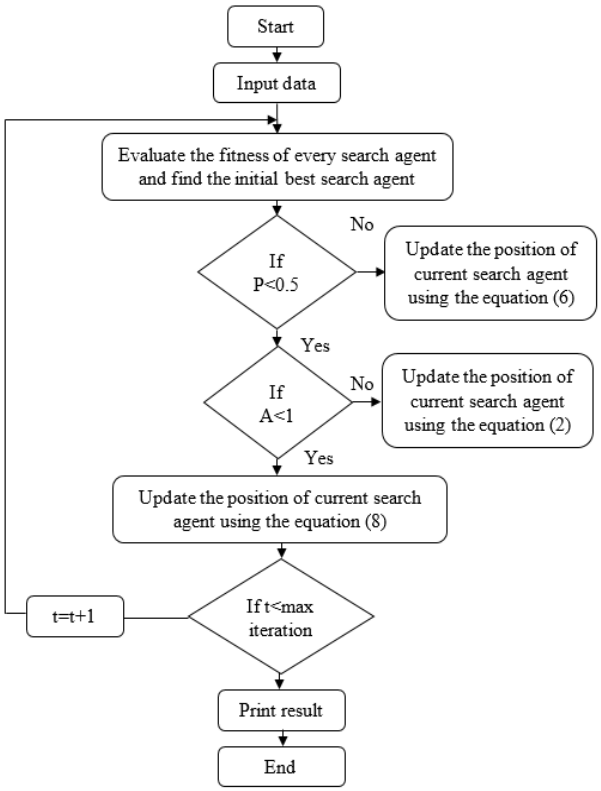

Fig. 2. Flow chart of WOA

D. Multi-class classification using MSVM

The optimized feature values are given as the input for multi-class classifier: MSVM to classify the different traffic congestion stages. The stages 0, 1, 2, and 3 denotes smooth traffic, re-current congestion, non-recurrent congestion of downstream, and non-recurrent congestion of upstream.

Generally, the conventional-SVM is a two class classification technique. Though, it is essential to modify the multi-binary classification problems in order to extend SVM to MSVM. In SVM, the multi-class classification is rehabilitated into n^{th}two class and i^{th} two class problems, class i is separate from the residual classes. In SVM, two most prominent methodologies are One-Against-All (1-a-a) and One-Against-One (1-a-1) [16]. In this scenario, the 1-a-a solution creates a binary classifier for every class that distinct the objects belong to the same class. In n^{th}class, the 1-a-a approach generates n^{th}binary classifiers, and the ith classifier is trained with the samples ini^{th} class with positive labels and the remaining samples are trained with negative labels. The outcome of n^{th}class in 1-a-a approach relates with the 1-a-1 approach for obtaining the highest output value. In addition, the 1-a-1 approach is the resultant of previous researches on two class classifier.

The idea behind MSVM is to generate all possible two class classifiers from a training set of n^{th}classes, each and every classifier trained only two out of nth classes, and there would be $n \times (n-1)/2$ classifiers. In MSVM, decision function is an effective way to moderate the multi-class problems, which is constructed by assuming all then^{th}

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

classes. The M-SVM classification technique is an extension of SVM, which is mathematically represented in the equations (9), (10), and (11).

$$min\Phi(w,\xi) = 1/2\sum_{m=1}^{k}(w_m.w_m) + c\sum_{i=1}^{l}\sum_{m\neq yi}\xi_i^m \quad (9)$$

Subjected to,

$$(w_{yi}.x_i) + b_{yi} \geq (w_{yi}.x_i) + b_m + 2 - \xi_i^m, \quad (10)$$

$$\xi_i^m \geq 0, i = 1,2,3 \ldots l, m, yi \in \{1,2,3 \ldots k\}, m \neq yi \quad (11)$$

Where, ξ_i^m is stated as slack variables, l is considered as the training data point, c is represented as the user's positive constant, yi is denoted as the class of training data vectors x_i, and k is stated as the number of classes. At last, the decision function is represented in the equation (12).

$$f(x) = arg\,max[(w_i.x) + b_i], i = 1,2,3,..k \quad (12)$$

IV. EXPERIMENTAL RESULT AND ANALYSIS

The proposed system was simulated by using MATLAB (version 2018b) with 3.0 GHz, Intel i5 processor and 1 TB hard disc. Here, the performance of proposed system was compared with an existing system (Gradient descent and particle swarm optimization algorithm (GPSO)-McMaster algorithm [17])) in order to evaluate the efficiency and effectiveness of the proposed system. In this experimental research, the proposed system performance was evaluated in light of classification accuracy, recall, and precision.

A. Performance measure

Generally, performance measure is defined as the procedure of collecting, investigating and reporting information on the basis of system performance. The performance measure also involves in studying the approaches with-in organizations or studying engineering procedures/ phenomena/parameters to see whether the output is in line or not. The mathematical formula of precision, accuracy, and recall are denoted in the equations (13), (14), and (15).

$$Precision = \frac{TP}{FP+TP} \times 100 \quad (13)$$

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN} \times 100 \quad (14)$$

$$Recall = \frac{TP}{FN+TP} \times 100 (15)$$

Where, FP is stated as false positive, TP is represented as true positive, FN is indicated as false negative and TN is specified as true negative.

B. Quantitative analysis using simulated dataset

In this sub-section, simulated dataset (collected from beibei tunnel (4.2km long road in yuwu free-way) is utilized for evaluating the performance of proposed system for classifying the traffic congestion stages; smooth traffic, re-current congestion, non-recurrent congestion of downstream, and non-recurrent congestion of upstream. Here, the performance evaluation is validated for optimized attributes, 52 pieces of recurrent congestion, 46 pieces of non-recurrent congestion, and 612 pieces of smooth traffic flow data with 60 % training and 40% testing of data.

Table 1 denotes the performance evaluation of the proposed system for classifying the traffic congestion stages by means of recall, classification accuracy and precision. In addition, table 1 evaluates the proposed system performance with optimization and without optimization algorithm. The recall, classification accuracy and precision of the proposed system (without WOA) is 89.77%, 93.02%, and 90.59%. Similarly, recall, classification accuracy and precision of the proposed system (with WOA) is 91.334%, 97.68%, and 92.34%. Table 1 clearly shows that the M-SVM classifier improves the classification accuracy in traffic congestion detection up to 4.5%, while using WOA. The graphical representation of recall, classification accuracy and precision of the proposed system is denoted in the Fig. 3.

Table 1. Performance evaluation of proposed system

Simulated dataset			
Optimization	**Precision (%)**	**Recall (%)**	**Accuracy (%)**
Without-WOA	90.59	89.77	93.02
With-WOA	92.34	91.334	97.68

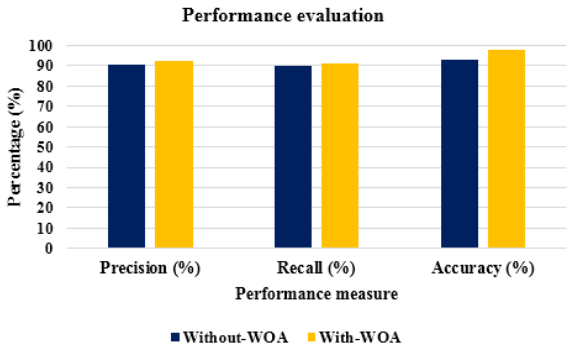

Fig. 3. Graphical evaluation of proposed system

C. Comparative analysis

The comparative analysis of proposed and existing system is detailed in the table 2.D. Sun, *et al*, [17] developed a new system (GPSO and McMaster algorithm) for forecasting the traffic congestion. In this comparative paper, McMaster algorithm parameters were validated with a real time data from Yuwu expressway in Chongqing. The experimental section shows that the GPSO helps the operators to identify the best parameters more effectively and makes the McMaster algorithm to yield a higher recall and classification accuracy rate. The developed system almost achieved 96% of classification accuracy and 82% of recall rate. However, the proposed system achieved 97.68% of classification accuracy and 91.334% of recall rate, which was higher compared to the existing paper.

Tables 1 and 2 clearly shows that the proposed system has achieved higher accuracy compared to the existing system. The simulated dataset has multiple attributes and the missing of attributes resulted in less accuracy and recall. In the case of WOA-MSVM, it is very much optimized for specific conditioning of distances regarding their placement of data samples. The graphical comparison of comparative analysis is denoted in the Fig. 4.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 2. Comparative analysis of proposed and existing system

Methodologies	Classification accuracy (%)	Recall (%)
GPSO-McMaster algorithm [17]	96	82
WOA-MSVM	97.68	91.334

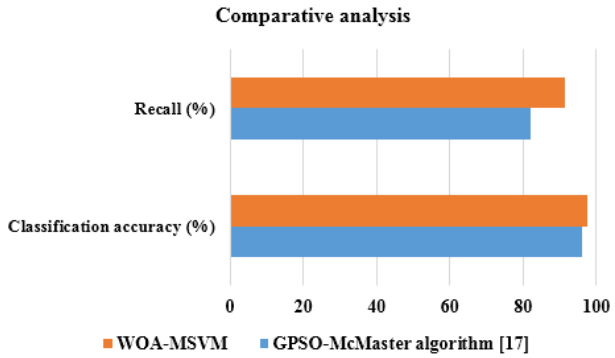

Fig. 4. Graphical comparison of comparative analysis

V. CONCLUSION

In this research study, an effective supervised system is developed for classifying the stages of traffic congestion. Here, the input data is collected for beibei tunnel (4.2km long road in yuwu free-way). The collected data is pre-processed using data normalization in order to reduce the complexity of the proposed system. Then, the optimal data attributes are selected by using WOA that effectively finds the attributes weights by resolving the convex optimization issues. These optimal attributes are utilized for classifying the stages of traffic congestion by applying MSVM classifier. Compared to other existing systems in traffic congestion detection, the proposed system accomplished an effective performance in light of classification accuracy, which showed 9.3% of enhancement in traffic congestion stage classification. In future work, a new optimization algorithm will be designed for analysing the sub-stages of traffic congestion detection.

VI. ACKNOWLEDGMENT

It is optional. The preferred spelling of the word “acknowledgment” in American English is without an “e” after the “g.” Use the singular heading even if you have many acknowledgments. Avoid expressions such as “One of us (S.B.A.) would like to thank ....” Instead, write “F. A. Author thanks “Sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page.

REFERENCES

1. S.B. Li, G.M. Wang, T. Wang, and H.L. Ren. (2017). Research on the Method of Traffic Organization and Optimization Based on Dynamic Traffic Flow Model.Discrete Dynamics in Nature and Society, pp. 1-9.
2. V. Cahill. (2010). Soilse: A decentralized approach to optimization of fluctuating urban traffic using reinforcement learning. In proceedings of IEEE 13th International Conference on Intelligent Transportation Systems (ITSC), pp.531-538.
3. F. Agyapong, and T.K. Ojo. (2018). Managing traffic congestion in the Accra Central Market, Ghana. Journal of Urban Management, 7(2). Pp.85-96. 2018.
4. F. Ahmad, S.A. Mahmud, and F.Z. Yousaf.(2017) Shortest processing
5. Time scheduling to reduce traffic congestion in dense urban areas.IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47(5). pp.838-855.
6. S.H. Melouk, B.B. Keskin, C. Armbrester and M. Anderson. (2011). A simulation optimization-based decision support tool for mitigating traffic congestion. Journal of the Operational Research Society, 62. Pp.1971-198.
7. S. Surya, and N. Rakesh. (2016). Flow based traffic congestion prediction and intelligent signalling using Markov decision process. In Proceedings of IEEE International Conference on Inventive Computation Technologies (ICICT), 3. Pp.1-6, 2016.
8. P. Lopez-Garcia, E. Onieva, E. Osaba, A. D. Masegosa, and A. Perallos. (2016). A Hybrid Method for Short-Term Traffic Congestion Forecasting Using Genetic Algorithms and Cross Entropy.IEEE Trans. Intelligent Transportation Systems, 17(2). Pp.557-569.
9. H. Zhang, X. Wang, J. Cao, M. Tang, and Y. Guo. (2018). A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics. Applied Intelligence, 48(8). Pp.2429-2440, 2018.
10. Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang. (2018).A hybrid deep learning based traffic flow prediction method and its understanding. Transportation Research Part C: Emerging Technologies, 90, pp.166-180.
11. H.Y. Liu. (2013). Utilize improved particle swarm to predict traffic flow. In Advanced Materials Research, Trans Tech Publications, 756. Pp.3744-3748.
12. M. Ma, and S. Liang. (2018). an optimization approach for freeway network coordinated traffic control and route guidance.PloS one, 13(9), pp.e0204255.
13. X. Zhang, E. Onieva, A. Perallos, E. Osaba, and V. C. Lee. (2014). Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. Transportation Research Part C: Emerging Technologies, 43. Pp.127-142.
14. E. Walraven, M.T. Spaan, and B. Bakker. (2016). Traffic flow optimization: A reinforcement learning approach. Engineering Applications of Artificial Intelligence, 52. Pp.203-212.
15. X. Yu, S. Xiong, Y. He, W. E. Wong, and Y. Zhao. (2016). Research on campus traffic congestion detection using BP neural network and Markov model. Journal of Information Security and Applications, 31. Pp.54-60.
16. D.B. Prakash, and C. Lakshminarayana. (2017). Optimal siting of capacitors in radial distribution network using whale optimization algorithm. Alexandria Engineering Journal, 56(4). Pp.499-509.
17. T.H. Le, H.S. Tran, and T.T. Nguyen. (2012). Applying Multi Support Vector Machine for Flower Image Classification. In proceedings of International Conference on Context-Aware Systems and Applications, Springer, Berlin, Heidelberg, pp.268-281.
18. D. Sun, C. Zhang, M. Zhao, L. Zheng, and W. Liu. (2017). Traffic congestion pattern detection using an improved McMaster algorithm. In Proceedings IEEE29th Chinese of Control and Decision Conference (CCDC), pp.2814-2819.