Received March 17, 2018, accepted April 19, 2018, date of publication April 26, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2830651

Effective Features to Classify Big Data Using Social Internet of Things

LAKSHMANAPRABU S. K.[1], SHANKAR K.[2], ASHISH KHANNA[3], DEEPAK GUPTA[3], JOEL J. P. C. RODRIGUES[4,5,6], (Senior Member, IEEE), PLÁCIDO R. PINHEIRO[7], (Member, IEEE), AND VICTOR HUGO C. DE ALBUQUERQUE[7], (Member, IEEE)

[1]Department of Electronics and Instrumentation Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai 600048, India
[2]School of Computing, Kalasalingam Academy of Research and Education, KrishnanKoil 626126, India
[3]Assistant Maharaja Agrasen Institute of Technology, GGSIP University, Delhi 110078, India
[4]National Institute of Telecommunications, Santa Rita do Sapucaí/MG 37540-000, Brazil
[5]Instituto de Telecomunicações, 1049-001 Lisbon, Portugal
[6]ITMO University, 197101 St. Petersburg, Russia
[7]Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza/CE 60811-905, Brazil

Corresponding author: Joel J. P. C. Rodrigues (joeljr@ieee.org)

This work was supported in part by the National Funding from the Fundação para a Ciência e a Tecnologia Project under Grant UID/EEA/50008/2013, in part by the Government of Russian Federation, under Grant 074-U01, in part by the Brazilian National Council for Research and Development under Grant309335/2017-5, Grant 301928/2014-2, and Grant 304315/2017-6, and in part by Finep, with resources from Funttel, through the Centro de Referência em adiocomunicações-CRR Project of the Instituto Nacional de Telecomunicações, Inatel, Brazil, under Grant 01.14.0231.00.

ABSTRACT Social Internet of Things (SIoT) supports many novel applications and networking services for the IoT in a more powerful and productive way. In this paper, we have introduced a hierarchical framework for feature extraction in SIoT big data using map-reduced framework along with a supervised classifier model. Moreover, a Gabor filter is used to reduce noise and unwanted data from the database, and Hadoop Map Reduce has been used for mapping and reducing big databases, to improve the efficiency of the proposed work. Furthermore, the feature selection has been performed on a filtered data set by using Elephant Herd Optimization. The proposed system architecture has been implemented using Linear Kernel Support Vector Machine-based classifier to classify the data and for predicting the efficiency of the proposed work. From the results, the maximum accuracy, specificity, and sensitivity of our work is 98.2%, 85.88%, and 80%, moreover analyzed time and memory, and these results have been compared with the existing literature.

INDEX TERMS Internet of Things, social Internet of Things, machine Learning, big data, feature selection.

I. INTRODUCTION

Developing IoT enabled technologies and their solutions are a major challenge. However, IoT is about the pervasive collection and sharing of data towards a common goal [1]. In IoT, the data refers to attribute values such as variables or integer values; and event refers to when certain conditions are met or when certain states are reached [2], [4]. IoT service allows certain functions to be carried out through a predefined interface [5], [6]. Some researchers are particularly interested in identifying risk issues arising during discovering and integrating data within IoT [7]. Recently, the SIoT have been developed; another utilization of Internet of Things (IoT). The SIoT is a larger social network, connecting people and people, people and objects, and objects and objects [8]. Thus creating the opportunities which bring several challenges to data processing systems for improving data collection, cleaning and storage, and performing real-time analytics [9]. Moreover, in the current scenario of Big Data, various standards and platforms have been introduced by relational database vendors and can be used for data aggregation as well as data analysis [10], [11]. The ‘‘big data and SIoT is the perfect representation of social systems and the IoT to characterize human progression [12].Various feature selection algorithms have been proposed; which can be classified into two broad categories, i.e., filter approaches, and wrapper approaches. In filter-based approach, the filtration process is performed before classification process because of the independent usage nature of classification algorithms [13].

Linear Kernel Support Vector Machine-based classifier is used to classify the data. LK-SVM has numerous benefits like resolving little specimen; nonlinear and high dimensional pattern recognition [14]. Subsequently, the advanced things can't specifically apply to SIoT. In addition, heterogeneous nature of SIoT, different things communicating with each other under its umbrella devours a high volume of memory, power processing, along with high bandwidth [15].

- The main contribution of our manuscript as SIoT structure to ensure system with big data based on data classification model help some features.
- For this analysis consider some social databases and finally performed classification model aid of LK-SVM classifier, moreover improve the performance of the system; optimized features are used and finally this approach results also.
- This work will be useful for the SIoT systems and will help to provide the guidance to the researchers of SIoT and big data.

Hussein *et al.* [17] have examined a novel administration structure in view of a reasoning approach for dynamic SIoT administrations revelation in smart spaces. Thus, reasoning about clients' situational needs, inclinations, and other social angles along with clients' encompassing condition were proposed for creating a file of situation-aware services which coordinates clients' needs. This reasoning approach is then actualized as a proof-of-idea model, specifically Airport Dynamic Social, inside a brilliant airport.

Nguyen *et al.* [18] investigated the issue in detail and propose a model-driven way to optimize an IoT application in regards to its non-practical prerequisites. A source code transformation, that updates the source code with the produced movable parameter values and then executes the compiler to make another binary image of the application. The investigational results shows that non-useful prerequisites, for example, power consumption and reliability can be enhanced generously at the time of optimization.

Multiple birth support vector machine is a novel machine learning calculation for multi-class characterization, which is considered as an augmentation of twin support vector machine by Zhang *et al.* [19]. To foresee another specimen initially decides an interim for each class depending on the separations between training tests and their hyper planes and after that groups the new example relying upon the separations amongst hyper planes and the new example which are in the relating interims. What's more, smoothing procedure is connected to the model, the first occasion when it was utilized as a part of multi-class twin support vector machine. The results on false datasets and UCI datasets demonstrated that the proposed algorithm was effective and has great order execution.

The rest of the paper described as follows, section 2 performs state of the art in SIoT, classification and big data models and section 3 presents the detailed proposed work with different layers function includes block diagram. Section 4 comprises of simulation results of proposed work and performs comparison with the existing works and finally concludes our research work with future enhancement.

II. BRIEF LITERATURE REVIEW

SIoTs alludes to the quickly developing system of associated protests and individuals that can gather and exchange information using embedded sensors proposed by Hasan and Al-Turjman [16]. The researchers proposed a bio-inspired particle multi-swarm enhancement (PMSO) directing calculation to build, recoup and select disjoint ways that endure the disappointment while fulfilling Quality of Service (QoS) parameters. Multi-swarm technique empowers deciding the optimal directions in choosing the multipath directing while at the same time trading messages from all positions in the system. Results demonstrated that the procedure utilizing the qualities of all individual best data is a substantial technique for the motivations behind enhancing the PMSO execution.

High-Performance Computing (HPC) solution has turned into a key issue and has pulled into consideration as of late been proposed by Ahmad *et al.* [20]. It presented a framework engineering that chooses Artificial Bee Colony (ABC). In addition, a Kalman filter was used as a part of Hadoop biological system that is utilized for evacuation of noise. A total four-level engineering was recommended that productively totals the information by removing superfluous information, and investigate the information using the proposed Hadoop-based ABC algorithm. To check the effectiveness of the proposed calculations in the proposed framework engineering, we have executed the proposed framework utilizing Hadoop and Map Reduce with the ABC calculation. ABC calculation was used to choose highlights, though; Map Reduce was supported by a parallel algorithm that proficiently forms an enormous volume of data.

The Internet of Things (IoT) is overpopulated by a huge number of objects along with millions of interactions and services [21]. Mardini *et al.* [21] have recommended the SIoT; where each question in the IoT can utilize its companions' or friends-of-friends' connections to look for a particular administration. This is generally a moderate procedure on the grounds that every hub (object) is required to deal with a substantial number of companions. This work tends to the issue of link determination of companions and scrutinizes five techniques in the literature. It has been proposed a link determination procedure utilizing the Genetic Algorithm (GA) to locate the close optimal solution. The outcomes demonstrated a change in the inspected methodologies regarding a few parameters.

Thus, the main challenges are:

- Internet of Things gadgets, to interface with each other in a proficient way, consequently creating Big Data not deliver better reduction.
- The SIoT Big Data, to decrease the vitality utilization engaged with the SIoT at the time of communicating Data over the Internet.

- Advanced data analysis and intelligent approaches will provide utility to the SIoT with optimization and classifier model [20].
- Develop a bio-inspired PSO routing algorithm achieve fast recovery from path failure by attempting to extend an existing approach. Optimization is required, in order to meet non-functional requirements in the application design phase and the problem of IoT is overcome by the proposed method [17], [18].

III. PROPOSED APPROACH

A. SIoT DATA COLLECTION

The developed things in the SIoT can able to detect the physical conditions, collects data, exchange or scatter information, process information for fitting applications, and communicate with different things. Henceforth, SIoT has been introduced with power innovation that aids in understanding the physical world and external jolts. While IoT can be considered of having ordinarily correspondence to the physical world by detecting or activating through huge numbers of various gadgets, SIoT worldview and this specimen appeared in figure 1, raises imperative issues regarding why and how to use these services as well as application.

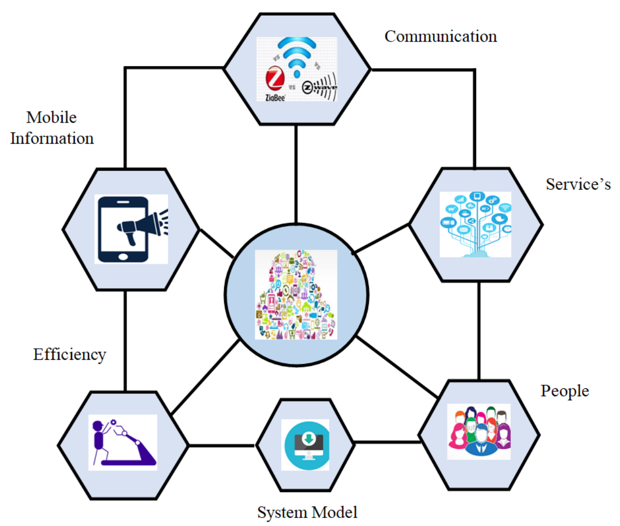

FIGURE 1. Sample social IoT.

B. FILTERING

Filtering is the key to the proposed approach, as for a particular dataset, some filter will be used to select the subset of dataset and, subsequently, the execution will remove noise and undesirable data.

1) GABOR FILTER

Gabor is Gaussian function, which have been revised by complex sinusoids. Under specific conditions, the period of the reaction of Gabor channels is roughly linear. Gabor wavelet symbolizes a class of functions. The utilization of the Gabor channel bank can viably diminish the calculation, decrease the measurement and remove noisy information. A Gabor base is a Gaussian function, in which the exponential is adjusted with respect to the result of a Gaussian and an exponential. The two-dimensional Gabor functions (u) can be described using equation 1 and 2.

$$u(a, b) = g(a, b)^{*}\exp^{-2\pi j f_{y} a} \tag{1}$$

Here

$$g(a, b) = \frac{1}{2\pi\sigma_a\sigma_b} * \exp^{\frac{-1}{2}\left(\frac{a^2}{\sigma_a^2}+\frac{b^2}{\sigma_b^2}\right)} \tag{2}$$

Gabor capacities are bandpass filters, which are Gaussians, focused on frequency in the spatial-area. Gabor channel with various frequencies and with introductions in various ways have been used to localize and extricate text area from data by removing noise.

C. DATA BASE REDUCTION

Hadoop MapReduce (HRM) comprises of a Job Tracker and a few Task Trackers. When a MapReduce task is executed, the Job Tracker divides it into smaller tasks (map as well as reduce), which can be processed by the task trackers and when the input dataset is given to a MapReduce data, it is split into autonomous data chunks using map task. The output of map task is the input of reduce task. The map function executes the technique of allocating each sample to the nearest focus while reduce task executes the new centers updating. As a result, the given data is mapped in a few groups and thus, the Map Reducer reduces SIoT database using a threshold value (Figure 2).

1) MAPPING PROCESS

Once the filtered data is converted into a set of data, a mapper can evaluate the nearest center point for every data. Maps are the individual tasks that change input data into the middle of the road information; such that the transformed intermediate data shouldn't have similar type as input data.

2) REDUCES PROCESS

This process is known as diminish mission, which inputs the productivity from a map and unites the data tuples into a slighter group of tuples and results into a reduced set of values. Now, the reduce process combines the data created during the mapping process.

D. FEATURE SELECTION

Now, we have used feature selection technique to select optimal subset of features out of all the features. Here, we have used EHO for feature selection and optimization. Elephant Herding Optimization searches the element space covetously and is constantly able to determine a better feature subset to optimize the given input data. It is used to improve classification.

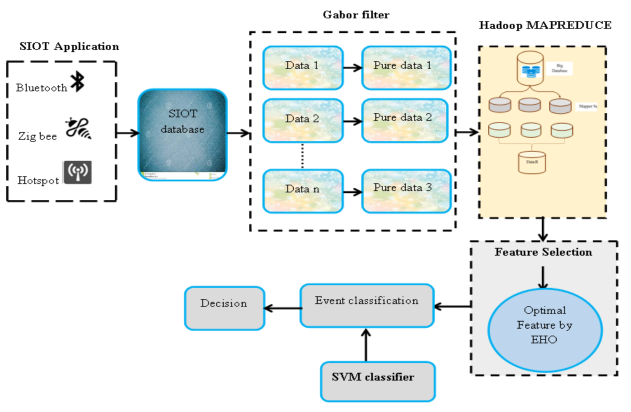

FIGURE 2. Proposed EH based classifying system for SIoT.

1) ELEPHANT HERDING OPTIMIZATION

In nature, the elephant is also considered as a social animal and the herd consists of several clans of female elephants and their calves. Under the influence of leader elephant or matriarch, the movement of every clan is defined. The Female Elephant (FE) use to live with their family groups whereas the Male Elephant (ME) separated when they grow up and live in contact with their family group utilizing low-frequency vibrations. Following assumptions are considered for herding:

- The population of elephants is divided into clans; each clan contains a fixed number of elephants.
- A fixed number of ME leaves their clan and live alone.
- Each clan moves under the leadership of a matriarch.

The group of matriarch holds the best solution in the herd of elephants while the worse solution is decoded from the position of the group of male elephants. Updating procedure of EHO has been shown below.

2) ELEPHANT POSITION

In this step, the position of each elephant in different clans can be determined except the matriarch and male elephant that holds the best and worse solution respectively. For each clan C_i elephants; each clan has 'P' elephants. The position of i^{th} elephant i = 1,2,...E also j^{th} clan j = 1,2,...l is represented by $G_{i,j}$.

a: Movement update of fittest elephant of each clan

Elephants that move far from the group can be used to model the investigation. Every clan have some number of elephants with the most noticeably bad estimations of the target work has been moved to the new position.

b: Separating worst elephants in clan

The worst elephant or male elephant will be separated from their family groups. During clan separation operation, the bits can be changed randomly since multiplication is performed with the random number. If the probability of randomness is given, then the number of bits to be changed can be estimated.

Where,

$G_{newli,j} \rightarrow$ Updated position,

$G_{li,j} \rightarrow$ Old position,

$G_{bestli,j} \rightarrow$ Position of best in the clan,

Algorithm 1 EHO Algorithm

(i) Initialization process

Choose feature attributes from the reduced database, EH parameters are scale factor and random values.

(ii) Fitness evaluation

Optimal feature selection consider accuracy as fitness function

(iii) Position of new elephant

$$G_{new,l_{i,j}} = G_{ci,j} + \alpha(G_{best,li,j} - G_{li,j}) \times r \quad (3)$$

(iv) Movement update

The position update for best fit in the clan is given by

$$G_{new,l_{i,j}} = \beta \times G_{center,l_j} \quad \text{And} \quad G_{center,l_j} = \sum_{i=j}^{n} G_{l_{i,j}}/n_l \quad (4)$$

(v) Separate worst position

The worst position updated as

$$G_{worst,l_{i,j}} = G_{\min} + (G_{\max} - G_{\min} + 1) \times r \quad (5)$$

(vi) Memorize the best probable results

(vii) Exit if

The optimal solution with maximum accuracy reached.

else

Iteration = Iteration +1.

α and β, And $r\ \varepsilon\ 0\ to\ 1$, n_l $\rightarrow$ The total number of elephants in each clan,

$G_{worst,ci,j} \rightarrow$ Worst male elephants in the clan and

$G_{\max}$ and $G_{\min}$ $\rightarrow$ Maximum and minimum allowable boundary limits for the clan elephants.

In a given scenario of EHO, each clan source is linked with a bit vector, where the vector depends on the overall number of features. All the features are then evaluated and the feature, which has its value '1', is in the subset of optimal features. These optimal features are fed into the classifier model to improve precision and performance rate of the proposed model.

E. CLASSIFICATION

For the data classification, the optimal feature attributes are given as an input to support vector machine (SVM). The SVM could perceive new data as members or non-members of the functional class based on the expression data. The classifier has two models one as training and another as testing as shown in figure 3.

1) TRAINING MODEL

The scaling is essential for the variable (traits) with bigger variance. Training data can be used for separating hyper plane and if the training data are directly divisible, then choose these hyper planes such that there are no focuses amongst them and to maximize their distance.

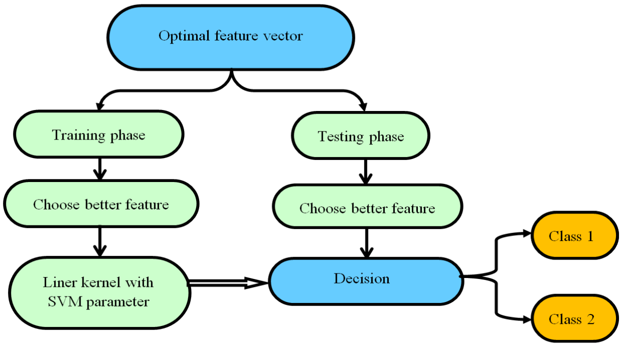

FIGURE 3. Linear kernel SVM-based classifier.

2) LINEAR KERNEL SVM

A supposition is made that bigger the margin or distance between these parallel hyperplanes; the better the classifier error. Before considering N-dimensional hyperplanes, let us consider a simple 2-dimensional example. The linear kernel function appears beneath condition.

$$\sum_{i=l}^{N} \alpha_i \cdot f_i \cdot K(g, g_i) = 0 \tag{6}$$

Here 'g' vector drawn from the input space, assumed to be in the dimension

p_0, α_i are the Lagrange coefficient,

f_i is the equivalent target output and

$K(g, g_i)$ denotes the inner product of two vectors convinced in the feature space by the input vector.

$$K(g, g_i) = g_i^T g_j + C \tag{7}$$

Now, we have to apply the set of above parameter to the training dataset and get the classifier. An SVM maps the input into a high-dimensional feature space and finds a separating hyper plane.

3) TESTING MODEL

Now, again apply the same set of parameters to the training dataset and get the classifier. Later on, use the classifier to classify the testing dataset to get the generalization accuracy.

IV. RESULTS AND DISCUSSIONS

The proposed SIoT data classification with optimal feature selection implemented in Java programming language with JDK 1.7.0 in a windows machine with the system setup, i5 processors with 4GB RAM and 64bit operating system platform.

The proposed model has been compared with the existing approaches like Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Genetic Algorithms (GA) based on the parameters like the total accuracy, the number of features, and classification results with performance metrics.

TABLE 1. Five database attribute information.

Database	Number of Instances	Class label 1	Class Label 2
GPS Trajectories	163	Car	Bus
Indoor User Movement Prediction from RSS	13197	The location changing movements (1)	Location preserving movements (-1)
Water Treatment Plant	527	Healthy	Faulty
Hepatitis	155	Die	Live
Twitter Dataset for Arabic Sentiment Analysis	20000	Positive polarity	Negative polarity

A. DATABASE DESCRIPTION

The proposed optimal feature based classification models the training and testing process and considers the five different databases from the UCI machine learning repository. Each database contains some features/attributes and different classes; this information has been demonstrated in table 1. Each dataset have been analyzed as per the number of features. Description of the five databases used in the work are-

B. GPS TRAJECTOIRES

This dataset contains trip information regarding taxi stand, call origin details, unique taxi id, unique trip id and the Polyline. The Polyline is a string of GPS coordinates with each coordinate being obtained after every 15 seconds of the trip beginning with the pickup and ending with the drop-off.

C. INDOOR USER MOVEMENT PREDICTION FROM RSS

This dataset represents a real-life benchmark in the area of Ambient Assisted Living applications. Input data contains temporal streams of radio signal strength (RSS) measured between the nodes of a WSN, comprising 5 sensors: 4 anchors deployed in the environment and 1 mote worn by the user.

D. WATER TREATMENT PLANT

This dataset comes from the daily measures of sensors in an urban waste water treatment plant. The objective is to classify the operational state of the plant in order to predict faults through the state variables of the plant at each of the stages of the treatment process.

E. HEPATITIS

The dataset contains 155 instances distributed between two classes die with 32 instances and live with 123 instances. There are 19 features or attributes, 13 attributes are binary while 6 attributes with 6-8 discrete values. The goal is to forecast the presence or absence of hepatitis virus.

F. TWITTER DATASET FOR ARABIC SENTIMENT ANALYSIS

The chosen tweets express some kind of emotions (positive or negative) and the goal of our model is to extract valuable information from such tweets with the end goal to determine the sentiment orientation of the inputted text. The months-long comment procedure of the tweets is physically led principally by two human specialists (local speakers of Arabic). On the off chance that the two specialists agree on the label of a specific tweet, then the tweet is doled out this label. Something else, a third master is counseled to break the tie.

Performance Metrics:

$$Sensitive = \frac{TP}{TP + FN} \tag{8}$$

$$Sensitive = \frac{TN}{TN + FP} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

Throughput: The throughput is the quantity of data (bits) conveyed within a time component (sec).

Energy Consumption (EC): The energy consumption is the standard of the energy devoted in data conveying from basic data to the target data within a time component.

TABLE 2. Optimal features with accuracy for proposed work.

Database	Total Number of Features	Optimal Features	Accuracy
GPS Trajectories	15	6	78.21
Indoor User Movement Prediction from RSS	4	3	84.56
Water Treatment Plant	38	19	96.45
Hepatitis	19	7	92.1
Twitter Dataset for Arabic Sentiment Analysis	2	1	95.14

Five datasets such as GPS Trajectories, Indoor User Movement Prediction from RSS, Water Treatment Plant, Hepatitis and Twitter Dataset for Arabic Sentiment Analysis and their optimal set of selected features are detailed in the table 2. For example, water treatment plant has 38 numbers of features, among that, 19 optimal features are selected using the optimization algorithm with 96.45% accuracy.

Figure 4 explains the comparative analysis of optimization algorithm in the feature selection process. For each database, the accuracy of optimal feature selection is compared with the algorithm like EHO, PSO, ACO, GA and ABC.

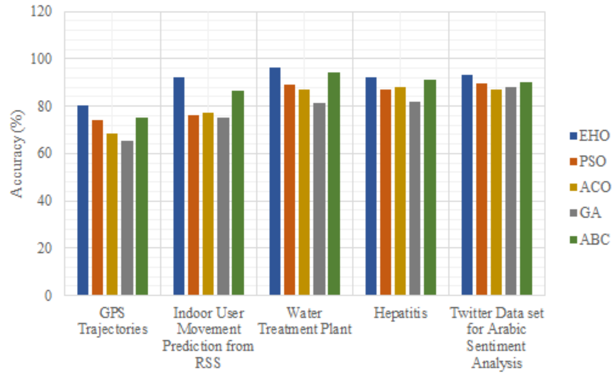

FIGURE 4. Comparative analysis for feature selection.

The optimization process by choosing an appropriate set of adjustable parameters, we practical that our approach is able to achieve a good rough calculation From the overall analysis, the proposed EHO based feature selection attains maximum accuracy in all the five databases compared to other related algorithms.

TABLE 3. Computational time and memory for proposed work.

Data Size (Mb)	Time (Sec)	Memory (byte)
50	55	1209421
100	58	1324581
150	120	1425131
200	180	1485402
250	200	1528231
300	280	1651941

Table 3 describes the time required for the data computation and its memory allocation for different data size. For example, 100 Mb data size needs 58 seconds for data computation and its corresponding memory allocation is about 1324581 bytes. Similarly, for other data set, the computation time and memory size are revealed in the above table.

Figure 5 represents the throughput (Kbps) for various data size (Mb) such as 50, 100, 150, 200, 250 and 300. Throughput is analyzed with two conditions namely, Hadoop with optimization and Hadoop without optimization. On comparing both the analysis, Hadoop with optimization accomplishes maximum throughput in the range of 6000 to 12000.

The energy consumption (EC) of different data size is depicted in figure 6 under two conditions, namely Hadoop with optimization and Hadoop without optimization. The amount of energy consumed during the data convey from basic data to the target data is measured and compared. The energy consumption of wireless statement primarily is governed by on the numeral of bits of data to be

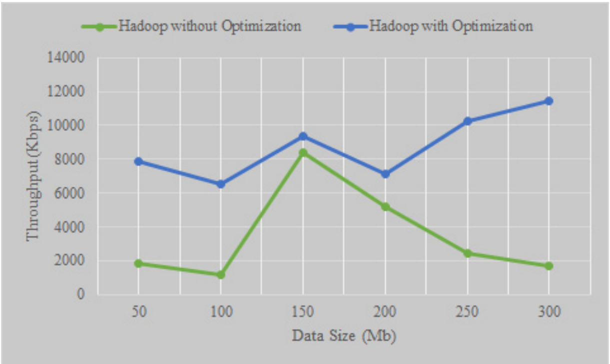

FIGURE 5. Throughput Vs data size.

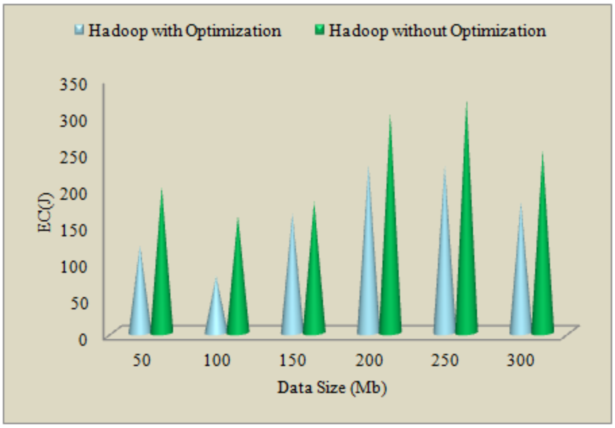

FIGURE 6. Energy consumption Vs data size.

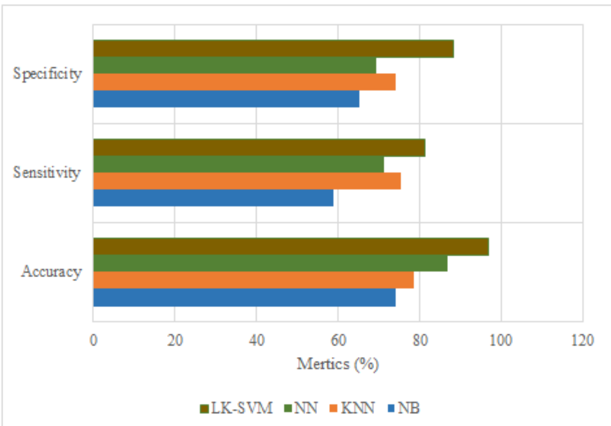

FIGURE 7. Performance metrics for various classifiers.

communicated with sensor network. The graph clearly shows that the energy consumption is high in the Hadoop without optimization compared to Hadoop with optimization.

Figure 7 illustrates the performance metrics of various classifiers such as LK-SVM, Neural Network (NN), K-Nearest Neighbour (KNN), and Naïve Bayes (NB). Compared to the existing classifiers, the proposed LK-SVM classifier attains more specificity and sensitivity. The SVM could classify the data as members or non-members of the class based on the expression data. The highest accuracy of data classification is 98.86%, which is achieved, in the proposed LK-SVM.

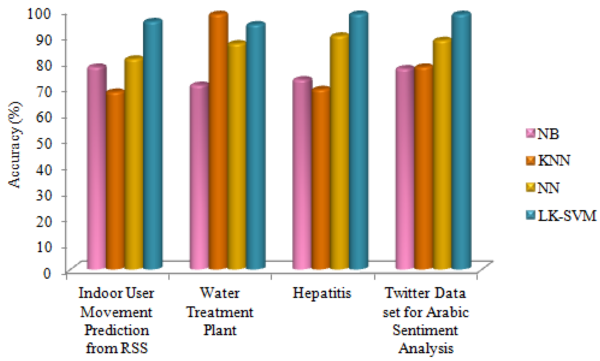

FIGURE 8. Comparative analysis for classification Performance.

Figure 8 represents the accuracy of the different classifier for the four different data sets i.e. Indoor User Movement Prediction from RSS, Water Treatment Plant, Hepatitis and Twitter Dataset for Arabic Sentiment Analysis. For the four datasets, the maximum accuracy is accomplished in the proposed LK-SVM compared to other classifiers.

V. CONCLUSION

The study investigated the classification of SIoT with big data map-reduce system. Here, Gabor filter has been used to reduce the redundant data from the database. Furthermore to improve the efficiency of the proposed work, Hadoop Map Reduce has been analyzed for mapping the data. The optimal features were selected from the database using the algorithm for data classification. In this research work, the EHO has been used to select the optimal subset of features from each dataset. Finally, LK-SVM classifier was proposed to classify the data as class 1 and class 2 based on the expression data. The result shows that the proposed model attains a maximum accuracy of 98.86% compared to other existing approaches. The researcher can use the proposed work to improve the performance measures of big data classification based on SIoT by using advanced classifier with the hybrid optimization algorithms.

An application of the Internet of things in a social network that includes a smart device that can impulsion its status messages on Twitter and in few areas like telecommunication, marketing and finance have been early adopters of big data analytics.

REFERENCES

[1] G. Han, L. Zhou, H. Wang, W. Zhang, and S. Chan, "A source location protection protocol based on dynamic routing in WSNs for the social Internet of Things," *Future Generat. Comput. Syst.*, vol. 82, pp. 689–697, May 2018.

[2] J. Lopez, R. Rios, F. Bao, and G. Wang, "Evolving privacy: From sensors to the Internet of Things," *Future Generat. Comput. Syst.*, vol. 75, pp. 46–57, Oct. 2017.

[3] O. Johny, S. Sotiriadis, E. Asimakopoulou, and N. Bessis, "Towards a social graph approach for modeling risks in big data and Internet of Things (IoT)," in *Proc. Int. Conf. Intell. Netw. Collaborative Syst. (INCoS)*, Sep. 2014, pp. 439–444.

[4] A. Paul, A. Ahmad, M. M. Rathore, and S. Jabbar, "Smartbuddy: Defining human behaviors using big data analytics in social Internet of Things," *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 68–74, Oct. 2016.

[5] H.-J. Yim, S. Dongmin, H. Jung, M.-K. Back, I. Kim, and K.-C. Lee, "Description and classification for facilitating interoperability of heterogeneous data/events/services in the Internet of Things," *Neurocomputing*, vol. 256, pp. 13–22, Sep. 2017.

[6] Z. Song, Y. Sun, J. Wan, L. Huang, Y. Xu, and C.-H. Hsu, "Exploring robustness management of social Internet of Things for customization manufacturing," *Future Generat. Comput. Syst.*, to be published.

[7] S. F. Ochoa, G. Fortino, and G. Di Fatta, "Cyber-physical systems, Internet of Things and big data," *Future Generat. Comput. Syst.*, vol 75, pp. 82–84, Oct. 2017.

[8] M. Wasi-ur-Rahman *et al.*, "High-performance RDMA-based design of Hadoop MapReduce over InfiniBand," in *Proc. IEEE 27th Int. Parallel Distrib. Process. Symp. Workshop Ph.D. Forum (IPDPSW)*, May 2013, pp. 1908–1917.

[9] E. Ahmed *et al.*, "The role of big data analytics in Internet of Things," *Comput. Netw.*, vol. 129, pp. 459–471, Dec. 2017.

[10] M. M. Rathore, A. Ahmad, and A. Paul, "IoT-based smart city development using big data analytical approach," in *Proc. IEEE Int. Conf. Autom. (ICA-ACCA)*, Oct. 2016, pp. 1–8.

[11] K. Guo, Y. Tang, and P. Zhang, "CSF: Crowdsourcing semantic fusion for heterogeneous media big data in the Internet of Things," *Inf. Fusion*, vol. 37, pp. 77–85, Sep. 2017.

[12] F. Alam, R. Mehmood, I. Katib, and A. Albeshri, "Analysis of eight data mining algorithms for smarter Internet of Things (IoT)," *Proc. Comput. Sci.*, vol. 98, pp. 437–442, Jan. 2016.

[13] C.-M. Huang, C.-H. Shao, S.-Z. Xu, and H. Zhou, "The social Internet of Thing (S-IOT)-based mobile group handoff architecture and schemes for proximity service," *IEEE Trans. Emerg. Topics Comput.*, vol. 5, no. 3, pp. 425–437, Jul./Sep. 2017.

[14] A. Jian, X. Gui, W. Zhang, and X. He, "Social relation predictive model of mobile nodes in Internet of Things," *Elektronika Elektrotechnika*, vol. 19, no. 4, pp. 81–86, 2013.

[15] A. Rasooli and D. G. Down, "A hybrid scheduling approach for scalable heterogeneous hadoop systems," in *Proc. SC Companion High Perform. Comput., Netw., Storage Anal. (SCC)*, Nov. 2012, pp. 1284–1291.

[16] M. Z. Hasan and F. Al-Turjman, "SWARM-based data delivery in Social Internet of Things," *Future Generat. Comput. Syst.*, to be published.

[17] D. Hussein, S. N. Han, G. M. Lee, N. Crespi, and E. Bertin, "Towards a dynamic discovery of smart services in the social Internet of Things," *Comput. Electr. Eng.*, vol. 58, pp. 429–443, Feb. 2017.

[18] X. T. Nguyen, H. T. Tran, H. Baraki, and K. Geihs, "Optimization of non-functional properties in Internet of Things applications," *J. Netw. Comput. Appl.*, vol. 89, pp. 120–129, Jul. 2017.

[19] X. Zhang, S. Ding, and Y. Xue, "An improved multiple birth support vector machine for pattern classification," *NeuroComputing*, vol. 225, pp. 119–128, Feb. 2017.

[20] A. Ahmad *et al.*, "Toward modeling and optimization of features selection in big data based social Internet of Things," *Future Generat. Comput. Syst.*, vol. 82, pp. 715–726, May 2017.

[21] W. Mardini, Y. Khamayseh, M. B. Yassein, and M. H. Khatatbeh, "Mining Internet of Things for intelligent objects using genetic algorithm," *Comput. Electr. Eng.*, vol. 66, pp. 423–434, Feb. 2017.

LAKSHMANAPRABU S. K. was born in Karaikudi, Tamil Nadu, India, in 1988. He received the B.E. degree in electronics and instrumentation engineering from the R.M.K. Engineering College, Chennai, in 2009, and the M.E. degree in industrial engineering from the Sudharsan Engineering College, Pudukkottai, India, in 2011. He is currently pursuing the Ph.D. degree in multivariable process control with the Department of Electronics and Instrumentation Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology. His area of interests includes multivariable control, evolutionary algorithm, fuzzy logic control, artificial intelligence, Internet of Things, model based development, and hardware in the loop testing. He is a Senior Research Fellow with the B.S. Abdur Rahman Crescent Institute of Science and Technology. He was a recipient of the Award of Rajiv Gandhi Nation Fellowship from University Grand Commission, Government of India, Delhi, for doing the Ph.D. degree from 2013 to 2018.

SHANKAR K. received the M.C.A., M.Phil., and Ph.D. degrees in computer science from Alagappa University, Karaikudi, India. He is currently an Assistant Professor with the School of Computing, Kalasalingam Academy of Research and Education, Krishnankoil, India. He has several years of experience in research, academia, and teaching. His current research interests include cryptography, data mining, big data, and Internet of Things.

ASHISH KHANNA received the B.Tech. and M.Tech. degrees from GGSIP University, Delhi, in 2004 and 2009, respectively, and the Ph.D. degree from the National Institute of Technology, Kurukshetra. He is a highly qualified individual with around 15 years of rich expertise in teaching, entrepreneurship, and research and development with specialization in computer science engineering subjects. He has been a part of various seminars, paper presentations, research paper reviews, and conferences as a convener and a session chair, a guest editor in journals and has co-authored several books and papers in national, international journals and conferences. He has authored or co-authored many research papers in reputed journals and conferences. He also has papers in SCI-indexed and Springer journals. He was a Guest Editor in the IEEE conference-IC3TSN-2017 and managed a special session on Parallel and Distributed Network-Based Computing Systems and is a Guest Editor in Springer Conference at ICDMAI-2018 and is managing a special session on computational intelligence for data science. He was a Reviewer in some SCI indexed Journals like *Cluster Computing* (Springer) and the IEEE conferences. He displayed vast success in continuously acquiring new knowledge and applying innovative pedagogies and has always aimed to be an effective educator and have a global outlook. He has co-authored ten text books and edited books i.e., *Distributed Systems*, *Java Programming and Website Development*, *Java Programming*, *Computer Graphics*, *Computer Graphics and Multimedia*, *Computer Networks*, *Computer Networks and Data Communication Networks*, *Success Mantra for IT interviews*, and *Fundamental of Computing*. He has also an edited book in Lambert publication. He is also a Reviewer and a Session Chair of the IEEE International Conference ICCCA 2016 and 2017. He has designed the syllabus for cloud computing, Java Programming and Distributed Systems for GGSIP University, Delhi. He recently successfully managed Smart India Hackathon at MAIT GGSIP University in 2017, with teams under him winning prices at their respective events. His research areas include distributed computing, distributed systems, cloud computing, vehicular ad hoc networks, and opportunistic networks. He is currently associated with some Springer and the IEEE conferences and managing special sessions for them and looking forward for some more challenging tasks.

DEEPAK GUPTA received the M.E. degree from the Delhi College of Engineering in 2010 and the Ph.D. degree from Dr. APJ Abdul Kalam Technical University in 2017. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, GGSIP University, Delhi, India. He has authored or co-authored many scientific papers in SCI journals, like JoCS, IAJIT, and FGCS. In addition, he has authored/edited over 25 books (Elsevier, Springer, IOS Press, and Katson) and six journal special issues. His research area includes software usability, human-computer interaction, nature-inspired computing, machine learning, and soft computing. He is a convener and an organizer of ICICC-2018 Springer conference. He has also started a research unit under the banner of Universal Innovator. He is also associated with various professional bodies, like ISTE, IAENG, IACSIT, SCIEI, ICSES, UACEE, Internet Society, SMEI, IAOP, and IAOIP.

JOEL J. P. C. RODRIGUES (S'01–M'06–SM'06) received the five-year B.Sc. degree (licentiate) in informatics engineering from the University of Coimbra, Portugal, the M.Sc., Ph.D. degrees, and Aggregation in informatics engineering from UBI, and the Habilitation degree in computer science and engineering from the University of Haute Alsace, France. He is currently a Professor and a Senior Researcher with the National Institute of Telecommunications, Brazil, and also a Senior Researcher with the Instituto de Telecomunicações, Portugal. He is also the Leader of the Internet of Things Research Group, the Director for Conference Development–IEEE ComSoc Board of Governors, the IEEE Distinguished Lecturer, the President of the scientific council, ParkUrbis–Covilhã Science and Technology Park. He has authored or co-authored over 600 papers in refereed international journals and conferences, three books, and two patents. He is the Past-Chair of the IEEE ComSoc Technical Committee on eHealth, the Past-chair of the IEEE ComSoc Technical Committee on Communications Software, a Steering Committee Member of the IEEE Life Sciences Technical Community, a Publications Co-Chair, and a Member Representative of the IEEE Communications Society on the IEEE Biometrics Council. He received the Academic Title of Aggregated Professor in informatics engineering from UBI. He is the Editor-in-Chief of the *International Journal on E-Health and Medical Communications*, the Editor-in-Chief of the *Recent Advances on Communications and Networking Technology*, the Editor-in-Chief of the *Journal of Multimedia Information Systems*, and editorial board member of several high-reputed journals. He had been awarded several Outstanding Leadership and Outstanding Service Awards by the IEEE Communications Society and several best papers awards. He is a Licensed Professional Engineer (as a Senior Member), a member of the Internet Society, and a Senior Member of ACM.

PLÁCIDO R. PINHEIRO (M'07) received the B.S. and Licentiate degrees in mathematics from the Federal University of Ceara, Ceara, Brazil, in 1979 and 1984, respectively, the degree in electrical engineering from the University of Fortaleza, Fortaleza, CE, Brazil, in 1983, the M.S. degree in mathematics from the Federal University of Ceara, Ceara, Brazil, in 1990, and the Ph.D. degree in systems engineering and computing degree from the Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil, in 1998. He is currently a Titular Professor with the University of Fortaleza and also an Associate Professor of the State University of Ceara. He has experience in industrial processes modeling applying mathematical programming and multi criteria. His academic formation allows publishing in applied mathematics area, with an emphasis on discrete and combinatorics mathematics, working mainly in mathematical programming and multicriteria.

VICTOR HUGO C. DE ALBUQUERQUE (M'17) received the degree in mechatronics technology from the Federal Center of Technological Education of Ceará, in 2006, the M.Sc. degree in teleinformatics engineering from the Federal University of Ceará, in 2007, and the Ph.D. degree in mechanical engineering with emphasis on materials from the Federal University of Paraíba, in 2010. He is currently an Assistant VI Professor of the Graduate Program in Applied Informatics, and also a Coordinator of the Laboratory of Bioinformatics, University of Fortaleza. He has authored or co-authored over 160 papers in refereed international journals, conferences, four book chapters, and four patents. He has experience in computer systems, mainly in the research fields of: applied computing, intelligent systems, visualization and interaction, with specific interest in pattern recognition, artificial intelligence, image processing and analysis, and automation with respect to biological signal/image processing, image segmentation, biomedical circuits and human/brain-machine interaction, including augmented and virtual reality simulation modeling for animals and humans. Additionally, he has research at the micro structural characterization field through the combination of non-destructive techniques with signal/image processing and analysis, and pattern recognition. He is the Leader of computational methods with the Bioinformatics Research Group. He is the Editorial Board Member of the IEEE Access, *Computational Intelligence and Neuroscience*, the *Journal of Nanomedicine and Nanotechnology Research*, and the *Journal of Mechatronics Engineering*. He has been a lead guest editor of several high-reputed journals and a TPC member of many international conferences.