**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.

**Published in:**[IEEE Access](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6287639) ( Volume: 5 )

**Page(s):**15650 - 15666

**Date of Publication:** 10 February 2017

**Electronic ISSN:** 2169-3536

**INSPEC Accession Number:**17124287

**DOI:**[10.1109/ACCESS.2017.2666785](https://doi.org/10.1109/ACCESS.2017.2666785)

**Publisher:**IEEE

**Sponsored by:**[IEEE](http://www.ieee.org/index.html) 

**Funding Agency:**

Taxonomy of mobile SMS spamming techniques.

Hide Full Abstract

**SECTION 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]–[12][13]. However, there are next to no literature reviews that summarize the currently available methods, challenges 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.

**SECTION 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.

**SECTION 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 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 1.**

Online academic database used for searching the literature.

[View All](https://ieeexplore.ieee.org/document/7851079/all-figures)

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.

**FIGURE 2.**

Search procedure and study selection diagram.

[View All](https://ieeexplore.ieee.org/document/7851079/all-figures)

**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:

1. True Positive (TP) - the amount of samples that are properly classified;
2. True Negative (TN) - the amount of samples that are properly rejected from the class.
3. False Positive (FP) - the amount of samples that are wrongly rejected from the class;
4. False Negative (FN) - represents the amount of samples wrongly classified to the correct class;
5. 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=TP+TNTP+TN+FP+FN(1)

View SourceRight-click on figure for MathML and additional features.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=TPTP+FP(2)

View SourceRight-click on figure for MathML and additional features.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=TPTP+FN(3)

View SourceRight-click on figure for MathML and additional features.The F1 score is a valuable and efficient metric for unbalanced data [26].

F1score=2Precision×RecallPrecision+Recall(4)

View SourceRight-click on figure for MathML and additional features.Spam Precision constitutes the quotient of correct to incorrect classifications among filtered spam messages [24].

SpamPrecision(SP)=Amount of SPAM messageTotal Amount of SPAM message(5)

View SourceRight-click on figure for MathML and additional features.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=Amount of correct Non−SPAM messageTotal Amount of Non−SPAM message(6)

View SourceRight-click on figure for MathML and additional features.The FAR is to compute the false classifying rate of the non-spam messages [27].

FAR=TP∩FNTP×100%(7)

View SourceRight-click on figure for MathML and additional features.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(TP×TN)−(FP×FN)(TP+FP)×(TP+FN)×(TN+FP)×(TN+FN)−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−√(8)

View SourceRight-click on figure for MathML and additional features.

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.

**FIGURE 3.**

Performance Metric.

[View All](https://ieeexplore.ieee.org/document/7851079/all-figures)

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.

**SECTION 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 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.

**FIGURE 4.**

Architecture of the SMS spam transmission line.

[View All](https://ieeexplore.ieee.org/document/7851079/all-figures)

**FIGURE 5.**

Taxonomy of mobile SMS spamming techniques.

[View All](https://ieeexplore.ieee.org/document/7851079/all-figures)

**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.

**TABLE 1**Summary of the Major Research Findings on Mobile SMS Spam

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 ≥ 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.

**SECTION 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.

**TABLE 2**Spam SMS Research Datasets

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 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.

**SECTION 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.

**SECTION 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]–[106][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]–[110][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.

**SECTION 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.