IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 11, 2017, accepted February 7, 2017, date of publication February 13, 2017, date of current version August 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2666785

A Review on Mobile SMS Spam Filtering Techniques

SHAFI'I MUHAMMAD ABDULHAMID[1], (Member, IEEE), MUHAMMAD SHAFIE ABD LATIFF[2], HARUNA CHIROMA[3], (Member, IEEE), OLUWAFEMI OSHO[1], GADDAFI ABDUL-SALAAM[5], ADAMU I. ABUBAKAR[6], (Member, IEEE), AND TUTUT HERAWAN[4]

[1]Department of Cyber Security Science, Federal University of Technology, Minna PMB 65, Nigeria
[2]Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[3]Federal College of Education (Technical), Gombe P. M. B 60, Nigeria
[4]AMCS Research Center, Yogyakarta 1003, Indonesia
[5]Kwame Nkrumah University of Science and Technology, Kumasi 3419, Ghana
[6]Department of Information Systems, International Islamic University of Malaysia, Kuala Lumpur 50728, Malaysia

Corresponding author: Tutut Herawan (tutut@uty.ac.id)

This work was supported by International Islamic University Malaysia Research under Grant RIGS16-364-0528.

ABSTRACT Under short messaging service (SMS) spam is understood the unsolicited or undesired messages received on mobile phones. These SMS spams constitute a veritable nuisance to the mobile subscribers. This marketing practice also worries service providers in view of the fact that it upsets their clients or even causes them lose subscribers. By way of mitigating this practice, researchers have proposed several solutions for the detection and filtering of SMS spams. In this paper, we present a review of the currently available methods, challenges, and future research directions on spam detection techniques, filtering, and mitigation of mobile SMS spams. The existing research literature is critically reviewed and analyzed. The most popular techniques for SMS spam detection, filtering, and mitigation are compared, including the used data sets, their findings, and limitations, and the future research directions are discussed. This review is designed to assist expert researchers to identify open areas that need further improvement.

INDEX TERMS Review, spam, mobile SMS, access layer, service provider layer.

I. INTRODUCTION

Globally, short messaging service (SMS) is one of the most popular and also most affordable telecommunication service packages. However, mobile users have become increasingly concerned regarding the security of their client confidentiality. This is mainly due to the fact that mobile marketing remains intrusive to the personal freedom of the subscribers [1]. SMS spamming has become a major nuisance to the mobile subscribers given its pervasive nature. It incurs substantial cost in terms of lost productivity, network bandwidth usage, management, and raid of personal privacy [2]. Thus, in short spamming threatens the profits of the service providers [3], [4]. Mobile SMS spams frustrate the mobile phone users, and just like e-mail spams, they cause new societal frictions to mobile handset devices [5]. Email spam is sent or received via the World Wide Web, while the SMS mobile spam is typically broadcasted via a mobile network.

Spam can be described as unwanted or unsolicited electronic messages sent in bulk to a group of recipients. The messages are characterized as electronic, unsolicited, commercial, mass constitutes a growing threat mainly due to the following factors: 1) the availability of low-cost bulk SMS plans; 2) reliability (since the message reaches the mobile phone user); 3) low chance of receiving responses from some unsuspecting receivers; and 4) the message can be personalized. Mobile SMS spam detection and prevention is not a trivial matter. It has taken on a lot of issues and solutions inherited from relatively older scenarios of email spam detection and filtering [8]. Unsolicited SMS text messages are a common occurrence in our daily life and consume communication time, bandwidth and resources. Although the existing spam filters provide some level of performance, the spams misinform receivers by manoeuvring data samples [9].

The existing studies show that mobile SMS spam filtering techniques have remained at their initial stage of classification, for example the simple character string similarity or explicit number blocking [8], [10]. Traditional filtering techniques such as Bayesian classification filter, logistic regression and decision tree algorithm for mitigating SMS spam messages are still highly time consuming [1]. Studies have been completed on the different types of propose techniques for the filtering and mitigation of mobile SMS spam [11]–[13]. However, there are next to no literature reviews that summarize the currently available methods, chal-

2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

lenges and future research directions for the mobile SMS spam detection, filtering and mitigation techniques.

In this paper, we present a summary of the currently available methods, challenges and future research opportunities on the mobile SMS spam detection, filtering and mitigation methods. Firstly, it provides a taxonomy of the techniques of mobile SMS spam detection and filtering; secondly, it offers an indepth analysis of these techniques with respect to the performance evaluation metrics; thirdly, it examines the available research datasets relevant in the current and future research; and finally, it identifies the limitations of the existing studies and points out the future research directions. This study can be used by young researchers as a starting point and used by the experts in the field as a benchmark for further improvements.

The subsequent sections of the paper are organized as follows: Section II presents related surveys and highlights their limitations. Section III outlines the method used for the literature search including the performance metrics most commonly used in assessing the effectiveness of the proposed techniques. Section IV offers an overview of the existing research in mobile SMS spam detection techniques. Section V contains the description of the benchmark dataset including its sources. Section VI discusses legal laws against spam SMS in some countries. Section VII describes limitations and future research directions. Finally, Section VIII sums up the paper with concluding remarks.

II. RELATED SURVEYS

In this section are examined similar surveys that have been conducted by other researchers. This approach is followed in order to point out what issues have yet to be addressed and to highlight the differences with our current analysis. A survey on the filtering of mobile SMS spam and developments was done by Delany *et al*. [14]. The authors addressed the problems of collecting the research dataset and its accessibility. The article advanced future research in this domain. Subsequently, a preliminary benchmark experiment was conducted which indicated a lack of consensus on the best methods for mobile SMS spam detection and filtering. Furthermore, it showed what methods were being applied in the text classification of extensive SMS filtering. However, not taken into consideration were the explicit features of SMS. In general, the methods applied in the paper were basic. For the preliminary benchmark experiment, the authors compared the fingerprint of every newly received SMS to the fingerprints of all identified spam texts. Those related to any of the already identified spam fingerprints were classified as spam. However, the survey was limited to the publications prior to 2012 and thus, more recent methods and benchmark datasets were left out.

Bantukul and Marsico [15] conducted a survey of methods and applications for detecting and filtering unsolicited advertising messages or spam in a telecommunication network. The result showed that if the message passed the spam screening, the original mail would be delivered to its intended destination. However, the survey concentrated more on the techniques used for e-mail spam detection and excluded other forms of mobile SMS spam techniques such as the artificial immune system.

Camponovo and Cerutti [6] offered an overview on regulatory frameworks of spam in the mobile SMS business in Switzerland, the European Union and the USA. The paper also investigated the expected inferences for the commercial mobile industry.

Wang *et al*. [3] completed a survey on SMS anti-spam systems that combined both behaviour-based social network and temporal or spectral testing to identify spam with very high accuracy and recall. The authors explained the classification infrastructure and presented a fairly accurate neighbourhood index solution that addressed the scalability issue of social networks.

Chou and Lien [16] investigated the 'mobile teaser ads' by conducting two separate experiments on brand awareness, representative friendliness, and representative expertise and the ways in which they could sway brand interest in subscribers with diverse SMS mind-sets. The result showed that for teaser ads featuring highly awareness products, a friendly and well-known representative reduced the users' inquisitiveness.

Jindal and Liu [17] reviewed spam and spam recognition on products advertisement blogs. However, the review only covered spams related to product advertisement blogs and made no mention of SMS spam.

Similarly, Web [18] appraised many algorithms for filtering distrustful behavior over the period of ten years and categorized suspicious behavior into the four classes of traditional spam, fake reviews, social spam, and link farming. However, this appraisal only covered e-mail spams, fake blog reviews and social media spam and lacked an in-depth analysis of SMS spam. A word attack approach that takes advantage of the control of classifier with the lowest amount of introduced characters using the weight values combine with the length of words in the SMS was described by Chan *et al*. [9].

The feature reweighting technique was proposed together with an innovative rescaling method that reduced the significance of the element signifying a small word so as to necessitate more inserted characters for a successful avoidance. This approach was appraised experimentally by using the text messages and the sample comment spam bank. The outcome of the experiment showed that word length constituted an essential feature of the robustness of SMS spam filtering to good word attack. In the next section, we present the methods used in reviewing the existing studies.

III. REVIEW METHODS

In this section are discussed the methods used in the existing review works.

A. LITERATURE SEARCH

The preliminary search terms were comprehensively evaluated to identify the most suitable search terms. Based on the

stated objectives, the following terms were used to search the relevant literature in the established academic databases: 'SMS spam', 'mobile spam', 'mobile SMS spam', 'SMS spam filtering', 'SMS spam detection', 'SMS spam mitigation', and etc. All articles were identified and retrieved from the academic online databases via the ACM Digital Library, Emerald, DBLP Computer Science Bibliography, IEEE Xplore™, SAGE Journals, ScienceDirect®, Scopus™, SpringerLink, Taylor & Francis Online and Web of Science.

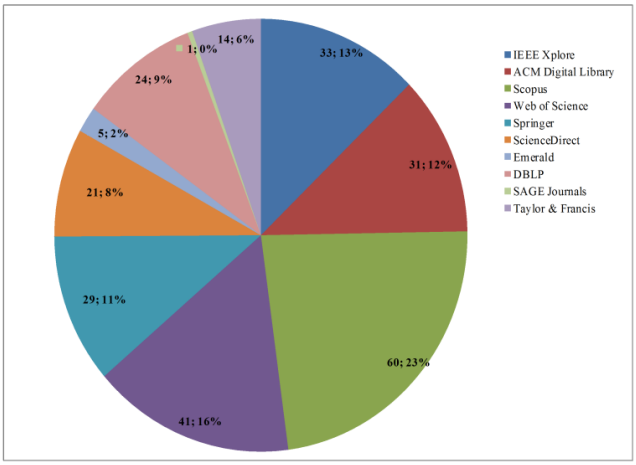

FIGURE 1. Online academic database used for searching the literature.

Figure 1 shows the representation in percentage of the different databases used. The articles searched on the databases were published in the period 2009 to 2016.

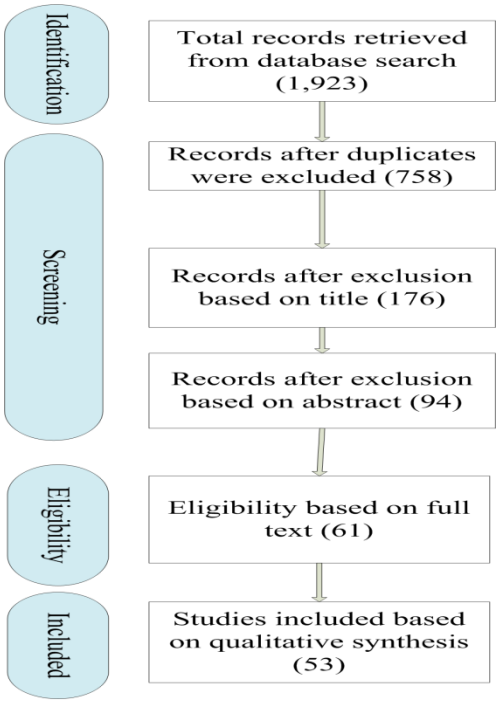

FIGURE 2. Search procedure and study selection diagram.

The search strategy applied at this stage were based on the work of Liberati *et al*. [19]. The references that are found to match the search terms proposed in our study were scanned to identify studies cited in the articles selected for inclusion in the study. A screening was done by sorting the relevant titles related to the objectives of this paper following the example of Hartling *et al*. [20]. The screening was based on the titles, abstracts and conclusions of the selected papers. The papers that fulfilled the screening criteria were used in our study. Finally, only relevant articles that contained experimental descriptions of their investigations were selected for this study. Data extraction was done on the sorted papers and subsequently tabulated and Figure 2 is created. The articles that were returned and identified from the online databases amounted to a total of 1,923.

B. PERFORMANCE METRICS

In order to evaluate or determine the accuracy of the mobile SMS spam filtering techniques, certain performance evaluation metrics were applied to the selected papers. The following parametrics were found to be prominent:

a. True Positive (TP) - the amount of samples that are properly classified;
b. True Negative (TN) - the amount of samples that are properly rejected from the class.
c. False Positive (FP) - the amount of samples that are wrongly rejected from the class;
d. False Negative (FN) - represents the amount of samples wrongly classified to the correct class;
e. Word attributes (WA) - information bit that establishes the characteristics of a field or tag in a database or a string of characters in a display.

From the parametrics definitions, the following metrics were deduced: accuracy, precision, recall, F1 score, spam precision, non-spam precision, False Accept Rate (FAR) and Matthews Correlation Coefficient (MCC). The deduced metrics have been defined as follows:

Percentage Accuracy (PA) determines the percentage spam SMS classified accordingly [21], [22].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision can be understood as the extended edition of accuracy. It is a simple metric that calculates the fraction of cases for which the accurate outcome is returned [23].

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall is the quotient of accurate to inaccurate forecasts within real spam texts [23], [24]. It is also called Detection Rate (DR) [25].

$$DR = Recall = \frac{TP}{TP + FN} \tag{3}$$

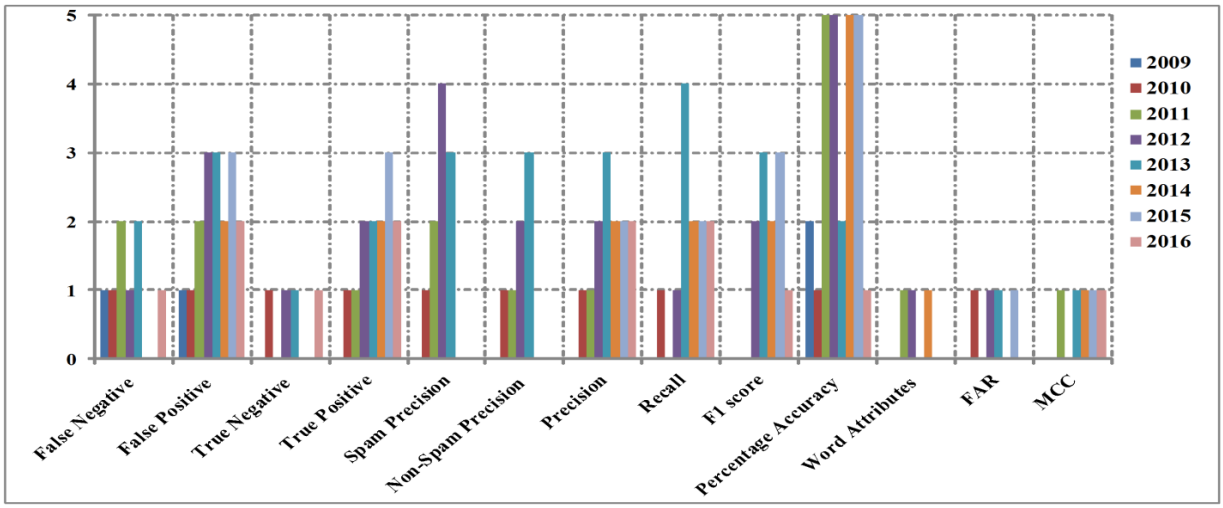

FIGURE 3. Performance Metric.

The F1 score is a valuable and efficient metric for unbalanced data [26].

$$F1score = 2\frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Spam Precision constitutes the quotient of correct to incorrect classifications among filtered spam messages [24].

$$SpamPrecision(SP) = \frac{Amount\ of\ SPAM\ message}{Total\ Amount\ of\ SPAM\ message} \tag{5}$$

Non-Spam Precision (NSP) is the quotient of real to not real non-spam texts within those texts filtered as non-spam texts [24].

$$NSP = \frac{Amount\ of\ correct\ Non-SPAM\ message}{Total\ Amount\ of\ Non-SPAM\ message} \tag{6}$$

The FAR is to compute the false classifying rate of the non-spam messages [27].

$$\text{FAR} = \frac{TP \cap \text{FN}}{TP} \times 100\% \tag{7}$$

The MCC is used in machine learning evaluation as a determinant of the value of binary classifications. It computes and returns a real value within the range $[-1, +1]$. A coefficient of $+1$ signifies a perfect prediction; 0 signifies a normal arbitrary prediction; and -1 signifies an inverse prediction [28].

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \tag{8}$$

In Figure 3, the metrics used for evaluating the accuracy of the filtering techniques found in the publications spanning over the years 2009 to 2016 are shown.

Percentage Accuracy (PA%) and Spam Precision appear to be most widely used. Other popular ones are FN, FP, TP, Precision, Non-Spam Precision, Recall and F1 Score. The TP and WA follows in line. FAR and MCC are the least used probably due to their mathematical complexity. Choosing the right evaluation metrics for an experiment in order to obtain maximum accuracy and to derive relevant information is important for validating any propose approach. In the following section, we present a taxonomy of research directions based on the identified research articles obtained in the review.

IV. TAXONOMY OF RESEARCH DIRECTION

Recent studies on mobile SMS spam show that several techniques are used to detect, filter or classify spam text messages. The solutions are designed to work either in the Access Layer (AL) or Service Provider Layer (SPL). The AL is the user-end layer mostly utilized in the form of light-weight

FIGURE 4. Architecture of the SMS spam transmission line.

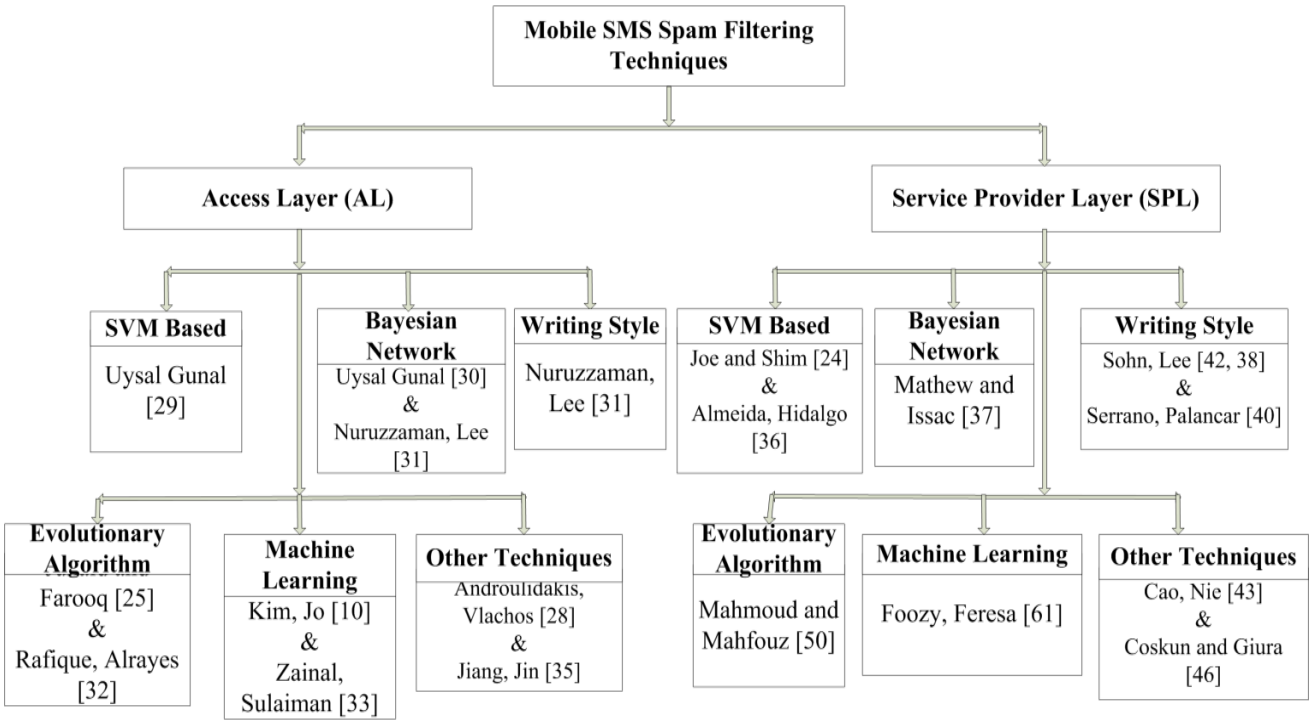

FIGURE 5. Taxonomy of mobile SMS spamming techniques.

software on the Android platform. A number of spamming techniques are designed to work directly on the mobile phone. Other techniques are designed to be deployed at the SPL. Figure 4 shows the basic architecture of the SMS spams transmission line, and Figure 5 shows the taxonomy of researches in mobile SMS spam.

A. ACCESS LAYER (AL)

Uysal *et al*. [29] suggested a *k*-nearest neighbor (kNN) and Support Vector Machine (SVM) classification of real-time mobile application for Android based mobile phones. Different permutations of the Bag-of-Word (BoW) and structural features (SF) are fed into widely used pattern classification algorithms in order to classify the SMS messages. In the simulation, collection of BoW features are based on Chisquare (CHI2) and information gain (IG) methods, where the number of certain features ranges from 1 to 100% of the entire BoW features. The experimental results and analysis on the relevant test sets show that the mixture of BoW and SFs (instead of BoW characteristics alone) allows for a more effective and precise performance classification on both test sets. It is also found that the efficiency of utilizing characteristics selection processes varies in each language.

Uysal *et al*. [30] proposed a Bayesian-based filtering framework consisting of the features derived from the BoW model together with the collection of SF explicit to spam. When the presence of a new SMS message is detected, the recommended model determines whether the SMS is spam or not. When it is identified as a legitimate message, it is kept in the 'Inbox' and the user is alerted of an incoming legitimate text message. If not, the alert is deactivated and the text is silently sent into the 'Spam Box'. However, the spam text can be traced back if needed. The performance of the framework is experimentally assessed on a bulk message collection which includes spam and non-spam messages. The results show that a considerably high level of precision in terms of the classification is achieved for both spam and non-spam SMSs. Nuruzzaman *et al*. [31] offered a text classification technique using Naïve Bayes and word occurrences tabling. The technique contributes to filter SMS spam on an independent mobile phone based on Naïve Bayes and word occurrences table. The technique is applied to a Google Android HTC Nexus One with Android™ 2.1 (Éclair) Operating System, Qualcomm® QSD8250™, 1 GHz Processor and a MicroSD™ memory card. Two experiments are carried out with the new technique. The first simulation depicts a scenario where the applicability is low since the user needs to have a huge amount of data during the initialization of the training data. Therefore, a second simulation is run as the subscriber needs about 10 SMS spam and 10 SMS ham as training data. The results show that the proposed spam filtering scheme on an autonomous mobile phone achieves outstanding precision with low storage consumption and satisfactory execution time.

Junaid and Farooq [25] applied an evolutionary learning classifier to create a detection system that filters spam SMS at the access layer of a mobile phone. It investigates

a SMS message in the hexadecimal system and mines out two features from this format, octet1 bigrams and frequency distribution of bytes. They evaluate the practicability of a number of evolutionary and non-evolutionary classifiers (functional on the exceeding feature sets) for the sieving system. The outcome of the experiments suggest that the Supervised Classifier System (SCS) working on the characteristic set achieves a detection rate above 89% and an almost 0% false alarm rate.

Rafique *et al*. [32] utilized certain evolutionary algorithms and a structural learning algorithm in vague environment (SLAVE). The study contributes a real-time spam detection architecture that models the byte-level transitions of SMS for the diverse classification algorithms. Also, the four evolutionary and four non-evolutionary classifiers are examined in detail. The algorithm is implemented as a classifier based on an open source software called KEEL. For cANT-Miner, the implementation is based on an unbiased evaluation that uses standard metric for values of all the classifiers. The classifiers are independently trained on SMS messages that are in 7-bit, 8-bit, and 16-bit encoding systems. Then, a stratified 10-fold cross validation technique on the dataset of every encoding system is done. In this procedure, every dataset is divided into 10 segments where 9 of them are used for preparing the classifiers and the remaining segment used for analysis. This procedure is repeated with all the other segments, the recounted outcome constituting the average for all the segments. The experimental result proves that the SLAVE technique achieves a detection precision of at least 93% with a false alarm rate of about 0.13% in filtering spam messages.

Kim *et al*. [10] developed a keyword frequency ratio and WEKA 3.7 machine learning tool simulation scheme for the light and fast system. The mobile message filter can be executed within the phones autonomously through the use of keyword frequency ratio (FR). Each message is broken down into a set of keyword components by utilizing the function 'string to word vector' in a WEKA interface. Then pre-processing is completed on 5,574 SMS messages.

The WEKA 3.7 intelligent tool utilized to perform the experiment carries out the feature selection technique by using Naive Bayes, J-48 Decision Trees, and Logistic. The algorithms are selected with a full training set and 10-fold cross justification for the testing determinations. The 10-folds are based on haphazardly selected data which are distributed into 10 disjoint subsets of almost identical dimension. Each subset is used as the justification set while the remainder are used to form the classifier. The justification set is subsequently used to appraise the accuracy. The accuracy estimate is the mean of the estimates for each of the classifiers. The outcome of processing the algorithms with the FR shows that the Naive Bayes returns 0.01 seconds of CPU Time and 94.70% of Accuracy, the J-48 algorithm 0.02 and 94.82%, and Logistic algorithm 0.1 and 94.71% respectively.

Zainal *et al*. [33] proposed a Bayesian technique developed on RapidMiner and Weka simulators. In order to execute the experiment, two freeware tools are utilized, RapidMiner and Weka interface using a dataset downloaded from the UCI Machine Learning Repository. The dataset contains 5,572 occurrences consisting of 4,825 messages labelled as ham and 747 as spam. The results show that both tools perform similar using the same data with the same classification and clustering technique.

Androulidakis *et al*. [34] described a technique called Filtering Mobile External SMS Spam (FIMESS). The FIMESS is designed to perform simple yet effective tests on the SMS headers so as to classify them as spam or non-spam. The technique is used on the mobile phone's control of storing and forwarding SMS texts and is capable of performing checks against schemes used by spammers. The prototype is implemented on the android operating system (OS) yet can easily be ported or installed on other mobile phone OS platforms. Primarily, the scheme screens incoming SMS texts and registers them in a lightweight database the Short Message Service Center (SMSC) of each SMS sender. It is able to utilize the important information in the SMS headers and identifies the SMS spam.

Jiang *et al*. [35] used the Grey Phone Space (Greystar) technique designed to monitor phone numbers from the grey phone space and employs a new statistical model to detect spam numbers based on their footprints on the grey phone space. The experimental results show that the Greystar performs much faster in detecting spam messages than the existing spam report techniques. The Greystar also minimizes the spam traffic to almost 75% at peak periods.

B. SERVICE PROVIDERS LAYER (SPL)

Joe and Shim [24] used SVM for spam filtering the mobile system by applying experience-based learning to identify spam SMSs. The terms contained in a SMS text are mined out while passing through a pre-processor and dictionary. If the homogenized term is contained in the feature list, the word catalogue is set to 1 or 0. The produced vector values are utilised as learning data to change the SVM hyperplane. After every feature vector is marked 0 or 1, a learning process is concluded from the side-to-side SVM classifier. A Gaussian Radial Basis Function (RBF) is used as the kernel function. The constant value is set as 10, 20, 40, and gamma values are set at 0.01, 0.05, and 0.1. The technique shows its performance with a feature vector rate of 150, a constant rate of 20 and a gamma rate of 0.01. The detection rate is significantly reduced when the pre-processing device cannot separate word lines appropriately.

Almeida *et al*. [36] applied the SVM to the then newest and largest mobile phone spam compilation. The method consists of setting up a ceiling of 20 iterations and utilizes the rest of the default values for Expectation-Maximization (EM) parameters in WEKA. It then uses the following well known performance measures of Spam Caught (SC%), Blocked Hams (BH%), Accuracy (Acc%), and Matthews Correlation Coefficient (MCC). The resulting corpus is presented in the form of token frequencies. The results show that the

SVM-based scheme performs better than other comparative filters. As such, the SVM can be used as a sound basis for further assessment.

Mathew and Issac [37] compared the variety of intelligent Bayesian classifiers with other classifier techniques for mobile spam filtering in mobile SMS. The WEKA does not read strings and therefore all strings are converted into data in the form of feature vectors. This is achieved by using the Weka 'StringToWordVector' function for this transformation. The variety of the Bayesian methods proves to be very efficient with a success rate of about 98%.

Sohn *et al*. [38] used a stylistic feature-based shallow linguistic as the new feature that indicates the writing style of SMSs for content-based mobile spam filtering. The simulations are performed by means of 10-fold cross support, and the results show that the content-based stylistic techniques outperform the comparative Bayesian scheme earlier proposed by Gómez Hidalgo *et al*. [39].

Serrano *et al*. [40] designed a writing style-based spam filter using extrinsic information, sequential labeling extraction and term clustering to store the writing style of spam and non-spam SMS. This is achieved by preserving the order of the content in the feature space. All the classifiers use the 10-fold cross validation in WEKA. The experimental results show that the technique produces a low dimensional feature space that shows competitive classification precision for the tested schemes.

Table 1 presents the summary of the review papers. The first, second, third, fourth, fifth and sixth columns represent serial number (S/N), references, the method or techniques proposed by the researchers, description of the data set used, method or technique used for evaluation of the proposed methods or techniques, and major findings or contribution of the study, respectively.

Mahmoud and Mahfouz [50] created an Artificial Immune System (AIS) SMSs classification scheme for filtering SMS spam. The AIS system uses a set of features to serve as an input spam filter. It categorizes text messages by using a trained dataset which consists of Phone Numbers, Spam Words and Detectors. The experimental results are obtained using the iPhone Operating System (iOS). The outcome of this experiment shows that the proposed scheme is able to classify messages either as spam or non-spam with more accuracy and convergence speed than the Naive Bayesian algorithm.

Foozy *et al*. [61] used Naive Bayes and J48 machine learning techniques to classify the spam and ham SMS. The paper also contains a collection of SMS phishing in Malay language as alternative mitigation to overcome SMS spam and phishing. The Malay SMS collection is tested using the Naïve Bayes algorithm and J48 and compares them with unsupervised Machine Learning techniques. The results show that both schemes were relatively similar in their performance.

Cao *et al*. [43] proposed an ontology-based spam message detection system that improves the spam classification accuracy and viability. When a suspicious message is received, the sending number information and the network information are obtained and used for the ontology mapping and classification. As a result of the logical and authorized detection of spam, the mobile network operating quality and commerce level are enhanced.

Coskun and Giura [46] examined a network-based online detection technique that swiftly classifies all SMS spamming activity by identifying unusual number of related SMS via a network over a short period. The central design plan of this technique is to retain a fairly accurate count of message information to check if there is an unusually high amount of similar messages sent within a very short interval of time. For this reason, an efficient data structure known as Counting Bloom Filter (BF) is utilized. The experimental results show that in order to achieve high detection rates in all cases, there huge Bloom filters (i.e. BFSize $\geq$ 500, 000) are used. The method shows high precision in detection and filtering unsolicited spam messages sent through the SPL servers.

The TPA method [67] is proposed and tested to normalize original short and cluttered text messages and thus obtain better attributes and enhance the categorization performance. This approach for text processing is based on lexicographic and semantic dictionaries along with modern methods for semantic analysis and context discovery.

Another SVM was also presented in Reaves *et al*. [72]. The outcome of the reported experiment shows that the proposed classification techniques and labeled clusters, precision and recall rise to 100% and 96.8%. The GrumbleText dataset is also used to test and compare the efficiency of the proposed technique.

The HMM method is discussed in [73]. It utilizes the GrumbleText dataset for a rigours evaluation of the proposed method. The experimental results indicate that it achieves about 97% detection rates with a near zero false alarm rate during SMS spam classification.

V. DATASETS

Accessibility to a requisite dataset constitutes one of the challenges researchers often face in successfully carrying out research on filtering or classifying SMS spam messages. This also applies to the evaluation of newly proposed SMS spam filtering and detection methods [55], [74]. In this review, we have summarized and present credible research datasets used by previous researchers as shown in Table 2.

It provides easy access to credible sources of datasets for the benefit of the researchers. The table contains four columns representing the serial number, name of the dataset, URL address and reference respectively. The different types of the mobile SMS spam datasets are described as follows:

A. GRUMBLETEXT

GrumbleText is a website that collects MS spam texts which can be extracted for experiments. The site is a United Kingdom's online forum that allows mobile customers to

TABLE 1. Summary of the major research findings on mobile SMS spam.

S/N	Reference	Proposed Technique/(s)	Dataset Description	Comparison Technique/(s)	Major Finding/(s)
1	Sohn *et al*. [38]	Stylistic content	Real world SMS from Korea with 18,000 (60%) non-spam SMS and 12,000 (40%) spam SMS.	Bayesian [39]	The content-based stylistic technique performs better than the Bayesian method in terms of FN and FP rates.
2	Joe and Shim [24]	SVM	Training: 200 non-spam messages and 100 spam messages. Testing: 80 spam messages and 80 non-spam messages.	No evaluation	The SVM based on SP, NSP and Recall achieves optimal performance with a feature vector value of about 150, a constant value of about 20 and a gamma value of about 0.01.
3	Adrian [41]	Challenge Response Based (Turing Test)	Not disclosed	Humans	The Challenge Response System based on the Turing test achieved a PA between 94% and 100% while that of the machine achieves 0%.
4	Sohn *et al*. [42]	Stylistic information	Real world Korea text messages	Baseline, Shallow, Structural and All style	The proposed technique is more efficient in Korean mobile spam detection compared to the other techniques.
5	Almeida *et al*. [36]	SVM	GrumbleText and NUS SMS Corpus based on 425 SMS extracted from a total of 3,375 SMS.	Naıve Bayes, kNN, Linear SVM,C4.5, Boosted and PART	The SP, PA and MCC show that the SVM based scheme performs better than other comparative techniques of spam SMS filtering.
6	Cao *et al*. [43]	Ontology	Not disclosed	No evaluation	The ontology shows an enhanced mobile network operating quality, commerce level, and etc.
7	Junaid and Farooq [25]	Evolutionary Learning Classifiers	Real world dataset of SMS with non-spam SMS of about 6,000 and about 2,000 spam SMS and GrumbleText	Fuzzy Ada Boost, GAssist-ADI, Naïve Bayes, JRip SVM, XCSC4.5 and UCS.	The results of the experiments based on DR, FAR, TP, TN, FP and FN show that the UCS when operated on the set features achieves 89% above the detection rate performance and zero per cent false alarm rate.
8	Mathew and Issac [37]	Bayesian Classifiers	Over 5,000 SMS, about 15% spam messages and 5,574 English, real and non-encoded SMS tagged according to legitimacy (ham) or spam.	Logistic, Simple Logistic, Decision Table, Conjunctive Rule, Ada Boost Ml, SPegasos, Filtered Classifier, Logic Boost, Dragging, Nnge, Lazy KStar, and etc.	Evaluation of the Bayesian classifier according to FP, FN and PA proved to be very efficient with a success rate of 98% compared to the other techniques.
9	Nuruzzaman *et al*. [31]	Naïve Bayes and Word Occurrences table (Writing Style).	English dataset with 425 SMS spam data from Grumble Text and 450 SMS non-spam from Caroline Tag's PhD thesis	No evaluation	The spam filter based on the Naïve Bayes and word occurrences table improve precision by lowering storage consumption and fasten execution time measured in terms of WA and PA.
10	Rafique *et al*. [32]	SLAVE	5 000 real world SMS text messages with 800 spam SMS and Grumble Text website	Naive Bayes, RIPPER, cAnt-Miner, SVM, UCS and XCS.	The SLAVE accomplishes a detection correctness of over 93% with a false alarm rate of about 0.13% in filtering spam messages measured in FAR and PA.
11	Vural and Venter [44]	Artificial Immune Systems (AIS)	Real world SMS text messages, 60 non-spam and 6 spam	AIS, Threshold and AIS, Affinity	It is found that the botnet detection method correctly filters spam up to 86% with threshold and 93% with affinity measured in PA.
12	Yadav *et al*. [45]	SMSAssassin and Bayes	Real world SMS, 2 195 non-spam and 2 123 spam	Bayesian learning and SVM	The technique proposed in the study produces 97% classification accuracy in non-spam and around 72.5% classification accuracy in spam SMS detection and outperforms the other techniques measured in PA.
13	Coskun and Giura [46]	Network based online detection technique	104,809 comments from YouTube converted to 160 characters	Bloom filter	Experimental results show that to achieve high detection rates in all cases, there is the need to utilize huge bloom filters (i.e. BFSize≥ 500, 000). This is measured in terms of FP and detection rate.

TABLE 1. ***(Continued.)*** **Summary of the major research findings on mobile SMS spam.**

14	Androulidakis *et al*. [47]	SMSC black listing scheme	DIT SMS Spam Dataset and NUS SMS Corpus (NSC)	No evaluation	The outcomes of the proposed technique showed that it is quite a valuable system in filtering SMS spam.
15	Hidalgo *et al*. [48]	Near duplicate detection approach	GrumbleText Website, Tag's PhD Thesis, NUS SMS Corpus and SMS Spam Corpus v.0.1 Big	No evaluation	The findings suggest that the proposed method does not lead to near-duplicates. The proposed test set is consistent to be used in performance evaluation by different spam classifiers. Average spam is used as the assessment metric.
16	Liu *et al*. [26]	kNN	600,000 SMS by 180,000 different users during a 3-day period	SVM and AdaBoost	The proposed kNN performs better than the SVM and AdaBoost in terms of Precision, Recall and F1 score.
17	Longe *et al*. [49]	Bayes	Not disclosed	No evaluation	The Bayesian filter rejects scam messages while taking measure on the sender.
18	Mahmoud and Mahfouz [50]	AIS	Not disclosed	Bayes	The AIS system can classify the SMS spam with more accuracy and better convergence speed based on FP, FN, TP, TN and FAR compared to Naïve Bayesian.
19	Uysal *et al*. [30]	Bayes	Real world data:747 spam and 4,827 non-spam messages	No evaluation	The Bayesian filter achieves a remarkable precision of classification for both spam and non-spam SMS in terms of PA.
20	Uysal *et al*. [29]	kNN and SVM	Turkish SMS and English SMS containing 425 spam and 450 legitimate SMS	BoW ,SF, BoW + SF	The experiment results and analysis on the relevant test sets show that the mixtures of BoW and SFs allows for a more effective and precise performance classification. The evaluation is done in terms of F1 score.
21	Taufiq Nuruzzaman *et al*. [51]	Bayes	GrumbleText website, Tag's PhD thesis, 425 SMS spam messages, 450 SMS ham messages	No evaluation	The proposed Naïve Bayes classifier indicates that WA can be decreased by approximately 50% without significantly reducing the PA.
22	Vural and Venter [52]	AIS	Real world SMS, 60 non-spam and 6 spam	No evaluation	The AIS precisely detects 84% of ham SMS which translates into 67 of 80 valid messages. It also detects 65% of spam messages which represents 13 of 20 invalid SMS. Its total error is 20% measured in PA, SP and NSP.
23	Yadav *et al*. [53]	Bayes	Not disclosed	No evaluation	The Bayesian filter demonstrates good features and functionality interfaces yet without any metric performance evaluation.
24	Almeida *et al*. [28]	Near-duplicate detection +Top scoring N-grams and SVM	Tag's PhD Thesis, 450 SMS and NUS SMS Corpus, 3,375 SMS and Grumbletext 425 SMS spam	Bayesian,kNN,PART,C4.5, Boosted, Logistic Regression, Random Forest.	The proposed method performs better than the other comparative schemes in detecting spam SMS.
25	Androulidakis *et al*. [34]	FIMESS	DIT SMS Spam Dataset with 1,353 spams SMS and NUS SMS Corpus.	No evaluation	FIMESS is able to utilize the important information in the SMS headers and identifies the SMS spam. The performance of this scheme is measured using a FP: FN ratio.
26	Deng *et al*. [54]	Frequent Time location (FTL) Algorithm	Not disclosed	With FLT and Without FLT	The experiment data used in both spam and non-spam SMS messages indicate that utilizing the incorporated model leads to a similar precision and meets the real-time filtration constraint. The performance is measured using Precision, Recall and F1 score.
27	Jiang *et al*. [35]	Greystar	SMS Call Detail Records (CDRs) and Victim spam reports in North America	No evaluation	The experimental results show that the Greystar is faster in detecting spam messages than other victim spam reports. It minimizes the spam traffic by almost 75% at peak period. The performance indicators used are TP, FP, TN, PA and spamming rate.

TABLE 1. (Continued.) Summary of the major research findings on mobile SMS spam.

28	Narayan and Saxena [55]	Bayes and SVM	NUS SMS corpus and GrumbleText	SVM,SVM on Bayes, Bayes on Bayes,	The proposed method filters spam with a precision and PA of about 98% better than the compared techniques. All measurements are taken in terms of PA, DR, F1 Score, FP, FN, TP and TN.
29	Eshmawi and Nair [56]	Domain Knowledge based and CART	NUS SMS Corpus, GrumbleText, PhD thesis.with 5,574 English SMS	SVM, Random Forest, Bayesian, CART	The proposed method significantly improves the performance of the state of the art methods in spam SMS categorization.
30	Griesel and Fourie [57]	BoW	Real world20 HAM and 20 spam	Naive Bayes, SVM and Decision Trees.	The differences between the performance of BoW and the Naive Bayes and the Multinomial Naïve Bayes approaches are statistically considerable yet the difference between the performance of SVM and BoW is not statistically significant.
31	Xia *et al*. [27]	Self-feedback algorithm and FTL algorithm	Not disclosed	No evaluation	The application of the integrated framework leads to a comparable precision and meets the real time filtration condition in terms of precision, recall and FAR.
32	Murynets and Jover [58]	Machine to Machine (M2M) System	Call Detail Records (CDR) 9,000 spam and almost 17,000 non-spams.	Spammers Vs. Legitimate	The M2M systems display message profiles are comparable to the spammers that may misinform spam detectors.
33	Zhang *et al*. [59]	Bayes	NUS SMS Corpus (NSC) GrumbleText Jon Stevenson Corpus (JSC).10,202 non-spam 82 spam	SVM,C45 and PART	The SVM performs better than the other techniques. Given that the distribution of SMS is more impartial in the collection, it allows finding more analytical features. The PA can score over 95% if the False Positive rate is set to zero or over 99% if a few False Positives are allowed.
34	Alzahrani and Ghorbani [60]	Multi Agent System	SMS Spam Collection dataset	No evaluation	The proposed system filters spam message botnets and spots ways to detect attacks in order to avoid damages caused by these spammers.
35	Foozy *et al*. [61]	Naire Bayes	Malay SMS collection	J48	The performance of the Naïve Bayes scheme andJ48 classifiers are the same as measured in terms of TP, FP and PA.
36	Karami and Zhou [21]	New Content based Features	GrumbleText, NUS and Tag's PhD thesis	Random Forest and SVM	The categorization PA lies between 92% and 98% within diverse techniques. The boosting algorithm of Random Forest and SVM algorithm indicates the best performance measured in Recall, Precision, PA, F1 score and MCC.
37	Modupe *et al*. [62]	Latent Dirichl Allocation (LDA) and Social Network Analysis (SNA)	Manual collection by authors	No evaluation	The proposed technique is able to extract features from SMS that can automatically filter spam SMS.
38	Najadat *et al*. [63]	Naïve Bayes, Stochastic Gradient Descent (SGD) and SVM	Total dataset of 1,500 SMS collated manually	AdaBoostM1, Decision Table, J48, Random Forest, SVM, kNN, KStar, Naïve Bayes, SGD, Voted Perceptron, DMNB text, NB Multinomial	The Discriminative Multinomial Naïve Bayes (DMNB), Stochastic Gradient Descent, and SVM yield the optimum PA. The DMNB has the best PA of 96.46%. In the complete data set, the results of the classifiers are promising. The SVM achieves a PA of 98.6%.
39	Serrano *et al*. [40]	Writing style of spam using Extrinsic Information, Sequential Labeling Extraction and Term Clustering	SMS Spam Collection	DTNB, Bayesian, SVM, NB Tree, Random Tree, Simple Cart, Lazy Start, and etc.	Low dimensional feature space gives more precision of classification measured in terms of TP and FP rates.
40	Skudlark [64]	Content-based	Mobile Terminating (MT) International Mobile Equipment Identity (IMEI)	No evaluation	An insight into content categorization of spam SMS as well as spam characteristics based on sending model and geo-location measured in PA.

TABLE 1. (Continued.) Summary of the major research findings on mobile SMS spam.

41	Al-Hasan and El-Alfy [22] & El-Alfy and AlHasan [65]	Dendritic Cell Algorithm (DCA)	SMS Spam Corpus V.0.1 and NUS SMS corpus (10,000 legitimate SMSs) and GrumbleText and Tag's PhD Theses (450 SMS), National University of Singapore (3,375 SMS) and Jon Stevenson Corpus (1,324 SMS).	SVM, KNN and Naïve Bayes	The Dendritic Cell Algorithm (DCA) can enhance the PA, Recall and Precision of spam and non-spam messages compared to the other techniques.
42	Kim *et al*. [10]	Keyword FR technique	Not disclosed	Naive Bayes, J-48 and Logistic	The FR shows that the Naive Bayes returns 0.01 seconds of CPU Time and 94.70% of accuracy, the J-48 algorithm 0.02 and 94.82%, and Logistic algorithm 0.1 and 94.71%, respectively.
43	Zainal *et al*. [33]	Bayesian scheme	UCI Machine Langue Repository	SVM and kNN	The analysis of algorithms to filter spam SMS using RapidMiner and Weka indicate that the performance of the algorithms in both RapidMiner and Weka based on TP, FP and PA is the same.
44	Ahmed *et al*. [23]	Frequent Itemset and Ensemble Learning (FIEL)	UCI SMS spam and SMS spam collection Corpus v.0.1	SPY-EM and PEBL	The recall, F1 score and PA indicated that the proposed FIEL is more stable and robust than the compared methods.
45	Chan *et al*. [9]	SVM	SMS Spam Collection and Deceptive Opinion Spam Corpus v1.4	SVM−FR,SVM−FR−Len	The execution time of SVM-FR and SVM-FR-Len is nearly double than that of the SVM. This suggests that the SVM-FR-Len increases the cost, i.e. the amount of inserted characters for any attack plan. The performance is assessed based on PA.
46	Saeed and Waheeb [66]	SVM	UCI repository, NUS SMS corpus, DIT and Grumbletext SMS corpus.	Fuzzy Similarity, Artificial Neural Network (ANN)	The TP, FP, F1 score and MCC for SVM supersedes that of the compared techniques. Also, the results prove the efficiency of the SVM classifier for SMS spam filtering with a PA of 99.2%.
47	Almeida *et al*. [67]	Text Processing Approach (TPA)	SMS Spam Collection which is a public dataset composed of 5,574 English SMS	Bagging of Decision Trees, Binary Context Tree Weighting, Boosted C4.5, Naïve Bayes, Context Tree Weighting, KNN, SVM, Logistic regression, Markov Compression, Prediction by Partial Match and Probabilistic Suffix Trees Compression.	For the Wilcoxon Signed-Ranks Test, the null hypothesis is rejected with α=0.05, that is with a confidence level of 95%.
48	Xu *et al*. [68]	Non-Content Features (NCF)	Real world data set from a large telecommunications operator in China Telco data set	SVM and KNN	On a real SMS dataset the temporal features and network features can be efficiently integrated to assemble an SVM classifier, with an increase of around 8% in improvement on Area Under the ROC Curve (AUC) as compared to those that are only based on conventional static features.
49	Rafique and Abulaish [69]	Graph-based Learning Model (GLM)	NUS SMS Corpus, collection of more than 5,000 real world benign and 800 spam SMS and GrumbleText	Receiver Operating Characteristic (ROC) analysis	The GLM system produces a remarkable 98% detection rate with a false alarm rate of less than 0.08% during the classification of spam SMS.
50	Ezpeleta *et al*. [70]	Personality Recognition (PR)	SMS Spam Collection and British SMS dataset	J48.1, J48.2 and J48.3	The PR feature improves almost all the results. In terms of accuracy a 98.94% is reached, the best result obtained when applying the classifiers to the original dataset.
51	Chen *et al*. [71]	TruSMS system	NUS SMS Corpus	Not mentioned	The F measure can generally reach 1 within 370s with a spam dissemination speed '1 piece/200 ms', 410s with spam dissemination speed '1 piece/300 ms' and 420s with spam dissemination speed '1 piece/400 ms' in the same situation. TruSMS can afford hybrid and integrated attacks well.

TABLE 1. (Continued.) Summary of the major research findings on mobile SMS spam.

52	Reaves *et al.* [72]	SVM	GrumbleText	Sender Volume vs. Message Volume	The proposed labeled dataset of SMS is collected over 14 months; the precision and recall of past classifiers fall to 23.8% and 61.3% respectively. Based on the proposed classification techniques and labeled clusters, the precision and recall rises to 100% and 96.8%.
53	Rafique and Farooq [73]	Hidden Markov Models (HMMs)	GrumbleText	SMS in 7-bit and 8-/16-bit encoding	The simulations show that by analyzing the test scenarios, the proposed HMMs algorithm provides more than 97% detection rate with a 0% false alarm rate in the classification of SMS spam.

TABLE 2. Spam SMS research datasets.

No	Dataset	URL Address	Reference
1	Grumbletext	http://www.grumbletext.co.uk/	[25, 31, 32, 36, 55]
2	Caroline Tag's PhD thesis	http://etheses.bham.ac.uk/253/1/Tagg09PhD.pdf	[22, 31, 51, 74]
3	YouTube Comments	http://www.youtube.com	[28, 46]
4	TurkishSMS	http://ceng.anadolu.edu.tr/par/	[29, 30]
5	National University of Singapore (NUS) Corpus	https://www.comp.nus.edu.sg/~rpnlpir/downloads/corpora/smsCorpus/	[22, 36, 55, 59]
6	DIT SMS Spam Dataset	http://www.dit.ie/computing/research/resources/smsdata/	[14, 34]
7	Malay SMS corpus	http://eprints.uthm.edu.my/6494/1/Cik_Feresa_Mohd_Foozy_U.pdf	[61]
8	SMS Spam Corpus Collection	http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/ and http://www.esp.uem.es/jmgomez/smsspamcorpus/	[22, 40]
9	UCI SMS spam	https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection	[33]
10	Deceptive Opinion Spam Corpus v1.4	http://myleott.com/op_spam/	[9, 75, 76]
11	Manual Real World SMS Collection	*Not applicable*	[24, 25, 32, 38, 42]

report public complains about SMS spam texts. To effectively use this dataset, it involves a lot of manual screening of the messages for spam, which is a very monotonous and time-consuming task as it involves carefully screening a huge number of web pages.

B. PhD THESIS

Caroline Tag's PhD thesis titled "*A Corpus Linguistics Study of SMS Text Messaging*" contains a highly popular research dataset [74]. It contains a list of about 450 SMS ham texts messages. The SMS Spam Corpus Big has an approximately 1,002 SMS ham text messages and 322 spam texts. It is a public dataset of mobile messages that is gathered for the usage of mobile SMS spam mitigation and filtering research. It also has a certain collection of SMSs composed of 5,574 English, real and non-encoded texts classified as being spam or non-spam.

C. YouTube COMMENTS

Coskun and Giura [46] created a dataset of SMS-like text remarks made by 104,809 YouTubers commenting on several YouTube videos. The data have been generated by crawling YouTube comments via YouTube public Application Program Interface, primarily from a small number of videos.

D. TurkishSMS

TurkishSMS is a Turkish SMS text dataset generated purely for the purpose of academic research. It consists of both spam and non-spam text messages. The TurkishSMS dataset is an open source. The dataset is freely available for use by scientists for educational purposes with the condition that the source of the data must be acknowledge as in Uysal *et al.* [29].

E. NATIONAL UNIVERSITY OF SINGAPORE SMS CORPUS

The National University of Singapore (NUS) SMS Corpus (NSC) constitutes a research dataset consisting of tens of thousands of both ham and spam text messages gathered at the Faculty of Computer Science at the NUS. The text messages are mostly derived from Singaporean users and predominantly from academicians.

F. DUBLIN INSTITUTE OF TECHNOLOGY

The Dublin Institute of Technology (DIT) SMS Spam text messages dataset consists of a corpus of 1,353 distinctive spam messages gathered by scouring messages using two UK community user complaint sites. All messages are tagged together with the time they are reported and the corpus covers 2003 to mid-2010. The data have been collected in the same linguistic area, which means all texts were originally received by UK mobile customers [14].

G. MALAY SMS CORPUS

Foozy *et al*. [61] created a Malay SMS corpus and created a dataset based on SMS collections from public websites, personal SMS forwarding and online forums. The dual classification of the SMS includes ham and scam SMS, while second class is based on the scam SMS that is further classified according to SMS spam and SMS phishing.

H. SMS SPAM CORPUS COLLECTION

The SMS Spam Collection has been created by Tiago A. Almeida in collaboration with José María Gómez Hidalgo. The SMS Spam Collection is an open access dataset of text messages which have been gathered for mobile spam research. It has an English collection of about 5,574 messages, real and non-encoded messages, labelled as spam and non-spam [28], [36], [48].

I. UNIVERSITY OF CALIFORNIA, IRVINE (UCI) SMS SPAM

The UCI SMS Spam Collection constitutes another free dataset of spam text messages composed and gathered for mobile SMS spam studies. The compilation is collected in one text file, where each text- row has the accurate class followed by the raw text message [36].

J. DECEPTIVE OPINION SPAM CORPUS

The Deceptive Opinion Spam Corpus v1.4 is a dataset corpus that is made of spam and non-spam hotel review comments of about twenty Chicago hotels. This dataset is freely available to researchers if they appropriately acknowledge their source [75], [76].

K. MANUAL REAL-WORLD SMS COLLECTION

Any real-world SMS collection is collected manually by individuals. A sample area and a specific language is identified and considered over a certain period of time. The collection procedures differ depending on the study and also on what the corpus is expected to use.

VI. SMS ANTI-SPAM LAWS ACROSS COUNTRIES

In the U.S. the SMS anti-spam laws are heavily dependent on two important legal provisions; the Telephone Consumer Protection Act (TCPA) [77] and the U.S. CAN-SPAM act of 2003 [78]. The TCPA regulates any means "*to communicate with or try to get into communication with a person by telephone*" while the CAN-SPAM Act established the major national standards for sending and receiving spam SMS required of the Federal Trade Commission (FTC). In Canada, the legal provision is the Canada's Anti-Spam Legislation (CASL) [79] with a broader scope covering all e-messages. The EU member countries are governed by an anti-spam provision called the Contact Network of Spam Enforcement Authorities (CNSA) [80]. The legal provision is applied by each member state independently which gives room for local adaptation.

In Italy, one can even be imprisoned for propagating SMS spam under the Italian Personal Data Protection Code (legislative decree no. 196/2003) [81]. The Unsolicited Electronic Messages Act 2007 and the Department of Internal Affairs provides strict legal guidelines for SMS anti-spam laws in New Zealand [82]. Others include The Privacy and Electronic Communications (EC Directive) Regulations [83] in the UK, the Regulation of Spam in South Africa - South African Law [84], the Information Technology Act of 2000 [85] in India and the Commission Nationale de l'Informatique et des Libertés (CNIL) [86] in France. However, even with the provision of many legal SMS anti-spam laws across many countries, there is no factual indication of its effective mitigation.

VII. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Table 1 shows that researchers mostly depend on SVMs and the Bayesian nework for designing classifiers that filter spam SMS. However, these applications have some limitations [87]. The number of support vectors (SV) is directly proportional to the size of the training dataset which forces SVMs to use unnecessary basis functions. SVMs are not suitable for the prediction of class labels because they are not based on probability. In addition, it is necessary for the kernel function in SVMs to fulfil Mercer's condition which means that they must have a continuous symmetric kernel of a positive integral operator. In contrast, relevance vector machine (RVM) prediction is based on probability. Here, the sensitivity of the RVM to free parameters is lower than that of the SVMs, and the selection of arbitrary kernel functions is probabilistic [88]. Also, fewer relevance vectors (RV) are used in RVMs compared to the SVs of a SVM. Thus, time computational complexity (TCC) for classification using RVMs is less than that of SVMs [89]. Based on our review observations, RVM is not common among researchers in proposing methods for the detection and classification of spam SMS. Future researchers can apply the RVM to construct the classifiers for the detection and classification of SMS spam and compare their performance with that of the SVM. Similarly, to improve the accuracy and convergence of speed of SVM, powerful bio-inspired algorithms such as the cuckoo search algorithm can be used for optimizing the parameters of the SVM [90]. Thus, an SVM optimized with the Cuckoo search algorithm is recommended for constructing classifier for the filtering of SMS spam.

The Bayesian network is still being used, despite the availability of improved and more powerful algorithms. It seems that the newly improved algorithms have not received sufficient attention in this research domain. For example, the 'Particle Swarm Optimization' [91] which is known for its fast convergence rate, the 'Artificial Bee Colony' [92] which is known to be effective in local search, the 'Bat' algorithm [93] and the 'Bees' algorithms [94] which are prominent for global search, the 'Fish Swarm' algorithm [95], and the 'Cat Swarm' [96], among others. A brief review of the biologically inspired algorithms can be found in [97]. More recently, Approximate Muscle Guided Beam Search [98] has proven to be efficient in classification problems, the

League Championship Algorithm [99] summarised in [100] and [101] has also proven to be effective in local and global settings, the Magnetotactic Bacteria Optimization based on Moment Migration [102] and the Symbiotic Organism Search Algorithm [103] as discretely used in [104] are also efficient in both discrete and continuous problems. The biologically inspired algorithms can be used in three different ways when combined with SVM: by optimizing the SVM weights, by automatically determining the SVM structure and its internal parameters without the SVM designer efforts, and lastly, by adapting the SVM learning rules. Empirical evidence indicates that the evolutionary SVM typically advances the efficiency, effectiveness and robustness of the SVM [105]–[107]. As such, we recommend that future work should utilize evolutionary algorithms in building classifiers for the detection and classification of SMS spam. This will likely produce enhanced, more effective, efficient and robust SMS spam filtering algorithms. In that way, it will support the arguement by Yang and Deb [108] that the optimal algorithm for application in solving real world problem needs to be robust, accurate and fast. The Bayesian network should be hybridized with the evolutionary algorithm for creating SMS spam detection classifier to improve its effectiveness and efficiency in view of the fact that hybrid algorithms are known to be more effective and efficient than single algorithms due to the fact that the weaknesses of the constituent hybrid algorithms are eliminated while their strengths are improved [109]–[111].

The evaluation of the commonly used techniques of filtering SMS spam messages calls for their comparison based on certain performance metrics (see section IV) with the currently employed techniques. In this way, the improvements achieved by each reviewed technique can be clearly demonstrated. However, many researchers such as Joe and Shim [24]; Cao *et al*. [43], Nuruzzaman *et al*. [31], Androulidakis *et al*. [47], Hidalgo *et al*. [48], Longe *et al*. [49], Uysal *et al*. [30], Taufiq Nuruzzaman *et al*. [51], Vural and Venter [52], Yadav, *et al*. [53], Androulidakis *et al*. [34], Jiang *et al*. [35], Xia *et al*. [27], Alzahrani and Ghorbani [60], Modupe *et al*. [62], and Skudlark [64] as listed in the table do not compare their proposed methods with the already established techniques. It therefore becomes difficult to assess their level of improvement. We therefore strongly recommend that all the proposed methods are to be evaluated in the light of the established techniques. For a method to be effective in filtering SMS spam using the Bayesian classification filter, logistic regression and decision tree algorithm for mitigating SMS spam messages, high level performance from the system resources, and loads of SMS test sets, are usually required. Presently, these weaknesses of spam filtering or detecting SMS spam messages have remained unresolved.

Currently, most spam filtering systems lack functionality support for secret and anonymous feedback. Thus, extending the spam filtering systems by adding functionality support for secret and anonymous feedback is important in order to assemble a dataset of the diverse SMSCs that are liable for sending malicious and spammed messages. The activities of cyber criminals frequently start with the occasional providers, registrars and hosting services. Whether or not mobile SMS spammers exploit the same infrastructure still remains an open question.

The limitation of the available research datasets is that they are valuable only for the study of content classification. The research related to general methodology is more general and relies on a variety of non-linguistic characteristics such as SMSC originator, Reply Path, HTTP links, Mobile Station International ISDN Number (MSISDN), and Protocol Identifiers such as TP-PID of the mobile text messages in order to decide if a message is spam or non-spam. Therefore, as a future research direction, it is recommended to create a more general and standard research dataset.

VIII. CONCLUSION

In this paper, we have summarized the recent advances in SMS spam filtering, mitigation and detection techniques as well as their limitations and future research direction. Many different SMS spam techniques, used datasets and comparisons are discussed. We have also developed a taxonomy of the techniques and identified the established results. The review discloses that most research is based on the support vector machine and the Bayesian network to construct SMS spam classifiers. This study highlights the fact that many powerful bio-inspired algorithms such as Monkey Search, Cat Swarm, Magneto Tactic Bacteria Optimization based on Moment Migration, Chicken Swarm Optimization, the Bat algorithm, the Cuckoo search algorithm, the Bees algorithms, and Particle Swarm Optimization are not being used in the creation of SMS spam classifiers. In summary, bio-inspired evolutionary algorithms have so far received little attention in this type of research. The sources of credible experimental benchmark datasets are revealed in the study. Novice researchers can use this study as a starting point while expert researchers can utilize it as a benchmark for further advancement. Also, the paper can serve as a source of information in the exploration of evolutionary algorithms that so far have received little or no attention from the research community.

REFERENCES

[1] E. B. Cleff, "Privacy issues in mobile advertising," *Int. Rev. Law Comput. Technol.*, vol. 21, no. 3, pp. 225–236, 2007.
[2] A. Lambert, "Analysis of SPAM," M.S. thesis, Dept. Comput. Sci., Univ. Dublin, Trinity College, Republic of Ireland, 2003, pp. 1–100.
[3] C. Wang *et al.*, "A behavior-based SMS antispam system," *IBM J. Res. Develop.*, vol. 54, no. 6, pp. 3:1–3:16, Nov./Dec. 2010.
[4] B. Reaves, N. Scaife, D. Tian, L. Blue, P. Traynor, and K. R. Butler, "Sending out an SMS: Characterizing the security of the SMS ecosystem with public gateways," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 339–356.
[5] T. Yamakami, "Impact from mobile SPAM mail on mobile internet services," in *Parallel and Distributed Processing and Applications*. Berlin, Germany: Springer, 2003, pp. 179–184.
[6] G. Camponovo and D. Cerutti, "The spam issue in mobile business: A comparative regulatory overview," in *Proc. 3rd Int. Conf. Mobile Bus.*, New York, NY, USA, 2004, pp. 1–17.
[7] J. Fu, P. Lin, and S. Lee, "Detecting spamming activities in a campus network using incremental learning," *J. Netw. Comput. Appl.*, vol. 43, pp. 56–65, Aug. 2014.

[8] J. Hua and Z. Huaxiang, "Analysis on the content features and their correlation of Web pages for spam detection," *China Commun.*, vol. 12, no. 3, pp. 84–94, Mar. 2015.
[9] P. P. K. Chan, C. Yang, D. S. Yeung, and W. W. Ng, "Spam filtering for short messages in adversarial environment," *Neurocomputing*, vol. 155, pp. 167–176, May 2015.
[10] S.-E. Kim, J.-T. Jo, and S.-H. Choi, "SMS spam filterinig using keyword frequency ratio," *Int. J. Secur. Appl.*, vol. 9, no. 1, pp. 329–336, 2015.
[11] O. Osho, O. Y. Ogunleke, and A. A. Falaye, "Frameworks for mitigating identity theft and spamming through bulk messaging," in *Proc. IEEE 6th Int. Conf. Adapt. Sci. Technol. (ICAST)*, Oct. 2014, pp. 1–6.
[12] R. Islam and J. Abawajy, "A multi-tier phishing detection and filtering approach," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 324–335, Jan. 2013.
[13] O. Osho, V. L. Yisa, O. Y. Ogunleke, and S. I. M. Abdulhamid, "Mobile spamming in Nigeria: An empirical survey," in *Proc. Int. Conf. Cybersp. (CYBER-Abuja)*, Nov. 2015, pp. 150–159.
[14] S. J. Delany, M. Buckley, and D. Greene, "SMS spam filtering: Methods and data," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9899–9908, Aug. 2012.
[15] A. Bantukul and P. J. Marsico, "Methods, systems, and computer program products for short message service (SMS) spam filtering using E-mail spam filtering resources," U.S. Patent 7 751 836 B2, Jul. 6, 2010.
[16] H.-Y. Chou and N.-H. Lien, "Effects of SMS teaser ads on product curiosity," *Int. J. Mobile Commun.*, vol. 12, no. 4, pp. 328–345, Jul. 2014.
[17] N. Jindal and B. Liu, "Review spam detection," in *Proc. 16th Int. Conf. World Wide Web*, 2007, pp. 1189–1190.
[18] M. Jiang, P. Cui, and C. Faloutsos, "Suspicious behavior detection: Current trends and future directions," *IEEE Intell. Syst.*, vol. 31, no. 1, pp. 31–39, Jan./Feb. 2016.
[19] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration," *Ann. Internal Med.*, vol. 151, pp. W-65–W-94, Jun. 2009.
[20] L. Hartling, C. Spooner, L. Tjosvold, and A. Oswald, "Problem-based learning in pre-clinical medical education: 22 years of outcome research," *Med. Teacher*, vol. 32, no. 1, pp. 28–35, 2010.
[21] A. Karami, A. Amir, and L. Zhou, "Improving static SMS spam detection by using new content-based features," in *Proc. AISeL*, 2014, pp. 1–9.
[22] A. A. Al-Hasan and E.-S. M. El-Alfy, "Dendritic cell algorithm for mobile phone spam filtering," *Procedia Comput. Sci.*, vol. 52, no. 1, pp. 244–251, 2015.
[23] I. Ahmed, R. Ali, D. Guan, Y.-K. Lee, S. Lee, and T. Chung, "Semi-supervised learning using frequent itemset and ensemble learning for SMS classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1065–1073, 2015.
[24] I. Joe and H. Shim, "An SMS spam filtering system using support vector machine," in *Future Generation Information Technology*. Berlin, Germany: Springer, 2010, pp. 577–584.
[25] M. B. Junaid and M. Farooq, "Using evolutionary learning classifiers to do MobileSpam (SMS) filtering," in *Proc. 13th Annu. Conf. Genet. Evol. Comput.*, 2011, pp. 1795–1802.
[26] J.-Y. Liu *et al.*, "Spam short messages detection via mining social networks," *J. Comput. Sci. Technol.*, vol. 27, no. 3, pp. 506–514, Jan. 2012.
[27] H. Xia, Y. Fu, J. Zhou, and Q. Xia, "Intelligent spam filtering for massive short message stream," *COMPEL-Int. J. Comput. Math. Elect. Electron. Eng.*, vol. 32, no. 2, pp. 586–596, 2013.
[28] T. Almeida, J. M. G. Hidalgo, and T. P. Silva, "Towards SMS spam filtering: Results under a new dataset," *Int. J. Inf. Secur. Sci.*, vol. 2, no. 1, pp. 1–18, 2013.
[29] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "The impact of feature extraction and selection on SMS spam filtering," *Elektron. Elektrotechn.*, vol. 19, no. 5, pp. 67–72, 2012.
[30] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "A novel framework for SMS spam filtering," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Jul. 2012, pp. 1–4.
[31] M. T. Nuruzzaman, C. Lee, and D. Choi, "Independent and personal SMS spam filtering," in *Proc. IEEE 11th Int. Conf. Comput. Inf. Technol. (CIT)*, Aug./Sep. 2011, pp. 429–435.
[32] M. Z. Rafique, N. Alrayes, and M. K. Khan, "Application of evolutionary algorithms in detecting SMS spam at access layer," in *Proc. 13th Annu. Conf. Genet. Evol. Comput.*, 2011, pp. 1787–1794.
[33] K. Zainal, N. Sulaiman, and M. Jali, "An analysis of various algorithms for text spam classification and clustering using RapidMiner and Weka," *Int. J. Comput. Sci. Inf. Secur.*, vol. 13, no. 3, pp. 66–74, 2015.
[34] I. Androulidakis, V. Vlachos, and A. Papanikolaou, "FIMESS: Filtering mobile external SMS spam," in *Proc. 6th Balkan Conf. Inform.*, 2013, pp. 221–227.
[35] N. Jiang, Y. Jin, A. Skudlark, and Z.-L. Zhang, "Greystar: Fast and accurate detection of SMS spam numbers in large cellular networks using grey phone space," in *Proc. USENIX Secur.*, 2013, pp. 1–16.
[36] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to the study of SMS spam filtering: New collection and results," in *Proc. 11th ACM Symp. Document Eng.*, 2011, pp. 259–262.
[37] K. Mathew and B. Issac, "Intelligent spam classification for mobile text message," in *Proc. Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Dec. 2011, pp. 101–105.
[38] D.-N. Sohn, J.-T. Lee, and H.-C. Rim, "The contribution of stylistic information to content-based mobile spam filtering," in *Proc. ACL-IJCNLP Conf. Short Papers*, 2009, pp. 321–324.
[39] J. M. G. Hidalgo, G. C. Bringas, E. P. Sánz, and F. C. García, "Content based SMS spam filtering," in *Proc. 2006 ACM Symp. Document Eng.*, 2006, pp. 107–114.
[40] J. M. B. Serrano, J. H. Palancar, and R. Cumplido, "The evaluation of ordered features for SMS spam filtering," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Cham, Switzerland: Springer, 2014, pp. 383–390.
[41] A. M. Adrian, "A challenge response system for filtering automated SMS spam," M.S. thesis, Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ. Sci. Technol., Taipei, Taiwan, 2010.
[42] D.-N. Sohn, J.-T. Lee, S.-W. Lee, J.-H. Shin, and H.-C. Rim, "Korean mobile spam filtering system considering characteristics of text messages," *J. Korea Acad.-Ind. Cooper. Soc.*, vol. 11, no. 7, pp. 2595–2602, 2010.
[43] L. Cao, G. Nie, and P. Liu, "Ontology-based spam detection filtering system," in *Proc. Int. Conf. Bus. Manage. Electron. Inf. (BMEI)*, May 2011, pp. 282–284.
[44] I. Vural and H. Venter, "Detecting mobile spam botnets using artificial immune systems," in *Advances in Digital Forensics VII*. Springer, 2011, pp. 183–192.
[45] K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik, "SMSAssassin: Crowdsourcing driven mobile-based system for SMS spam filtering," in *Proc. 12th Workshop Mobile Comput. Syst. Appl.*, 2011, pp. 1–6.
[46] B. Coskun and P. Giura, "Mitigating SMS spam by online detection of repetitive near-duplicate messages," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 999–1004.
[47] I. Androulidakis, V. Vlachos, and A. Papanikolaou, "Spam goes mobile: Filtering unsolicited SMS traffic," in *Proc. 20th Telecommun. Forum (TELFOR)*, Nov. 2012, pp. 1452–1455.
[48] J. M. G. Hidalgo, T. A. Almeida, and A. Yamakami, "On the validity of a new SMS spam collection," in *Proc. 11th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2012, pp. 240–245.
[49] O. B. Longe, K. Adegoke, A. Abdulganiyu, and F. A. Longe, "A prototype scalable system for secured bulk SMS delivery on mobile networks," *Int. J. Adv. Res. Comput. Sci.*, vol. 3, p. 43, Jan. 2012.
[50] T. M. Mahmoud and A. M. Mahfouz, "SMS spam filtering technique based on artificial immune system," *Int. J. Comput. Sci. Issues*, vol. 9, no. 2, p. 589, 2012.
[51] M. T. Nuruzzaman, C. Lee, M. F. A. bin Abdullah, and D. Choi, "Simple SMS spam filtering on independent mobile phone," *Secur. Commun. Netw.*, vol. 5, no. 10, pp. 1209–1220, 2012.
[52] I. Vural and H. S. Venter, "Combating mobile spam through botnet detection using artificial immune systems," *J. Universal Comput. Sci.*, vol. 18, no. 6, pp. 750–774, 2012.
[53] K. Yadav, S. K. Saha, P. Kumaraguru, and R. Kumra, "Take control of your SMSes: Designing an usable spam SMS filtering system," in *Proc. IEEE 13th Int. Conf. Mobile Data Manage. (MDM)*, Jul. 2012, pp. 352–355.
[54] A. Narayan and P. Saxena, "The curse of 140 characters: Evaluating the efficacy of SMS spam detection on Android," in *Proc. 3rd ACM Workshop Secur. Privacy Smartphones Mobile Devices*, 2013, pp. 33–42.
[55] A. Eshmawi and S. Nair, "Feature reduction for optimum sms spam filtering using domain knowledge," in *Proc. Int. Conf. Secur. Manage. (SAM)*, 2013, pp. 1–7.
[56] M. Griesel and W. Fourie, "Choosing the best classifier for the job: Mobile Filtering for the South African context," *Comput. Linguistics Netherlands J.*, vol. 2, pp. 23–33, Dec. 2013.
[57] I. Murynets and R. P. Jover, "Analysis of SMS spam in mobility networks," *Int. J. Adv. Comput. Sci.*, vol. 3, no. 1, pp. 1–8, 2013.

[58] L. Zhang, J. Ma, and Y. Wang, "Content based spam text classification: An empirical comparison between english and Chinese," in *Proc. 5th Int. Conf. Intell. Netw. Collaborative Syst. (INCoS)*, Sep. 2013, pp. 69–76.
[59] A. J. Alzahrani and A. A. Ghorbani, "SMS mobile botnet detection using a multi-agent system: Research in progress," in *Proc. 1st Int. Workshop Agents CyberSecur.*, 2014, p. 2.
[60] C. F. M. Foozy, R. Ahmad, and M. F. Abdollah, "A framework for SMS spam and phishing detection in malay language: A case study," *Int. Rev. Comput. Softw.*, vol. 9, no. 7, pp. 1248–1254, 2014.
[61] A. Modupe, O. O. Olugbara, and S. O. Ojo, "Filtering of mobile short messaging service communication using latent Dirichlet allocation with social network analysis," in *Transactions on Engineering Technologies*. Springer, 2014, pp. 671–686.
[62] H. Najadat, N. Abdulla, R. Abooraig, and S. Nawasrah, "Mobile SMS spam filtering based on mixing classifiers," *Int. J. Adv. Comput. Res.*, vol. 1, no. 1, pp. 1–7, 2014.
[63] A. Skudlark, "Characterizing SMS spam in a large cellular network via mining victim spam reports," in *Proc. 20th ITS Biennial Conf.*, Rio de Janeiro, Brazil Nov./Dec. 2014, pp. 1–23.
[64] E.-S. M. El-Alfy and A. A. AlHasan, "Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm," *Future Generat. Comput. Syst.*, vol. 64, pp. 98–107, Nov. 2016.
[65] H. Saeed and W. Waheeb, "The performance of soft computing techniques on content-based SMS spam filtering," M.S. thesis, Dept. Elect. Eng., Univ. Tun Hussein Onn Malaysia, Johor, Malaysia, 2015.
[66] T. A. Almeida, T. P. Silva, I. Santos, and J. M. G. Hidalgo, "Text normalization and semantic indexing to enhance instant messaging and SMS spam filtering," *Knowl.-Based Syst.*, vol. 108, pp. 25–32, Sep. 2016.
[67] Q. Xu, E. W. Xiang, Q. Yang, J. Du, and J. Zhong, "SMS spam detection using noncontent features," *IEEE Intell. Syst.*, vol. 27, no. 6, pp. 44–51, Nov./Dec. 2012.
[68] M. Z. Rafique and M. Abulaish, "Graph-based learning model for detection of SMS spam on smart phones," in *Proc. 8th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2012, pp. 1046–1051.
[69] E. Ezpeleta, U. Zurutuza, and J. M. G. Hidalgo, "Short messages spam filtering using personality recognition," in *Proc. 4th Spanish Conf. Inf. Retr.*, 2016, p. 7.
[70] L. Chen, Z. Yan, W. Zhang, and R. Kantola, "TruSMS: A trustworthy SMS spam control system based on trust management," *Future Generat. Comput. Syst.*, vol. 49, pp. 77–93, Aug. 2015.
[71] B. Reaves, L. Blue, D. Tian, P. Traynor, and K. R. B. Butler, "Detecting SMS spam in the age of legitimate bulk messaging," in *Proc. 9th ACM Conf. Secur. Privacy Wireless Mobile Netw.*, 2016, pp. 165–170.
[72] M. Z. Rafique and M. Farooq, "SMS spam detection by operating on byte-level distributions using hidden Markov models (HMMS)," in *Proc. 20th Virus Bull. Int. Conf.*, 2010, pp. 1–7.
[73] C. Tagg, "A corpus linguistics study of SMS text messaging," Ph.D. dissertation, Faculty College Arts Law, Univ. Birmingham, Birmingham, U.K., 2009.
[74] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Human Lang. Technol.*, vol. 1. 2011, pp. 309–319.
[75] M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," in *Proc. HLT-NAACL*, 2013, pp. 497–501.
[76] D. E. Sorkin, "Unsolicited commercial E-mail and the telephone consumer protection act of 1991," *Buffalo Law Rev.*, vol. 45, pp. 1001–1032, Jan. 1997.
[77] J. D. Sullivan and M. B. De Leeuw, "Spam after can-spam: How inconsistent thinking has made a hash out of unsolicited commercial E-mail policy," *Santa Clara Comput. High Technol. Law J.*, vol. 20, no. 1, p. 887, May 2003.
[78] E. Crowne and S. Provato, "Canada's anti-spam legislation: A constitutional analysis," *John Marshall J. Inf. Technol. Privacy Law*, vol. 31, no. 1, pp. 1–22, 2014.
[79] E. Moustakas, C. Ranganathan, and P. Duquenoy, "Combating spam through legislation: A comparative analysis of US and European approaches," in *Proc. CEAS*, 2005, pp. 1–8.
[80] F. Massacci, M. Prest, and N. Zannone, "Using a security requirements engineering methodology in practice: The compliance with the Italian data protection legislation," *Comput. Standards Int.*, vol. 27, no. 5, pp. 445–455, Jun. 2005.
[81] *Unsolicited Electronic Messages Act*, New Zealand Laws, The Department of Internal Affairs, New Zealand, 2007, pp. 1–6.
[82] E. Riach, "The privacy and electronic communications directive," *New Law J.*, vol. 1, no. 7071, pp. 379–380, 2003.
[83] B. D. Mdluli, "Online consumer protection: An analysis of the nature and extent of online consumer protection by South African legislation," M.S. thesis, Dept. Commercial Law, Univ. Cape Town, Cape Town, South African, 2014, pp. 1–109.
[84] S. Basu, "E-government and developing countries: An overview," *Int. Rev. Law, Comput. Technol.*, vol. 18, no. 1, pp. 109–132, 2004.
[85] M. Bernard, "Commission Nationale de l'Informatique et des Libertés," in *Bildiri, Bağimiz Ðdari Otoriteler, Türk-Fransiz Ortak Kolokyumu*. Paris, France, 1996, pp. 21–22.
[86] A. Tolambiya, S. Venkatraman, and P. K. Kalra, "Content-based image classification with wavelet relevance vector machines," *Soft Comput.*, vol. 14, no. 2, pp. 129–136, Jan. 2010.
[87] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.
[88] L. Wei, Y. Yang, R. M. Nishikawa, M. N. Wernick, and A. Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications," *IEEE Trans. Med. Imag.*, vol. 24, no. 10, pp. 1278–1285, Oct. 2005.
[89] D. Giveki, H. Salimi, G. Bahmanyar, and Y. Khademian. (2012). "Automatic detection of diabetes diagnosis using feature weighted support vector machines based on mutual information and modified Cuckoo search." [Online]. Available: https://arxiv.org/abs/1201.2173
[90] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, Nov./Dec. 1995, pp. 1942–1948.
[91] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Nov. 2007.
[92] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature Inspired Cooperative Strategies for Optimization (NICSO)*. Berlin, Germany: Springer, 2010, pp. 65–74.
[93] D. T. Pham, A. Ghanbarzadeh, E. KoçS. Otri, S. Rahim, and M. Zaidi, "The bees algorithm—A novel tool for complex optimisation problems," in *Proc. 2nd IPROMS Virtual Int. Conf. Intell. Prod. Mach. Syst.*, Jul. 2006, pp. 454–459.
[94] X. L. Li, Z. J. Shao, and J. X. Qian, "Optimizing method based on autonomous animats: Fish-swarm algorithm," *Xitong Gongcheng Lilun Shijian/Syst. Eng. Theory Pract.*, vol. 22, no. 11, p. 32, 2002.
[95] S.-A. Chu, P.-W. Tsai, and J.-S. Pan, "Cat swarm optimization," in *PRICAI: Trends in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 4099. 2006, pp. 854–858.
[96] I. Fister, Jr., X.-S. Yang, J. Brest, and D. Fister, "A brief review of nature-inspired algorithms for optimization," *Elektrotehn. Vestnik/Electrotechn. Rev.*, vol. 80, no. 3, pp. 116–122, 2013.
[97] H. Jiang, S. Zhang, Z. Ren, X. Lai, and Y. Piao, "Approximate muscle guided beam search for three-index assignment problem," in *Advances in Swarm Intelligence*. Berlin, Grmany: Springer, 2014, pp. 44–52.
[98] A. H. Kashan, "League championship algorithm: A new algorithm for numerical function optimization," in *Proc. Int. Conf. Soft Comput. Pattern Recognit. (SOCPAR)*, Dec. 2009, pp. 43–48.
[99] S. M. Abdulhamid, M. S. A. Latiff, S. H. H. Madni, and O. Oluwafemi, "A survey of league championship algorithm: Prospects and challenges," *Indian J. Sci. Technol.*, vol. 8, no. 3, pp. 101–110, 2015.
[100] S. M. Abdulhamid, M. S. A. Latiff, G. Abdul-Salaam, and S. H. H. Madni, "Secure scientific applications scheduling technique for cloud computing environment using global league championship algorithm," *PLoS ONE*, vol. 11, no. 7, p. e0158102, 2016.
[101] H. Mo, L. Liu, and M. Geng, "A magnetotactic bacteria algorithm based on power spectrum for optimization," in *Advances in Swarm Intelligence*. Springer, 2014, pp. 115–125.
[102] M.-Y. Cheng and D. Prayogo, "Symbiotic organisms search: A new metaheuristic optimization algorithm," *Comput. Struct.*, vol. 139, pp. 98–112, Jul. 2014.
[103] M. Abdullahi, M. A. Ngadi, and S. M. Abdulhamid, "Symbiotic organism search optimization based task scheduling in cloud computing environment," *Future Generat. Comput. Syst.*, vol. 56, pp. 640–650, Mar. 2016.
[104] X.-S. Yang, S. Deb, and S. Fong, "Accelerated particle swarm optimization and support vector machine for business optimization and applications," in *Networked Digital Technologies*. Berlin, Germany: Springer, 2011, pp. 53–66.
[105] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo, and W.-C. Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 397–408, Feb. 2007.

[106] S.-H. Min, J. Lee, and I. Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert Syst. Appl.*, vol. 31, no. 3, pp. 652–660, Oct. 2006.
[107] X.-S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 169–174, Jan. 2014.
[108] A. Abraham, E. Corchado, and J. M. Corchado, "Hybrid learning machines," *Neurocomputing*, vol. 72, nos. 13–15, pp. 2729–2730, Aug. 2009.
[109] H. Chiroma, N. L. M. Shuib, S. A. Muaz, A. I. Abubakar, L. B. Ila, and J. Z. Maitama, "A review of the applications of bio-inspired flower pollination algorithm," *Procedia Comput. Sci.*, vol. 62, no. 1, pp. 435–441, Jun. 2015.
[110] S. H. H. Madni, M. S. A. Latiff, and Y. Coulibaly, "An appraisal of meta-heuristic resource allocation techniques for IaaS cloud," *Indian J. Sci. Technol.*, vol. 9, no. 4, pp. 1–14, 2016.

SHAFI'I MUHAMMAD ABDULHAMID (M'13) received the B.Tech. degree in mathematics/computer science from the Federal University of Technology Minna, Nigeria, the M.Sc. degree in computer science from Bayero University Kano, Nigeria, and the Ph.D. degree in computer science from Universiti Teknologi Malaysia. He is currently a Lecturer with the Department of Cyber Security Science, Federal University of Technology Minna, Nigeria. He has authored many academic papers in reputable journals internationally. His current research interests are in cyber security, cloud computing, soft computing, and big data. He is a member of the International Association of Computer Science and Information Technology, the Computer Professionals of Nigeria, the International Association of Engineers, the Internet Society, the Cyber Security Experts Association of Nigeria, and the Nigerian Computer Society.

MUHAMMAD SHAFIE ABD LATIFF received the Ph.D. degree from Bradford University, U.K., in 2000. He is currently an Associate Professor and a member of the Pervasive Computing Research Group at the Faculty of Computing, Universiti Teknologi Malaysia. His research interests are in computer networks with focus generally on routing protocol, specifically for optical burst switching, grid and cloud computing, wireless sensor networks. He also involved in medical dataset visualization and analysis techniques using high performance computing. He has supervised eight Ph.D. students in related fields. He currently leads the research project on Cloud Computing Systems.

HARUNA CHIROMA (M'14) received the B.Tech. degree in computer science from Abubakar Tafawa Balewa University, Nigeria, the M.Sc. degree in computer science from Baero University Kano, and the Ph.D. degree in artificial intelligence from the University of Malaya, Malaysia. He is currently a Faculty Member with the Federal College of Education (Technical), Gombe, Nigeria. He has published articles relevance to his research interest in international referred journals, edited books, conference proceedings, and local journals. His main research interest includes metaheuristic algorithms in energy modeling, decision support systems, data mining, machine learning, soft computing, human computer interaction, social media in education, computer communications, software engineering, and information security. He is a member of the ACM, the NCS, the INNS, and the IAENG. He is currently serving on the Technical Program Committee of several international conferences.

OLUWAFEMI OSHO received the B.Tech. degree in mathematics/computer science and the M.Tech. degree in mathematics. He served as the Head of the IT Department of one of the leading mortgage banks in Nigeria. He is currently a Lecturer with the Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria. His current research interests include cybersecurity, mobile security, and security analysis. He is a Certified Ethical Hacker, a member of the Cyber Security Experts Association of Nigeria, and a host of other professional associations.

GADDAFI ABDUL-SALAAM received the B.Sc. degree in computer engineering from the Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in 2005, and the M.Sc. degree in advanced ICT studies from the Institute for Advanced ICT Studies, Ghana, in 2009. He is currently pursuing the Ph.D. degree with the Pervasive Computing Research Group Laboratory, Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia. He was the Head of the ICT Centre, Kwame Nkrumah University of Science and Technology (KNUST), Ghana, from 2010 to 2013, where he is currently an Academic Staff. His research interests include wireless sensor network, hybrid WSN, and energy efficient data collection protocols in WSN.

ADAMU I. ABUBAKAR (M'09) received the B.Sc. degree in computer science from Bayero University Kano Nigeria in 2002, and the M.Sc. and Ph.D. degrees in computer science from the International Islamic University of Malaysia in 2006 and 2010, respectively. He is currently an Assistant Professor with the International Islamic University of Malaysia, Kuala Lumpur, Malaysia. He has authored over 80 articles relevant in international journals, proceedings, and book chapters. His current research interests include information science, information security, intelligent systems, human computer interaction, and global warming. He is currently a member of the Technical Program Committee of several international conferences. He is a member of the ACM. He received several medals in research exhibitions. He served in various capacities in international conferences.

TUTUT HERAWAN received the Ph.D. degree in information technology from Universiti Tun Hussein Onn Malaysia in 2010. He is currently a Principal Researcher with the AMCS Research Center, Indonesia. He has over 12 years experiences as academic and also successfully supervised five Ph.D. students. He currently supervises 15 master's and Ph.D. students and has examined master's and Ph.D. Theses. His research area includes soft computing, data mining, and information retrieval. He is the Executive Editor of the *Malaysian Journal of Computer Science* (ISI JCR with IF 0.405). He has also guest edited many special issues in many reputable international journals. He has edited five many books and published extensive articles in various book chapters, international journals and conference proceedings. He is an active Reviewer of various journals. He delivered many keynote addresses, invited workshop and seminars and has been actively served as the Chair, the Co-Chair, a Program Committee Member, and a Co-Organizer of numerous international conferences/workshops.