Iraqi Journal of Science, 2022, Vol. 63, No. 1, pp: 409-419
DOI: 10.24996/ijs.2022.63.1.37

ISSN: 0067-2904

A Survey on Arabic Text Classification Using Deep and Machine Learning Algorithms

Farah A. Abdulghani*, Nada A.Z. Abdullah
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 20/3/2021 Accepted: 27/6/2021

Abstract
Text categorization refers to the process of grouping text or documents into classes or categories according to their content. Text categorization process consists of three phases which are: preprocessing, feature extraction and classification. In comparison to the English language, just few studies have been done to categorize and classify the Arabic language. For a variety of applications, such as text classification and clustering, Arabic text representation is a difficult task because Arabic language is noted for its richness, diversity, and complicated morphology. This paper presents a comprehensive analysis and a comparison for researchers in the last five years based on the dataset, year, algorithms and the accuracy they got. Deep Learning (DL) and Machine Learning (ML) models were used to enhance text classification for Arabic language. Remarks for future work were concluded.

Keywords: Arabic Text Classification, Neural Networks, Deep Learning, Machine Learning.

دراسة بحثية: تصنيف النص العربي باستخدام خوارزميات التعلم العميق والآلي

فرح علاء عبدالغني* , ندا عبدالزهرة عبدالله
قسم الحاسوب ، كلية العلوم ، جامعة بغداد ، بغداد ، العراق

الخلاصة
يشير تصنيف النص إلى عملية تجميع النصوص أو المستندات الى أصناف وفقا لمحتواها. تتكون عملية تصنيف النص من ثلاث مراحل هي: المعالجة المسبقة واستخراج مزايا النص والتصنيف. بالمقارنة مع اللغة الإنجليزية، لم يتم إجراء سوى القليل من الدراسات لتصنيف اللغة العربية. بالنسبة لمجموعة متنوعة من التطبيقات، مثل تصنيف النصوص وتتجمعها، فإن تمثيل النص العربي مهمة صعبة لأن اللغة العربية تشتهر بثرائها وتنوعها والمورفولوجيا المعقدة. تقدم هذه الورقة تحليلاً شاملاً ومقارنة للباحثين في السنوات الخمس الماضية استناداً إلى البيانات والسنة والخوارزميات والدقة التي حصلوا عليها. استعملت نماذج التعلم العميق (DL) والتعلم الآلي (ML) لتعزيز تصنيف النص للغة العربية. كما نختتم بعض الملاحظات للعمل في المستقبل.

1. Introduction

Finding useful knowledge on a given subject in a vast volume of online textual data that is rapidly growing is a difficult challenge. To solve this issue, organize data into predetermined categories could help. Algorithms of text classification are the basis of many applications for

*Email: farah.ahmed1201@sc.uobaghdad.edu.iq

natural language processing, such as text description, query response, detection of spam, and visualization of text. [1].
While Arabic language on the internet is rising increasingly, its content is still as poor as 3 percent. For researchers and developers, the recent rapid growth is a convincing incentive to develop successful frameworks and tools to advance study in Arabic NLP. The automated mapping of texts to predefine marks or classes is text categorization [1]. Methods of text classification are used in several applications, like e-mail search, filtering of spam and classification of news [2].
The main role of text classification can be described as follows: offered a D document, locate zero or many groups to where the document D belong. The process of binary classification requires a collection of two classes where as a multi classification process operates on more than two types of data gathered for assigning them to an unseen text. Categorization of text may be manual or automated. Since the early days, manual text classification has been the central role of classifying library meaning. Automatic text classification is performed primarily by computing device using classification algorithms [1].
For several persons and applications, classifying Arabic documents into particular groups is of high significance. In this survey, we are proposing a groundbreaking approach for deep learning to identify Arabic text documents by using new technologies of deep learning and algorithms to produce better outcomes. Deep learning has made extraordinary strides in speech recognition and machine vision [3].
The reset of this paper is as follows: In section two we present a review of the most important researches in the last five years. Section three illustrates the challenges of Arabic language. Classification model is presented in section four. In section five a brief comparison for Arabic text classification researches is illustrated. We conclude from the comparison some conclusions which presented in section 6, at last we present recommendations for future work.

2. Researches Used ML and DL for Arabic Text Classification in [2016-2020]

Several papers addressed the issue of automated categorization of text that proposes numerous techniques and solutions.
El-Alami et al. (2016) [4], for Arabic Text Categorization (ATC), they suggested an effective approach based on deep learning, using a deep stacked autoencoder that has word-count vectors as input. They used Restricted Boltzmann Machines (RBM) in the pre-training stage, then to make the deep network, they unrolled the model and backpropagation is used during the fine-tuning stage. They used decision tree, support vector machine and naïve bayes, their result showed that deep autoencoder worked good in Arabic text classification specifically for support vector machine.
Altaher (2017) [5], proposed a mixed approach focused on deep learning for sentiment analysis of Arabic tweets using features weighting. They used Term Frequency and Inverse Text Frequency (TF-IDF) as a feature selection to pick the most terms occurred in tweets and then they used features weighting to pick the most significant features. The deep leaning is used to examine the sentiment of Arabic tweets based on the chosen features as a strong emerging technique. The outcomes demonstrated the feasibility of their hybrid approach based upon deep learning with feature weighting (information gain and chai-square) approach and regarding accuracy and precision shown that the hybrid approach has outperforms the SVM, DT and NN classifiers and achieved the best efficiency.
Al-khurayji and Sameh (2017) [6], proposed a new method using kernel naïve bayes for Arabic text classification. At first, they preprocessed documents such as tokenize the Arabic words then removing the stop word and using word stemming. They used Term Frequency and Inverse Text Frequency (TF-IDF) as a feature extraction, they transformed those terms into vectors. Third, to solve the non-linearity issue of Arabic text classification, they suggested a successful solution based on the Kernel Naive Bayes (KNB) classifier. Finally,

Experimental findings on the collected dataset revealed that their methodology regarding precision and time against other baseline classifiers obtained excellent results upon the proposed classifier.
Boukil et al. (2018) [7], they suggested simple and precise technique for categorizing Arabic data set, to isolate, choose and decrease features they used an Arabic stemming algorithm. Then for feature weighting they used Term Frequency Inverse Document Frequency (TF-IDF). With CNN model and other standard machine learning methods, they analyzed their dataset as a benchmark. They argued that the CNN model performs well on the Arabic text classification challenge. Standard methods, such as SVM, do not do as well as the CNN model at the stage where the dataset is large and big.
Galal et al. (2019) [8], They concentrated on classifying Arabic Text using convolution neural network (CNN), as it achieved an excellent result in various processes of natural language (NLP), they also implemented a new algorithm focused on extra Arabic letters and word embedding distances to group related Arabic words. The algorithm name is Gstem. Their studies have shown that it improves the accuracy of the CNN model by using GStem as a preprocessing stage, since the number of separate terms has been decreased.
Elnagar et al. (2019) [9], for Arabic text classification they introduced large corpora. Named them as SANAD and NADiA. In order to enhance the efficiency of the classification tasks, they researched the effect of using word2vec embedding models. The result showed that convolutional-Gated recurrent unit (GRU) performed the lowest effectiveness and the attention-GRU preformed highest effectiveness, their experimental findings showed solid performance of both SANAD corpus models. Also, the attention-GRU achieved the highest effectiveness for NADiA.
Sundus et al. (2019) [1], they introduced a supervised feed forward Deep learning. The input of the first layer of deep learning is the term frequency inverse document frequency of the common words of datasets used. The first layer's output was used as the input to the next layer. They used supervised logistic regression of machine learning model. Compared to logistic regression, experimental studies demonstrated a substantial increase in efficiency of classification and time of building the model of deep learning model. The findings showed that Arabic text classification issue is very promising with deep learning classification models.
Alhawarat and Aseeri (2020) [3] implemented a CNN multi-kernel architecture with word embedding and specifically n-gram to classify Arabic documents of news and they named the model as a Superior Arabic Text Categorization Deep Model (SATCDM). regarding to the current studies on Arabic text classification, their approach achieves very high precision using 15 of the publicly available datasets.
Hazım et al. (2020) [10] they used a common type of Recurrent Neural Network (RNN) which is the Long Short-Term Memory (LSTM). They evaluate Arabic user comments on twitter, they showed that LSTM has more accurate performance with regarding to lower calculation of parameters, reduced working time and better efficiency compared to traditional pattern recognition techniques.
El-Alami et al. (2020) [11] they proposed an Arabic text categorization method based on Bag-of-Concepts and deep Autoencoder representations. It incorporates explicit semantics relying on Arabic WordNet and exploits Chi-Square measures to select the most informative features. successive stacks of Restricted Boltzmann Machines (RBMs) were applied to text vectors to produce high-level representations. The learned features were fed to another deep Autoencoder for categorization. An exhaustive set of experiments was carried out and has shown that using the Autoencoder as text representation model combined with Chi-Square and classifier outperformed state-of-the-art techniques which achieved the efficient results.

3. Arabic Language Characteristics

Arabic features are assorted in abundance aspects compared to English language [12]. Arabic is a global language that is commonly used and has considerable variations compared to the most common, such as Spanish, English and Chinese. There are several forms of grammatical, variations of synonyms word, and numerous meanings of word in the Arabic language, which differ based on factors such as order of the word. although such difficulties, the work on natural language processing (NLP) with Arabic has been minimal especially compared to the English language [13].

Arabic language is the fifth most commonly spoken language in the world and the fifth most frequently used on the internet. More than 422 million speak Arabic language (by more than 6.0% of the global population [14]. The requirements of Arabic language are not resolved by several tools, packages and APIs in information retrieval and natural language processing applications. To make these tools, software packages to handle Arabic language data, modifications and additional work are necessary. Arabic language written from right to left and it includes 28 different characters for the same letter, with varies formulations depending on position of the letter in word. In addition, there are diacritics for example small characters that may be added to a letter either as subscript or superscript to add distinct spelling, grammatical formulation, and these diacritics are widely found in formal Arabic, often indicating the letter as well as the whole word [13].

4. Text Classification Model

The goal of text classification is to create a model that used to classify different text documents to its predefined classes. Figure 1 represents the classification model phases.

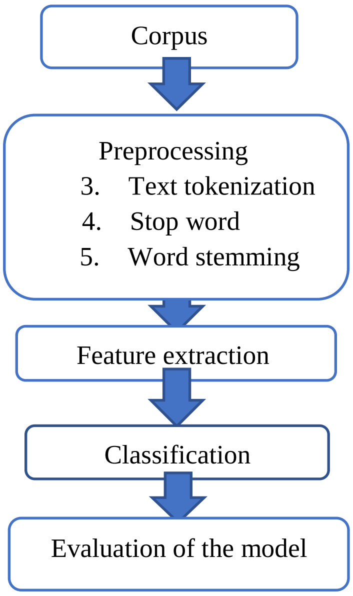

Figure 1-Classification General phases

4.1 Text Classification Datasets

There are several reference data sets for the processing of data that are publicly usable for English text classification. Unfortunately, an open access of standard dataset for the Arabic language are not aware to us. The Arabic Corpus Open-Source is freely available, but not organized. The bulk of researchers in Arabic text classification assembled their test corpora from the online dataset of Arabic news [14].

Any of commonly datasets used for text classification analysis are the following:

Masrawi is a very large data collection includes 451230 stories of news obtained by masrawy site. Articles are labelled with up to ten labels. Up to six tags per article are richly expressed in the dataset [8,9]. Assabah is a semi-automatic web crawling method of Arabic online newspapers. The records grouped into 5 classes in the dataset. The classes are: politics, society, sport, diversity and economy [7]. Hespress is semi-automatic web crawling processes for Arabic online newspapers. The records are grouped into 5 classes in the dataset. The classes: politics, society, sport, diversity and economy [7].

Akhbarona is semi-automatic web scraping technique for An Arabic internet newspaper. It comprises 78428 of Arabic documents. The records are grouped into 7 classes in the dataset. The classes are: Sports, Medical, Culture, Finance, Politics, Religion and Tech [3,7,9,15].

Khaleej includes 5690 documents in Arabic with 4 classes: Local news, Economy, sport, and International News [1,3,9,15]. In alarabiya.net all articles are divided into 7 categories with regard to the others, 2 of them did not provide sufficient data (culture and Iran News). "Iran News" was then combined into the group "Politics" and thereby provided a good dataset of training sizes. As a result, after removing the "Culture" category, the current categories are restricted to 5 categories [3,9,15].

OSAC corpus is a free corpus of public Arabic texts. The records are grouped into 10 classes in the dataset [3,11]. CNN Arabic news is made up of 5070 documents and is divided into 6 classes: sport, SciTech, entertainment, middle east, business and world [2-4].

4.2 The Preprocessing

Some preprocessing is required to deal with text data to select features which are semantically represent the document and remove other features that are not. This process which extracts important features that represent training dataset is named Feature Extraction (FE) [8]. The primary goal of the preprocessing phase is to minimize the space of testing and to decrease the rate of error [1]. Data preprocessing involves tokenization of text, removal of Stop-words, and term stemming. After preprocessing, the dataset is presented in a shape appropriate for the feature selection stage.

4.3 Feature Extraction

This stage includes taking stemming words and transform them into features to be used by the classifier. In the section below, we are giving a summarized introduction to text classification features used in this survey.

- Chi Square

Is a common method of collection of features that can be independently evaluated with regard to categories by computing the statistics of chi square. This suggests that the chi squared value analyze relationship between word and category. If the word is distinct from the category, the score would then be equal to 0, else it is 1. If the word has a higher chi-square value, that means it is more informative [5,11,16,17].

- Information Gain

Information gain method could be easier than the chi square. The fundamental principle is that for each feature that can represent discrimination between categories, we just have to determine the score, the features are then categorized according to this value and then only certain top-ranking ones are preserved [5,16,17].

- Term Frequency Inverse Document Frequency (TF-IDF)

It works by measuring how many times the word (relative frequency) in a text compared to the inverse ratio of a word over the whole corpus. This measure, of course, decide how relevant a given word is in a specific text. It is expected that words exist in single or multiple documents would have higher TF-IDF numbers than prepositions which are the common words [1,5-7,17-19].

- Word Embedding

It is a text representation that convert text into a numerical vector in vector space which represents both of syntactic and semantic characteristics of text. The word embedding models which recently provide enhanced results compared to bag of words which still used in some of natural language processing tasks. Bag of words model represents count of tokens in the text in which the location of the word is ignored in context of others [3,8,10,18,20-22]. Word2vec and Glove are the most popular models for word embedding.

4.4 Machine Learning Algorithms

In standard machine learning applications, each text document in an annotated of dataset training is then converted into a numerical representation of vectors related to text categorization. Upon the text vectorization techniques, like the Term Frequency Inverse Document Frequency (TF-IDF) algorithm, the words and phrases inside the documents are known as variables or features and scores are allocated [22].

The mostly used machine learning algorithms are: Naive Bayes Classifier [2-4,6,10,23-27], Support Vector Machines Classifier [3-5,10,17,19,21-25], Logistic Regression [1,19,21,25], Decision Trees [4,5,10,19,25], Rule Induction Classification [25], and K-Nearest Neighbor (KNN) [10,19,23-25]

4.5 Deep Learning Algorithms

Deep learning refers to a vast number of machine learning approaches and frameworks that have the advantage of employing multiple levels of hierarchical nonlinear data processing. Based on the intended application of the architectures and techniques, such as synthesis/generation or identification/classification [28].

Comparing deep learning and traditional machine methods it could be instructive to suggest parallels to a regression model in learning algorithms. Users should define a specific model when running a regression as logistic regression or linear for example. To optimize accuracy and supply data for an outcome (dependent) variable and related data for predictor (independent) variables, the regression algorithm will then fit parameters into the model. The regression algorithm would then fit the parameters to the model to optimize accuracy and supply data for the result [22].

Typically, neural networks consist of neurons operating together to form a layer. To form the network, multiple layers are then connected. The neural networks with hidden layers are Deep Neural Networks (DNN) that are deep and rich. The hidden layers are extra layers that are applied to the network to add additional processing, when the task is very difficult for a tiny network. The number of hidden layers will reach a one hundred or more. DNN are known to be creative and have excellent precision. There are several forms of DNN, many of them are alerted to function on image data, and many of them are texts sources of data [9].

The mostly used deep learning algorithms are: Multilayer Perceptron Network (MLP) [19,22], Convolution Neural Networks (CNN) [3,7,8,20-22,26,27], Recurrent Neural Networks (RNN) [27,29-31], Long Short-Term Memory (LSTM) [10,18,22,26,27,29,31], Capsule Neural Networks [27,32-34], Gated Recurrent Unit (GRU) [26,31], Bidirectional Long Short-Term Memory Networks (BiLSTM), CNN–BiLSTM Networks, and BiLSTM–CNN Networks [22].

5. Brief Comparison of Arabic Text Classification Researches

This section shows a description for this survey work. Table 1 gives comparison of different text classification methods used and displays the selected dataset by various authors for testing.

Table 1- Comparison of Arabic text classification researches

Authors	TC Technique	Year	Dataset	Accuracy
[1]	a feed-forward neural network model and Logistic Regression	2019	Two readily available databases have been used the dataset1 is Khaleej-2004 and in dataset2 they used in-house corpus from online Arabic news such as Al-hayat, Al-Jazeera, Al-Ahram, Al-Nahar and Al-Dostor)	A feed forward: 97.5% in dataset1 while in dataset2 reached 99.4% while in logistic regression model in dataset1 reached 93% and dataset2 reached 91%.
[2]	C4.5, Naïve Bayesian and, Discriminative parameter learning for Bayesian networks for text (DMNBtext)	2018	(BBC and CNN) dataset	The DMNBtext obtained 99 % accuracy on the dataset of BBCs, while with datasets of CNN, obtained an accuracy of more than 93%. In comparison, C4.5 provided better efficiency when using light stem and Boolean raw text. They noticed that the best one between the models used in their work is DMNBtext then Naïve Bayesian.
[3]	SATCDM (CNN with n- gram word embedding). And traditional machine learning algorithms	2020	Their study uses 15 freely available Arabic datasets (SANAD and NADA) usually used for TC	They compared their model with machine learning model. The result of their SATCDM model on SANAD dataset are with lowest of 98.44% and a highest of 99.49%. And the result on their linear kernel SVM classifier achieves accuracy among 96% and 97% on Abuaiadah dataset, their CNN model achieved higher accuracy outcome 98.50% on Abuaiadah dataset.
[4]	Naïve Bayesian (NB), Decision Tree (DT), Support Vector Machine (SVM) and deep Autoencoder with Restricted Boltzmann Machines (RBM)	2016	CNN Arabic news dataset	The results showed that Arabic text classification method, particularly for the SVM classifier, exploring deep autoencoder representation performed well.
[5]	H2O deep learning with features weighting and other classifiers such as support vector machine, neural network, and decision tree.	2017	They collected a dataset consists of 500 Arabic tweets	The results confirmed the efficiency of their proposed hybrid approach based on deep learning with TFIDF and the feature weighting which are (information gain and chai-square) and showed that their approach outperforms the DT, SVM and NN classifiers and achieved the highest achievement regarding the precision and accuracy that are 93.7% and 90% respectively.

[6]	Naive Bayes (NB), Proposed Kernel Naïve Bayes (KNB), HMM, KNN, J48, SVM, BN	2017	The dataset is collected from multiple online newspapers	Their proposed Kernel Naive Bayes classifier obtained the maximum accuracy 91.2281% and showed improved efficiency than other baseline used classifiers in machine learning.
[7]	CNN, SVM, LR	2018	Assabah, Hespress, Akhbarona	they proposed an easy and powerful way to differentiate Arabic text from a broad dataset. They compare their dataset with the CNN model and some traditional SVM and LR machine learning models as benchmarks. They argued that the CNN model worked best on the Arabic text classification task. Traditional approaches such as SVM do not operate well as CNN at the level where the dataset is big and large. The best results for both sizes of dataset used were recognized by CNN model. The accuracy over 92 % is more than satisfying with CNN.
[8]	CNNNorm, CNNGStem	2019	Author collected an Arabic news dataset formed from four categories	they introduced a new technique (GStem) to group similar words that share the same root based on the Arabic extra letters as a preprocessing layer for the CNN. CNN-Norm which all word vectors learned from scratch using SG word2vec then learn document classes using the CNN without using GStem CNN-NORM: 87.75% CNN-Gstem: 87.50% -92.42%
[9]	CNN, LSTM CGRU, BiLSTM, BiGRU, CLSTM, HANLSTM, GRU, HANGRU	2020	The have used Masrawy, SANAD and NADiA	All 9 models perform well on SANAD dataset. CNN achieved the highest performance on NADiA dataset. HANGRU achieved the highest performance on Masrawy dataset.
[11]	Restricted Boltzmann Machines autoencoder, MLP, SVM	2020	OSAC corpus	the learned features were fed to another deep Autoencoder for categorization. An exhaustive set of experiments was carried out and has shown that using the Autoencoder as text representation model combined with Chi-Square and classifier outperformed state of the art techniques which achieved the best results by 94% of precision and 93% of f-measure.

6. Conclusions

The process of classifying texts into categories by subject, author or title is called Arabic text classification. The core of this systemic analysis was reporting regarding various algorithms and datasets.

The databases used in the presented work are constructed using websites of Arabic news, while other studies used datasets created by other researchers such as open source Arabic corpus.

It criticized the classification for corpus and the approaches create the model, either they included deep learning or machine learning technique. The types of deep learning used were also listed such as RNN, MLP, CNN, GRU, LSTM, FFNN and others. In addition, the attention was upon publication year and the datasets of which the articles were written. Furthermore, it reviewed the performance metrics used to compare the built models.

In addition, in this systematic analysis we have found for several reasons we cannot generalize one kind of deep learning as the efficient one in Arabic text classification because in each study the neural networks used were distinct. There was other missed information in situations where the researchers used the same kind of NN. But after deep analysis, we noticed that LSTM is more appropriate than other because for text classification tasks, such networks are an attractive solution since word order in text can be essential.

The researchers did not demonstrate what parameters they used in these networks in depth and how the parameters are tuned. Typically, by adjusting parameters and rerunning the tests to produce significant effects, the machine learning algorithms are tuned. This made it impossible to compare or make sharp choices on which neural networks were the strongest.

The majority of the work, if the data size is huge, showed a better output measurement of the deep learning technique over machine learning. But traditional Machine Learning algorithms are superior to limited data sizes. So, we used deep learning because the dataset SANAD size is large enough and specifically we used akhbarona of this dataset which it contains 7 categories [Medical, Sports, Finance, Religion, Culture, Politics, and Tech].

Our direction will be toward the deep learning algorithms because in machine learning the testing accuracy will reach a certain limit and cannot increase while the deep learning algorithms increases more in testing accuracy whenever the dataset is large.

7. Future work

To promote more study on the Arabic language and to help create benchmarks, we proposed the importance of developing a qualified and diverse Arabic corpus. Also, we suggest using the word embedding techniques, and multiple features to enhance the classification performance. More study might be done on employing semi-supervised machine or deep learning approaches to reduce the necessity for a large training dataset created with human participation, which is prone to mistakes.

References

[1] K. Sundus, F. Al-Haj, and B. Hammo, “A Deep learning approach for Arabic text classification,” 2019 2nd Int. Conf. New Trends Comput. Sci. ICTCS 2019 - Proc., pp. 1–7, 2019, doi: 10.1109/ICTCS.2019.8923083.

[2] R. Alshammari, M. Syiam, Z. Fayed, and M. Habib, “Arabic Text Categorization using Machine Learning Approaches” (IJACSA) *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, 2018.

[3] M. Alhawarat and A. O. Aseeri, “A Superior Arabic Text Categorization Deep Model (SATCDM),” *IEEE Access*, vol. 8, pp. 24653–24661, 2020, doi: 10.1109/ACCESS. 2020.2970504.

[4] F.-Z. El-Alami, S. Ouatik, and E. Alaoui, “An Efficient Method based on Deep Learning Approach for Arabic Text Categorization,” Int. Arab Conf. Inf. Technol., 2016.

[5] A. Altaher, “Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting,” *Int. J. Adv. Appl. Sci*., vol. 4, no. 8, pp. 43–49, 2017, doi: 10. 21833/ijaas.2017.08.007.

[6] R. Al-khurayji and A. Sameh, “An Effective Arabic Text Classification Approach Based on Kernel Naive Bayes Classifier,” *Int. J. Artif. Intell. Appl*., vol. 8, no. 6, pp. 01–10, 2017, doi: 10.5121/ijaia.2017.8601

[7] S. Boukil, M. Biniz, F. El Adnani, L. Cherrat, and A. E. El Moutaouakkil, “Arabic text classification using deep learning technics,” *Int. J. Grid Distrib. Comput*., vol. 11, no. 9, pp. 103–114, 2018, doi: 10.14257/ijgdc.2018.11.9.09

[8] M. Galal, M. M. Madbouly, and A. El-Zoghby, "Classifying Arabic text using deep learning," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 23, pp. 3412–3422, 2019.

[9] A. Elnagar, R. Al-Debsi, and O. Einea, "Arabic text classification using deep learning models," Inf. Process. Manag., vol. 57, no. 1, p. 102121, 2020, doi: 10.1016/j.ipm.2019.102121.

[10] W. Hazım, G. Gwad, I. Mahmood, I. Ismael, and Y. Gültepe, "Twitter Sentiment Analysis Classification in the Arabic Language using Long Short-Term Memory Neural Networks," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 3, pp. 235–239, 2020, doi: 10.35940/ijeat.b4565.029320

[11] F. El-Alami, A. El Mahdaouy, S. O. El Alaoui, & N. En-Nahnahi, "A deep autoencoder-based representation for Arabic text categorization.," vol. 3, no. 3, pp. 381–398, 2020.

[12] M. A. Otair, "Comparative Analysis of Arabic Stemming Algorithms," *Int. J. Manag. Inf. Technol.*, vol. 5, no. 2, pp. 1–12, 2013, doi: 10.5121/ijmit.2013.5201.

[13] T. Kanan and Edward "A. Fox Automated Arabic Text Classification with P-Stemmer, Machine Learning, and a Tailored News Article Taxonomy" *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. July, pp. 1852–1863, 2013, doi: 10.1002/asi.11

[14] M. A. R. Abdeen, S. AlBouq, A. Elmahalawy, and S. Shehata, "A closer look at arabic text classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 11, pp. 677–688, 2019, doi: 10.14569/IJACSA.2019.0101189.

[15] Akhbarona Arabic corpus https://data.mendeley.com/datasets/57zpx667y9/1

[16] M. Zareapoor and S. K. R, "Feature Extraction or Feature Selection for Text Classification: A Case Study on Phishing Email Detection," *Int. J. Inf. Eng. Electron. Bus.*, vol. 7, no. 2, pp. 60–65, 2015, doi: 10.5815/ijieeb.2015.02.08.

[17] Y. Liu, S. Ju, J. Wang, and C. Su, "A New Feature Selection Method for Text Classification Based on Independent Feature Space Search," *Math. Probl. Eng.*, vol. 2020, pp. 1–14, 2020, doi: 10.1155/2020/6076272.

[18] A. Grünwald and S. Rauf Ahmad, "Applications of Deep Learning in Text Classification for Highly Multiclass Data," 2019. UPPSALA UNIVERSITET, Sweden.

[19] L. Al Qadi, H. El Rifai, S. Obaid, and A. Elnagar, "Arabic text classification of news articles using classical supervised classifiers," 2019 2nd Int. Conf. New Trends Comput. Sci. ICTCS 2019 - Proc., pp. 1–6, 2019, doi: 10.1109/ICTCS.2019.8923073.

[20] H. Xu, A. Kotov, M. Dong, A. I. Carcone, D. Zhu, and S. Naar-King, "Text classification with topic-based word embedding and Convolutional Neural Networks," ACM-BCB 2016 - 7th ACM Conf. Bioinformatics, Comput. Biol. Heal. Informatics, no. October, pp. 88–97, 2016, doi: 10.1145/2975167.2975176.

[21] J. Wang, Z. Wang, D. Zhang, and J. Yan, "Combining knowledge with deep convolutional neural networks for short text classification," IJCAI Int. Jt. Conf. Artif. Intell., vol. 0, pp. 2915–2921, 2017, doi: 10.24963/ijcai.2017/406.

[22] A. Varghese, G. Agyeman-Badu, and M. Cawley, "Deep learning in automated text classification: a case study using toxicological abstracts," *Environ. Syst. Decis.*, vol. 40, no. 4, pp. 465–479, 2020, doi: 10.1007/s10669-020-09763-2.

[23] M. Sayed, R. Salem, and A. E. Khedr, "Accuracy evaluation of Arabic text classification," Proc. ICCES 2017 12th Int. Conf. Comput. Eng. Syst., vol. 2018-Janua, pp. 365–370, 2018, doi: 10.1109/ICCES.2017.8275333.

[24] A. M. F. Al Sbou, "A survey of arabic text classification models," *Int. J. Informatics Commun. Technol.*, vol. 8, no. 1, p. 25, 2019, doi: 10.11591/ijict.v8i1.pp25-28.

[25] M. Thangaraj, M. Sivakami, "Text Classification Techniques: A Literature Review," *interdisciplinary journal of information knowledge*[and management., vol. 13, 2018, doi: 10.28945/4066.12

[26] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "A comparative review on deep learning models for text classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 325–335, 2020, doi: 10.11591/ijeecs.v19.i1.pp325-335.

[27] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review," arXiv, vol. 1, no. 1, pp. 1– 43, 2020.

[28] H. M. Ahmed and H. H. Mahmoud, "Effect of Successive Convolution Layers to Detect Gender," *Iraqi J. Sci.*, vol. 59, no. 3C, pp. 1717–1732, 2018, doi: 10.24996/ijs.2018.59.3c.17.

[29] 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," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 6, pp. 6629–6643, 2020, doi: 10.11591/IJECE.V10I6.PP6629- 6643.
[30] H. Sak, A. Senior and F. Beaufays, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling"arXiv preprint arXiv:1402.1128, 2014.
[31] S. K. Abbas and L. E. George, "The performance differences between using recurrent neural networks and feedforward neural network in sentiment analysis problem," *Iraqi J. Sci.*, vol. 61, no. 6, pp. 1512–1524, 2020, doi: 10.24996/ijs.2020.61.6.31.
[32] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao, "Investigating capsule networks with dynamic routing for text classification," *arXiv preprint arXiv*:1804.00538, 2018.
[33] W. Zhao, H. Peng, S. Eger, E. Cambria, and M. Yang, "Towards scalable and reliable capsule networks for challenging NLP applications," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 1549–1559, 2020, doi: 10.18653/v1/p19-1150.
[34] M. Yang, W. Zhao, L. Chen, Q. Qu, Z. Zhao, and Y. Shen, "Investigating the transferring capability of capsule networks for text classification," *Neural Networks*, vol. 118, pp. 247–261, 2019, doi: 10.1016/j.neunet.2019.06.014.