DETECTION OF SUSPICIOUS TERRORIST EMAILS USING TEXT CLASSIFICATION: A REVIEW

Ghulam Mujtaba[1, 4*], Liyana Shuib[1*], Ram Gopal Raj[2], Roshan Gunalan[3]

[1]Department of Information Systems
Faculty of Computer Science and Information Technology
University of Malaya
50603 Kuala Lumpur
Malaysia
[2]Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
50603 Kuala Lumpur
Malaysia
[3]Department of Orthopaedic Surgery
Faculty of Medicine
University of Malaya
50603 Kuala Lumpur
Malaysia
[4]Department of Computer Science
Sukkur IBA University
65200 Sukkur
Pakistan

Corresponding Authors: Ghulam Mujtaba and Liyana Shuib
Email: mujtaba@siswa.um.edu.my, liyanashuib@um.edu.my, ramdr@um.edu.my, roshan@um.edu.my

DOI: https://doi.org/10.22452/mjcs.vol31no4.3

ABSTRACT

This paper provides a comprehensive review and analysis of the detection of suspicious terrorist electronic mails (e-mails) using various phases and methods of text classification. We explored, analyzed, and compared different datasets, features, feature extraction techniques, feature representation techniques, feature selection schemes, text classification techniques, and performance measurement metrics used in the detection of suspicious terrorist e-mails. 30 articles were retrieved from 6 well-known academic databases after rigorous selection. From the study, we found that researchers often generate their own e-mails dataset since there is no public dataset is available in the research area of detecting suspicious terrorist e-mails. In most of the studies, researchers used content and context-based features to detect terrorist e-mails. Our findings also show that the most commonly used feature extraction techniques are the bag of words and n-gram, the most typically applied feature representation schemes are binary representation and term frequency, the most usually adopted feature selection method is information gain,, the most common and most accurate text classification algorithms are naïve bayes, decision trees, and support vector machines, and the widely employed performance measurement metrics are accuracy, precision, and recall. Open research challenges and research issues that involve significant research efforts are also summarized in this review for future researchers in the area of suspicious terrorist e-mail detection using text classification techniques where the critical analysis presented in this paper also provides valuable insights to guide these researchers. Finally, the indicated issues and challenges presented in this paper can be used as future research directions in this area.

Keywords: Text Classification, Cyber Terrorism, Suspicious Emails, Text Representation, Feature Selection, Performance Measures

1.0 INTRODUCTION

The Internet is currently popular for satisfying individuals with numerous services related to various areas. Almost everything is accessible over the Internet with the progressive advancement in technology. The Internet also provides different services, such as electronic mails (e-mails), File Transfer Protocol (FTP), World Wide Web (WWW), chat rooms, social networking sites, discussion forums, and blogs. E-mail is the best platform to connect with family, business contacts, and friends because it is fast, simple, and asynchronous in nature. Such platform does not incur the stamp cost, and one does not need to wait for a long time to receive a reply. In business organizations, government institutions, and the corporate world, using e-mail is a simple and fast approach to communicate with colleagues and employers/employee. It is being used to prevent from squandering profitable time when urgent or important assignment can be delivered via e-mails. According to the previous study [1], approximately 196.3 billion e-mails are sent and received per day. Therefore, observing all messages and characterizing them into categories are practically impossible for human specialists.

Terrorism is a genuine threat to worldwide peace and security, and no country can consider itself as resistant from the threats of terrorism. After the 9/11 terrorist attacks, national security concern has significantly increased and intelligent agencies are vigorously gathering local and overseas intelligence reports to prevent such attacks in future. Moreover, the intelligent agencies are looking for some intelligent tools to automatically detect suspicious activities with the help of machine learning tools and techniques. Recently, machine learning has been considerably employed in various application areas including, banking [2], health bioinformatics [3-6], and e-mail classification [7-9]. Many terrorists use e-mail as a means to communicate [10, 11]. After the 9/11 incident, many researchers have developed new approaches [12-14] and software tools [15, 16] to analyze and mine terrorism-related information to prevent terrorism-related activities. Such research has provided beneficial contributions to the succeeding generation of counterterrorism tools.

Numerous researchers are working in the field of automatic e-mail classification, such as to categorize an e-mail as ham or spam [7, 8], a phishing e-mail [17-19], or an e-mail in a specific directory [20-22]. A few researchers have also been actively working in the area of classifying threatening terrorist e-mails since it was found that the terrorists used the e-mails to exchange information among each other in such tragic terrorism incidents as the 9/11 attacks [23]. Researchers have published review articles on spam and phishing e-mail classification in well-known academic journals from the perspective of automatic text classification to summarize e-mail classification works. For example, Blanzieri *et al.* [24] presented a review of existing machine learning approaches for spam e-mail filtering. In addition, a survey of publicly available datasets for spam e-mail classification, image and text-based features, evaluation metrics, and spam-filtering techniques was also presented. Guzella *et al.* [25] surveyed various datasets, feature selection algorithms, and classification techniques to detect spam e-mails. Besides than that, they also surveyed the existing literature on spam e-mail classification through images-based features. Almomani *et al.* [26] presented a comprehensive survey of phishing e-mail-filtering techniques and indicated the types of phishing attacks, phishing e-mail classification, and evaluation methods. The existing reviews have reported the works either on spam or phishing e-mail detection. However, to the best of our knowledge, no studies have reviewed the work on detection of suspicious terrorist e-mails in well-known databases, such as the Web of Science, Scopus, IEEE Xplore, ACM, Google Scholar, Science Direct, and Springer. Hence, a comprehensive review is needed to investigate the current state-of-the-art research work in the field of detection of suspicious terrorist e-mails.

In this survey, our aim was to present a comprehensive review and analysis of the detection of suspicious terrorist e-mails on articles published from 1998 to 2015 by exploiting and critically reviewing terrorist e-mail classification research based on the following aspects: publicly available datasets, features, feature extraction techniques, feature representation techniques, feature selection techniques, text classification techniques, and performance measures. After rigorous selection, 30 academic articles [10, 11, 27-54] were selected from 6 well-known academic databases to perform the review. This review was conducted to help researchers working in the area of detecting suspicious terrorist e-mails by answering the 8 following research questions:

Research Question 1: Are there any e-mail datasets available in the area of detecting suspicious terrorist activities using text classification? If yes, which is the best one to use?
Research Question 2: What features are mostly used by researchers to detect suspicious terrorist activities and why?

Research Question 3: Which feature extraction technique is the most commonly used in the area of detecting suspicious terrorist activities using text classification and why?
Research Question 4: Which feature representation technique is the most commonly applied in the area of detecting suspicious terrorist activities using text classification and why?
Research Question 5: Which feature selection technique is the most commonly adopted in the area of detecting suspicious terrorist activities using text classification and why?
Research Question 6: Which text classification algorithm is the most accurate at classifying in the area of detecting suspicious terrorist activities using text classification and why?
Research Question 7: Which performance measures are the most widely employed in the area of detecting suspicious terrorist activities using text classification?
Research Question 8: What are the future research directions and challenges in the area of detecting suspicious terrorist activities using text classification?

The remainder of this paper is divided into 7 sections. Section 2 presents the taxonomy related to cybercrime which emphasize on the differences among cybercrime, cyberattack, cyberthreats, cyberterrorism, cyberbullying, and cyber harassment. Section 3 discusses text classification, its phases, and techniques. Section 4 provides the research methodology for this review. Section 5 indicates a critical review on the selected articles based on five different aspects, namely, (1) datasets, (2) feature set, (3) feature extraction, representation, and selection techniques, (4) text classification techniques, and (5) performance measures. Section 6 focuses on the research challenges and future research directions in detecting suspicious terrorist e-mails classification domain. Finally, Section 7 concludes this paper by summarizing our findings on the review.

2.0 CYBERCRIME AND ITS TAXONOMY

Many buzzwords are used interchangeably with the term cybercrime, such as cyberterrorism, cyberharassment, cyberattack, cyberbullying, and cybergrooming. However, all these terms have different interpretations. Cybercrime is an umbrella term for all these terms. It is any unlawful activity, which makes through a computer the fundamental means of commission [55, 56]. Cybercriminals can be categorized into two groups. The first group comprises the people who are expert in computer software, hardware, and networks. These people utilize their technical expertise to commit a cybercrime, such as to illegally access any information system, computer network, network infrastructure, or database, to hack credit card information. The second group consists of the people who may not expert users of computer but utilize computers and the Internet to harass other users, such as to bully someone using social media, to threaten national security using Internet services, and to send threatening e-mails. Given these characteristics, cybercrime is classified into two broad categories, namely, (i) *technology-based cybercrimes* and (ii) *content-based cyber* [57], as shown in Fig. 1. The following sections present the details of these two categories of cybercrime.

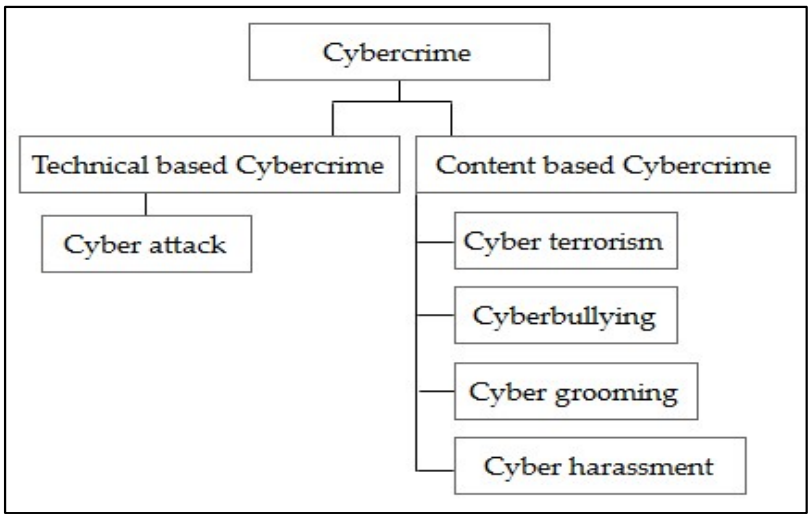

Fig. 1: Classification of Cybercrime

2.1 Technology-based Cybercrime

This type of cybercrime is performed by technical experts and involves hacking, injection of malicious codes, and incidents of espionage [57]. Cyberattack, a type of technology-based cybercrime, is any sort of offensive act utilized by people or entire associations that targets information systems, critical infrastructure, computer networks, and/or personal computers via different methods for pernicious acts normally starting from an unknown source that either takes, changes, or decimates a predefined focus by hacking into a susceptible system [58]. Cyberattacks can range from installing spyware on a computer to attempts to demolish the infrastructure of entire states.

2.2 Content-based Cybercrime

This cybercrime type is related to threatening national security by terrorist organizations, child pornography, and sexual harassment [57]. Content-based cybercrimes include cyberterrorism, cyberbullying, cyberharassment, and cybergrooming. Cyberterrorism is the use of cyberspace to commit terrorist acts. In other words, cyberterrorism is the utilization of information and communication technology by terrorist groups to cause fear and/or physical harm to the population [59]. Numerous terrorist associations utilize e-mail as essential source of communication for terrorism because of its worldwide reach at no cost and it is perfect for group communication [10]. Terrorists may prefer e-mail correspondence for illicit exercises in light of the fact that it permits a single person to stay mysterious because a fictitious name can be utilized when subscribing for a new e-mail account. The mix of such peculiarities makes e-mail an advantageous medium for correspondence that can be utilized to direct illicit exercises by terrorist groups.

Cyberbullying is a type of bullying that is usually taken place over the Internet or using another technological gadget, such as mobile phone to exchange some dangerous and hostile messages by e-mail, social media, or text messages to threaten the victim [60]. Cyberbullying usually happens amongst children and youth. Bullying, whether verbal, physical, or cyber, often depresses the victims, thereby decreasing their confidence and making them stop from going to school or socializing with new people. A worse case occurs when the victim commits suicide to stop being bullied [61]. Cyberharassment is the same as cyberbullying, with the only difference being that cyberharassment is usually done among adults [62]. Cybergrooming includes an individual attempting to set up a sexually damaging circumstance through utilizing digital advances, such as the Internet and cellular telephones [63]. However, this review focuses only on cyberterrorism.

3.0 SUPERVISED TEXT CLASSIFICATION

Supervised text classification utilizes the labelled training set of text to learn and construct the text classifier and then automatically classifies the unlabelled test set of text using the constructed classifier [64]. These days, most text categorization is performed by human observers. Every day, hundreds of files, email messages, and web addresses are stored in the folders. Human categorization is very slow and expensive and its application is a hurdle for large or rapidly changing collections. Moreover, inconsistency in category assignment and adapting changing category structures are also some of the limitations of human categorization. Hence, a growing interest has been witnessed in developing technologies for text categorization. There are several approaches to classify unstructured text documents such as, supervised text classification, unsupervised text classification, and semi-supervised text classification. Of all these approaches, supervised text classification is the most widely used [9]. Fig. 2 shows the general supervised text classification process. As shown here, this process mainly comprised of 7 main steps. These seven steps are discussed briefly in the subsequent sections.

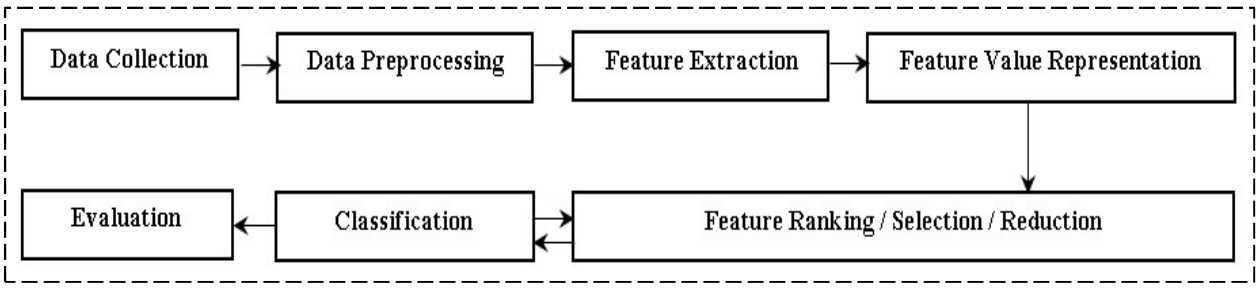

Fig. 2: Process of supervised text classification

3.1 Data Collection and Data Pre-Processing

The first step in text classification process is data collection. In this step, we collect the dataset related to area under consideration. For instance, in the case of detecting suspicious terrorist emails, the email dataset is collected that has two types of emails, namely, suspicious terrorist emails and non-suspicious or legitimate emails. Once the dataset is collected, the next step is to pre-process the collected dataset. This step involves the preparation of data in a format that is suitable for a classifier to classify. In this step, the training set and testing set are prepared from collected email dataset. In training set, each email is labelled as suspicious or legitimate. This training set is required in classifier construction step (discussed in section 3.4). In testing set, the emails are not labelled with any category and this set is used to evaluate the effectiveness of constructed classifier (as discussed in section 3.5). Furthermore, in pre-processing step, each email is tokenized into tokens or words. From these words, all stopwords such as *a, an, of,* and *the* are removed. Besides than tokenization and stopwords removal, stemming or lemmatization is also performed to convert the words to their root forms, such as *retrieving* to *retrieve*.

3.2 Feature Extraction

The third step in supervised text classification process is the feature extraction. Here, the useful features are discovered from the collected email dataset. These features may be the content based features, context based features, header features, etc. These features are further discussed in section 5.2. The most commonly used feature extraction method is *n*-gram where *n* represents the number of word features. If the value of *n* is 1, it is called unigram, if the value of *n* is 2, it is called bigram, if the value of *n* is 3, it is called trigram and so on. Another feature extraction method is the Bag of Words (BoW), which is also known as the Vector Space Model (VSM), one of the basic and most common methods of text representation. The BoW or VSM has been a standard model for representing documents in information retrieval for almost three decades [65].

3.3 Feature Representation

Feature representation, or term-weighting is responsible for representing the extracted features into numeric form [66, 67]. There are three main feature value representation techniques namely, Binary Representation (BR), Term Frequency (TF), and Term Frequency with Inverse Document Frequency (TFiDF) [67]. In BR, the feature value can be either '0' or '1', where '1' symbolizes the occurrence of a feature in the document and '0' represents its non-occurrence [66]. In TF, the feature value symbolizes the number of times that particular feature appeared in a document [68]. However, if that particular feature is very frequent in all documents with relatively the same frequency, then that feature is not a result-oriented feature [69]. Therefore, to address this issue, TFiDF was introduced. The crux of TFiDF is that the feature *f* can be a result-oriented feature if *f* repeatedly appears in the documents belonging to same class. However, if same feature *f* repeatedly appears across different classes, then *f* is not a result-oriented feature [68].

3.4 Feature Selection

The complete feature set often contains extraneous features, which causes some of the limitations in mining or classification tasks. These limitations reduce the accuracy of the text classification algorithm, slow down the classification process, produce problems in storing and retrieving the information and make it difficult to infer the classification results. To overcome these limitations, feature selection techniques are often used to extract the most result-oriented subset of features from a feature set [70]. The feature selection methods can be further categorized into the filter, wrapper, and embedded Methods [70]. Filter methods employ the statistical measures to assign a score to each feature. Based on these scores, the features are either selected or eliminated from the feature set. The examples of such methods include, Information Gain (IG), and Chi-Squared (χ^2). Wrapper methods apply searching techniques to select the best features where various combinations of features are prepared and evaluated with other combinations. The example of wrapper methods includes recursive feature elimination technique. Finally, the embedded methods learn the importance of features during the construction of classification model. Regularization or penalization methods including LASSO, Elastic Net, and Ridge Regression are examples of embedded feature selection methods. Of these three feature selection methods, the most commonly used feature selection methods are filter methods [71].

3.5 Classifier Construction

In this step, a machine learning algorithm is employed on training set to construct a decision or classification model. The constructed model has the capability to classify new incoming and unlabeled emails into a specific category such as suspicious or legitimate. Numerous machine learning algorithms (namely, Naive Bayes, Decision Tree, Random Forest, Support Vector Machines, Artificial Neural Network, and Genetic Algorithm) have been employed to classify text documents. The brief introduction of some of these algorithms are presented in subsequent sub-sections. Some good overviews of the different machine learning algorithms can also be found in [72] and [73].

3.5.1 Naive Bayes

Naïve Bayes (NB) is a simple probabilistic classifier. The NB method is responsible for making a probabilistic model of data within each class. It is a statistical analysis algorithm that works on numeric data [74]. It has a capability to predict the parameters essential for classification using a small amount of training data. It is a simple and fast classification algorithm. It works well with text representations, such as BoW. The detail discussion on Naïve Bayes classifier can be found in [75].

3.5.2 Decision Tree

Decision tree (DT) is the widely used algorithm in text classification domain. It denotes the classification rules in the tree-shaped diagram that is used to determine every possible outcome of a decision. The decision tree comprised of decision node, leaf node, edge and path [76]. There are variety of DT classifiers such as ID3, C4.5, C5.0, etc. One of the key drawback of DT classifier is that they are prone to overfitting. This is because, the trees if grown deeper, are able to fit all kinds of variations in the data, including the noise. In addition, a minor change in data can drastically affect the

performance of a DT. Therefore, to enhance the performance of individual DTs, ensemble methods such as random forest (discussed in 3.5.3) were proposed, in which many trees are constructed and trained by splitting the training set and final predictions are aggregated across the trees. The details of DT classifiers are discussed in [77].

3.5.3 Random Forest

Random forest (RF) was developed by Leo Breiman in 1999 [78]. It divides the training set into different sub-training sets. From each sub-training set a DT classifier is constructed. Each input vector will be classified by all the decision trees in a forest and the forest chooses the classifier having the most votes [79]. In the literature, RF shows better performance than single DT [80, 81]. In addition, it also resolves the overfitting problem.

3.5.4 Support Vector Machines

Support vector machine (SVM) algorithm employs the statistical learning theory to construct a classification model. SVM learns the classification rules from the training data. In SVM, an optimum hyperplane separates the two classes (such as ham and spam) by minimizing the distance between the classes. Such hyperplanes are termed as support vectors. The right side of the hyperplane contains the *ham* class data points and the left one contains the *spam* class data points. This separation of classes is performed with the help of training examples [82]. SVM has been proven successful in several applications including, image classification, handwriting recognition, and bioinformatics [81].

3.5.5 Artificial Neural Network

Artificial neural network (ANN) consists of input, hidden, and output layers. The input layer and hidden layer comprise many nodes, while the output layer has only one node. The nodes in a neural network contain an activation function. With the help of the input layer, patterns are provided to the neural network, which interacts with hidden layers. The actual processing is done in hidden layers by allocating random weights to edges. The hidden layer is further connected to an output layer where the final answer is computed. Mostly, ANNs use learning rules for modifying the weights of the connections as per input patterns. One of the popular learning rules is Delta rule, which is often utilized by Backpropagation Neural Networks (BPNNs). Delta rule is a supervised learning rule that occurs with each cycle and the initial pattern is determined by a random guess [83].

3.5.6 Genetic Algorithm

Genetic algorithm (GA) was first proposed by John Holland in the early 1970s [84]. The idea behind the implementation of GA is to use the process like natural evolution to resolve the issue of optimization. In GA, a gene is comprised of a string of bits. Generally, the preliminary genes population is generated randomly. The bit string length depends on the nature of problem to be resolved. From the initial population, a subset of genes is extracted based upon some quality fitness measurement. After the selection, the next step is mating and crossover of which there are different types. Random mating with crossover is one of the easy type, where, the genes in selected population are randomly selected and mated in pairs. Usually, a point for crossover is chosen for each selected pair. After crossover, the information is swapped between two pairs. In final mutation step, each gene bit has a definite Probability P to get inverted. The more details on GA can be found in [85].

3.6 Classifier Evaluation

In this step, constructed classifier predicts the class of unlabelled text documents such as (spam or ham emails) using test set. The accuracy of the classifier can be evaluated by calculating the quantity of correctly predicted class cases (true positives), the quantity of accurately predicted class cases that do not belong to the class (true negatives), and the class cases that were either inaccurately predicted to the specific class (false positives) or that were not predicted as the class samples (false negatives). These four numbers constitute a confusion matrix as in Table 1 for the case of binary classification. Various performance measures are used to evaluate the performance of constructed classifier. Some common performance measures in text categorization are discussed briefly below. The more details of performance metrics can be found in [86].

Table 1: Confusion Matrix

		Actual Class	
		Yes	**No**
Predicted class	**Yes**	TP	FN
	No	FP	TN

3.6.1 Precision

Precision is also known as positive predicted value. It is the proportion of predictive positives which are actual positive (as shown in Equation 1).

$$Precision = \frac{TP}{(TP + FP)}$$

Equation 1: Formula for calculating precision

3.6.2 Recall

It is the proportion of actual positives which are predicted positive (as shown in Equation 2).

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

Equation 2: Formula for calculating recall

3.6.3 F-Measure

It is the harmonic mean of recall and precision (as shown in Equation 3). The standard F-measure is F1, which gives equal importance to precision and recall.

$$F - measure = \frac{2 \times (precision \times recall)}{(precision + recall)}$$

Equation 3: Formula for calculating f-measure

3.6.4 Accuracy

Classification accuracy is the number of samples correctly classified (true positive and true negative) and is evaluated using Equation 4.

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN}$$

Equation 4: Formula for calculating accuracy

4.0 RESEARCH METHODOLOGY

The flow chart of the research methodology that we followed in this survey is shown in Fig. 3. First, the survey objectives were identified. Search keywords were then formulated to search and retrieve the literature from 6 academic databases, namely, Web of Science, IEEE Xplore, Science Direct, Scopus, Google Scholar, and Springer (more details

in Subsection 4.1). After selecting the appropriate articles, various bibliometric information of selected articles was presented (more details in Subsection 4.1). Afterwards, a critical review was performed on all selected articles via five different aspects, namely, (1) dataset analysis, (2) feature set analysis, (3) feature extraction, representation, and selection technique analysis, (4) text classification technique analysis, and (5) performance measure analysis, as indicated in Section 5. Finally, the current trends, issues, challenges, and future research directions in the area of terrorist e-mail classification for future researchers are discussed in Section 6.

4.1 Search and Selection of Articles

The academic literature for this review was searched from 6 aforementioned databases. The articles published from 1998 to 2015 were considered. Various relevant keywords were formulated to search the literature on "automatic detection of terrorist e-mails" from the selected databases. These keywords were selected from the e-mail classification and terrorist e-mail classification domain to search the literature from the selected databases. Initially, the keywords "Email Classification" or "E-mail Classification" or "Email Filtering" or "E-mail Filtering" in the title field were used for searching literature from all the selected databases. Afterwards, these keywords were combined with some other keywords, such as "Threatening email," "Suspicious email," or "Terrorist email" to narrow down the search with the use of "AND" Boolean operator. Finally, article-type filter and language filter were used to further narrow down the search. Table 2 shows these three search queries.

Table 2: List of 3 search queries for selecting the relevant papers; No. =query number, Query= search string.

No.	Query
Q1	TITLE: (E-mail or Email or Email Classification or E-mail Classification or Email Filtering or E-mail Filtering or Detecting Email or E-mail detecting or E-mail Detection or Email Detection)
Q2	TITLE: (E-mail or Email or Email Classification or E-mail Classification or Email Filtering or E-mail Filtering or Detecting Email or E-mail detecting or E-mail Detection or Email Detection) AND TITLE: (Suspicious or Terrorist or Terrorism or Threatening or Illegitimate or Cyber Terrorism)
Q3	TITLE: (E-mail or Email or Email Classification or E-mail Classification or Email Filtering or E-mail Filtering or Detecting Email or E-mail detecting or E-mail Detection or Email Detection) AND TITLE: (Suspicious or Terrorist or Terrorism or Threatening or Illegitimate or Cyber Terrorism) Refined by: DOCUMENT TYPES: (ARTICLE OR PROCEEDINGS PAPER OR BOOK CHAPTER) Refined by: Language: (English)

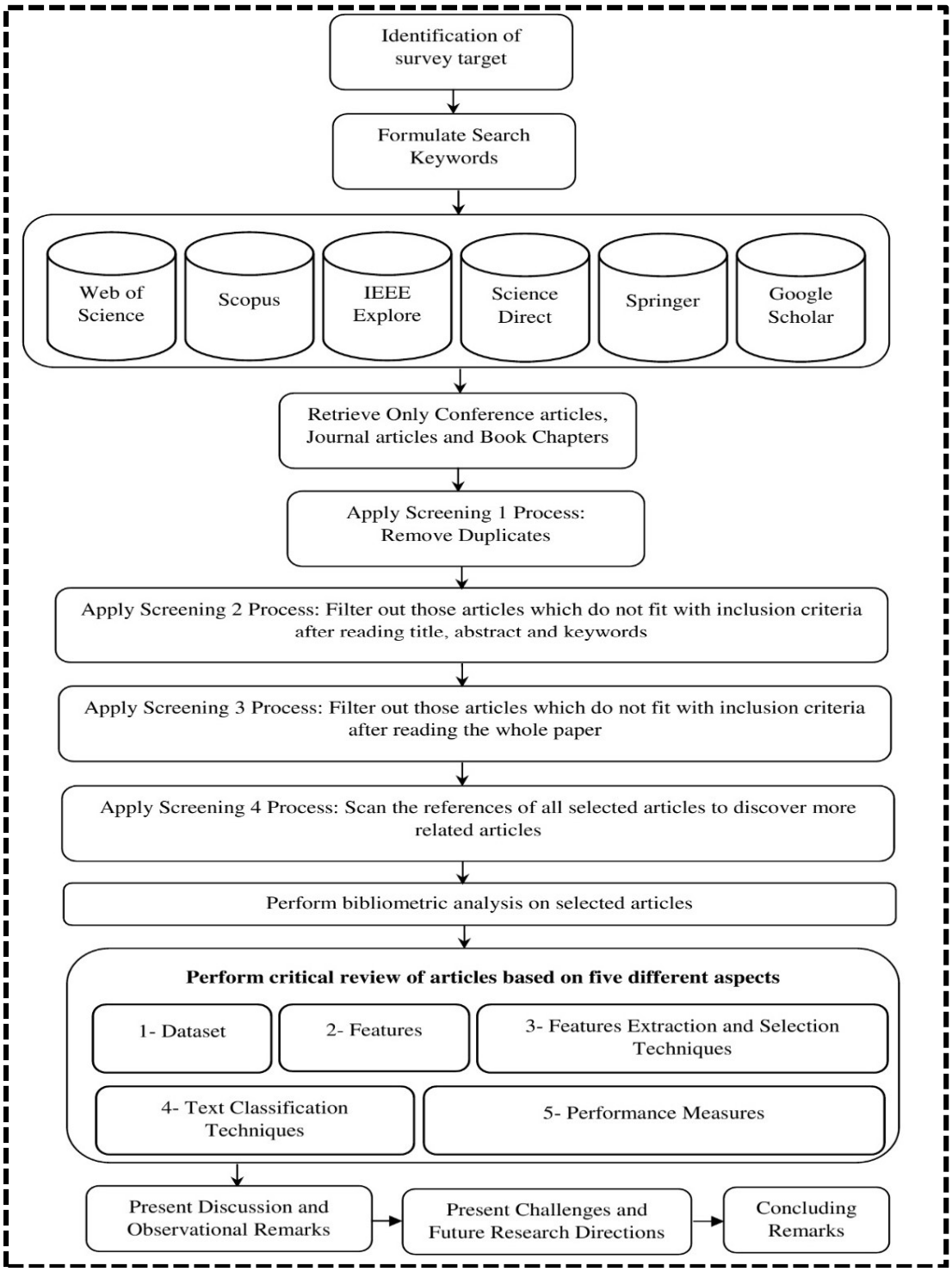

Fig. 3: Flow chart of research methodology

All these 3 queries (Table 2) were applied on all 6 selected databases to retrieve the academic articles. Table 3 shows the results obtained from these 3 queries when applied on all 6 selected databases. When query 1 (Q1) is applied on the Web of Science, 3544 articles were retrieved. When this result was further narrowed by applying query 2 (Q2), 16 articles were returned. When the third filter (Q3) was applied, 5 articles from the Web of Science database were retrieved.

Likewise, after applying all these 3 queries on Scopus, IEEE Xplore, Science Direct, Springer, and Google Scholar, 9, 6, 5, 4, and 17 articles, respectively, were retrieved. Hence, a total of 46 articles were retrieved after applying all the queries on all databases. The duplicate articles were removed from the list of collected articles. We called this process "screening 1." In screening 1, 8 articles were removed because of their availability in more than one database. The title, abstract, and keywords of all 38 retrieved articles were read to see whether these articles fit our inclusion criteria. We called this process "screening 2." The list of our inclusion criteria is shown in Table 4. After this process, 7 articles that did not match with our inclusion criteria were dropped. All remaining 31 articles were read thoroughly to see whether they match our inclusion criteria. We called this process "screening 3." After screening 3, 6 more articles were removed on the basis of our inclusion criteria. 25 articles were retained after screening 3. Finally, all the references of all 25 selected articles were scanned to find any suitable articles that fulfill our inclusion criteria. We called this process "screening 4." This process found 5 new articles for our survey. A total of 30 articles were selected for this survey after the rigorous article selection from the 6 well-known databases, as shown in Table 3.

Table 3: Search and screening results from all six databases

Database	Query	Results	Screening 1	Screening 2	Screening 3	Screening 4
Web of Science	Q1	3544	5	3	3	3
	Q2	16				
	Q3	5				
Scopus	Q1	5046	7	5	4	4
	Q2	11				
	Q3	9				
IEEE Xplore	Q1	153	6	5	5	6
	Q2	6				
	Q3	6				
Science Direct	Q1	47	4	3	3	3
	Q2	7				
	Q3	5				
Springer	Q1	10950	3	3	3	4
	Q2	4				
	Q3	4				
Google Scholar	Q1	35200	13	12	7	10
	Q2	17				
	Q3	17				
Total			**38**	**31**	**25**	**30**

Table 4: List of inclusion criteria

S. No.	Inclusion Criteria
1	Machine Learning is used.
2	Email dataset is used.
3	Article is either conference article or journal article or book chapter.
4	Article must be published in 'English' language.
5	Article must be published in between 1998 to 2015.
6	It should propose new feature set, feature extraction scheme, feature representation scheme, feature selection scheme, propose new classification or clustering technique, or apply existing classification or clustering technique.

Fig. 4 shows the academic database wise distribution of the selected 30 articles. Among the 30 articles, 3 were selected from the Web of Science, 4 from the Scopus, 6 from the IEEE Xplore, 3 from the Science Direct, 4 from the Springer, and 10 from the Google Scholar.

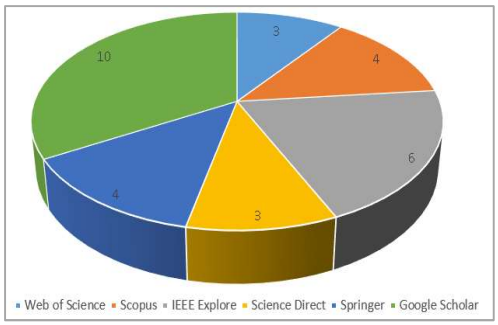

Fig. 4: Academic database wise distribution of selected articles

Fig. 5 shows the distribution of selected studies based on article-type. Among the 30 articles, 16 articles are conference-proceeding articles, 11 are journal articles, and 3 are book chapters. Fig. 6 shows the yearly publication count and yearly citation count of articles. The horizontal axis shows the year, and the vertical axis shows the number of articles published in the year and citation count received in a year. A fluctuating trend in publication rate and citation rate was found in the targeted area. The highest number of articles was published in year 2009, followed by 2005, 2007, and 2010. The highest citation count was found in 2005, followed by 2009, 2010, and 2007. A decreasing trend in publication count and citation count was observed from 2010 to 2014. No any publication on the targeted topic was identified in 2015. Fig. 7 shows the country wise distribution of selected articles. The highest number of articles on the selected topic was published from India, followed by Denmark and the USA and finally by Pakistan.

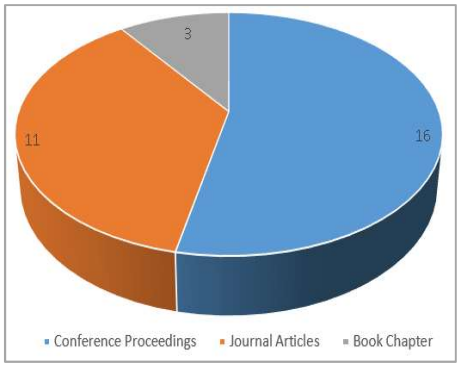

Fig. 5: Article type wise distribution of selected articles

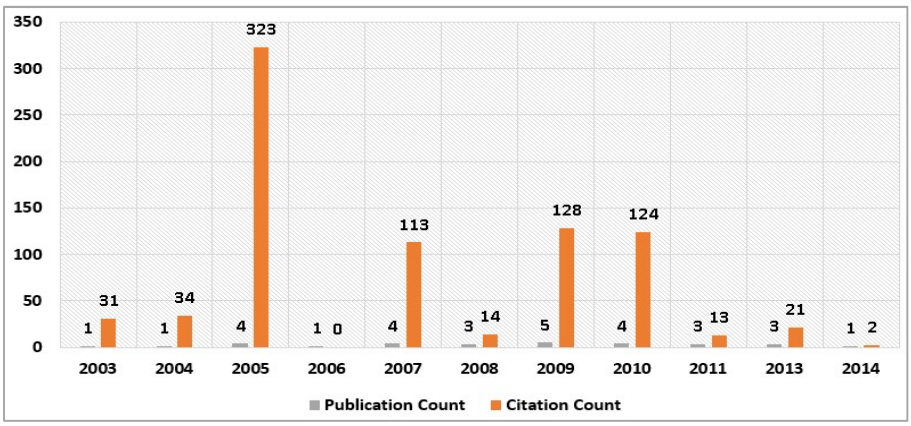

Fig. 6: Year wise publication count and citation count of selected articles

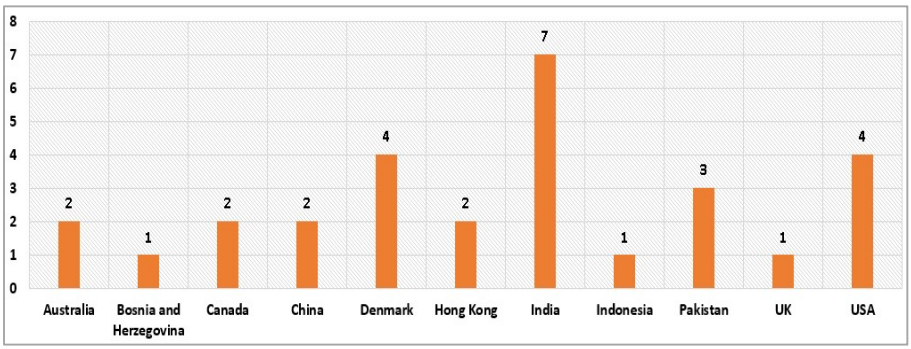

Fig. 7: Country wise distribution of selected articles

In all 30 selected studies, authors have either proposed new feature sets, new feature extraction techniques, new feature representation techniques, new feature selection schemes, or new classification or clustering techniques or applied existing classification or clustering techniques. Hence, we classified all selected studies into 6 major objectives. One study may be categorized into more than one objective. Table 5 shows the list of all these objectives and related studies. For example, Sun [27] proposed content and contextual features for terrorism information extraction. Drozdova *et al.* [44] proposed technology-based features rather than content-based features to predict terrorist activities. They concluded that terrorists most often use low-technology communication for suspicious activities compared with high-technology communications. Sun [27] proposed a template feature extraction method to detect suspicious terrorist activities. Nizamani *et al.* [53] proposed an enhanced feature selection scheme for terrorist e-mail classification. In other study, Nizamani *et al.* [52] proposed a new cluster-based text classification technique to detect suspicious terrorist e-mails. Appavu *et al.* [39] proposed a new text classification technique called "Ad infinitum" to detect suspicious terrorist e-mails. This technique is an extension of the existing DT algorithm. Appavu *et al.* [34] also compared various text classification techniques to identify suspicious e-mails.

Table 5: Typography of 30 selected studies based upon common objectives

S. No.	Objectives	References	Study Count
1	To propose new feature sets	[27, 29, 30, 33, 36, 40, 42, 44, 45, 51-54]	19
2	To propose new feature extraction technique	[27-29, 53, 54]	5
3	To propose feature representation technique	[53, 54]	2
4	To propose feature selection technique	[33, 38, 45, 46, 53, 54]	6
5	To propose novel classification or clustering technique	[11, 32, 38, 39, 43, 49, 50, 52]	8
6	To apply existing text classification techniques	[10, 27, 31, 33-35, 37-39, 41, 47, 48, 54]	13

5.0 ANALYSIS OF TEXT CLASSIFICATION TECHNIQUES

This section reviews the text classification techniques for detecting suspicious terrorist e-mails in the 30 selected studies from 5 different aspects, namely, (1) datasets, (2) feature set, (3) feature extraction, representation, and selection techniques, (4) text classification techniques, and (5) performance measures. Subsection 5.1 presents the review of various datasets used for detecting suspicious terrorist e-mails. Subsection 5.2 reviews various features used to detect suspicious terrorist e-mails. Subsection 5.3 presents a review on feature extraction, feature representation, and feature selection techniques. Subsection 5.4 indicates a review on various classification techniques used to classify suspicious terrorist e-mails. Finally, subsection 5.5 reviews various performance metrics used to evaluate classification model performance for detecting suspicious terrorist e-mails.

5.1 Review and Analysis of Datasets

Table 6 shows the datasets that were utilized in each selected study. The list comprises the name of the datasets, a brief description of the datasets, the frequency counts of studies in which such particular datasets were used, and the references of those studies in which the datasets were used. We provide the availability links of various datasets used in the selected studies as shown in Table 7. Our survey results indicated that no standardized public dataset is available for detecting suspicious terrorist e-mails. However, in the past, some intelligent agencies had disclosed some terrorist e-mails on the Internet, in which terrorists planned terrorist activities. However, the number of such e-mails is limited. Hence, most researchers in this domain have used those real e-mails and prepared some dummy suspicious e-mails through brainstorming [10, 11, 34, 35, 37, 39]. They took "nonsuspicious" e-mails from the existing e-mail corpuses, such as LingSpam, PU, SpamBase, and Enron. For example, Appavu *et al.* [37] prepared a customized dataset to detect suspicious terrorist e-mails. They created a dataset of 500 e-mails, of which 160 e-mails were "suspicious," which they prepared by themselves, and 340 e-mails were "nonsuspicious," which they took from the LingSpam dataset.

Moreover, Appavu *et al.* [10, 11, 34, 35, 37, 39] prepared and used TC-Threatening-I and TC-Threatening-II datasets to detect suspicious terrorist e-mails. TC-Threatening-I contains 2099 e-mails in total, including 500 threatening e-mails, 481 spam e-mails, and 1118 legitimate e-mails where the spam and legitimate e-mails were taken from the LingSpam corpus. LingSpam is a well-known corpus having the 481 spam e-mails and 28931 legitimate e-mails. TC-Threatening-II contains 3893 e-mails in total, including 500 threatening e-mails, 481 spam e-mails, and 2912 legitimate e-mails. In TC-Threatening-II, the spam and legitimate e-mails were taken from the LingSpam corpus. Sun [27] used 1700 news articles about terrorist incidents that happened in Latin America to develop a classifier. Drozdova *et al.* [44] used the "Al-Qaeda" dataset to discover the types of communication technologies used by terrorists. The authors discovered that terrorists prefer to use low-technology communication channels over high-technology communication channels. Nizamani *et al* [50, 52, 53] prepared a dataset to detect suspicious terrorist e-mails. They prepared 800 suspicious e-mails and took 4000 nonsuspicious e-mails from the Enron e-mail dataset. The preceding discussion confirmed that no standardized public dataset is available for detecting suspicious terrorist e-mails. Hence, publicly available datasets for classifying terrorist e-mails are strongly needed.

Table 6: Datasets Analysis

S. No.	Dataset	Description	No. of Studies	Reference
1	LingSpam	This dataset was collected from a mailing list on linguistics. This dataset contains 481 spam emails and 2893legitimate emails.	3	[36-39]
2	Enron	The most popular email corpus for research has been the Enron dataset. It has 13496 spam emails and	5	[42, 46, 48, 52, 53]

S. No.	Dataset	Description	No. of Studies	Reference
		16545 legitimate emails.		
3	SpamBase	This dataset contains 4601 emails messages. Out of 4601, 39% marked as spam and 61% non-spam.	1	[86]
4	SpamAssasin	This dataset contains total 10744 emails (spam = 3793 and ham = 6951)	1	[30]
5	PU Dataset	This dataset contains total 7101 email (spam = 3020 and ham = 4081)	3	[36, 38, 39]
6	Phishing Corpus	This dataset contains total of 11,501 emails (phishing emails = 4550 , ham emails from SpamAssasin = 6951)	1	[54]
7	TC-Threatening-I	This dataset contains 2099 emails in total. This includes 500 threatening emails, 481 spam emails and 1118 legitimate emails.	6	[32-34, 36, 37, 39]
8	TC-Threatening-II	This dataset contains 3893 emails in total. This includes 500 threatening emails, 481 spam emails and 2912 legitimate emails.	6	[32-34, 36, 37, 39]
9	Customized Terrorist Emails	Researchers in many studies created their own dataset for suspicious terrorist email detection. They did not name that dataset. We call such datasets as customized datasets. The number of suspicious terrorist emails and legitimate emails vary from study to study	13	[26-29, 31, 35, 40-43, 45, 47, 51]
10	Reuters-21578	This dataset is famous for single labeled text categorization. This dataset has 20 news groups' text documents. The total number of documents are 18821. Of these, 11293 are used for training set and 7528 are used for testing set.	3	[49, 50, 52]
11	MUC-4	This dataset contains 1,700 news articles about terrorist incidents happened in Latin America.	1	[25]

S. No.	Dataset	Description	No. of Studies	Reference
12	Al-Qaeda Data	The dataset contains four hundred and ninety six specific communication technology use instances (Total N = 496)	1	[44]

Table 7: Availability Links of datasets

S. No.	Dataset	Availability Link
1	LingSpam	http://www.csmining.org/index.php/ling-spam-datasets.html
2	Enron	http://www.aueb.gr/users/ion/data/enron-spam/
3	SpamBase	http://archive.ics.uci.edu/ml/datasets/Spambase
4	SpamAssasin	http://spamassassin.apache.org/publiccorpus
5	PU Dataset	http://www.csmining.org/index.php/pu1-and-pu123a-datasets.html
6	Phishing Corpus	http://monkey.org/*jose/wiki/doku.php?id=PhishingCorpus
7	TC-Threatening-I	Authors have not made the data available publicly
8	TC-Threatening-II	Authors have not made the data available publicly
9	Customized Terrorist Emails	Authors have not made the data available publicly
10	Reuters-21578	https://datahub.io/dataset/reuters-21578
11	MUC-4	http://www-nlpir.nist.gov/related_projects/muc/muc_data/muc_data_index.html
12	Al-Qaeda Data	https://knoema.com/atlas/topics/Al-Qaeda/datasets

5.1.1 Discussion

Researchers have produced their own datasets to develop classifiers due to the lack of standardized public dataset for detecting suspicious terrorist e-mails. However, the findings of their studies may be dubious and biased because the training and testing classifiers are self-created. Another issue is the data imbalance, where the number of class examples is not equally distributed. For example, Appavu *et al.* [10, 11, 34, 35, 37, 39] used the TC-Threatening-II dataset to detect suspicious terrorist e-mails. This dataset is imbalance in nature because the "suspicious" e-mails are 500, the "spam" e-mails are 481, and the "legitimate" e-mails are 2912. This dataset comprises an unequal distribution of class instances, which can produce the biasness in classification results. A standardized publicly available e-mail dataset for classifying suspicious terrorist e-mails is thus necessary.

5.2 Review and Analysis of Features

In suspicious terrorist e-mail detection algorithms, features may include the presence or absence of frequency of some specific terms and the grammatical correctness of the text. The effective selection of feature set is essential to make the classification task efficient. In the selected literature, various researchers have proposed and used numerous features to

detect suspicious terrorist e-mails. These features are e-mail header, e-mail body, context-based, behavioral, stylometric, technological, and time-series features and the descriptions of these features are as follows:

E-mail Header Features: These features are selected from the header part of the e-mail which includes from, to, bcc, and cc fields.
E-mail Body Features: In e-mail body features, features are extracted from the e-mail body part which includes the main content of an e-mail.
Contextual Features: These features are defined by a portion of syntactic structure, in which a possible combination of features with the predefined syntactic structure is found.
Behavioral Features: These features can be useful in detecting abnormal sending behavior.
Time-Series Features: These features include the frequency of e-mail message sends over time, communication pattern with a specific user or group of users over time, and any particular time of sending e-mails.
Stylometric Features: These features include the unique linguistic style and the writing behavior of an individual to determine authorship.
Technological Features: These features include the type of technology used in communication, such as low-technology or high-technology communication device.

Table 8 shows the various features used to detect suspicious terrorist e-mails in all selected studies. 15 studies out of 30 used content-based features to detect suspicious terrorist e-mails. 6 studies used content-based and context-based features. Stylometric and content-based features were used in 4 studies. In one study, e-mail header, behavioral, and time-series features were used. Only e-mail header features were used in one study to detect suspicious e-mails, whereas technological features were used in another study. E-mail body, e-mail header, and contextual features were utilized in 2 studies and e-mail body and e-mail header features in one study.

Table 8: Features used in selected studies

Study	Features used in the study						
	Email Body	**Context**	**Behavioral**	**Time series**	**Stylometric**	**Technological**	**Email Header**
[27]	✓	✓	✗	✗	✗	✗	✗
[28]	✓	✗	✗	✗	✗	✗	✓
[29]	✓	✓	✗	✗	✗	✗	✗
[30]	✗	✗	✓	✓	✗	✗	✓
[31]	✗	✗	✓	✗	✗	✗	✗
[32]	✓	✗	✗	✗	✗	✗	✗
[33]	✓	✗	✗	✗	✗	✗	✗
[35]	✓	✓	✗	✗	✗	✗	✓
[10]	✓	✓	✗	✗	✗	✗	✓
[36]	✓	✗	✗	✗	✗	✗	✗
[34]	✓	✗	✗	✗	✗	✗	✗
[11]	✓	✗	✗	✗	✗	✗	✗
[37]	✓	✗	✗	✗	✗	✗	✗
[38]	✓	✗	✗	✗	✗	✗	✗
[39]	✓	✗	✗	✗	✗	✗	✗
[40]	✓	✗	✗	✗	✗	✗	✗
[41]	✓	✗	✗	✗	✗	✗	✓
[42]	✓	✗	✗	✗	✓	✗	✗
[43]	✓	✗	✗	✗	✗	✗	✗
[44]	✗	✗	✗	✗	✗	✓	✗
[45]	✓	✗	✗	✗	✓	✗	✗
[47]	✓	✗	✗	✗	✗	✗	✗

Study	Features used in the study						
	Email Body	Context	Behavioral	Time series	Stylometric	Technological	Email Header
[46]	✓	✗	✗	✗	✗	✗	✗
[48]	✓	✗	✗	✗	✗	✗	✗
[49]	✓	✗	✗	✗	✗	✗	✗
[50]	✓	✗	✗	✗	✗	✗	✗
[51]	✗	✗	✗	✗	✗	✗	✓
[53]	✓	✗	✗	✗	✗	✗	✗
[52]	✓	✓	✗	✗	✓	✗	✗
[54]	✓	✓	✗	✗	✓	✗	✗

5.2.1 Discussion

Most of the studies have adopted e-mail body, contextual, e-mail header, and stylometric features to identify suspicious terrorist e-mails. They reported that these features are suitable for finding suspicious terrorist e-mails. For example, Lim *et al.* [30] used e-mail body, e-mail header, and time-series features to detect suspicious terrorist e-mails. Regarding e-mail body features, they reported some specific word features that classified an e-mail as terrorist e-mail. They used sender e-mail address, receiver e-mail address, and subject field as e-mail header features and the frequency of sending e-mails over, the frequency of sending e-mail to a particular receiver, and the pattern of sending e-mails over time as time-series features. The authors reported that these features increased the classifier performance up to 83%. Nizamani *et al.* [53] used e-mail body and contextual features to detect suspicious terrorist e-mails. They reported that if the e-mail body contains such features as "*bomb," "kill*," "*bomb blast*," and future tense helping verbs, such as "*shall*" or "*will*," then the e-mail may be a suspicious terrorist e-mail. However, if these features appear with the past tense, then the e-mail may not be a suspicious terrorist e-mail. The authors achieved 87% accuracy. In another study, Nizamani *et al.* [52] used e-mail body, behavioral, and stylometric features, such as write prints, to identify the writing styles of e-mail authors. They reported that the e-mail body, stylometric, and behavioral features improve the classification accuracy of suspicious e-mail detection classifier up to 93%. Hence, when body, header, stylometric, contextual, and behavioral features are combined, they often achieve high accuracy for suspicious terrorist e-mail decision models.

5.3 Review and Analysis of Feature Extraction, Representation, and Selection

Table 9 shows the various feature extraction, representation, and selection techniques used in the selected studies. Fig. 8 (a), Fig. 8 (b), and Fig. 8 (c) show the frequency of feature extraction, feature representation, and feature selection schemes, respectively. Our findings indicate that the mostly used feature extraction techniques are BoW and *n*-gram. Out of 30 studies, 13 used BoW for feature extraction, 12 used *n*-grams, and 5 have not mentioned the feature extraction techniques. BR and TF feature representation techniques are mostly used to transform the extracted features into numeric values. Among the 30 studies, 11 used BR, 9 used TF, 5 used TF–inverse document frequency, and 5 did not mention any feature representation techniques they used. The most adopted feature selection technique to select discriminative features is IG. Out of the 30 studies, 12 used IG, 4 used gain ratio, 3 used chi-square, and 17 studies did not use any feature selection scheme to select the most powerful features.

5.3.1 Discussion

Feature extraction, feature representation, and feature selection are important steps in e-mail classification task. According to our survey results, most researchers used *n*-gram and BoW feature extraction schemes to extract features from e-mail content due to their simplicity and efficacy. They mostly used unigram approach in *n*-gram, where only one unique word is treated as a feature. For representing those extracted features, most researchers employed the BR scheme to determine whether suspicious terrorist e-mail features occurred in the e-mail. Some researchers also used TF representation scheme to identify the occurrences and frequency of suspicious features in an e-mail. Appavu *et al.* [34]

and Nizamani *et al.* [53] represented the e-mail features using BR and TF schemes. Both schemes approximately produced the same results and are suitable for representing e-mail features. Nizamani *et al.* [53] compared the performance of IG and gain ratio to select the most result-oriented features from the VSM. The authors concluded that IG is more suitable than gain ratio in selecting the powerful subset of features from VSM. In addition, Nizamani *et al.* [54] compared the performance of IG and chi-square and concluded that IG is more suitable than chi-square in selecting the powerful subset of features from the VSM. To summarize, the best feature extraction technique to represent an e-mail into the VSM is either *n*-gram or BoW; the best feature representation technique is BR, followed by TF; and the best feature selection method is IG.

Table 9: Feature extraction, representation and selection techniques in selected studies

Study	Feature Extraction	Feature Representation	Feature Selection
[27]	N-Gram, Template based feature extraction	Term Frequency	Information Gain
[28]	NA	NA	NA
[29]	N-Gram	Term Frequency	Information Gain
[30]	NA	NA	NA
[31]	NA	NA	NA
[32]	N-Gram	Binary Representation	Information Gain
[33]	Bag of Words	Binary Representation	NA
[35]	Bag of Words	Binary Representation	Chi-Square
[10]	Bag of Words	Binary Representation	Chi-Square, Information Gain
[36]	Bag of Words	Binary Representation	NA
[34]	N-Gram	Binary Representation	NA
[11]	N-Gram	TFiDF	Information Gain
[37]	N-Gram	Term Frequency	NA
[38]	Bag of Words	Binary Representation	Information Gain
[39]	Bag of Words	Binary Representation	Information Gain
[40]	Bag of Words	TFiDF	NA
[41]	N-Gram	Term Frequency	NA
[42]	N-Gram	Term Frequency	NA
[43]	Bag of Words	Term Frequency	NA
[44]	NA	NA	NA
[45]	N-Gram	Term Frequency	NA
[47]	Bag of Words	TF	NA
[46]	N-Gram	Binary Representation	Information Gain and Gain Ratio
[48]	N-Gram	Binary Representation	NA
[49]	Bag of Words	Binary Representation	Information Gain and Gain Ratio
[50]	N-Gram	Term Frequency, TFiDF	Information Gain and Gain Ratio
[51]	NA	NA	NA
[53]	N-Gram	Term Frequency	Information Gain and Gain Ratio
[52]	Bag of Words	TFiDF	NA
[54]	Bag of Words	TFiDF	Chi-Square and Information Gain

5.4 Review and Analysis of Classification Techniques

According to our findings in this study, various text classification algorithms have been employed to detect suspicious terrorist e-mails. Table 10 shows the classification techniques used in each selected study while Fig. 9 shows the frequency of classification techniques used in the studies. As shown in Table 10, each study used more than one classifier to compare which one will perform better with their proposed methods. As presented in Fig. 9, the most common classification techniques used in suspicious terrorist e-mail classification are DT, SVM, and NB. Out of the 30 selected studies, 17 used DT, 15 used SVM, 11 used NB, 4 used ANN, and 4 used cluster-based classification models (CCMs). The rest of the classifiers, such as k-nearest neighbor, rough set theory, fuzzy set, and k-means, were used in one or two studies only.

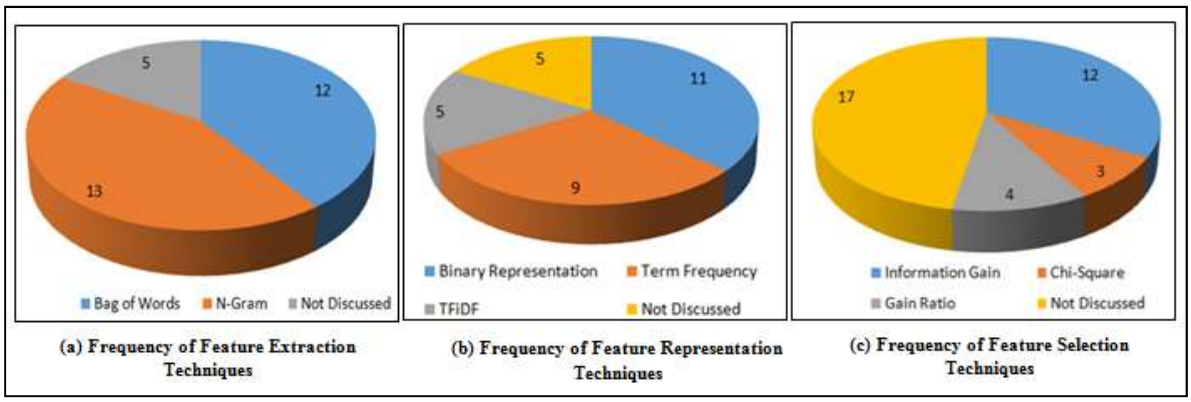

Fig. 8: Frequency of Feature Extraction, Representation and Selection Techniques used in selected studies

Table 10: Classification Techniques in the selected studies

Study	SVM	DT	LR	NB	ANN	KNN	RST	FL	APR	KM	AB	AI	KMP	CPM	CCM
[27]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[28]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[29]	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[30]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[31]	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
[32]	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
[33]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[35]	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
[10]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[36]	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[34]	✓	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[11]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
[37]	✓	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[38]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[39]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
[40]	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[41]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
[42]	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[43]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
[44]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗
[45]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗
[47]	✓	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
[46]	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[48]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
[49]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[50]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓
[51]	✗	✗	✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[53]	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[52]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[54]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Total	**15**	**17**	**2**	**11**	**4**	**1**	**1**	**1**	**1**	**2**	**2**	**2**	**1**	**1**	**4**

SVM = Support Vector Machine, **DT = Decision Tree, **LR** = Logistic Regression, **NB** = Naive Bayes, **ANN** = Artificial Neural Network, **KNN** = K-Nearest Neighbor, **RST** = Rough Set Theory, **FL** = Fuzzy Logic, **APR** = Apriori, **KM** = K-Means, **AB** = AdaBoost, **AI** = Ad Infinitum, **KMP** = Knuth-Morris-Pratt string matching algorithm, **CPM** = Crime Prediction Model, **CCM** = Cluster Based Classification Model

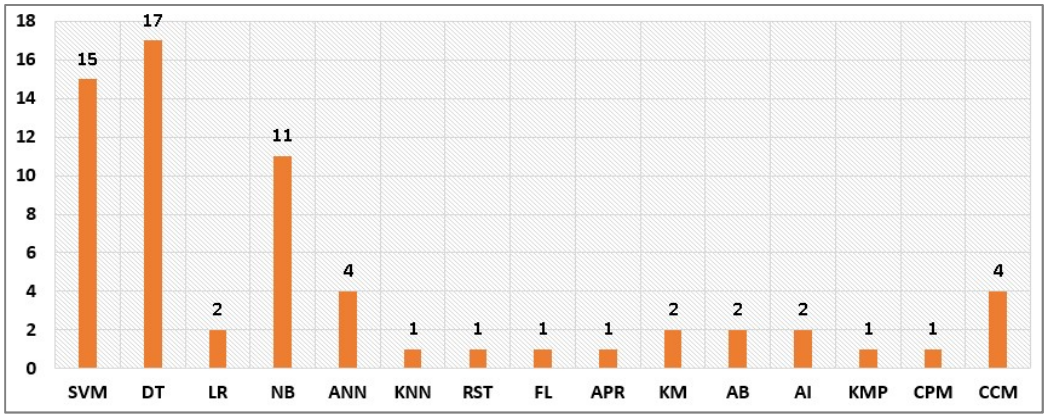

Fig. 9: Frequency of classification techniques used in the selected studies

5.4.1 Discussion

Various classification algorithms have been used in the literature to discover the suspicious e-mails. Sun [27] used SVM for terrorist e-mail classification and reported 86% F-measure. Abbasi *et al.* [29] used SVM and DT to detect extremist group messages where SVM obtained 97% accuracy, while DT obtained 90% accuracy. Negnevitsky *et al.* [31] used ANN and fuzzy logic, and they achieved 87% accuracy. Rajaram *et al.* [10] used DT classifier to detect suspicious terrorist e-mails and obtained 95% accuracy. Appavu *et al.* [34, 38] compared ANN, DT, SVM, and NB for detecting suspicious terrorist e-mails. Their findings implied that DT and SVM achieved the highest accuracy. Appavu *et al.* [39] proposed a modified version of DT, which referred to as Ad Infinitum algorithm, to detect suspicious terrorist e-mails and compared the proposed classifier with DT, SVM, and NB. The authors achieved 93% accuracy with the proposed algorithm, followed by 92% accuracy with SVM and DT and 74% accuracy with NB. Nizamani *et al.* [48] proposed a boosting algorithm called AdaBoost to detect suspicious e-mails and obtained 98.90% accuracy. In another study, Nizamani *et al.* [52] proposed a CCM to detect suspicious terrorist e-mails and obtained 81.30% accuracy.

From the review, it is found that DT and SVM obtain the highest accuracy for detecting suspicious terrorist e-mails. In other similar domains, namely, spam e-mail detection and phishing e-mail detection, these classifiers also outperform other classifiers. The DT classifier does not require any domain knowledge. In addition, it has capability to handle high dimensionality data with missing values. Finally, it is a non-parametric algorithm and thus it does not employ any suppositions for space distribution. The SVM shows best performance for detecting suspicious emails because the problem of classifying suspicious terrorist e-mails is linearly divisible, and it utilizes threshold functions to linearly split classes with margins. Besides than that, SVM is also susceptible to over-fitting and its performance does not suffer from the higher number of features [87].

5.5 Review and Analysis of Performance Measures

Table 11 shows the various performance measures used in selected studies. Most of the studies used accuracy to measure the classifier performance. Out of the 30 studies, 26 used accuracy, one study used accuracy, precision, and recall, two used precision and recall, and one study used F-measure performance metric to evaluate the performance of classifiers.

5.5.1 Discussion

From the review, we found that most researchers used accuracy as the performance measure to evaluate classifier performance. However, this metric alone is not sufficient to accurately evaluate the performance. The dataset adopted in

most studies is imbalance in nature, in which suspicious e-mails are less in number, whereas nonsuspicious e-mails are more in number. In such cases, the most useful metric is area under curve (AUC) [88, 89]. In general, this metric is beneficial in examining the decision model performance pertaining to a specific class. In [27, 38, 45, 46, 52, 54], the researchers developed a ternary classifier to classify an e-mail into "suspicious," "nonsuspicious," or "spam" and to measure the classifier performance. They used simple precision, recall, or accuracy. However, for multiclass classifiers using an imbalance dataset, the suitable measures are macro precision, macro recall, and overall accuracy [86].

Table 11: Performance Measures used in selected studies

Study	Performance measures used in each study								
	Precision	Recall	F-measure	Accuracy	Study	Precision	Recall	F-measure	Accuracy
[30]	✓	✓	✓	✗	[22]	✗	✗	✗	✓
[16]	✗	✗	✗	✓	[23]	✗	✗	✗	✓
[17]	✗	✗	✗	✓	[4]	✗	✗	✗	✓
[18]	✗	✗	✗	✓	[24]	✓	✓	✗	✗
[19]	✗	✗	✗	✓	[31]	✓	✓	✗	✗
[10]	✗	✗	✗	✓	[25]	✗	✗	✓	✗
[20]	✗	✗	✗	✓	[26]	✗	✗	✗	✓
[13]	✗	✗	✗	✓	[5]	✗	✗	✗	✓
[14]	✗	✗	✗	✓	[6]	✗	✗	✗	✓
[21]	✗	✗	✗	✓	[28]	✗	✗	✗	✓
[12]	✗	✗	✗	✓	[29]	✗	✗	✗	✓
[1]	✗	✗	✗	✓	[27]	✗	✗	✗	✓
[15]	✗	✗	✗	✓	[8]	✗	✗	✗	✓
[2]	✗	✗	✗	✓	[7]	✗	✗	✗	✓
[3]	✗	✗	✗	✓	[11]	✗	✗	✗	✓

6.0 OPPORTUNITIES, CHALLENGES, AND FUTURE RESEARCH DIRECTIONS

This section highlights the several research gaps that needs considerable research efforts to develop more accurate classification models in the area of detecting suspicious terrorist e-mail classification. The research challenges that are yet to be resolved are discussed below.

1) Availability of standardized dataset: The main issue in the area of finding suspicious terrorist e-mails is the public availability of dataset. The researchers in this area have produced their own datasets. Hence, the main issue in the existing studies is the data biasness because the training and testing sets were both developed by the researchers themselves and there is no any benchmark dataset exists to compare the performance of their proposed techniques for unbiased evaluation. In addition, the datasets used in the selected studies were not balance. In those datasets, a few suspicious e-mails existed, whereas the nonsuspicious e-mails were many. Such datasets are called imbalance datasets because the number of class examples is distributed unequally. For such datasets, researchers should use either machine learning techniques, such as smoothing, to make the dataset balance or performance metrics that are suitable for evaluating the performance of classifiers on imbalance datasets, such as AUC. Most of the selected studies suffer from the data imbalance problem, but none of them used machine learning techniques or suitable performance metric for imbalance dataset. Hence, the development of benchmark dataset will be a solid contribution in the area of suspicious terrorist e-mail detection, such as the LingSpam and Enron in spam e-mail detection.
2) Proposal of novel features: In most of the existing studies, the researchers used content-based features by exploiting the e-mail body content. Simple keyword matching and tenses were used in content-based features to classify the e-mails as suspicious or nonsuspicious. Only a few studies exploited the use of contextual and behavioral features to detect suspicious terrorist e-mail s. In future, researchers can explore various other features, such as phrase-based, lexical, structural, and time-series features, to accurately detect suspicious terrorist e-mails.
3) Ontology-based terrorist e-mail classification: In future, researchers can also focus on classifying suspicious terrorist e-mails using ontology. An adaptive ontology can be created as suspicious terrorist e-mail filtering. This

type of ontology will be extended and customized when a user reports a suspicious e-mail. A filter will be personalized, modularized, robust, and scalable by creating a suspicious terrorist e-mail filter using adaptive ontology.

4) Terrorist e-mail classification by applying deep learning concepts: The accuracy of suspicious terrorist e-mail detection classifier can be enhanced using various deep learning techniques. For example, numerous alternative approaches, including word to vector, gloVe, skip gram, and continuous BoW, can be used for effective feature extraction. For classification, various deep learning classifiers, such as convolution neural network (CNN), recursive neural network, and recurrent neural network, can be applied. Deep learning process has been used to improve classification performance in a number of application areas, such as web mining and text analysis [90].

5) Image-based and text-based classification: According to our survey, most of the researchers used textual features for terrorist e-mail classification. However, terrorists can send the terrorism activity-related information in various forms, such as images. They embed text in an image and send it as an e-mail. Hence, such suspicious e-mails need to be detected. Providing useful image-based features and their representation and image-based classification techniques that can enhance the significant accuracy and performance of terrorist e-mail classifiers will be valuable contribution.
6) Real-time evaluation: Most of the existing research effort on terrorist e-mail classification is based on offline datasets. The classifier performance is evaluated using offline test datasets. Such evaluation usually does not include the elements of online streaming environmental. The online streaming factors would deeply influence the performance of suspicious terrorist e-mail classifiers. The performance of constructed classifiers with online e-mail streaming should thus be evaluated because the e-mail traffic would be more complex than the case with an offline test dataset. For researchers, evaluation of e-mail classifiers in a real practical environment is yet another potential research direction. Given that users are the essential actors who ultimately use the services of e-mail classification, involving them in classifier evaluation will further assure the usability of e-mail classifiers.
7) Language-based barriers: In the existing literature on suspicious terrorist e-mail classification, researchers developed classifiers that can classify an e-mail written in English language. Hence, identifying and developing features and modifying classifiers that can be useful in detecting the suspicious terrorist e-mails written in other languages such as Chinese and Arabic call for work.

7.0 CONCLUSION

This paper presents a comprehensive review of text classification techniques for detecting suspicious terrorist e-mails. A quantitative analysis of the use of different datasets, features, feature extraction techniques, feature representation approaches, feature selection schemes, text classification algorithms, and performance measures was conducted. This comprehensive survey determined that no dataset for detecting suspicious terrorist e-mails is publicly available for researchers. Most of the researchers in this domain prepared their own datasets. The most widely used features were content-based features, in which the content of the e-mail body was exploited to detect suspicious terrorist e-mails. The most common feature extraction techniques were BoW and *n*-gram. These techniques were widely used because they are simple, easy to implement, and obtained significant results in previous literature. The most typical feature representation techniques were BR and TF. The BR technique was widely used because, in most of the studies, researchers have exploited the word features, in which they checked the occurrence of suspicious words. If suspicious words were present, they were represented by "1," otherwise "0." The TF technique was also widely used because selecting frequent words would increase the likelihood that the features will be occurred in future test cases. The IG was the most widely used feature selection method because it computes the reduction in entropy when the features are given versus absent. Furthermore, IG has a generalized form for nominal valued attributes. The review confirmed that the top three classification algorithms for binary class classification were DT, SVM, and NB. SVM achieved lower false positive rates and worked better than BR than TF. Accuracy, precision, recall, and F-measure were used as performance measures to evaluate the performance of terrorist e-mail classifiers. In most of the studies, researchers have used accuracy as the performance metric. Accuracy is a very basic and common performance measure, which calculates the performance of the correctly classified e-mails out of all e-mails. However, accuracy is not a good measure for imbalanced datasets. For imbalanced datasets, AUC is the most suitable performance metric to use. Precision, recall, and

F-measure are not suitable to measure the performance of multiclass classifiers. The suitable multiclass classifiers are macro precision, macro recall, and macro F-measure.

ACKNOWLEDGEMENT

This research work was supported by Faculty of Computer Science and Information Technology, University of Malaya under a special allocation of Post Graduate Fund. This research was also partially funded by the University Malaya Research Grant (No: RP028F-14AET).

REFERENCES

[1] S. Radicati, "Email Statistics Report, 2014-2018", The Radicati Group, INC., A Technology Market Research Firm, Palo Alto, CA, USA April, 2014.

[2] T. Mahmood and G. M. Shaikh", Adaptive Automated Teller Machines," *Expert Systems with Applications,* Vol. 40, 2013, pp. 1152-1169.

[3] M. A. Al-garadi, M. S. Khan, K. D. Varathan, G. Mujtaba, and A. M. Al-Kabsi, "Using online social networks to track a pandemic: A systematic review", *Journal of biomedical informatics,* Vol. 62, 2016, pp. 1-11.

[4] G. Mujtaba, L. Shuib, R. G. Raj, R. Rajandram, K. Shaikh, and M. A. Al-Garadi, "Automatic ICD-10 multi-class classification of cause of death from plaintext autopsy reports through expert-driven feature selection", *Plos One,* Vol. 12, Feb 2017, p. 27.

[5] G. Mujtaba, L. Shuib, R. G. Raj, R. Rajandram, and K. Shaikh, "Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study", *Journal of Forensic and Legal Medicine,* Vol. 57, 2018, pp. 41-50.

[6] G. Mujtaba, L. Shuib, R. G. Raj, R. Rajandram, K. Shaikh, and M. A. Al-Garadi, "Classification of forensic autopsy reports through conceptual graph-based document representation model", *Journal of biomedical informatics,* Vol. 82, 2018, pp. 88-105.

[7] I. Idris, A. Selamat, N. T. Nguyen, S. Omatu, O. Krejcar, K. Kuca*, et al.*, "A combined negative selection algorithm-particle swarm optimization for an email spam detection system", *Engineering Applications of Artificial Intelligence,* Vol. 39, Mar 2015, pp. 33-44.

[8] I. Idris and A. Selamat, "Improved email spam detection model with negative selection algorithm and particle swarm optimization", *Applied Soft Computing,* Vol. 22, Sept 2014, pp. 11-27.

[9] G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M. A. Al-Garadi, "Email Classification Research Trends: Review and Open Issues", *IEEE Access,* 2017.

[10] R. Rajaram and A. Balamurugan, "Suspicious E-mail detection via decision tree: A data mining approach", *CIT. Journal of computing and information technology,* 2007, Vol. 15, pp. 161-169.

[11] S. Appavu, R. Rajaram, M. Muthupandian, and G. Athippan, "Automatic mining of Threatening e-mail using Ad Infinitum algorithm", *International Journal of Information Technology,* Vol. 14, 2008, pp. 81-108.

[12] H. Chiroma, S. Abdul-Kareem, and A. Abubakar, "A Framework for Selecting the Optimal Technique Suitable for Application in a Data Mining Task", in *Future Information Technology*, ed: Springer, 2014, pp. 163-169.

[13] A. I. Abubakar, H. Chiroma, S. A. Muaz, and L. B. Ila, "A Review of the Advances in Cyber Security Benchmark Datasets for Evaluating Data-Driven Based Intrusion Detection Systems", *Procedia Computer Science,* Vol. 62, 2015, pp. 221-227.

[14] A. Zeki, A. Abubakar, and H. Chiroma, "An intermediate significant bit (ISB) watermarking technique using neural networks", *SpringerPlus,* Vol. 5, 2016, pp. 1-25.

[15] C. Haruna, M. Abdulhamid, Y. Abdulsalam, M. Ali, and U. Timothy, "Academic community cyber cafés-A Perpetration point for cyber crimes in Nigeria", *International Journal of Information Sciences and Computer Engineering,* Vol. 2, 2011, pp. 7-13.

[16] A. M. Zeki, E. E. Elnour, A. A. Ibrahim, C. Haruna, and S. Abdulkareem, "Automatic interactive security monitoring system", in *2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*, 2013, pp. 215-220.

[17] O. A. Adewumi and A. A. Akinyelu, "A hybrid firefly and support vector machine classifier for phishing email detection", *Kybernetes,* Vol. 45, 2016, pp. 977-994.

[18] R. Verma and N. Hossain, "Semantic Feature Selection for Text with Application to Phishing Email Detection", in *Information Security and Cryptology - Icisc 2013*, Vol. 8565, H. S. Lee and D. G. Han, Eds., ed Cham: Springer Int Publishing Ag, 2014, pp. 455-468.

[19] I. R. A. Hamid and J. Abawajy, "Hybrid Feature Selection for Phishing Email Detection", in *Algorithms and Architectures for Parallel Processing, Pt Ii*. Vol. 7017, Y. Xiang, A. Cuzzocrea, M. Hobbs, and W. Zhou, Eds., ed Berlin: Springer-Verlag Berlin, 2011, pp. 266-275.

[20] B. Zhou, Y. Y. Yao, and J. G. Luo, "Cost-sensitive three-way email spam filtering", *Journal of Intelligent Information Systems,* Vol. 42, Feb 2014, pp. 19-45.

[21] A. Chakrabarty, S. Roy, and Ieee, "An Optimized k-NN Classifier based on Minimum Spanning Tree for Email Filtering", in *2014 2nd International Conference on Business and Information Management (Icbim),* 2014.

[22] M. T. Banday and S. A. Sheikh, "Realization of Microsoft Outlook (R) Add-in for Language Based E-mail Folder Classification", in *2013 International Conference on Machine Intelligence and Research Advancement (Icmira 2013),* 2013, pp. 279-284.

[23] E. Kaplan, "Terrorists and the Internet", *Council on Foreign Relations,* Vol. 8, 2009.

[24] E. Blanzieri and A. Bryl, "A survey of learning-based techniques of email spam filtering", *Artificial Intelligence Review,* Vol. 29, 208, pp. 63-92.

[25] T. S. Guzella and W. M. Caminhas, "A review of machine learning approaches to spam filtering", *Expert Systems with Applications,* Vol. 36, 2009, pp. 10206-10222.

[26] A. Almomani, B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A survey of phishing email filtering techniques", *Communications Surveys & Tutorials, IEEE,* Vol. 15, 2013, pp. 2070-2090.

[27] A. Sun, "Using Support Vector Machine for Terrorism Information Extraction. 1st NSF," in *NIJ Symposium on Intelligence and Security Informatics*, 2003.

[28] J. Allanach, T. Haiying, S. Singh, P. Willett, and K. Pattipati, "Detecting, tracking, and counteracting terrorist networks via hidden Markov models", in *Aerospace Conference, 2004. Proceedings. 2004 IEEE*, Vol. 5, 2004, pp. 1-3257.

[29] A. Abbasi and H. Chen, "Applying authorship analysis to extremist-group Web forum messages", *IEEE Intelligent Systems,* Vol. 20, 2005, pp. 67-75.

[30] M. Lim, M. Negnevitsky, and J. Hartnett, "Tracking and monitoring e-mail traffic activities of criminal and terrorist organisations using visualisation Tools", *Journal of Information Warfare,* Vol. 5, 2005, pp. 112-224.

[31] M. Negnevitsky, M. J. Lim, J. Hartnett, and L. Reznik, "Email communications analysis: how to use computational intelligence methods and tools?", in *CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2005.*, 2005, pp. 16-23.

[32] W. Zhao and Y. Zhu, "An Email Classification Scheme Based on Decision-Theoretic Rough Set Theory and Analysis of Email Security", in *TENCON 2005 - 2005 IEEE Region 10 Conference*, 2005, pp. 1-6.

[33] B. Galitsky and B. Kovalerchuk, "Mining emotional profiles using e-mail messages for earlier warnings of potential terrorist activities", in *Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2006*, Kissimmee, FL, 2006.

[34] S. A. alias Balamurugan, D. R. Rajaram, G. Athiappan, and M. Muthupandian, "Data mining techniques for suspicious email detection: A comparative study", in *IADIS European Conference Data Mining*, 2007.

[35] S. Appavu Alias Balamurugan, M. Pandian, and R. Rajaram, "Association rule mining for suspicious email detection: A data mining approach", in *ISI 2007: 2007 IEEE Intelligence and Security Informatics*, New Brunswick, NJ, 2007, pp. 317-324.

[36] C. C. Yang and T. D. Ng, "Terrorism and Crime Related Weblog Social Network: Link, Content Analysis and Information Visualization", in *Intelligence and Security Informatics, 2007 IEEE*, 2007, pp. 55-58.

[37] S. A. A. Balamurugan, G. Athiappan, M. M. Pandian, and R. Rajaram, "Classification methods in the detection of new suspicious emails", *Journal of Information and Knowledge Management,* Vol. 7, 2008, pp. 209-217.

[38] S. A. A. Balamurugan and R. Rajaram, "Learning to Classify Threaten E-mail," in *2008 Second Asia International Conference on Modelling & Simulation (AMS)*, 2008, pp. 522-527.

[39] S. Appavu, R. Rajaram, M. Muthupandian, G. Athiappan, and K. S. Kashmeera, "Data mining based intelligent analysis of threatening e-mail," *Knowledge-Based Systems,* Vol. 22, Jul 2009, pp. 392-393.

[40] W. Eberle, L. Holder, and D. Cook, "Identifying threats using graph-based anomaly detection", in *Machine Learning in Cyber Trust*, ed: Springer, 2009, pp. 73-108.

[41] M. Fanlin, S. Wu, J. Yang, and Y. Genzhen, "Research of an E-mail forensic and analysis system based on visualization", in *Computational Intelligence and Industrial Applications, 2009. PACIIA 2009. Asia-Pacific Conference on*, 2009, pp. 281-284.

[42] R. Hadjidj, M. Debbabi, H. Lounis, F. Iqbal, A. Szporer, and D. Benredjem, "Towards an integrated e-mail forensic analysis framework", *Digital Investigation,* Vol. 5, 2009. pp. 124-137.

[43] F. Ozgul, Z. Erdem, and C. Bowerman, "Prediction of Unsolved Terrorist Attacks Using Group Detection Algorithms," in *Pacific-Asia Workshop on Intelligence and Security Informatics*, 2009, pp. 25-30.

[44] K. Drozdova and M. Samoilov, "Predictive analysis of concealed social network activities based on communication technology choices: early-warning detection of attack signals from terrorist organizations", *Computational and Mathematical Organization Theory,* Vol. 16, 2010, pp. 61-88.

[45] F. Iqbal, H. Binsalleeh, B. C. M. Fung, and M. Debbabi, "Mining writeprints from anonymous e-mails for forensic investigation", *Digital Investigation,* Vol. 7, 2010, pp. 56-64.

[46] S. Nizamani, N. Memon, and U. Wiil, "Detecting suspicious emails using improved features", in *IEEE International conference on modeling and simulation control*, 2010, pp. 232-236.

[47] D. A. Simanjuntak, H. P. Ipung, and A. S. Nugroho, "Text classification techniques used to faciliate cyber terrorism investigation", in *Advances in Computing, Control and Telecommunication Technologies (ACT), 2010 Second International Conference on*, 2010, pp. 198-200.

[48] S. Nizamani, N. Memon, and U. K. Wiil, "Detection of illegitimate emails using Boosting algorithm", in *Counterterrorism and Open Source Intelligence*, ed: Springer, 2011, pp. 249-264.

[49] S. Nizamani, N. Memon, and U. K. Wiil, "Cluster Based Text Classification Model", in *Counterterrorism and Open Source Intelligence*, ed: Springer, 2011, pp. 265-283.

[50] S. Nizamani, N. Memon, U. K. Wiil, and P. Karampelas, "CCM: a text classification model by clustering", in *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, 2011, pp. 461-467.

[51] A. Butkovic, S. Mrdovic, and S. Mujacic, "IP geolocation suspicious email messages", *2013 21st Telecommunications Forum (Telfor),* 2013, pp. 881-884.

[52] S. Nizamani and N. Memon, "CEAI: CCM-based email authorship identification model", *Egyptian Informatics Journal,* Vol. 14, 2013, pp. 239-249.

[53] S. Nizamani, N. Memon, U. K. Wiil, and P. Karampelas, "Modeling suspicious email detection using enhanced feature selection", *International Journal of Modeling and Optimization,* 2013, Vol. 2.

[54] S. Nizamani, N. Memon, M. Glasdam, and D. D. Nguyen, "Detection of fraudulent emails by employing advanced feature abundance", *Egyptian Informatics Journal,* Vol. 15, 2014, pp. 169-174.

[55] E. G. Amoroso, *Fundamentals of computer security technology*: Prentice-Hall, Inc., 1994.

[56] N. Martin and J. Rice, "Cybercrime: Understanding and addressing the concerns of stakeholders", *Computers & Security,* Vol. 30, 2011, pp. 803-814.

[57] M. Thangiah, S. Basri, and S. Sulaiman, "A framework to detect cybercrime in the virtual environment", in *Computer & Information Science (ICCIS), 2012 International Conference on*, 2012, pp. 553-557.

[58] E. H. Spafford, "Computer viruses as artificial life", *Artificial life,* Vol. 1, 1994, pp. 249-265.

[59] D. Verton and J. Brownlow, *Black ice: The invisible threat of cyber-terrorism*: Osborne, 2003.

[60] R. M. Kowalski, S. P. Limber, and P. W. Agatston, *Cyberbullying: Bullying in the digital age*: John Wiley & Sons, 2012.

[61] J. W. Patchin and S. Hinduja, *Cyberbullying prevention and response: Expert perspectives*: Routledge, 2012.

[62] H. Haron and F. B. M. Yusof, "Cyber stalking: The social impact of social networking technology", in *Education and Management Technology (ICEMT), 2010 International Conference on*, 2010, pp. 237-241.

[63] D. Michalopoulos, I. Mavridis, and M. Jankovic, "GARS: Real-time system for identification, assessment and control of cyber grooming attacks", *Computers & Security,* Vol. 42, 2014, pp. 177-190.

[64] K. Nithya, P. Kalaivaani, and R. Thangarajan, "An enhanced data mining model for text classification", in *Computing, Communication and Applications (ICCCA), 2012 International Conference on*, 2012, pp. 1-4.

[65] G. Salton and M. J. McGill, *Introduction to modern information retrieval*, McGraw-Hill, Inc, 1986.

[66] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval", *Information processing & management,* Vol. 24, 1988, pp. 513-523.

[67] F. Debole and F. Sebastiani, "Supervised term weighting for automated text categorization", in *Text mining and its applications*, ed: Springer, 2004, pp. 81-97.

[68] J. Ramos, "Using tf-idf to determine word relevance in document queries", in *Proceedings of the first instructional conference on machine learning*, 2003.

[69] A. Aizawa, "An information-theoretic perspective of tf–idf measures", *Information Processing & Management,* Vol. 39, 2003, pp. 45-65.

[70] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection", *The Journal of Machine Learning Research,* Vol. 3, 2003, pp. 1157-1182.

[71] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Icml*, 1997, pp. 412-420.

[72] P. J. Hayes and S. P. Weinstein, "CONSTRUE/TIS: A System for Content-Based Indexing of a Database of News Stories", in *IAAI*, 1990, pp. 49-64.

[73] Y. Yang, "An evaluation of statistical approaches to text categorization", *Information retrieval,* Vol. 1, 1999, pp. 69-90.

[74] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk e-mail", in *Learning for Text Categorization: Papers from the 1998 workshop*, 1998, pp. 98-105.

[75] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval", in *European conference on machine learning*, 1998, pp. 4-15.

[76] J. R. Quinlan, "Induction of decision trees", *Machine learning,* Vol. 1, 1986, pp. 81-106.

[77] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21 No. 3, 1990, pp. 660-674.

[78] L. Breiman, "Random forests", *Machine learning,* Vol. 45, 2001, pp. 5-32.

[79] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* Vol. 2, 2002, pp. 18-22.

[80] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network", *Computers in Human Behavior,* Vol. 63, 2016, pp. 433-443.

[81] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems", *J. Mach. Learn. Res,* Vol. 15, 2014, pp. 3133-3181.

[82] N. Cristianini and J. Shawe-Taylor, "An introduction to SVM," ed: Cambridge University Press, Cambridge, 1999.

[83] X. He and S. Xu, *Process neural networks: Theory and applications*: Springer Science & Business Media, 2010.

[84] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing,* Vol. 4, 1994, pp. 65-85.

[85] M. Mitchell, *An introduction to genetic algorithms*: MIT press, 1998.

[86] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management,* Vol. 45, 2009, pp. 427-437 .

[87] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, 1998, pp. 137-142.

[88] F. J. Provost and T. Fawcett, "Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions," in *KDD*, 1997, pp. 43-48.

[89] F. J. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," in *ICML*, 1998, pp. 445-453.

[90] S. Dumais and H. Chen, "Hierarchical classification of Web content," in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, 2000, pp. 256-263.