A Proposed Deep Learning based Framework for Arabic Text Classification

Mostafa Sayed[1]
Faculty of Computers and Artificial Intelligence, Beni-Suef University
Beni-Suef, Egypt

Hatem Abdelkader[2]
Faculty of Computers and Information, Menoufia University
Menoufia, Egypt

Ayman E. Khedr[3]
Information Systems Department
Faculty of Computers and Information Technology, Future University in Egypt (FUE)
Cairo, Egypt

Rashed Salem[4]
Faculty of Computers and Information, Menoufia University
Menoufia, Egypt

***Abstract*—Deep learning has become one of the crucial trends in the modern era due to the huge amount of data that has become available. This paper aims to investigate and improve a generic framework for Arabic Text Classification (ATC) with different deep learning techniques. Besides, it deals directly with a word in its original style as a basic unit of modern Arabic sentence and on a different level of N-grams versus a combination of Intersected Consecutive Word proposed method (ICW). However, it aimed to discuss the results of the different experiments for the enhancements of the proposed method on different deep learning algorithms such as Scaled Conjugate Gradient (SCG) and Gradient descent with momentum and adaptive learning rate backpropagation (GDX) on ATC. The results showed that the proposed framework applied with the SCG algorithm and TF-IDF outperforms the GDX algorithm with an accuracy ratio of 90.65%.**

***Keywords*—Text classification; arabic text classification; scaled conjugate gradient; TF-IDF; GDX; ICW**

I. Introduction

The classification of texts is becoming more crucial every day due to the tremendous diversity in the use of different human knowledge sources. This usage of cognitive resources resulted in the momentum and abundance of information and data circulating between many devices with large volumes, rapid and remarkable development of artificial intelligence. Therefore, it was necessary to work effectively in containing this momentum in order to classify these texts effectively, not only to facilitate the retrieval of information but also for machine learning uses.

Text classification (TC – also known as text categorization, or topic spotting) is the task of automatically sorting a set of documents into categories (or classes, or topics) from a predefined set. This assignment falls at the intersection of Information Retrieval (IR) and Machine learning (ML). It can be defined as the process of classifying or structuring documents into a predefined set of categories according to a group structure that is known in advance [1]. Also, Khorsheed defined it as "The assignment of free-text documents to one or more predefined categories based on their content" [2].

Valuable information can be elicited out of organized/unorganized textual resources like documents and classified into a stable number of predefined categories that are established in advance, they express the fundamental idea of text categorization. Each document can be in multiple, solo, or no class at all [3],[4].

The problems of classifying texts are not only represented in the tremendous diversity in the use of cognitive sources or in the momentum but also the abundance of information and circulating data, which sometimes reach millions of terabytes. Moreover, there is a crucial factor which is the language in which these texts are written [28].

In the field of Natural Language Processing (NLP), different languages were interested in researches development more than the others. Whatever, some of these researches are concerned with the Arabic language as it has a gorgeous impact even it became one of the most commonly used languages all over the world; despite it considers the fifth spoken one. It uses profusely in many of the different Arab countries as it is the main language of the Holy Quran. Moreover, various applications are still limited for the Arabic language owing to its enormous variation in shape, structure, and component, although different studies were carried out for text classification using the English language [5], [26], [27].

There are many challenges concerning the documents written in the Arabic language. From these challenges the common characteristic of the language, e.g., richness of vocabulary, the complexity of grammar, combinations of orthography, the existence of short vowels, …, etc. These problems are particularized in detail in [6], [7]. Besides, the algorithms that were developed for English perform unwell for Arabic [8].

Different researchers interested in the field of NLP are concerned with applying and studying deep learning algorithms in order to explore many results for the reasons of development and improvement [31]. Deep learning is a coherent and integrated set of algorithms that interpret and link data to each other in order to achieve the greatest degree of accuracy in order to identify and extract new information that was previously unknown [33]. However, the method of learning these algorithms is a representation of the way human brain cells work in transmitting and interacting signals.

Machine learning techniques can be classified for ATC under two categories. The first category is the classical machine learning techniques which contain approved algorithms such as SVM, KNN, NB, and others. Whereas, the second category, is modern machines learning techniques which contain algorithms such as stochastic gradient descent (SGD), convolutional neural network (CNN), and bi-directional long-short term memory (BLSTM).

The contribution of this paper is presenting a novel framework for handling the binary classification problem in Arabic text by employing a new proposed ICW method as a feature representation. Also, the proposed framework handles the effects of deep learning techniques in binary classification problem in Arabic text.

The rest of the paper is organized as follows: The second section mentions the Arabic text classification phases. The third section mentions a literature study of deep learning algorithms and the related works in ATC. The fourth section, introduces some crucial background to facilitate the understanding the following sections. Section five presents the proposed framework based on deep learning techniques. Section six presents the experimental study and discusses the experimental results obtained. Finally, the seventh section concludes the paper's contributions and future work.

II. ARABIC TEXT CLASSIFICATION PHASES

Building a generic framework depends on previous phases of text classification. Fig. 1 shows the generic framework extracted based upon [9], [10] was able to take a step forward towards applying deep learning techniques in ATC. Six phases are considered as the main phases for achieving the text classification and dimensionality curse problems.

A. Data Collection Phase

According to the huge usage of the Internet and social media, there are different types of data which are differed in shape and volume. For that, data can be collected through the Internet or detected system in a represented shape such as text, documents, web pages, videos, spreadsheets, or database files. Data can be retrieved through different resources which can be categorized under defined symmetric or asymmetric data. Also, it can be categorized into three categories i.e., structured, unstructured, and semi-structured in the same group.

B. Pre-Processing Phase

Different steps are made for cleaning and preparing the collected data as a result of the last preprocessing phase. Preparing the different data collected from different resources in a homogeneity manner is a prerequisite objective for the classifiers in the classification phase [30]. However, there are general steps used in the Arabic text classification phase such as excluding stop words which include pronouns, conjunctions, and prepositions. Also, exclude digits, punctuation marks, Latin alphabet, removal of isolated letters, and non-Arabic words.

The text prepared before is tokenized and divided by representing it in different manners under the conditions of usage purposes. There are two popular models for the reason of tokenizing text N-gram and bag of words (BOW) models.

C. Representing Phase

After preparing the data, it is represented through an indexed matrix vector space. Indexing is a crucial process in (IR) systems [25], [29]. It reduces the documents into the informative terms contained in them.

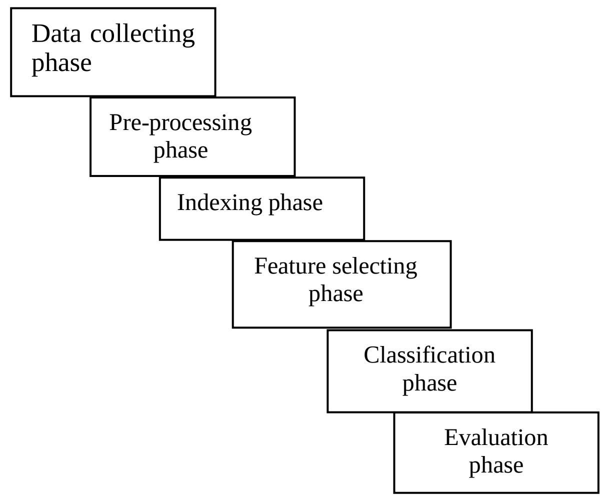

Fig. 1. Conceptual View of Text Classification Steps.

It provides a mapping from the terms to the respective documents containing them [11]. The produced indexed matrix-vector can be represented as equation number (1).

$$M = X * Y \quad (1)$$

Where X = (D_0, D_1, D_2,…….,D_{n-1},D_n) where D_n represent documents on data set

Y = (w_0, w_1, w_2,…….,w_{n-1},w_n) where w_n represent the words in the document and $Y \subset D_n$.

When training a machine or a deep learning classification model on a dataset, a matrix of vector space is prepared considerably. The rows of the matrix represent the actual documents that were contained in the training dataset, and the columns represent the features represented in each document.

D. Feature Selecting Phase

In this phase, a matrix of vector space is represented among all extracted features of documents represented in the data set. A massive number of features extracted through the data-set crash with the “Curse of dimensionality” problem. Features reduction algorithms are adaptable techniques for solving that problem that reduces the low priority of features in consideration of the high quality of the text classification process.

E. Classification Phase

It's the important phase through all the previous processes. In this phase, the classifier is trained as a model based on previous train data-set. The objective of this phase is to train the classifier to generalize the method of classifying data on another symmetric data-set.

F. Evaluation Phase

The last phase is evaluating the generated classification model from the previous phases. A symmetric test data-set is prepared for the purpose of measuring the accuracy of the classifier according to the confusion matrix.

III. Related Work

Different studies concerned with machine learning techniques were interested in defining Neural Networks as a biased term for complicated deep learning techniques. CNNs are very similar to ordinary Neural Networks, they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, executes a dot product, and may follow it with a non-linear function. The entire system still expresses a single differentiable score function: from the inputs to the classes in the outputs. Moreover, they still have a loss function on the last layer. CNN is typically a sequence of layers, and the outputs of every layer are the inputs of the next layer. Those layers are convolutional, pooling layers, and fully connected layers [12], [13].

Sagheer et al, discussed applying deep learning techniques for Arabic sentence classification. They presented a CNN model to achieve the purpose of classifying the text in the used dataset into three labels. The used dataset contained the Arabic sentence withdraw from Essex Arabic Summaries Corpus (EASC). They used different techniques for preparing text for the deep learning algorithm, they first tokenized the words of text and then transformed them to sequences represented by word indices. According to the variation of the sequence length of words, they used a padding technique to trim the sequence length into a unified length. They used the word embedding layer for comparing three models of CNN. For avoiding the overfitting issues through the training phase of the algorithm, they used dropout layers and weighting regularization functions. The results showed that CNN models in top of word embedding layer achieved high-performance accuracy in the NLP task, Arabic sentences classification [14].

In other work, Helmy et al., proposed a deep learning-based approach for keyphrase extraction of Arabic text. As mentioned before a shortage in Arabic text datasets was found for adapting deep learning models for Arabic keyphrase extraction. However, they established a new dataset that was prepared to consist of 6,000 abstracts of scientific Arabic documents. They apply a word embedding representation method for representing the tokens of the documents which are divided into sentences. The embedding layer works as a lookup table that transforms discrete features such as the words of Arabic text into continuous real-valued vector representations, which are then concatenated and provided to the neural network. Also, they proposed applying the BLSTM network for utilization instead of a feed-forward network. They used two hidden layers forward and backward hidden sequence to generate the output. Then, a cunctation layer is connected to a softmax output layer with three neurons for each word. A dropout technique was implemented between Bi-LSTM and the dense layer to prevent overfitting. The evaluation results showed that the proposed approach achieves state-of-the-art performance in the Arabic KPE domain [15].

Also, Samir Boukil et al., proposed a method for Arabic text classification. The proposed method that followed depends mainly on the known steps for preparing the Arabic text as preprocessing steps. Also, they used a stemmer for purpose of extracting and selecting the features. They used TF and TF-IDF techniques as feature weighting techniques for representing the text. For the classification phase, they compared three classifiers CNN, SVM, and Logistic Regression. In the CNN experiment, they employed stochastic gradient descent (SGD) to train the network and use a backpropagation algorithm to calculate the gradients. Besides, they used a learning rate of 0.001 and a dropout ratio with a value of 0.5 to enhance the classifier performance. They argued that the CNN algorithm achieved high results in large and big datasets versus the traditional algorithms [16].

Whatever, deep learning can be a coherent and integrated set of algorithms that interpret and link data to each other in order to achieve the greatest degree of accuracy in order to identify and extract new information that was previously unknown. The method of learning these algorithms is a representation of the way human brain cells work in transmitting and interacting signals. Because of the lack of research in the field of ATC, the previous methodologies used in the mentioned research are considered important points. Also, these researches are characterized as the basic steps in the direction of building generally proposed frameworks to improve deep learning algorithms to classify texts in the Arabic language. Moreover, as mentioned in the literature review several deep learning techniques differ between them in architectures and performance and not all of them have been applied to the ATC problem.

The previous issues have motivated us to propose a large and accessible benchmark dataset of binary-label Arabic texts classified under a legal text. Besides, it has motivated us for exploring in a comparative manner the effect between two deep learning algorithms, i.e., GDX and SCG for ATC.

IV. Background

A. Data Preprocessing based Techniques

In the following, we demonstrate the proposed combination of words for representing the Arabic text based on the structure of the Arabic sentence. Arabic sentence contains in its normal structure two types of sentences, i.e., noun sentence and verbal sentence. The noun sentence consists of two parts or tokens "mobtada" and "khaber" whereas the verbal sentence consists of three parts or tokens subject, verb, and object. For extracting all possible sentences from the text the next two definitions were formulated for that purpose.

Definition 1: Arabic text can be represented by T= (w_0, w_1, w_2, …, w_{n-1}, w_n) Noun Sentence (NS) can represented by NS =

($w_0+w_1,w_1+w_2,\ldots\ldots\ldots, w_{i-1}+w_i$) where $w_{i-1}+w_i \subset T$ and $w_{i-1}+w_i \leq w_{n-1}+w_n$ and i=n .

According to the last definition, several noun sentences NS will be represented by several tokens each one consisting of (wi-1+wi) Ɛ T.

Definition 2: Arabic text can be represented by T= ($w_0, w_1,w_2,w_3\ldots,w_{n-2},w_{n-1},w_n$) Verbal Sentence (VS) can represented by VS = ($w_0+w_1+w_2,w_1+w_2+w_3,\ldots\ldots\ldots, w_{i-2}+w_{i-1}+w_i$) where $w_{i-2}+w_{i-1}+w_i \subset T$ and $w_{i-2}+w_{i-1}+w_i \leq w_{n-2}+w_{n-1}+w_n$ and i=n .

ICW Proposed method

Input: file of text contains a number of lines

Output: separated files each one contains one of a line of words equals to the token value

1: For each (line in lines)

2: If the line is not empty then

3: Read the words in each line

4: For each line do

5: Read the token value

6: Divide the words equals to token value

7: Create a separated file for each line;

8: Write to file the words equals to token value;

9: Exit

B. Feature Selection Phase

The feature selection phase is crucial in our proposed framework as it's the last step in drawing the features vector space for each class. However, two methods are proposed for applying to establish the features vectors are TF and TF-IDF [32]. Term frequency is concerned with how frequently a word or a combination of words occurs in a detected one document. Where TF-IDF is concerned with how frequently a word or a combination of words occurs within the overall document [28].

For calculating TF-IDF assuming that a given a group of documents D, a word w, and an individual document d є D, we calculate *wd* the weight of the word w by applying (1) and (2).

$$wd = TF * IDF \tag{2}$$

$$wd = fwd * log(|D|/fwd) \tag{3}$$

Where fwd is the number of times the word w appears in the document d, |D| is the size of the corpus, and fwD is the number of documents in which w appears in D [17].

C. Classification Phase

In this phase, the algorithm is trained on the data-set for purpose of achieving the classification task. Whatever the well-defined deep learning classification algorithms we choose SCG and GDX.

1) Scaled conjugate gradient back propagation algorithm: This algorithm was developed by (Moller,1990). It was built based upon a network training function that updates weight and bias values according to the scaled conjugate gradient method. It depends on conjugate directions, though it does not perform a line search at each iteration for avoiding the time-consuming linear search of conjugate and optimal direction that occurs with other algorithms.

The scaled conjugate gradient method relies on fast strategy search using information from the second-order approximation [18]. The mathematical equations used for that algorithm can be summarized as follows:

$$E(w + y) \approx E(w) + E'(w)Ty + 1/2yTE(w) \tag{4}$$

The quadratic approximation to E for the point w can be achieved through Eqw(y) in eq (5)

$$Eqw(y) = E(w) + E'(w)Ty + 1/2yTE''(w)y \tag{5}$$

For determining minima to Eqw(y) the critical points must be detected in equation (6). The critical points are the primitive keys for linear systems [19].

$$Eqw(y) = E''(w)y + E'(w) = 0 \tag{6}$$

Assume that conjugate systems with start point Y_1, and $P1\ldots PN$. We can consider a linear combination of the points from Y_1 to Y^* till PN. Where Y^* is a critical point.

$$Y^* - Y1 = \sum_{i=1}^{n} \propto ipi \; where \; \propto_1 \in R \tag{7}$$

$$P_j^T\left(-E'(w) - E''(W)y1\right) = \propto jP_j^T E''(w)P_j \tag{8}$$

$$\propto = \left(P_j^T\left(-E'(w) - E''(w)y1\right)\right)/\left(jP_j^T E''(w)P_j\right) \tag{9}$$

Using Eqs (7), (8), and (9), we can iteratively determine the value of the critical point which is Y*. Eqw(Y), is given by the equation,

$$E_{qw}(Y) = E_{qw}(Y^*) + 1/2(Y - Y^*)TE^N(W)(Y - Y^*) \tag{10}$$

2) Gradient descent with momentum and adaptive learning rate back propagation: The algorithm can be defined as it updates weight and bias values according to gradient descent momentum and an adaptive learning rate. Momentum factors can be accomplished by adding a fraction of the previous weight change to the current weight change. This term encourages movement in the same direction on successive steps. The addition of such a term can help smooth out the descent path by preventing extreme changes in the gradient due to local anomalies. Therefore, it is likely to suppress any oscillations that result from changes in the slope of the error surface [20].

Back propagation is used to calculate derivatives of performance *perf* with respect to the weight and bias variables X. Each variable is adjusted according to gradient descent with momentum,

$$dX = mc * dXprev + lr * mc * dperf/dX \tag{11}$$

where, *dXprev* is the previous change to the weight or bias.

D. Evaluation Measures

Precision and recall are widely used for evaluation measures in information retrieval and machine learning [21]. Precision is the fraction of retrieved documents that are relevant to the query. In other words, it concerns how useful the search results are. Recall is the segment of the documents which is exactly related to the inquiry that is absolutely recalled. However, it concerns with how complete results are. F-measure is approximately the average of precision and recall.

TABLE I. DOCUMENTS POSSIBLE SETS BASED ON A QUERY IN IR

Iteration	Relevant	Irrelevant
Documents Retrieved	true positives (tp)	false positives (fp)
Documents not Retrieved	false negatives (fn)	true negatives (tn)

According to Table I, precision, recall, and (macro average) measures can be computed as the following equations:

$$Precision = \frac{tp}{(tp+fp)} \quad (12)$$

$$Recall = \frac{tp}{(tp+fn)} \quad (13)$$

$$F-measure = \frac{2*Precision*Recall}{Recall+Precision} \quad (14)$$

V. PROPOSED FRAMEWORK BASED ON DEEP LEARNING TECHNIQUES

For building a generic framework, it builds based upon the conceptual view of text classification. Fig. 2 shows the generic framework which extracted based upon [9], [10]. Moreover, it is able to take a step forward towards for applying deep learning techniques in ATC. In addition, a new legal dataset for requests of prosecutors was presented for the purpose of testing and training the proposed framework in different experiments.

In the data collection phase, data from two courts of the council state of Egypt are collected and reviewed by three technical reviewers. The reviewing process was performed according to the spatiality of each document of the previously detected two classes.

Different steps have been made for cleaning and preparing the collected data in preprocessing phase such as excluding stop words which include pronouns, conjunctions, and prepositions. Also, exclude digits, punctuation marks, Latin alphabet, removal of isolated letters, and non-Arabic words [24].

Besides, the adding point of the proposed method for representing the text is in the features representation phase. The proposed method was built based upon definition [1, 2] and the ICW algorithm. Two combination forms for representing the text were proposed for evaluating the proposed framework. Firstly, it evaluates the combination between the uni-word and noun sentence. Secondly, it evaluates the combination between noun sentences and verbal sentences.

A comparison between two representation methods in the features representation phase showed. On one hand, the N-gram model with three aspects of a word-level uni-gram, bi-gram, and tri-gram is established. On the other hand, the proposed method for representing the text with its two combinations of uni-word, noun sentence, and verbal sentence is presented.

In classification phase two deep learning algorithms are applied for the purpose of classifying data into its class SCG and GDX. They applied for the purpose of measuring their accuracy with the previous steps of preparing the data of two classes with TF representing features vector in one side. In the other side, they also applied for measuring their accuracy in classification with TF-IDF. Besides, deep learning architectures changed in a manner for evaluating and exploring the effects of changing the number of both hidden learning layers and neurons of each layer in the accuracy of the classifiers. Fig. 2 shows the different steps from collecting the data from documents to the evaluation step of the classifier.

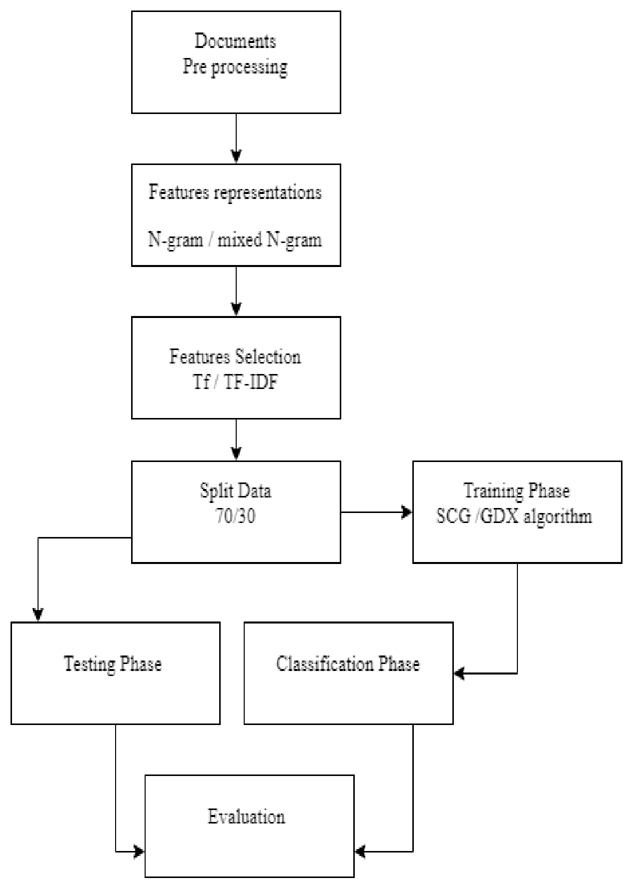

Fig. 2. Proposed Framework for ATC with Deep Learning Techniques.

VI. EXPERIMENTS AND RESULTS

A Matlab software with deep learning toolbox version 2018b was used for implementing different steps of the proposed framework of Fig. 2. Besides, the hardware machine used for implementing the different experiments was a server machine with a specification of Xeon processor 1.96 and memory 32 gigabytes.

A. Dataset

To evaluate any categorization system, text collection which consists of different categories must be available for training and testing purposes. There is no standard Arabic text collection as a benchmark dataset to the best of our knowledge for researchers who work on classifying Arabic text in the legal environment [22], [23]. Wherefore, one of the main purposes of this study is to contribute to the research community by representing a legal dataset written with Arabic text. Its value stems from the real addition to the trends of digital transformation and digitization of the work of the Egyptian judiciary. In addition to establishing the principle of complete justice, transparency, and facilitating the work of the judiciary, within the framework of Egypt's plan for sustainable development 2030.

We collected a subset of the Arabic text for requests of Prosecutors covering two topics from an Arabic dataset built locally at the council state of Egypt. The reasons for choosing these Prosecutors' requests are firstly, each Prosecutors request has a unique label. This makes it easier for us to access and fetch the files by their label. Secondly, each request has been reviewed by two different phases of technical reviewing.

Two categories were collected of this dataset containing 500 text documents for each class the total size is 1000 text documents. The total features extracted from these data exceed thousands of words beginning from a least 2000 words and in some cases exceed 16000 features. For more details of the dataset, description sees Table II.

The following statement is an example of ATC problem which labeled as class one:

اولا: قبول الدعوى شكلا. ثانيا: وفي الموضوع الحكم بالزام المدعى عليهم بصفتهم بان يؤدوا للطالبه اجرا مضاعفا عن عملها ايام الراحات الاسبوعية والعطلات والاجازات والبالغ قدرها 326 يوم ثلاثمائة وستة وعشرون يوما مع الزام الجهة الادارية بالمصروفات والاتعاب على ان ينفذ الحكم بمسودته دون اعلان.

Moreover, the following statement is labeled as class two.

اولا: قبول الدعوى شكلا. ثانيا: وفى الموضوع الحكم بالزام المدعى عليه بان يؤدى للطالب اجر مضاعف عن جميع ايام الراحات الاسبوعية والعطلات الرسمية واجازات الاعياد منذ التحاقه بالعمل حتى الان مع الزام المدعى عليه بالمصروفات والاتعاب مع حفظ كافة حقوق الطالب الاخرى.

Different preprocessing steps have been made for handling the last data such as removing the duplicated words and digits etc.…another example for tokenizing the text into the different methods such as N-gram or ICW method is described below.

Example one: tokenizing the text into a bi-gram method.

باحقية المدعى # المدعى المعاملة # المعاملة المالية # المالية طبقا # طبقا لقرار # لقرار رئيس # رئيس مجلس # مجلس الوزراء # الوزراء رقم # رقم لسنة # لسنة فترة # فترة ابتعاثة # ابتعاثة لدولة

Example two: tokenizing the text into a tri-gram method.

باحقية المدعى المعاملة # المدعى المعاملة المالية # المعاملة المالية طبقا # المالية طبقا لقرار # طبقا لقرار رئيس # لقرار رئيس مجلس # رئيس مجلس الوزراء # مجلس الوزراء رقم # الوزراء رقم لسنة # رقم لسنة فترة # لسنة فترة ابتعاثة # فترة ابتعاثة لدولة

TABLE II. DATASET DESCRIPTION ACCORDING TO N-GRAM AND PHRASE STRUCTURE

	Features of Class one	**Features of Class two**	**Before remove duplicate**	**After remove duplicate**
uni-gram	1331	1489	2820	2182
bi-gram	3826	3734	7560	6831
tri-gram	5220	4910	10430	9582
one word and noun phrase	5115	5198	10313	8952
noun phrase and verbal phrase	8997	8613	17610	16339
one word, noun phrase and verbal phrase	10286	10077	20363	18460

B. Experimental Configuration

The experimental configuration was built based on illustrating the effect of changing the learning layers architecture for the algorithm on the classifier's accuracy. Also, it was built based on illustrating the effect of using the proposed method for representing Arabic text based on noun and verbal sentences with various combinations versus N-gram. Different types of experiments with detailed sub-experiments were configured for achieving the last two objectives. First, experiments were set up for comparing the accuracy of the classifier between representing the uni-gram with TF and representing the uni-gram with TF-IDF with the increasing number of layers respecting two a constant of the number of neurons in each layer. Second, experiments were set up for comparing the accuracy of the previous deep learning algorithm represented in SCG algorithm with (TF) and (TF-IDF) features selection method with a proposed ICW method of a uni-bi gram and bi-tri gram.

Other experiments were set up for comparing the accuracy of the SCG algorithm versus the GDX algorithm with a two and three hidden layer with 100 neurons for each layer. However, according to the main objective of the comparisons (TF-IDF) and TF features selection method is still used for unifying the configuration of the experiments.

C. Evaluating the Accuracy of SCG Algorithm through Changing the Architecture Layers:

The experiment was executed based upon the framework in Fig. 2 with the detailed steps which have been discussed before. This experiment was established for the purpose of exploring the effects of changing the architecture layers in the accuracy level of the classifier. The first factor is increasing the number of learning hidden layers whereas the other factor is fixing the number of neurons. We measure the accuracy level of the classifier with equation number based on the confusion matrix and equation (14). Comparisons have been made between representing the uni-gram with TF and the uni-gram with TF-IDF with the increasing number of layers.

$$Layer\ size = nl * N \qquad (15)$$

Where nl >= 2 and N is a constant equal 100.

The factor nl is the number of layers and N in the number of neurons.

The experiments are executed till the stopping condition is achieved. The stopping condition is the accuracy level is lower than the first experiment with the number of layers being 2.

TABLE III. ACCURACY OF SCG ALGORITHM WITH CHANGING THE ARCHITECTURE LAYERS

Number of layers	Uni-gram TF	Uni-gram TF-IDF
LayerSize = [2*100]	88.02	86.29
Layer Size = [3*100]	88.74	87.18
Layer Size = [4*100]	88.77	89.67
Layer Size = [5*100]	88.53	88.69
Layer Size = [6*100]	88.08	86.90
Layer Size = [7*100]	88.89	87.21
Layer Size = [10*100]	86.98	84.85
Layer Size = [20*100]	89.61	87.41
Layer Size = [30*100]	89.19	90.07
Layer Size = [50*100]	87.12	86.86
Layer Size = [80*100]	85.1	86.11

The results in Table III showed that increasing the number of hidden layers with a fixed number of neurons affects the accuracy of the classifier with TF and TF-IDF representation methods. Also, Fig. 3 showed a variance of accuracy ratios changed with increasing the number of hidden layers. However, it has been detected that the SCG algorithm achieved the best accuracy ratio with 90.07% with the number of hidden layers equaling 30 layers.

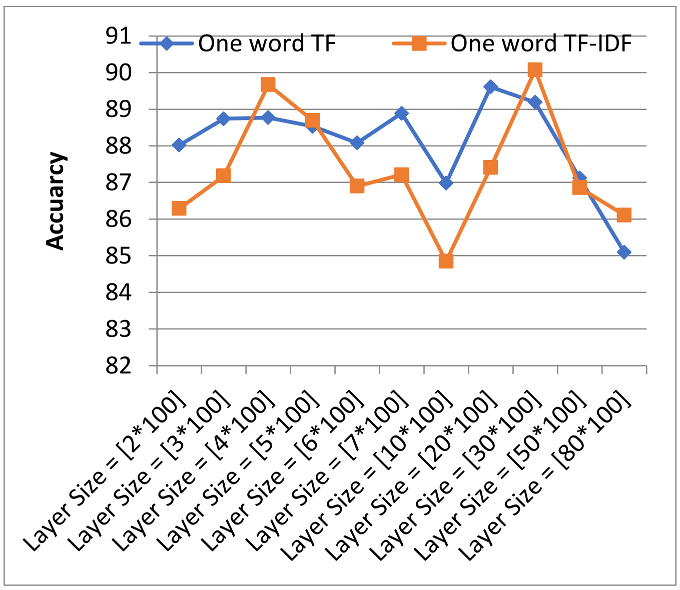

Fig. 3. The Effect of Accuracy According to the Change of Layers Number with Fixed Neurons.

D. Accuracy Ratios of SCG Algorithm According to TF and TF-IDF Features Selection with Fixed Layers and Neurons:

The experiment was executed based upon the framework in Fig. 2 with the detailed steps that have been discussed before. This experiment was established for the purpose of exploring the effects of two factors in the accuracy level of the classifier. The first factor is a fixed number of learning hidden layers whereas the other factor is a fixed number of neurons for each layer. Table II showed the number of words represented as features used in this experiment. We measure the accuracy level of the equation number based on the confusion matrix. Comparisons have been made between two different methods on one hand a combination between (uni-gram and bi-gram) and (bi-gram and tri-gram) on the other hand the typical N-gram model. The last two combinations are compared with TF and TF-IDF representing methods with a fixed number of layers. The fixed number of layers is represented with an assumed number of neurons which equals 100 neurons for each one of the layers.

Table IV shows the results for discussing the factors that affects the accuracy of the classifier with both representation methods TF and TF-IDF. The main purpose of the experiment stills the highest accuracy for text classification that achieved with two detected layers and 100 of neurons. It has been showed that representing the text with TF-IDF representation method outperform TF representation method in bi-gram, tri-gram, and (uni-bi) gram with ratios 90.34%,88.36%, and 90.65% respectively.

TABLE IV. ACCURACY OF SCG ALGORITHM ACCORDING TO TF AND TF-IDF FEATURES SELECTION WITH FIXED LAYERS AND NEURONS

	Tf	Tf-idf
Uni-gram	88.02	86.29
Bi-gram	89.78	90.34
Tri-gram	87.79	88.36
(Uni,Bi)-gram	89.55	90.65
(Bi,Tri)-gram	88.76	88.55

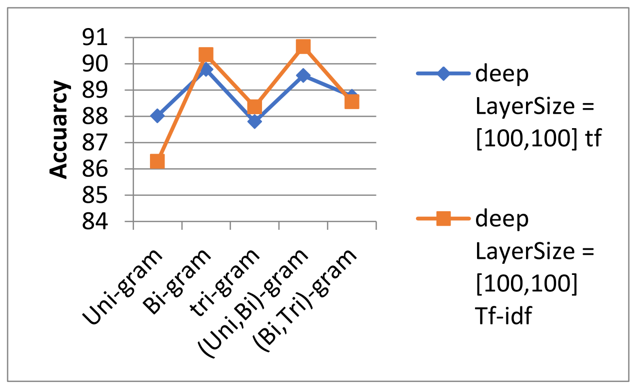

Fig. 4. The Change of Accuracy for Proposed Method versus N-Gram Model.

Moreover, the best result achieved absolutely in this experiment was in the proposed combination of (uni-bi) gram with ratio 90.65%. Fig. 4 shows the highest of accuracy ratios between the N-gram and the proposed combination of words with TF and TF-IDF representation methods.

E. Comparing of the SCG and the GDX According ICW Proposed Method Versus N-Gram Method

The experiment was executed based upon the framework in Fig. 2 with the detailed steps which have been discussed before. This experiment was established to explore the effects of both fixed number of hidden learning layers and fixed number of neurons for each layer in the accuracy level of the classifier. Besides, it compares different combination methods such as uni-bi gram, bi-tri gram versus the N-gram with different representation methods. We measure the accuracy level of the equation number based on the confusion matrix.

Tables V and VI shows the results for discussing the factors that affect the accuracy of the classifier with both representation methods TF and TF-IDF. The main purpose of the experiment stills the highest accuracy for text classification that achieved with a detected number of layers and 100 neurons between the proposed combination method (ICW) and the N-gram. It has been shown that the proposed combination uni-bi gram with the SCG algorithm outperforms the other technique with three layers and 100 neurons even with the TF representation method or with the TF-IDF representation method with ratios of 89.45% and 89.60%, respectively.

TABLE V. COMPARING OF THE SCG VERSUS THE GDX ACCORDING N-GRAM METHOD

Feature selections	Layer size	Algorithm	Uni-gram	Bi-gram	tri-gram
			ACC	ACC	ACC
Tf	Ls[2*100]	scg	88.02	89.78	87.79
		gdx	85.82	82.23	80.12
	Ls[3*100]	scg	88.74	88.35	87.86
		gdx	87.35	81.98	79.52
Tf-idf	Ls[2*100]	scg	86.29	90.34	88.36
		gdx	86.07	81.95	79.68
	Ls[3*100]	scg	87.18	87.86	87.55
		gdx	84.22	83.94	79.90

TABLE VI. COMPARING OF THE SCG VERSUS THE GDX ACCORDING ICW PROPOSED METHOD

Feature selections	Layer size	Algorithm	(Uni,Bi)-gram	(Bi,Tri)-gram
			ACC	ACC
Tf	Ls[2*100]	scg	89.55	88.76
		gdx	84.93	80.52
	Ls[3*100]	scg	89.45	86.62
		gdx	83.93	83.20
Tf-idf	Ls[2*100]	scg	90.65	88.55
		gdx	83.62	80.27
	Ls[3*100]	scg	89.60	87.66
		gdx	83.99	80.31

Moreover, the best result achieved absolutely in this experiment was in the proposed combination of (uni-bi) gram with a ratio of 90.65% with the TF-IDF representation method, two layers, and 100 neurons. In addition, the last comparison mentioned that the SCG algorithm outperforms the GDX algorithm in all experiments.

VII. CONCLUSION

This paper aims to investigate and to improve a generic framework for Arabic text classification. It deals directly with a word in its original style as a basic unit of modern Arabic sentence and on different levels of N-grams versus ICW proposed method. However, it aimed at discussing the results of the different experiments for studying the effect of changing the architecture concerning learning layers of different deep learning algorithms on ATC as a case study. In addition, a new legal dataset for requests of Prosecutors was presented for the purpose of testing and training the proposed framework in different experiments.

The main results that are drawn from this work showed that with increasing the number of hidden layers with a fixed number of neurons affects the accuracy of the SCG classifier with TF and TF-IDF representation methods respecting to uni-gram. However, it has been detected that the SCG algorithm achieved the best accuracy ratio with 90.07% with the TF-IDF representation method; also, in comparing the accuracy of the SCG algorithm between N-gram and the proposed method (ICW) with a fixed number of layers and neurons. It has been shown that representing the text with the TF-IDF representation method and the proposed method (ICW) (uni-bi) gram outperforms TF representation method with ratio 90.65%. Moreover, it has been shown that the proposed method (ICW) (uni-bi) gram with the SCG algorithm with TF-IDF representation method outperforms the GDX algorithm with a ratio of 90.65%.

For future work, the proposed model needs to be tested with different large datasets as a benchmark for generalizing and extracting a lot of results. Besides, the proposed model needs to integrate an optimization technique as feature reduction for enhancing the “curse of dimensionality” problem.

REFERENCES

[1] Khreisat, L., A machine learning approach for Arabic text classification using N-gram frequency statistics, Journal of Informatics, Volume 3, 2009.

[2] Khorsheed, Mohammad S., and Abdulmohsen O. Al-Thubaity, Comparative evaluation of text classification techniques using a large diverse Arabic dataset, Language Resources and Evaluation, 2013.

[3] Joachims.T, Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Proceedings of the European Conference on Machine Learning (ECML), Springer, 1998.

[4] Elhassan.R, Ahmed.M, Arabic Text Classification on Full Word, International Journal of Computer Science and Software Engineering (IJCSSE), Volume 4, Issue 5, 2015.

[5] Abd El-Monsef .M. E, Amin.M, Atlam.E, El-Barbary.O, Arabic Document Classification: A Comparative Study,Journal of Computing, Volume 3 ,Issue 4, April 2011.

[6] Abbès, R., Dichy, J, AraConc, An Arabic Concordance Software Based on the DIINAR.1 Language Resource. In: The 6th International Conference on Informatics and Systems, 2008.

[7] Darwish. K., Building a shallow Arabic morphological analyzer in one day. In Proceedings of the ACL 2002 Workshop on Computational Approaches to Semitic Languages, Stroudsburg, PA, USA, pages 1–12, 2002.

[8] Abuaiadah. D, Arabic Document Classification Using Multiword Features , International Journal of Computer and Communication Engineering, Vol. 2, No. 6, 2013.

[9] F. Al-Zaghoul and S. Al-Dhaheri, Arabic Text Classification Based on Features Reduction Using Artificial Neural Networks, Proceedings of the 15th International Conference on Computer Modelling and Simulation (UKSim), Cambridge University, United Kingdom, , pp. 485-490, April 2013 .

[10] Sayed. M, Salem. R and Khedr. AE, A survey of Arabic text classification approaches, International Journal of Computer Applications in Technology 59(3):236 – 251, March 2019 .

[11] H. Kaur and V. Gupta, Indexing process insight and evaluation, International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7830087.

[12] A. Krizhevsky , I. Sutskever and G. E. Hinton, ImageNet Classification with Deep Convolution Neural Networks, Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12), Lake Tahoe, Nevada, (2012) December, vol 1, pp. 1097-1105.

[13] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, (1998) November, pp. 2278-2324.

[14] Dania Sagheer, Fadel Sukkar, "Arabic Sentences Classification via Deep Learning ", International Journal of Computer Applications (0975 – 8887) Volume 182 – No.5, July 2018.

[15] Muhammad Helmy, R. M. Vigneshram , et al ," Applying Deep Learning for Arabic Keyphrase Extraction "The 4th International Conference on Arabic Computational Linguistics (ACLing 2018), 2018, Dubai, United Arab Emirates.

[16] Samir Boukil, Mohamed Biniz, Fatiha El Adnani3, Loubna Cherrat and Abd Elmajid El Moutaouakkil. Arabic Text Classification Using Deep Learning Technics. International Journal of Grid and Distributed Computing Vol. 11, No. 9 (2018), pp.103-114.

[17] G. Salton and C. Buckley, "Term-weighing approache in automatic text retrieval", Information Processing & Management Journal., vol. 24, no. 5, (1988), pp 513-523.

[18] M.F. Moller.: A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Vol. 6, pp. 525–533, 1993.

[19] Maryam Habibie and Andrei Propescu-Belis.: Keyword Extraction and Clustering for Document Recommendation in Conversations, IEEE/acm Transactions on Audio, Speech, and Language Processing, vol. 23, no. 4, pp. 746, April 2015.

[20] M.Z. Rehman and N.M. Nawi : ICSECS 2011, Springer ,Part I, CCIS 179, pp. 380–390, 2011.

[21] Thabtah, F., Eljinini, M., Zamzeer, M., & Hadi, W. Naı¨ve Bayesian based on Chi Square to categorize Arabic data. In Proceedings of The 11th International Business Information Management Association Conference (IBIMA) Conference on Innovation and Knowledge Management in Twin Track Economies, pp. 930–935, 2009.

[22] Alharbi, F.R., Khan, M.B. Identifying comparative opinions in Arabic text in social media using machine learning techniques. SN Appl. Sci. 1, 213, 2019.

[23] Al-Taani A.T., Al-Sayadi S.H. Classification of Arabic Text Using Singular Value Decomposition and Fuzzy C-Means Algorithms. In: Johri P., Verma J., Paul S. (eds) Applications of Machine Learning. Algorithms for Intelligent Systems. Springer, Singapore,2020.

[24] El-Fishawy, N., Hamouda, A., Attiya, G. M., & Atef, M. Arabic summarization in Twitter Social Network. Ain Shams Engineering Journal, 5(2), 411–420. https://doi.org/10.1016/j.asej.2013.11.002, 2014.

[25] M Attia, MA Abdel-Fattah, Ayman E. Khedr," A proposed multi criteria indexing and ranking model for documents and web pages on large scale data ". Journal of King Saud University-Computer and Information Sciences,2021.

[26] Boudad, N., Faizi, R., Oulad Haj Thami, R., & Chiheb, R. Sentiment Analysis in Arabic: A review of the literature. Ain Shams Engineering Journal, 9(4), 2479–2490. https://doi.org/10.1016/j.asej.2017.04.007, 2018.

[27] Al-Anzi, F. S., & AbuZeina, D. Synopsis on Arabic speech recognition. Ain Shams Engineering Journal, 13(2), 101534. https://doi.org/10.1016/j.asej.2021.06.020, 2021.

[28] Amira M. Idrees, Essam M. Shaaban," Building a Knowledge Base Shell Based on Exploring Text Semantic Relations from Arabic Text". International Journal of Intelligent Engineering and Systems, Vol.13, No.1, 2020.

[29] Mohamed, M. A., Abdel-Fattah, M. A., & Khedr, A. E. (2021). Challenges and Recommendations in Big Data Indexing Strategies. International Journal of e-Collaboration (IJeC), 17(2), 22-39. http://doi.org/10.4018/IJeC.2021040102

[30] Khedr, A. E., Idrees, A. M., & Alsheref, F. K. (2019). A Proposed Framework to Explore Semantic Relations for Learning Process Management. International Journal of e-Collaboration (IJeC), 15(4), 46-70. http://doi.org/10.4018/IJeC.2019100104.

[31] Mostafa, A. M., Idrees, A. M., Khedr, A. E., & Helmy, Y. M. (2020). A proposed architectural framework for generating personalized users' query response. Journal of Southwest Jiaotong University, 55(5). https://doi.org/10.35741/issn.0258-2724.55.5.3

[32] Amr Mansour Mohsen, Hesham Ahmed Hassan, Amira M. Idrees.(2016). Documents Emotions Classification Model Based on TF-IDF Weighting Measure. International Journal of Computer and Information Engineering Vol:10, No:1.

[33] Afify, E. A., Sharaf, A., & Khedr, A. E. (2020). Facebook profile credibility detection using machine and deep learning techniques based on user's sentiment response on status message. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/10.14569/ijacsa.2020.0111273