Detecting Arabic textual threats in social media using artificial intelligence: An overview

Article *in* Indonesian Journal of Electrical Engineering and Computer Science · March 2022

DOI: 10.11591/ijeecs.v25.i3.pp1712-1722

4 authors, including:

Hossam Elzayady
Military Technical College
10 PUBLICATIONS **23** CITATIONS
SEE PROFILE

Gouda I. Salama
Military Technical College
63 PUBLICATIONS **634** CITATIONS
SEE PROFILE

Indonesian Journal of Electrical Engineering and Computer Science
Vol. 25, No. 3, March 2022, pp. 1712~1722
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v25.i3.pp1712-1722

Detecting Arabic textual threats in social media using artificial intelligence: An overview

Hossam Elzayady, Mohamed S. Mohamed, Khaled M. Badran, Gouda I. Salama
Department of Computer Engineering and Artificial Intelligence, Military Technical College, Cairo, Egypt

Article Info

Article history:

Received Sep 8, 2021
Revised Dec 18, 2021
Accepted Jan 7, 2022

Keywords:

Artificial intelligence
Deep learning
Machine learning
Natural language processing
Offensive language

ABSTRACT

Recent studies show that social media has become an integral part of everyone's daily routine. People often use it to convey their ideas, opinions, and critiques. Consequently, the increasing use of social media has motivated malicious users to misuse online social media anonymity. Thus, these users can exploit this advantage and engage in socially unacceptable behavior. The use of inappropriate language on social media is one of the greatest societal dangers that exist today. Therefore, there is a need to monitor and evaluate social media postings using automated methods and techniques. The majority of studies that deal with offensive language classification in texts have used English datasets. However, the enhancement of offensive language detection in Arabic has gotten less consideration. The Arabic language has different rules and structures. This article provides a thorough review of research studies that have made use of artificial intelligence (AI) for the identification of Arabic offensive language in various contexts.

Corresponding Author:

Hossam Elzayady
Department of Computer Engineering and Artificial Intelligence, Military Technical College
Cairo, Egypt
Email: hossamelzaiade@gmail.com

1. INTRODUCTION

Individuals are getting increasingly engaged with one another as a result of the growth of social networks during the last few decades [1]. People from all over the globe were given the chance to communicate on a massive scale and in real-time using microblogging technologies [2]. Humans now have the ability to communicate freely, allowing them to share a wide range of ideas, feelings, and information. Furthermore, users of these platforms may prefer to remain anonymous, raising the risk of technical misuse [3]. As a result, offensive languages of diverse kinds, such as hate speech and cyberbullying, have become more widespread on social media [4].

According to legislation, hate speech on social networking platforms is prohibited in certain nations. In Germany, for example, the Network Enforcement Act was issued in 2017 [5]. Moreover, legislative amendments currently attempt to combat offensive language. Advanced technical approaches that can aid social media platforms and others in implementing these laws [6]. Online offensive language spotting has been used in multiple languages, such as English, German, Turkish, Hindi, Chinese, and Arabic [7]–[10].

Working with Arabic may be difficult because of morphological complexity and the lexical ambiguity of Arabic [11], [12]. Another issue is that the Arabic language includes a wide range of dialects [13]. In this article, we focus on the implemented artificial intelligence approaches applied, quality measurement performance, and dataset details (source, dialect, annotation methodology) used for offensive

detection in Arabic language. Future studies will be guided by this, since it will provide researchers with a more uniform and compatible viewpoint on the issue.

The rest of the article is structured as follows. Offensive language types, Arabic language issues, data preparation steps, feature representation techniques, AI approaches and related work are presented in section 2. In section 3, we look into Arabic datasets that have been used in previous studies. Section 4 discusses significant works and ongoing research in the area of Arabic social threat detection, as shown in section 5, which comprises an evaluation of the results and a discussion. Finally, in section 6, the conclusion is demonstrated.

2. BACKGROUND

2.1. Offensive language

It is a complicated undertaking to provide a precise definition of offensive language [14], [15]. Actually, personal knowledge and culture are both crucial elements in defining what is offensive and what is not. A word used in a written or spoken communication is offensive if it contains conduct meant to cause hurt, pain, or anger [10], [16]. Hate speech and cyberbullying are two significant forms of offensive language, their prevalence on social media has recently risen.

2.1.1. Hate speech

Hate speech is text directed against a number of individuals with the purpose of hurting people, leading to violence, or social upheaval [17]. It is described as "any use of modern digital technology to propagate racial, religious, extremist, or terrorist ideas" [18]. Hate speech can be classified into the following categories: gendered and religious [10], [19].

2.1.2. Cyberbullying

Cyberbullying is defined as an online assault directed at a specific person [20]. Due to the extreme nature of online resources, which can circulate harassment quickly and make it available to a larger audience, cyberbullying can have more serious consequences than physical and verbal abuse [21]. Cyberbullying may be classified in to nine types: flame, masquerade, impersonation, harassment, outing, deceit, exclusion, and cyberstalking [22].

2.2. Arabic language

Arabic is a Semitic language that is strongly tied to Islam and Muslim culture, and it is the language of the Quran, used by all Muslims (over 1.62 billion people) [23]. It is also the mother tongue of over 422 million people [24]. There are 28 alphabets in this language, and lines are expressed from right to left [25]. Arabic is spoken in a variety of dialects, including classical, modern standard, and numerous local dialects.

2.2.1. Dialects of Arabic languages

The most famous Arabic dialects used are the following [12]: Egyptian Arabic is spoken in both Egypt and Sudan.Gulf Arabic encompasses United Arab Emirates, Bahrain, and Saudi Arabia, Kuwait, and Qatar. North African Arabic encompasses Morocco, Algeria, Tunisia, and Mauritania. Levantine Arabic covers Palestine, Lebanon, Syria, and Jordan.

2.2.2. Challenges regarding to Arabic

Arabizi or Franco-Arab is a contemporary social media fad in which a person expresses Latin letters using Arabic characters. This problem has several negative impacts on Arabic categorization, yet it has received little attention in research [23], [25]. Arabic has a huge variety of dialects. There are various dialects of Arabic that vary between areas and even within the same nation. Furthermore, due to the numerous dialects, the Arabic data available online may contain terms with diverse meanings [26]. According to where the letters are in the word, the same letters may assume several different shapes. For instance, the letter "ق /kaf" can take the forms "ق/ قـ/ـقـ " depending on whether it's at the start, middle, or end of a word [13], [23]. Based on diacritical marks and punctuation, many words with the same spelling can have various pronunciations and meanings. The majority of existing Arabic materials are written without these markings, resulting in lexical confusion [23], [25].

2.3. Arabic text preprocessing

The data obtained via the internet is unstructured and must be preprocessed before being used in later stages [24], [27]. A great deal of work is required before preprocessing Arabic content on social media, since most of it will be informal (not standard) and may include dialects, misspellings, characters with diacritical marks, and elongations [12]. Therefore, additional processing, including removing elongations,

diacritical markings, and extra characters, is required when handling Arabic language [11], [25]. For example, it converts every letter into its standard form; for example, “alif” has numerous forms: "أ","إ" and "ا". After the text cleaning step, text normalization is applied. Some Arabic terms on social media are spelled in unconventional manners, such as using repeated letters. The normalization algorithm substitutes a non-normal word with a normal one by eliminating repeated letters and employing a collection of commonly used non-normal terms [27], [28]. As a result of the limitations listed above, cleaning Arabic text may be done using Arabic natural language processing (NLP) tools like:

- Tokenization: To fragment text into words, the tokenization algorithm employs spaces between words as well as punctuation such as stop signs, commas, and semicolons. Then, each word is saved in its own database column.
- Lemmatization: The mapping of a word form to its matching lemma, the canonical representation of its lexeme, is known as lemmatization. Lemmatization is a subset of the broader process of lexeme identification in which ambiguous lemmas are resolved further [13].
- Stemming: the act of eliminating affixes and suffixes from a word to isolate the root. Because there are so many alternative ways to represent text in Arabic, three stemming approaches are frequently used [25].

2.4. Feature representation

The most essential component of NLP pipeline is feature representation, which is critical for identifying abusive speech. As a result, many different feature kinds and combinations have been extensively examined to find the best successful approach. This section will go through most relative types of feature representations that have lately been utilized in the area of offensive speech detection.

2.4.1. Languages models

Text representations based on probabilities are known as language models. Due to their effectiveness and simplicity, Bag-of-words is one of the most common methods to represent any text, even if word order is not considered [18], [19]. The N-gram is another popular way to represent a text. This method is regarded as superior than the others. In addition, in the literature, char n-gram is the standard feature representation method. Despite its simplicity, it beats the term n-gram in similar tasks [19]. The term frequency inverse document frequency (TF-IDF) statistical technique is extensively utilized in NLP classification problems. TF-IDF often employed with basic machine learning classifiers and yields somewhat worse results when compared to the cutting edge feature representations. Other traditional classifications, such as word n-grams and char n-grams, have worse statistical performance compared to TF-IDF [19], [29].

2.4.2. Word embedding

Word embedding (sometimes referred to as word vectors) are numerical representations of texts that assist in language understanding via mathematical methods [30], [31]. Word embedding uses a vector space model that takes into account the word vectors' correlation to one another, enabling them to elucidate word meanings [2]. One of the approaches for producing word embedding is Word2Vec [32], [33]. Two distinct kinds of word2vec models are the skip-gram and the continuous bag of words (CBOW). The first foresees context words from a certain source word, whereas the second reverse and forecasts a word from its context window [30]. Moreover, AraVec [34] is a pre-trained word embedding for the representation of Arabic words, with a total of 1,169,075,128 tokens of Arabic words. FastText is a Facebook developed research library for fast learning of word representations and sentence categorization. Like Word2vec, FastText handles each word in a corpus differently. FastText considers each word to be made up of character n-grams. As a result, a word's output vector is the sum of its character n-grams [35].

2.5. Artificial intelligence

Artificial intelligence (AI) is a field that highlights the invention of intelligent computers that function and respond like people. Today's digital world generates an incredible amount of data. To carry out our tasks, all of these many aspects, including the web, sensors, software, gadgets, and several other variables, all give birth to vast amounts of organized, unstructured, and semi-structured data [8]. Data is a new type of oil that is essential but requires more processing before it can be used. This data is available to be used in order to enable computers to learn and then transfer that information to humans. Artificial intelligence is divided into many sub-fields, which are shown as follows.

2.5.1. Machine learning algorithms

In the current technological context, machine learning (ML) is one of the most significant fields of artificial intelligence. Machines are developed in such a manner that they can learn and comprehend from the enormous quantities of data that are accessible to them [8]. ML begins with a training phase, followed by a

decision phase. Training the model and feeding it data are the first two steps. In the second phase, the system generates predictions and modifies itself [25]. For text classification, there are two major subcategories of machine learning approaches.

a. Supervised learning

Supervised machine learning methods rely largely on labeled training data. Using labeled data, the model is trained. Unknown input is fed into the system after the model has been trained to provide the desired result [36]–[38].

b. Unsupervised learning

In this kind of learning method, the model delivers data that does not have any labels attached to it. Unsupervised learning algorithms include clustering, anomaly detection, and neural networks. Through the use of clusters or grouping similar things together, the goal is to uncover certain patterns and information in the dataset [39], [40].

c. Deep learning (DL)

Due to the increasingly excellent performance in many fields such as voice and handwriting recognition, computer vision, and lately NLP including text classification, deep learning has received a lot of attention from researchers [11]. Deep learning (DL) is a branch of machine learning that uses multilayer neural networks to learn data representations with many levels of abstraction. Various deep neural network methods have been used to tackle the text classification problem. Including convolution neural network (CNN), recurrent neural network (RNN), and long short term memory (LSTM). Transformers have recently become popular in NLP and text categorization. The transformer model is a deep neural network architecture based completely on the attention mechanism, replacing the recurrent layers with auto encoder-decoder architectures with special called multi-head self-attention layers [41].

2.6. Related work

Currently, there are obvious difficulties in the research on abusive language identification in the Arabic language. All of the research takes into account dialectal Arabic, which is widely used on social media. Only two research papers look at dialect differences and their impact on detecting hateful content [42], [43]. Haidar *et al*. [44] establish particular dimensions for the Middle East as a criterion in choosing their dataset. They include Lebanon, Syria, the Gulf Area, and Egypt. When they annotated their data, they did not take into account the implications of this significant variety in dialects. Sap *et al*. [45] examined the impact of social context on inappropriate language detection research, citing racial and cultural biases as being the most prevalent obstacles in analyzing foul language found in dataset labelling.

Furthermore, the majority of datasets originate from a single source. Only two multiplatform datasets were discovered in the review of the literature [46], [47]. Previous research has shown the uniqueness of foul language on each platform. As a result, a model constructed using a Twitter dataset cannot be used to Instagram, Facebook, or other social media platforms. Some platforms are popular in some nations but not in others. As a result, foul language on one platform does not reflect the accent and culture of the people who live in the missing nations. Users on social media frequently utilize emoji and emoticons to communicate their sentiments and attitudes about persons, subjects, and things. Treating them with the same weight as other textual material can help discover offensive words. However, we can only locate a few studies that take into account emoji or emoticons in the literature; most of them delete them during the pre-processing stage. The primary purpose is to detect harmful language; however, this does not mean that non-offensive material is overlooked. All investigations concentrate solely on thoroughly studying the offending samples, with no additional examination of the non-offensive samples [10].

3. ARABIC DATASETS USED IN PREVIOUS RESEARCH

The lack of Arabic corpora, lexicons, and tools restricts study in this field [11]. Furthermore, the available Arabic language analysis tools have significant shortcomings when it comes to coping with the language's complexity [23]. This section gives an overview of Arabic offensive language datasets that are publicly accessible. The most essential information about the datasets is provided in Table 1.

Mubarak *et al*. [48] provided MSA dataset for the purposes of identifying racist, sexist, abusive attacks, instigating, and irrelevant comments from Aljazeera.net users. Only the shorter comments (3 to 200 characters) were retained, reducing the final dataset to 32,000 remarks. Professional annotators categorized the dataset into three categories: obscene, offensive, and clean. This research also offered another dataset, which included 1,100 Egyptian tweets. The use of a dialectal Arabic dataset for offensive language was unprecedented, despite the fact that the dataset contained just a limited number of tweets. Albadi *et al*. [49] utilized Twitter data to build the first Arabic religious hate speech corpus. The data was extracted via the use of Arabic keywords and includes the six most significant: Christian, Islamic, Sunni, Shia, Jewish, and Atheist views. The training dataset consisted of 6,000 tweets; 1,000 of them indicated each religion or belief,

whereas the testing dataset comprised 600 tweets; 100 of those represented each religion or belief. Every tweet is associated with two levels of labels. Initially, annotators were tasked with assigning tweets to one of three categories: hate, neutral, or non-relevant (later excluded). Next, the hateful tweets in the class were given seven different labels: Shia, Sunnis, Muslims, Jews, Christians, atheists, and others. The Levantine Twitter dataset for hate speech and abusive language (L-HSAB) was publicly available in [50]. It is thought to be the first Arabic hate speech dataset focusing on the Levantine area. Political issues were the main theme of the dataset. The L-HSAB included 5,846 tweets, 3,650 of which were classified as "normal," 1,728 of which were categorized as "abusive," and 468 of which were labeled as "hate". Three native Levantine speakers from the region provided annotations for the dataset.

Haddad *et al*. [51] presented Tunisian hate and abusive speech (T-HSAB) dataset. Many political, social, religious, women's rights, and immigration problems were addressed by T-HSAB. Unfortunately, the authors did not specify which online sources they selected as a data source, although they did make it clear that the data was collected from social media sites between October 2018 and March 2019. Number of rows of the dataset was 6,075, including 3,834 normal commentaries, 1,127 abusive commentaries, and 1,078 hate commentaries. Mubarak *et al*. [52] introduced open-source Arabic corpora and corpora processing tools (OSACT) dataset. Tweets were chosen based on 2 factors: Tweets containing the vocative particle ﻳ/yA/O and released between April 15 and May 6, 2019. This publicly available OSACT 2020 Arabic dataset consisted of 10,000 thoroughly annotated tweets. For each tweet, a 2-level hierarchy was used. The highest degree of labeling was binary: either offensive or not offensive. Only 1,900 tweets out of 10,000 were offensive, and only 95 of those were hate speech.

Omr *et al*. [53] developed the first multi-platform dataset for detecting Arabic hate speech. This dataset was gathered using Facebook, Twitter, YouTube and Instagram. To gather data from each website automatically, the web crawler is used to discover pages, and then the data is stored in a text file. Annotating the instances was the responsibility of three native Arabic. Some of the samples were classified as hate speech, while others were not. This method was repeated until a balanced collection of 20,000 samples was created. The first dataset in Arabic language and collected from more than platform and contains many dialectal is Arabic Multi-Platform Offensive Language Dataset (MPOLD), released by Chowdhury *et al*. [47]. The data comes from three social networking sites: YouTube (40%), Facebook (20%) and Twitter (40%). There were 4,000 total comments, with 84.13% being non-offensive and 16.88 % being offensive.

Alakrot *et al*. [54] built a YouTube dataset based on choosing channels with contentious celebrity videos. Their final datasets contained a total of 167,549 comments from 84,354 users and 87,388 responses from 24,039 people in 150 videos on YouTube. Each comment is labeled by three annotators of various Arabic nations who represented the same nationalities as the bloggers' comments. The inter-annotator agreement was then calculated. Offensive statements were labelled positively, whereas non-offensive remarks were labelled negatively, and unclear comments were left unidentified. A further Twitter hate speech dataset was developed for the Arabic language by Aljarah *et al.* [55], although it contains less samples than the OSACT dataset. The total number of tweets in the final dataset was 3,696. Religion, racism, and journalism were among the topics covered in the collection of tweets. Using three categories, two annotators categorized the data. Following the removal of redundant and unnecessary tweets, the dataset included 843 hate speech tweets, 790 non-hate speech tweets, and 2,061 neutral tweets. For the collection of cyberbullying dataset, Almutiry *et al*. [56] introduced AraBully-Tweets dataset. Dataset gathered using Twitter API and ArabiTools from MSA tweets. A total of 17,749 Arabic tweets were gathered, including 14,178 Cyberbullying and 3,570 Non-Cyberbullying. The authors utilized AraBully-Words to annotate the dataset, which was done via the use of Python code. If a tweet contains Ara-Bully words, in this is the case, the tweet will be labeled as cyberbullying, if it doesn't, then tweet is considered non-cyberbullying. To evaluate the efficacy of Python-based automatic annotation. Three Arabic native speakers performed a manual annotation process after the Python-based automatic annotation.

Table 1. Datasets information

REF	Dataset Name	Dialect	Platform	Dataset Size	Annotation method
[48]	-	MSA	Aljazeera.net, Twitter	32K comments	CrowdFlower
[49]	-	Unspecified	Twitter	6,600 tweets	CrowdFlower
[50]	L-HSAB	Levantine	Twitter	5,846 tweets	Manual labeling
[51]	T-HSAB	Tunisian	Unspecified	6,075 comments	Manual labeling
[52]	OSACT	MSA	Twitter	10,000 tweets	Manual labeling
[53]	-	Unspecified	Facebook, Twitter, YouTube, Instagram	20,000 comments	Manual labeling
[47]	(MPOLD).	different dialects	Facebook, Twitter, YouTube	4,000 comments	crowdsourcing platform
[54]	-	Unspecified	YouTube comments	167,549 comments	Manual labeling
[55]	-	Unspecified	Twitter	3,696 tweets	Manual labeling
[56]	AraBully-Tweets	MSA.	Twitter	17,748 tweets	Automatic annotation

4. ARABIC TEXTUAL SOCIAL THREATS CLASSIFICATION USING AI

This part provides a thorough overview of the major works and current research in the field of automated detection, with a focus on Arabic textual social threats. Studies were divided into three categories: hate speech, cyberbullying, and general offensive or abusive behavior (for research that did not concentrate on a specific kind). Several methodologies used by different researchers were explained in the following section, which applied to the three mentioned categories. Table 2 summarizes the Arabic textual social threat classification using machine learning (offensive language, hate speech and cyberbullying).

Table 2. Summarization studies of Arabic textual social threats

Paper	Preparation steps	Features	ML Algorithm	Dataset label	Best accuracy	Best F1-Score
[57]	Tokenization Filtering. Normalization.	combinations of word-level, and N-gram features	SVM	Offensive/not Offensive	SVM, 90.05%	SVM, 82%
[58]	Remove non-Arabic letters and special characters. Removal of Emoticons. Shortening some of the letters.	AraVec word embedding and skip-gram model	CNN Bi-LSTM Attention Bi-LSTM Combined (CNN-LSTM)	Offensive/not Offensive	CNN , 87.84%	CNN, 84.05%
[59]	Normalization Tokenization. Elongation removal Removing unknown characters, diacritics, punctuation, URLs.	skip-gram word2vec embeddings,	NB BiLSTM BERT CNN-BiLSTM MTL MTL-S MTL-S-N	Offensive/not Offensive	NA	MTL-S-N, 90.4%
[47]	Removing diacritics, punctuation, stopwords, and URLs.	TF-IDF	SVM	Offensive/not Offensive	NA	SVM 84%
[60]	Diacritical removal Fixed words elongated Tokenization	TF-IDF, AraVev, word embeddings, and Word2Vec	tfidf+ LR CNN + Arave BiLSTM Multi-lingual BERT AraBERT	Offensive/not Offensive hate speech/ not hate speech	AraBERT, 92.8%	NA
[61]	Normalizing Removing diacritics Handling elongated words Stop words removal lemmatization	n-gram, Aravec	LR SVM GRU + word embeddings + handcrafted	hate speech/not hate speech	GRU + word embeddings 79%	GRU + word embeddings 77%
[51]	Eliminating Rt, @, and # Eliminating Emoji icons, digits.	Ngram and TFIDF	NB SVM	hate speech/not hate speech	NB 92.9%	NB 92.3%
[55]	Filtering out non-Arabic characters. Removing numbers, symbols, punctuation, hashtags, web addresses.	BoW, TF, TF-IDF, profile features, emotion features.	SVM DT NB RF	hate speech/not hate speech	RF 91.3%	NA
[4]	Remove non-Arabic characters. Remove all diacritics. Remove all punctuation Replace repeated characters with only one	n-grams, Mazajak embeddings	SVM Bagged SVM CNN-BiLSTM M-BERT Ensemble Method	hate speech/not hate speech	SVM 97.1%	Ensemble Method 79.3%
[56]	Removing non-Arabic Letters Removing duplicate tweets, re-tweets, and pictures of the tweets. Applying Khoja Stemmer	TF-IDF	SVM	Cyberbullying/ Non-Cyberbullying	SVM 85.49%	NA
[44]	NA	TweetToSentiStrength Feature	NB SVM	Cyberbullying/ Non-Cyberbullying	NA	SVM 92.7%
[31]	Eliminating hyperlinks Eliminating non-Arabic characters	one hot encoding	FFNN	Cyberbullying/ Non-Cyberbullying	FFNN 92.53%	NA

4.1. General offensive classification

Alakrot *et al*. [57] used a dataset of Arabic YouTube comments [54] to perform machine learning experiments for the detection of offensive language. The NLTK's stop-word list was used to filter the comments. Pre-processing includes ignoring diacriticals, non-alphabetic characters (e.g., punctuation, numerals) and text normalization. Text is converted to tokenized stems of words using light Arabic stemmer ARLSTem. N-gram (n=1-5) and word-level features were investigated. After completely preprocessing texts, the best performance scores were obtained using support vector machine (SVM)-based classifier with 10-fold cross validation and word-level features, which achieved 90.05%. Mohaouchane *et al*. [58] attempted to tackle the problem of detecting offensive language on Arabic social networking sites automatically. They applied four deep learning models (CNN, Bidirectional LSTM, the bidirectional LSTM with attention mechanism, and the combined CNN-LSTM) on the Arabic dataset mentioned in [54]. The authors utilized Arabic word embedding to represent the comments after running the dataset through a number of pre-processing procedures. Overall, the findings showed that CNN had better accuracy, precision, and F1-score than other algorithms, while the combined CNN-LSTM model outperformed other models in terms of recall. OSACT dataset [52] was used by Abu Farha *et al*. [59] to compare several multitask deep learning models. The word vectors were created using the Arabic word embedding Mazajak, thus the researchers employed the identical preparation methods that were utilized to generate Mazajak on the tweets. Letter normalization for Ya, Ha, and Hamza, elongation elimination, and ordinary cleaning, such as removing unfamiliar letters, diacritics, punctuation, and URLs are all part of this preparation steps. Different classifiers were examined by authors: Convolutional Neural Networks and bidirectional long short term memory neural network (CNN-BILSTM), M-BERT, multinomial naive bayes (MNB), and BILSTM. Furthermore, the authors used CNN-BILSTM to create multitask learning classifiers that included three different versions. The third multitask learning model outperforms the first and second models in offensive language detection, achieving a 90.4% macro F1 score.

Chowdhury *et al*. [47] proposed a system that can detect abusive language using a multi-platform dataset. The authors evaluated the system using a variety of datasets mentioned in the previous section, including the MPOLD [47], the L-HSAB [50], the Egyptian tweets dataset, and the Aljazeera.net deleted comments datasets. Text was tokenized before it was preprocessed, and stop words, URLs, diacritics, and punctuation were filtered out. Emoji and hashtags were preserved as well, given their contexts. Using the Leave-One-Platform-Dataset-Out approach (LOPO), the authors used an SVM-based model and evaluated it using 5-fold cross-validation. To provide an example, utilizing Facebook, Aljazeera.net and YouTube datasets for model evaluation while Twitter dataset applied for model training. Findings demonstrated little generalization and poor model performance with one platform dataset compared to multi-platform dataset training. It was stated that the model was effective with a macro F1 score of 84%. Keleg *et al*. [60] utilized OSACT dataset [52] to test various classification models for offensive language detection: AraBERT, M-BERT, CNN, BiLSTM, and logistic regression (LR). For each classification model, several preprocessing methods were used, including diacritics removal, correcting elongated words, and tokenization. A variety of features, such as one to nine-character n-grams, TF-IDF, AraVev word embedding, and word2vec, have been implemented as well. The AraBERT model, which had a list of profane words added to it, had the best macro F1 score of 88%.

4.2. Hate speech classification

Albadi *et al*. [61] utilized dataset that depends on the Arabic language in its construction for hate speech classification [49]. The pre-processing operations of stemming and filtration texts were supported by SRIS-temmer and MADAMIRA 2.1. Their model was constructed on the lexicons arahate-bns, arahate-pmi, and arahate-chi, which gave each tweet a score depending on the lexicon terms to which the tweet matched. They also designed a (GRU- RNN) model based on the AraVec Twitter-CBOW 300-dimension embedding model with 32-batch size and Adam as the optimizer, as well as (LR and SVM-based models) with a character n-gram feature (n=1 to 4). Using the proposed GRU-based model, the highest performance of F-score is 77%. An 84% recall rate is achieved by adding temporal, user, and content features to the model.

Haddad *et al*. [51] used the T-HSAB dataset which was mentioned earlier. Unigrams, bigrams, and trigrams have all been used to build traditional machine learning classifiers, including SVM and NB. The term frequency weighting was employed to decrease the dimensionality of the features. In all evaluation metrics, the NB model outperformed the SVM. The NB model obtained an F1 score of 83.6%, a recall of 79.8%, a precision of 89.5%, and an accuracy of 87.9%. As previously mentioned, the Twitter hate speech dataset [55], authors in this paper did not include samples from the neutral class. This led to the use of binary classifiers. After eliminating non-Arabic letters, numerals, symbols and punctuation from the data and filtering stop words and negation words, the data was ready for analysis. Several features were examined in the research, including emotion features, TF vectors, BoW vectors, profile features, TF-IDF vectors. In the

course of designing the system, the authors examined the random forest model, the Gaussian NB, and SVM. Ten -fold cross-validation and Grid search both applied to assess the performance of the model.

The study found that the random forest model, which included features based on TF-IDF with additional profile features, had the best overall geometric mean of 91.2% and the greatest performance of 91.3%. The OSACT4 dataset [52] was utilized by Hassan *et al*. [4]. In order to remove non-Arabic letters, diacritics, and punctuation, and to restrict characters' recurrence to one, they conducted fundamental preprocessing operations. convolutional neural networks with image features (CNN-BILSTM) and an ensemble of SVM, bagged SVM, and CNN-BiLSTM were used to develop classifiers. SVM, bagged SVM, and CNN-BiLSTM are employed in a more complex ensemble classifier. Using 1 to 5 n-gram characters, 2 to 6 n-gram characters, 1 to 3 n-gram words, pre-trained Mazajak word embedding feature extractors, and CNN feature extractors, the ensemble classifier outperformed all other classifiers. The top accuracy, was 97.7%.

4.3. Cyberbullying classification

As previously mentioned, the AraBully-Tweets Dataset [56], basic preprocessing is carried out. Authors in this research eliminated non-Arabic letters, user mentions (@user), single Arabic characters, numerical special characters (%, &, +, /, %), duplicate tweets, re-tweets, and photos of tweets. The next stage, after data cleaning, is data normalization in which every Arabic word form is converted into a consistent form. Light Stemmer and ArabicStemmerKhoja are the two stemmers that are employed. During three experiments, the authors used the SVM method. They employed WEKA and Light Stemmer on the first one. Second, they applied ArabicStemmerKhoja with WEKA. Python is used to carry out the final experiment. The findings indicate that WEKA is more efficient in properly classifying the text, with a performance of 85.49%, while Python is more effective, taking 142.68 seconds to construct the model.

Haidar *et al*. [44], depending on the posts' geographic location, they able to gather a dataset (4.93 GB) from Twitter. The dataset included the bulk of posts from Egypt, Lebanon, Syria, and the Gulf area. After removing duplicates and samples in languages other than Arabic and English, they partitioned the dataset into two datasets: An Arabic dataset of 35,273 tweets and an English dataset of 91,431 tweets. The SentiStrength tool was used to produce the Tweet-To-SentiStrength-Feature-Vector. Classifiers were implemented utilizing NB and SVM-based learning techniques. Despite the system's aim to offer multilingual functionality, the researchers trained and tested the system on only Arabic datasets. The findings of NB revealed an F1 score of 90.5% without the use of any features on the model, while for the SVM the results were 92.7% with the use of a both Tweet-to-SentiStrength Vector and the transformation of strings to word vectors.

The prior investigation was extended by Haidar *et al*. [31] to incorporate deep learning models. For Arabic cyberbullying detection, authors created a feed forward neural network (FFNN), also known as a multilayer perceptron. The same dataset used in their prior research was utilized in this investigation, with minor changes. All hyperlinks, non-Arabic characters, emoticons were removed. The dataset was then subdivided into two datasets: one with 4,913 tweets (1,688 bullying) and the other with 34,890 tweets (3,015 bullying). FFNN using 2 epochs, 7 hidden layers, and a batch size of 16 provided the greatest outcome. This yielded a tested accuracy of 92.53%.

5. DISCUSSION AND LEARNED LESSONS

Unbalanced classes were a frequent issue in the majority of datasets. We observed that the OSACT dataset had become something of a standard in research of abusive language. When generating datasets for abusive language identification, only MSA, Levantine, and Tunisian dialects were considered. The methods used to preprocess Arabic text differ from those used to preprocess English text. Also, some methods utilized included Kashida removal, diacritics removal, ArabicStemmerKhoja, AraVec word embeddings, and Mazajak word embeddings. Across all of the literature, the SVM-based classifier was the most popular algorithm used in offensive language detection. The CNN, NB, and GRU-based classifiers were the next models used after the SVM. In the literature, a variety of assessment measures were employed, including accuracy and F1. Machine learning using an SVM-based classifier has achieved the greatest record accuracy of 97.1% in detecting hate speech. Deep learning's greatest accuracy is 92.8% by applying AraBERT. The F1 score takes into account the tradeoff between precision and recall, as a result, provides a more realistic assessment of categorization performance. In identifying cyberbullying, machine learning employing an SVM-based classifier has reached the highest record accuracy of 92.7%. Deep learning that employs multitask learning using (CNN-BiLSTM) recorded the greatest F-score of 90.4% when it comes to identifying offensive language.

6. CONCLUSION

Lately, many researchers have recently become interested in detecting Arabic offensive language on social networks using artificial intelligence. The suggested methods to identify the issue of Arabic offensive language are discussed in this article, which includes various forms of offensive language like hate speech and cyberbullying. Included are the techniques utilized, performance metrics, and dataset characteristics (dialect, annotation method, and platform). In the study's findings, it is shown that the topic study is in its initial stages, and most techniques have not yet been used to identify a practical classification system for Arabic text. Even yet, only a small number of Arabic datasets are available for offensive categorization. As a consequence, this work is challenging due to the restricted amount of datasets, complex pre-processing procedures, and a lack of publications in this area.

REFERENCES

[1] S. Rathore, P. K. Sharma, V. Loia, Y. S. Jeong, and J. H. Park, "Social network security: Issues, challenges, threats, and solutions," *Information Sciences*, vol. 421, pp. 43–69, 2017, doi: 10.1016/j.ins.2017.08.063.

[2] K. Cortis and B. Davis, *Over a Decade of Social Opinion Mining*, vol. 2. Springer Netherlands, 2020.

[3] I. Science and H. H. Universit, *Arabic Language Processing: From Theory to Practice*, vol. 782, no. April. 2018.

[4] S. Hassan, Y. Samih, H. Mubarak, A. Abdelali, A. Rashed, and S. A. Chowdhury, "{ALT} Submission for {OSACT} Shared Task on Offensive Language Detection," *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, no. May, pp. 61–65, 2020, [Online]. Available: https://www.aclweb.org/anthology/2020.osact-1.9

[5] Samantha Kent, "German Hate Speech Detection on Twitter," *Proceedings of GermEval 2018, 14th Conference on Natural Language Processing (KONVENS 2018)*, no. Konvens, pp. 120–124, 2018.

[6] F. Fortin, J. Delle Donne, and J. Knop, "The Use of Social Media in Intelligence and Its Impact on Police Work," in *Policing in an Age of Reform*, no. January, 2021, pp. 213–231.

[7] E. Hamdy, "Neural Models for Offensive Language Detection," *arXiv*, 2021.

[8] S. Mehra and M. Hasanuzzaman, "Detection of Offensive Language in Social Media Posts by Sidharth Mehra This thesis has been submitted in partial fulfillment for the," Thesis, Cork Institute of Technology, 2020.

[9] J. Risch, "Reader Comment Analysis on Online News Platforms," Thesis, Universität Potsdam, 2020.

[10] F. Husain and O. Uzuner, "A Survey of Offensive Language Detection for the Arabic Language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 1, pp. 1–44, 2021, doi: 10.1145/3421504.

[11] A. Wahdan, S. Hantoobi, S. A. Salloum, and K. Shaalan, "A systematic review of text classification research based on deep learning models in Arabic language," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 6, pp. 6629–6643, 2020, doi: 10.11591/IJECE.V10I6.PP6629-6643.

[12] S. M. Abdou and A. M. Moussa, "Arabic Speech Recognition: Challenges and State of the Art," *Computational Linguistics, Speech and Image Processing for Arabic Language*, pp. 1–27, 2018, doi: 10.1142/9789813229396_0001.

[13] N. Y. Habash, *Introduction to Arabic natural language processing*, Synthesis Lectures on Human Language Technologies, vol. 3, no. 1. 2010.

[14] J. Risch, R. Ruff, and R. Krestel, "Explaining Offensive Language Detection," *Journal for Language Technology and Computational Linguistics*, vol. 34, no. 1, pp. 1–19, 2020.

[15] L. Wu, F. Morstatter, K. M. Carley, and H. Liu, "Misinformation in Social Media: Definition, Manipulation, and Detection," *ACM SIGKDD Explorations Newsletter*, vol. 21, no. 2, pp. 80–90, 2019.

[16] R. Pradhan, A. Chaturvedi, A. Tripathi, and D. K. Sharma, "A review on offensive language detection," in *Lecture Notes in Networks and Systems*, vol. 94, no. January, 2020, pp. 433–439.

[17] G. Kovács, P. Alonso, and R. Saini, "Challenges of Hate Speech Detection in Social Media," *SN Computer Science*, vol. 2, no. 2, pp. 1–15, 2021, doi: 10.1007/s42979-021-00457-3.

[18] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in social networks: A survey on multilingual corpus," in *Computer Science & Information Technology(CS & IT)*, Feb. 2019, no. March, pp. 83–100, doi: 10.5121/csit.2019.90208.

[19] O. Istaiteh, R. Al-Omoush, and S. Tedmori, "Racist and Sexist Hate Speech Detection: Literature Review," *2020 International Conference on Intelligent Data Science Technologies and Applications, IDSTA 2020*, pp. 95–99, 2020, doi: 10.1109/IDSTA50958.2020.9264052.

[20] W. N. Hamiza Wan Ali, M. Mohd, and F. Fauzi, "Cyberbullying Detection: An Overview," *Proceedings of the 2018 Cyber Resilience Conference, CRC 2018*, no. November, pp. 1–3, 2019, doi: 10.1109/CR.2018.8626869.

[21] K. R. Talpur, S. S. Yuhaniz, N. N. B. A. Sjarif, B. Ali, and N. B. Kamaruddin, "Cyberbullying detection: Current trends and future directions," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 16, pp. 3197–3208, 2020.

[22] B. Haidar, M. Chamoun, and A. Serhrouchni, "A multilingual system for cyberbullying detection: Arabic content detection using machine learning," *Advances in Science, Technology and Engineering Systems*, vol. 2, no. 6, pp. 275–284, 2017, doi: 10.25046/aj020634.

[23] O. Oueslati, E. Cambria, M. Ben HajHmida, and H. Ounelli, "A review of sentiment analysis research in Arabic language," *Future Generation Computer Systems*, vol. 112, pp. 408–430, 2020, doi: 10.1016/j.future.2020.05.034.

[24] M. O. Hegazi, Y. Al-Dossari, A. Al-Yahy, A. Al-Sumari, and A. Hilal, "Preprocessing Arabic text on social media," *Heliyon*, vol. 7, no. 2, p. e06191, 2021, doi: 10.1016/j.heliyon.2021.e06191.

[25] M. Alruily, "Classification of Arabic Tweets : A Review," *Electronics*, vol. 10, no. 10, 2021, doi: 10.3390/electronics10101143.

[26] S. O. Alhumoud, M. I. Altuwaijri, T. M. Albuhairi, and W. M. Alohaideb, "Survey on Arabic Sentiment Analysis in Twitter," *International Journal of Social, Behavioral, Educational, Economic and Management Engineering*, vol. 9, no. 1, pp. 364–368, 2015.

[27] H. Elzayady, K. M. Badran, and G. I. Salama, "Arabic Opinion Mining Using Combined CNN - LSTM Models," *International Journal of Intelligent Systems and Applications*, vol. 12, no. 4, pp. 25–36, 2020, doi: 10.5815/ijisa.2020.04.03.

[28] B. Y. AlHarbi, M. S. AlHarbi, N. J. AlZahrani, M. M. Alsheail, and D. M. Ibrahim, "Using machine learning algorithms for automatic cyber bullying detection in Arabic social media," *Journal of Information Technology Management*, vol. 12, no. 2, pp. 123–130, 2020, doi: 10.22059/JITM.2020.75796.

[29] J. Salminen, M. Hopf, S. A. Chowdhury, S. gyo Jung, H. Almerekhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–34, 2020, doi: 10.1186/s13673-019-0205-6.
[30] A. A. Altowayan and L. Tao, "Word embeddings for Arabic sentiment analysis," *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, pp. 3820–3825, 2016, doi: 10.1109/BigData.2016.7841054.
[31] B. Haidar, M. Chamoun, and A. Serhrouchni, "Arabic Cyberbullying Detection: Using Deep Learning," *Proceedings of the 2018 7th International Conference on Computer and Communication Engineering, ICCCE 2018*, pp. 284–289, 2018, doi: 10.1109/ICCCE.2018.8539303.
[32] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in Arabic tweets using deep learning," *Multimedia Systems*, no. 0123456789, 2021, doi: 10.1007/s00530-020-00742-w.
[33] A. Alharbi, M. Taileb, and M. Kalkatawi, "Deep learning in Arabic sentiment analysis: An overview," *Journal of Information Science*, vol. 47, no. 1, pp. 129–140, 2021, doi: 10.1177/0165551519865488.
[34] A. B. Soliman, K. Eissa, and S. R. El-beltagy, "ScienceDirect ScienceDirect AraVec : A set of Arabic Word Embedding Models for use in Arabic NLP," *Procedia Computer Science*, vol. 117, pp. 256–265, 2017, doi: 10.1016/j.procs.2017.10.117.
[35] S. Engineering, "A Comparative Evaluation of Deep Learning based Transformers for Entity Resolution Master Thesis Mohammad Mohammadkhani," Otto-von-Guericke-University Magdeburg, 2020.
[36] P. Yang and Y. Chen, "A survey on sentiment analysis by using machine learning methods," *Proceedings of the 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2017*, vol. 2018-January, pp. 117–121, 2018, doi: 10.1109/ITNEC.2017.8284920.
[37] N. Benchettara, R. Kanawati, and C. Rouveirol, "Supervised machine learning applied to link prediction in bipartite social networks," *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*, pp. 326–330, 2010, doi: 10.1109/ASONAM.2010.87.
[38] M. Desai and M. A. Mehta, "Techniques for sentiment analysis of Twitter data: A comprehensive survey," *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, pp. 149–154, 2017, doi: 10.1109/CCAA.2016.7813707.
[39] S. Jaeger, S. Fulle, and S. Turk, "Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition," *Journal of Chemical Information and Modeling*, vol. 58, no. 1, pp. 27–35, 2018, doi: 10.1021/acs.jcim.7b00616.
[40] Y. Ko and J. Seo, "Automatic text categorization by unsupervised learning," pp. 453–459, 2000, doi: 10.3115/990820.990886.
[41] M. Abdul-Mageed, A. R. Elmadany, and E. M. B. Nagoudi, "ARBERT & MARBERT: Deep bidirectional transformers for Arabic," *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pp. 7088–7105, 2021, doi: 10.18653/v1/2021.acl-long.551.
[42] M. S. Jahan and M. Oussalah, "A systematic review of Hate Speech automatic detection using Natural Language Processing," *arXiv preprint arXiv:2106.00742*, no. May, May 2021, [Online]. Available: http://arxiv.org/abs/2106.00742.
[43] R. ALBayari, S. Abdullah, and S. A. Salloum, "Cyberbullying Classification Methods for Arabic: A Systematic Review," in *The International Conference on Artificial Intelligence and Computer Vision*, 2021, no. May, pp. 375–385, doi: 10.1007/978-3-030-76346-6_35.
[44] B. Haidar, M. Chamoun, and A. Serhrouchni, "Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content," *2017 1st Cyber Security in Networking Conference, CSNet 2017*, vol. 2017-Janua, pp. 1–8, 2017, doi: 10.1109/CSNET.2017.8242005.
[45] M. Sap, D. Card, S. Gabriel, Y. Choi, and N. A. Smith, "The risk of racial bias in hate speech detection," *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 1668–1678, 2020, doi: 10.18653/v1/p19-1163.
[46] S. A. Azzi and C. B. O. Zribi, *From Machine Learning to Deep Learning for Detecting Abusive Messages in Arabic Social Media: Survey and Challenges*. Springer International Publishing, 2021.
[47] S. A. Chowdhury, H. Mubarak, A. Abdelali, S. G. Jung, B. J. Jansen, and J. Salminen, "A multi-platform arabic news comment dataset for offensive language detection," *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, no. May, pp. 6203–6212, 2020.
[48] H. Mubarak, K. Darwish, and W. Magdy, "Abusive Language Detection on Arabic Social Media," in *Proceedings of the First Workshop on Abusive Language Online*, 2017, pp. 52–56, doi: 10.18653/v1/W17-3008.
[49] N. Albadi, M. Kurdi, and S. Mishra, "Are they our brothers? analysis and detection of religious hate speech in the Arabic Twittersphere," *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018*, pp. 69–76, 2018, doi: 10.1109/ASONAM.2018.8508247.
[50] H. Mulki, H. Haddad, C. Bechikh Ali, and H. Alshabani, "L-HSAB: A Levantine Twitter Dataset for Hate Speech and Abusive Language," in *Proceedings of the Third Workshop on Abusive Language Online*, 2019, pp. 111–118, doi: 10.18653/v1/W19-3512.
[51] H. Haddad, H. Mulki, and A. Oueslati, "T-HSAB: A Tunisian Hate Speech and Abusive Dataset," *Communications in Computer and Information Science*, vol. 1108, no. October, pp. 251–263, 2019, doi: 10.1007/978-3-030-32959-4_18.
[52] H. Mubarak, K. Darwish, W. Magdy, T. Elsayed, and H. Al-Khalifa, "Overview of {OSACT}4 {A}rabic Offensive Language Detection Shared Task," *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, no. May, pp. 48–52, 2020.
[53] A. Omar and T. M. Mahmoud, *Comparative Performance of Machine Learning and Deep Learning Algorithms for Arabic Hate Speech Detection in OSNs*, no. March. Springer International Publishing, 2020.
[54] A. Alakrot, L. Murray, and N. S. Nikolov, "Dataset Construction for the Detection of Anti-Social Behaviour in Online Communication in Arabic," *Procedia Computer Science*, vol. 142, pp. 174–181, 2018, doi: 10.1016/j.procs.2018.10.473.
[55] I. Aljarah *et al.*, "Intelligent detection of hate speech in Arabic social network: A machine learning approach," *Journal of Information Science*, 2020, doi: 10.1177/0165551520917651.
[56] S. Almutiry and M. A. Fattah, "Arabic CyberBullying Detection Using Arabic Sentiment Analysis," *The Egyptian Journal of Language Engineering*, vol. 8, no. 1, pp. 39–50, 2021.
[57] A. Alakrot, L. Murray, and N. S. Nikolov, "Towards Accurate Detection of Offensive Language in Online Communication in Arabic," *Procedia Computer Science*, vol. 142, pp. 315–320, 2018, doi: 10.1016/j.procs.2018.10.491.
[58] H. Mohaouchane, A. Mourhir, and N. S. Nikolov, "Detecting Offensive Language on Arabic Social Media Using Deep Learning," *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*, no. December, pp. 466–471, 2019, doi: 10.1109/SNAMS.2019.8931839.

[59] I. Abu Farha and W. Magdy, "Multitask Learning for {A}rabic Offensive Language and Hate-Speech Detection," *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, no. May, pp. 86–90, 2020.
[60] A. Keleg, S. R. El-Beltagy, and M. Khalil, "ASU OPTO at OSACT4-Offensive Language Detection for Arabic text," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 2020, no. May, pp. 11–16.
[61] N. Albadi, M. Kurdi, and S. Mishra, "Investigating the effect of combining GRU neural networks with handcrafted features for religious hatred detection on Arabic Twitter space," *Social Network Analysis and Mining*, vol. 9, no. 1, pp. 1–19, 2019, doi: 10.1007/s13278-019-0587-5.

BIOGRAPHIES OF AUTHORS

Hossam Elzayady is a Ph.D. candidate at the Department of computer engineering, received a Bachelor Degree in computer engineering and Masters of Science degree from the MTC, Cairo, Egypt, in 2005 and 2018, respectively. His research interests are in artificial intelligent, data science, machine learning. He can be contacted at email: hossamelzaiade@gmail.com.

Mohamed S. Mohamed received his M.S. degree (2011) and B.S (2004) in Computer Engineering from Military Technical College, Egypt (MTC). He also received his Ph.D. degree (2018) in electrical and computer engineering from University of Idaho, USA (UI). He is currently a faculty member at electrical and computer engineering department in MTC. His researches focus on cyber security, malicious act, variabilities issues related to connected vehicles, survivable systems and networks. He can be contacted at email: mohamedms@mtc.edu.eg.

Khaled M. Badran received a Bachelor Degree in computer engineering and Masters of Science degree from the MTC, Cairo, Egypt, in 1995 and 2000, respectively. He also received the Ph.D. degree in Electrical and Computer engineering from Sheffield University, UK, in 2009. He is currently a faculty member of the Department of Computer Engineering, MTC. His research interests are in artificial intelligent, data mining, semantic web and database security. He can be contacted at email: khaledbadran@mtc.edu.eg.

Gouda I. Salama received the Bachelor engineering and Masters' engineering degrees from MTC, Cairo, Egypt, in 1988 and 1994, respectively. As well, he received the Ph.D. degree in Electrical and computer engineering from Virginia Tech. University, U.S.A., in 1999. He is currently a faculty member with the Department of Computer Engineering, MTC. His research interests are in image and video processing, pattern recognition, and information security. He can be contacted at email: gisalama@mtc.edu.eg.