MULTITASKING DEEP NEURAL NETWORK MODELS FOR ARABIC DIALECT SENTIMENT ANALYSIS

By

MUATH MOHAMMAD OQLAH ALALI

Thesis Submitted to the School of Graduated Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

August 2022

FSKTM 2022 26

DEDICATION

In the name of Allah, Most Gracious, Most Merciful

This thesis is dedicated to:

My father

My late mother

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

MULTITASKING DEEP NEURAL NETWORK MODELS FOR ARABIC DIALECT SENTIMENT ANALYSIS

By

MUATH MOHAMMAD OQLAH ALALI

August 2022

Chairman : Associate Professor Ts. Nurfadhlina binti Mohd Sharef, PhD
Faculty : Computer Science and Information Technology

Polarity classification or sentiment analysis is considered one of the opinion mining tasks which distinguishes between the polarities categories (two, three, and five) of opinions which focus on the degree of the sentiment (such as positive and negative for two polarities; and positive, neutral and negative for three polarities) that the text may contain. Limited deep neural network approaches are applied to this task for Arabic dialects (AD). On the other hand, traditional machine learning algorithms (ML) that are based on manually extracted features are considered tedious and time dunting, as Arabic language contains multiple dialects and no word-based order. Therefore, the process of extracting features such as syntactic and lexical information is more challenging for AD.

According to the literature review, the best registered performance and most used deep learning model for Arabic sentiment analysis was Convolutional Neural Network (CNN). The existing convolutional network models are based on wide convolutional with shallow structure that represents less uniform importance to the features, which is not capable of representing the entire sentiment information in text sequence and leads to poor sentiment information detection. Therefore, a Narrow Convolutional Neural Network (NCNN) is proposed to extract comprehensive sentiment information of text sequence by maximizing the feature detection range, which gives large uniform importance to the words and improves the final performance for Arabic dialect classification tasks (two and three polarities). NCNN achieves its optimum performance when structured by three convolutional layers. Sensitivity analysis is conducted to evaluate the impact of various combinations of NCNN structural hyperparameters, such as the size of pooling, filters, and the number of convolutional filters on the classification performances. The proposed NCNN achieved a higher macro average recall (R) and outperforms Naive Bayes (NB) on task A (three polarities) and Voting model on task B (two polarities) on the SemEval-2017 Arabic dialect Twitter dataset. In addition, the NCNN model outperforms CNN-ASWAR on Arabic Sentiment Tweets Dataset (ASTD) with higher F1-score.

The negation words in the Arabic language plays a significant role in SA. Negation words may cause a sentence's context to be reversed. So far, there has been no effort to handle the negation context in Arabic using a deep neural network. The existing approaches are based on traditional machine learning algorithms, such as support vector machine (SVM). However, these approaches did not consider Arabic dialect negation words. In addition, these approaches are based on domain specific features and lexicons, which might not work with other domains.

Ordinal (five polarities) classification problem has received attention in Arabic sentiment analysis. Most of the applied approaches are based on single task learning (STL) using machine learning algorithms, such as Logistic Regression (LR) and Hierarchical Classifier (HC) based on the divide-and-conquer approach. However, these approaches are based on simple sentence representation. Moreover, these models are based on single task learning (STL) and lack the ability to learn the relativity between different tasks (cross-task transfer) and modelling several polarities jointly, such as three and five polarities.

Therefore, a model called Multi-Tasking Learning based on Convolutional Hierarchical Attention Neural Network (MTL-CHAN) is proposed, comprising of (i) shared word encoder and word attention networks across classification tasks, (ii) task-specific layers with convolutional neural network-based attention (CNNA) on sentence-level; to handle the Arabic explicit negation words and improve the classification performance by training Arabic classification tasks (binary, ternary, and five) jointly. The experimental results showed outstanding performance of the proposed MTL-CHAN model, with high accuracy of 89.85%, 84.69%, 85.90 on HARD, LABR, and BRAD datasets, respectively, and higher macro average recall (R) of 0.680% and 0.810% on Twitter Arabic dialects datasets task A and B respectively. Also, the proposed model achieved higher accuracy of 95.25%, 87.75%, 86.01%, 90.95% on Hotel, Product, Movie, and Restaurant datasets, respectively.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MODEL RANGKAIAN NEURAL MENDALAM PELBAGAI TUGASAN UNTUK ANALISIS SENTIMEN DIALEK ARAB

Oleh

MUATH MOHAMMAD OQLAH ALALI

Ogos 2022

Pengerusi : Profesor Madya Ts. Nurfadhlina binti Mohd Sharef, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Klasifikasi polariti atau analisis sentimen dianggap sebagai salah satu tugas perlombongan pendapat yang membezakan antara polariti yang mungkin dikandungi oleh teks. Jumlah pendekatan rangkaian neural mendalam yang digunakan untuk tugas ini untuk dialek Arab (AD) adalah terhad. Bahasa Arab mengandungi pelbagai dialek dan tiada susunan berasaskan perkataan, oleh itu, proses mengekstrak ciri-ciri dan melaksanakan analisis sentimen adalah lebih mencabar dan memakan masa.

Menurut kajian literatur, prestasi berdaftar terbaik model pembelajaran mendalam untuk analisis sentimen bahasa Arab ialah Rangkaian Neural Konvolusi (CNN). Model rangkaian konvolusi sedia ada adalah berdasarkan konvolusi luas dengan struktur cetek yang mempunyai kepentingan yang kurang seragam kepada ciri-ciri, yang juga tidak mampu mewakilkan keseluruhan maklumat sentimen dalam urutan teks, dan membawa kepada pengesanan maklumat sentimen yang lemah. Oleh itu, Rangkaian Neural Konvolusi Sempit (NCNN) dicadangkan untuk mengekstrak maklumat sentimen yang komprehensif bagi urutan teks dengan memaksimumkan julat pengesanan ciri-ciri, yang seterusnya memberikan kepentingan seragam yang besar kepada perkataan dan meningkatkan prestasi akhir untuk tugas klasifikasi dialek Arab (tiga dan dua kekutuban). NCNN mencapai prestasi optimumnya apabila distrukturkan oleh tiga lapisan konvolusi. Model ini dibangunkan tanpa menggunakan ciri leksikal dan leksikon atau menambah set data. Analisis sensitiviti dijalankan untuk menilai kesan pelbagai kombinasi hiperparameter struktur NCNN, seperti saiz pengumpulan, penapis, dan bilangan penapis konvolusi pada prestasi pengelasan. NCNN yang dicadangkan mencapai purata ingatan semula (R) makro yang lebih tinggi, dan mengatasi prestasi Naive Bayas (NB) pada tugas A (tiga kekutuban) dan model pengundian pada tugas B (dua kekutuban) pada dataset Twitter dialek Arab Semeval-2017. Selain itu, model NCNN mengatasi CNN-ASWAR pada Set Data Twitter Sentimen Arab (ASTD) dengan skor F1 yang lebih tinggi.

Perkataan-perkataan penafian dalam bahasa Arab memainkan peranan penting dalam SA. Perkataan-perkataan-perkataan penafian boleh menyebabkan konteks ayat menjadi terbalik. Setakat ini, belum ada usaha untuk mengendalikan konteks penafian dalam bahasa Arab menggunakan rangkaian neural yang mendalam. Pendekatan sedia ada adalah berdasarkan algoritma pembelajaran mesin tradisional, seperti mesin vektor sokongan (SVM). Walau bagaimanapun, pendekatan ini tidak mengambil kira perkataan-perkataan penafian dialek Arab. Selain itu, pendekatan ini adalah berdasarkan ciri-ciri dan leksikon khusus domain, yang mungkin tidak berfungsi dengan domain lain.

Masalah klasifikasi ordinal kurang mendapat perhatian dalam analisis sentimen bahasa Arab. Kebanyakan pendekatan yang digunakan adalah berdasarkan pembelajaran tugasan tunggal (STL) menggunakan algoritma pembelajaran mesin, seperti Regresi Logistik (LR) dan Pengelas Hierarki (HC) berdasarkan pendekatan dasar pecah dan perintah. Walau bagaimanapun, pendekatan ini adalah berdasarkan perwakilan ayat mudah. Selain itu, model ini adalah berdasarkan pembelajaran tugasan tunggal (STL) dan tidak mempunyai keupayaan untuk mempelajari relativiti antara tugasan yang berbeza, seperti tugasan klasifikasi ternari dan lima kekutuban (pemindahan silang tugas).

Oleh itu, model yang dipanggil Pembelajaran Pelbagai Tugas berdasarkan Rangkaian Neural Perhatian Hierarki Konvolusi (MTL-CHAN) dicadangkan, yang terdiri daripada (i) pengekod perkataan dikongsi dan rangkaian perhatian perkataan merentas tugasan pengelasan, (ii) lapisan khusus tugas dengan perhatian berasaskan rangkaian neural konvolusi (CNNA) pada peringkat ayat; untuk mengendalikan kata-kata penafian eksplisit Arab dan meningkatkan prestasi klasifikasi dengan melatih tugas-tugas klasifikasi bahasa Arab (Binari, Ternari, dan Lima) secara bersama. keputusan eksperimen menunjukkan prestasi cemerlang model MTL-CHAN yang dicadangkan, dengan ketepatan tinggi 89.85%, 84.69%, 85.90 pada set data HARD, LABR, dan BRAD, dan purata ingatan semula makro (R) yang lebih tinggi sebanyak 0.680% dan 0.810% di Twitter dialek Arab set data Tugas A dan B. Selain itu, model yang dicadangkan mencapai ketepatan yang lebih tinggi iaitu 95.25%, 87.75%, 86.01%, 90.95% pada set data Hotel, Produk, Filem dan Restoran.

ACKNOWLEDGEMENTS

“First and foremost, solemn praise and humble gratitude to Allah, the Almighty, for bestowing His blessings on me throughout the research work to its successful completion.

I would like to express my deepest and sincerest gratitude to my research supervisor, Assoc. Prof. Nurfadhlina Mohd Sharef for allowing me to conduct research and providing invaluable guidance throughout the study. Her dynamism, vision, sincerity, and motivation have deeply inspired me. She has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honour to work and study under her guidance. I am extremely grateful for what she has offered me. Besides my advisor, I would like to thank the rest of my supervisor committee: Assoc. Prof. Masrah Azrifah Azmi Murad, Dr Hazlina Hamdan, and Dr. Nor Azura Husain for their encouragement and insightful comments.

I am extremely grateful to my father for his care, love, prayers, sacrifices, support, motivation, and guidance to complete this work and prepare me for my future.

Also, I would like to express my gratitude to Dr. Safa at the University of Jordan and Dr Laith in Kyungpook National University, South Korea, for their help and support in providing the hardware resources that helped me do the experimental works and complete this work.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Nurfadhlina binti Mohd Sharef, PhD
Associate Professor Ts.
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Masrah Azrifah binti Azmi Murad, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Hazlina binti Hamdan, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Nor Azura binti Husin, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 12 January 2023

Declaration by Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and the copyright of the thesis are fully-owned by Universiti Putra Malaysia, as stipulated in the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from the supervisor and the office of the Deputy Vice-Chancellor (Research and innovation) before the thesis is published in any written, printed or electronic form (including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials) as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld in accordance with the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2015-2016) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: ______________________________ Date: __________________

Name and Matric No.: Muath Mohammad Oqlah Alali

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research and the writing of this thesis were done under our supervision.
- supervisory responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2015-2016) are adhered to.

Signature: ______________________________
Name of Chairman of Supervisory Committee: Associate Professor Ts. Dr. Nurfadhlina binti Mohd Sharef

Signature: ______________________________
Name of Member of Supervisory Committee: Associate Professor Dr. Masrah Azrifah binti Azmi Murad

Signature: ______________________________
Name of Member of Supervisory Committee: Dr. Hazlina binti Hamdan

Signature: ______________________________
Name of Member of Supervisory Committee: Dr. Nor Azura binti Husin

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Arabic dialect
ADSA	Arabic dialect sentiment analysis
ASA	Arabic sentiment analysis
CNN	Convolutional Neural Network
CNNA	Convolutional Neural Network-based Attention
DL	Deep Learning
DNN	Deep Neural Network
HAN	Hierarchical Attention Network
HC	Hierarchical classifier
MAE^M	Macro Mean Absolute Error
MSA	Modern Standard Arabic
MTL	Multitask learning
MTLCHAN	Multitask Learning based on Convolutional Hierarchical Attention Neural Network
NCNN	Narrow convolutional neural network
R	Macro Average Recall
STL	Single task learning

CHAPTER 1

INTRODUCTION

1.1 Background

Sentiment Analysis (SA) is a natural language processing (NLP) task that has become increasingly important in data analysis and information extraction fields over recent years (Pang & Lee, 2008). The primary purpose of SA is to detect the sentiments and to classify the text according to polarity categories such as binary (positive and negative), ternary (positive, neutral, negative) or five classes (strong positive, positive, neutral, negative, strong negative). Sentiment Analysis in the Arabic language had an interesting evolution recently. First, Arabic social media content has grown exponentially, and it is currently one of the most popular languages on social media. Second, there are a vast number of online users, estimated at 2.2 billion, and predicted to increase to 2.72 billion in 2020; therefore, it is difficult to manually collect and classify the comments written by such users, which should be automated instead. In addition, the enormous number of Arab users reveals that there are a lot of potential entities (both governmental and commercial) benefiting from a system that can extract and classify online comments according to their sentiments.

This work covers the sentiment analysis using textual data written in the Arabic dialect (AD). Arabic Dialects Sentiment analysis (ADSA) is considered a challenging task because of the language's morphology, orthography, and complex nature. Most of the approaches that have been applied to ADSA are based on conventional machine learning algorithms. In contrast, deep neural network approaches are still in the early stages of AD and have very limited applications. Most of these approaches are based on CNN (Al-Azani, Sadam, 2017; Dahou et al., 2016; Gridach et al., 2018; Alayba et al., 2017). These approaches typically comprise of shallow structures that are not capable of capturing the entire sentiment features articulated in text sequence for Arabic dialects. Moreover, these models are based on CNN structures applied to the English language (Kim, 2014). Besides, no attempt has been made to resolve the Arabic context negation using a deep neural network in ADSA. In addition, the five-polarity classification problem has less attention in ASA than binary and ternary classification problems. Most of the applied approaches are based on conventional machine learning (ML) algorithms that are based heavily on handcrafted features, which is considered time-consuming. Moreover, these approaches are based on single task learning and do not consider the related tasks. The reported performance of existing works on ADSA still has various areas to improve, which can be achieved using DNN approaches.

1.2 Problem Statements

The research on ADSA faces several challenges, which are listed below:

1.2.1 Effective classifier structure using CNN for ADSA

Modern Standard Arabic (MSA) language is different from AD, which people use daily on social media. Most educated Arabic-speaking people are supposed to understand MSA, but this is not always true for dialects. AD differs from MSA morphologically and syntactically, making the SA task quite challenging. Recently, in the Semeval-2017 challenge (Rosenthal et al., 2017), the Arabic dialect tweets Dataset that addresses three tasks has been released. The first Task (A) is to classify the tweets into three polarities: negative, neutral, and positive. The Second Task (B) is to classify the tweets into two polarities, i.e., negative and positive. The third Task (C) is to classify the tweets into five polarities: highly negative, negative, neutral, positive, and highly positive. Task C has not been considered in this study for the Arabic dialect tweets dataset because the Task C dataset is highly imbalanced. The number of samples for the high positive and high negative in the test dataset is one for both classes. Scale, polarity, and point are used interchangeably.

Most approaches that have addressed these tasks are based on hand-crafted features. In the best-conducted work identified, (El-Beltagy et al., 2017) have proposed a voting model for binary classification and a Naïve Bayes (NB) model for ternary classification. Their models are based on hand-crafted features to enrich the sentiment such as counts of exclamation marks, question marks, elongated words, negated contexts, positive emotions, negative emotions, part-of-speech tags, and positive and negative words. Besides lexicons, other hand-crafted features are also used to represent the syntactic and lexical based features such as word and character n-gram, translated lexicon, flags to indicate if the tweet starts with a link, positive and negative words, hashtag, and question marks. However, feature extraction is considered tedious and time daunting. These approaches can become obsolete over time, primarily if the content of the data diverges, such as when a different is involved.

On the other hand, the application of deep neural network approaches on Arabic colloquial are very limited. The most deep neural network models applied on ADSA are based on CNN; this is due to the convolutional filter operation it corresponds to a linguistic feature detector that learns to recognize a specific class of N-grams such as unigram, bigram and trigram. However, these CNN models are based on shallow CNN structures, consisting of one or two convolutional layers with various filter sizes applied on sentence matrix (Gridach et al., 2018). Thus, these models' structures have less uniform importance to the feature, which leads to poor sentiment information detection.

A CNN with single convolutional layer can extract simple features. On the other hand, a network with multiple convolutional layers helps in learning more complicated features by aggregating the simple features learned in previous convolutional layers. Therefore, every convolutional layer output has more uniform and important features than the previous convolutional layer. For example, the filters in the first convolutional layer identify only the edges or corners (connected dots or pixels) of the input image. Information like individual dots or small length connected dots are ignored. But these extracted features (edges or corners) are not sufficient to identify an object uniformly. Therefore, such features are known as "less uniform or less important features". When these corners and edges are input to the second convolutional layer, the output is the combination of these edges and corners extracted from the previous layer, which is a more powerful representation of the object, such as parts of the face, the paw of a dog, the hood of a car, etc. The output of the second convolutional layer is a feature map with "more uniform and important features" to identify the target object. Similarly, for text classification the single convolutional layer can extract simple features such as words. When these words used as input to the next convolutional layer, the CNN learns more complicated and important features such as words with its prefixes and suffixes. Therefore, shallow CNN structures have less uniform importance to the feature, while increasing the depth of CNN gives large uniform importance to the features.

1.2.2 Limitation in contextual sentence understanding of negation words

One of the biggest contextual challenges in AD that may mislead the opinion orientation is negation words. Negation words reverse the polarity of sentiment words that follow them. For example, the word “مهم” / (important) / mhm/ represents positive polarity but expresses a negative polarity in the following sentence “هذا الكتاب غير مهم” / (This book is not important) since it is preceded by the negation word “غير” / (isn’t) / gyr/. In AD, there are several forms of negation, for example {لا /lA/ (no)– ولا /wlA/ (nor)– فلا /flA/ (no)– بلا /blA/ (without)–اللا /AllA/ I'm not)– لم /lm/ (did not)– ولم /wlm/ (and did not)– فلم /flm/ (and did not)}.

The approaches that handled the explicit negation words in Arabic sentiment analysis are based on conventional machine learning algorithms such as SVM and utilized a list of negation words (Touahri & Mazroui, 2019). In addition, they flipped the term polarity if it happened to be negated for negation context detection. Moreover, these approaches do not consider the Arabic dialect's negation words. Furthermore, there has been no effort to address the Arabic negation context using a deep neural network in ADSA. However, handling the explicit negation context required a robust deep learning model to learn effective latent feature representations. Sentence representations, sentence feature representations, feature latent representation, and hidden representations are used interchangeably.

One of the robust deep learning models that the recent Arabic literature review (Nassif et al., 2020) emphasized on the need for a modernized deep learning model on Arabic sentiment analysis, such as the hierarchical attention network (HAN) model (Yang et al., 2016). The HAN model uses the recurrent neural networks (RNN) with Attention network on word and sentence levels. However, RNN only considers the global features by encoding the general structure, positional information, and long-term relationships among the text sequences, which encode the semantics of the view of text sequences and ignores the local features to some extent. Therefore, such features are called global features because these are obtained from the whole text (like sentence, paragraph, or a document). At the same time, CNN captures local relationships among the neighbour words in terms of context windows (filters). CNN applies multiple filters of varying sizes on the input text. By varying the size of the filters and concatenating their outputs, it allows to detect patterns of multiples sizes (2, 3, or 5 adjacent words). These patterns express (word N-gram). These features are called the local features because these are extracted from the fixed window size and not a whole comment. Also, these local features identify short range relationship among the words of a window (Jin et al., 2020). The N-gram features are considered the most informative features in ADSA literature (El-Masri et al., 2017). Therefore, incorporating the CNN in the HAN model is needed because the filtering mechanism in CNN acts as N-gram features extractor. Also, the attention network with CNN has been rarely explored in ADSA.

The success of the deep neural network model is primarily based on the availability of ample resources, such as a large collection of the training dataset (Ahmad et al., 2019). In addition, a deep neural network model such as RNN is often expensive and requires extensive training data (Ahmad et al., 2019). Moreover, the amount of training data used in training affects the quality of the deep learning model in learning a robust sentence representation (Hessel et al., 2019). Arabic is considered a low resource language, and most of the available datasets are small (Al-Ayyoub et al., 2019). Therefore, building and learning a robust deep learning model to learn effective latent feature representation over a small dataset is quite challenging. Multitask learning (MTL) (Thrun, 1997) is an effective method to learn robust sentence representation and improve the performance of sentiment analysis classification tasks, while there is not enough training data for any single task, and related tasks' datasets are available (Hessel et al., 2019). However, there has been no effort to use MTL in ASA so far.

1.2.3 Relativity learning limitation in ordinal classification

The five-polarity classification problem in ADSA has gained less popularity than other classification tasks (ternary and binary polarity) (Al-Ayyoub et al., 2019). The existing models that have tackled ordinal classification are based on single task learning using traditional ML algorithms (Al Shboul et al., 2015). The terms five-polarity, five-point, ordinal-scale, fine-grained are used interchangeably in this study.

The best work identified on the five-point classification problem on the LABR dataset is the hierarchical classifier (HC) (Nuseir et al., 2017). Their model is based on the divide-and-conquer approach in which the five classes are decomposed into nodes representing subproblems, where each node exemplifies a different classification subproblem. However, The HC model only selects core classifiers without considering sentence representation. The best works on the HARD and BRAD datasets are (Elnagar et al., 2017) and (Elnagar & Einea, 2016); both works are based on a logistic regression model with N-gram features. However, their approaches are based on ML algorithms which do not produce a robust sentence representation. Moreover, these approaches are based on single task learning and lack the ability to learn the relativity between different tasks, such as five-point and ternary tasks, which could be addressed through an MTL-based approach. However, no research has been found to use MTL for learning for learning a five-point ASA classification.

1.3 Research Objectives

The main objective of this study is to propose a multitasking model based on Convolutional Hierarchical Attention (MTL-CHAN) to learn more effective feature latent representation and improve the performance of sentiment classification tasks by learning these tasks (binary and three polarities) and (five and three polarities) jointly. This is achieved by the following sub-objectives:

1. To propose a narrow convolutional neural network (NCNN) to capture the sentiment information contained in text sequence by maximizing the features detection range, which gives large uniform importance to the words.

2. To empirically determine the best hyperparameter (the size of the filters, pooling, and the number of filters) for the proposed NCNN model.

3. To propose a convolutional hierarchical attention network (CHAN) based on hierarchical sentence representation (word and sentence levels) to produce effectives latent representation to handle Arabic negation context and improve the performance for five polarity classification tasks.

1.4 Research Scope

This research focuses on binary, ternary and five polarities ADSA problem that is solved by proposing a multitasking-based approach that exploits DNN structure and feature representation techniques. Nine datasets have been used to evaluate the proposed approaches. This study had to use different datasets. These datasets were used by the benchmark approaches and other researchers who focused on similar problems.

Two datasets were used for the first and second sub-objective which are Twitter Arabic dialect datasets called Semeval-2017 (Rosenthal et al., 2017), and the Arabic Sentiment Tweet Dataset (ASTD) (Nabil et al., 2015).

The third sub-objective is achieved by utilizing the Semeval-2017, ASTD, another four multi-domain datasets (product, hotel, restaurant, and movies reviews) (Elsahar & El-beltagy, 2015) datasets used for handling the Arabic explicit negation words, as well as the Large Scale Arabic Book Reviews (LABR) (Nabil et al., 2014), Book Reviews in Arabic Dataset (BRAD) (Elnagar & Einea, 2016), and the Hotel Arabic-Reviews Dataset (HARD) (Elnagar et al., 2017).

1.5 Thesis Contributions

The primary contributions of this research work are improving the prediction performance of CNN for ADSA classification tasks by proposing NCNN, which can maximize the features of the detector's range. However, the limitation of NCNN is in handling the negation context. The model could not detect the negation words in the text sequence, which misclassified the sentences containing a negation word. To overcome this limitation, MTL-CHAN has been proposed to handle the negation context and improve the final performance for ADSA classification tasks.

The main contributions of this research work are as follows:

1. A narrow convolutional neural network model (NCNN) for ADSA is introduced, and it is the first time such a model has been developed for ADSA. The proposed model structure improves the performance of classification tasks by maximizing the feature detector range, which gives large uniform importance to the features.

For example, the following tweet holds a positive polarity:

" التقى الطفل بابا نويل وتلقى هدية ...كانت أمنيته الأخيرة قبل المأسا "

"The Kid received the gift after he met Santa Claus, which was the last wish before the tragedy."

Figure 1.1 visualizes the features detected based on wide CNN, while NCNN output is illustrated in Figure 1.2. The CNN-ASWAR model (Gridach et al., 2018) consists of a single convolutional layer with max pooling. The range of feature detection of the model is very low, as it could only detect "the tragedy" (المأساة), "kid" (الطفل) as the informative features in the sentence and ignoring the sentiment feature in the sentence which then lead to misclassifying the sentence.

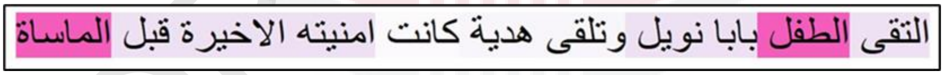

Figure 1.1: CNN-ASWAR features detection. Color gradient from light to dark, which the darker shades represent the selected features by CNN model (Gridach et al., 2018), which are: the tragedy (المأساة), Kid(الطفل)

Figure 1.2: NCNN Features Detection. Colour gradient from light to dark, which the darker shades represent the selected features by NCNN, which are: Kid(الطفل), Gift(هديه), Last Wish (أمنيته الأخيرة), Before the Tragedy (قبل المأساة), Santa Claus (بابا نويل)

2. The study redevelops and modifies the HAN model for ADSA. by proposing a convolutional neural network combined with an attention network, and the proposed model has been called a convolutional hierarchical attention network (CHAN). The proposed model was able to detect the explicit Arabic negation words.

3. MTL-CHAN is proposed to learn effective sentence representation from multiple sentiment classification tasks and improve the performance of Arabic sentiment classification tasks binary, ternary, and five polarities.

1.6 Thesis Outline

The rest of this thesis is organized as follows: Chapter 2 provides the background of the Arabic dialect and its challenges. Also, this chapter will give an overview of the classification approaches applied to ADSA, namely machine learning algorithms and deep neural networks, and explore the limitations of the current approaches on ADSA. Chapter 3 presents the review of the issues of current deep learning classification approaches and the proposed solution. The datasets and data preparation and metric evaluations will be described in this chapter. Chapter 4 presents both the proposed narrow convolutional neural networks in detail and the experimental results of the proposed model. Chapter 5 presents the proposed multitask learning model based on Convolutional Hierarchical Attention and the experimental results of the proposed model in detail. Chapter 6 presents the conclusion of all the experimental works and the recommendations for future work

REFERENCES

Abdelhade, N., Soliman, T. H. A., & Ibrahim, H. M. (2018). Detecting twitter users' opinions of arabic comments during various time episodes via deep neural network. *Advances in Intelligent Systems and Computing*, *639*, 232–246.

Abdul-Mageed, M., & Diab, M. (2012). *Toward building a large-scale Arabic sentiment lexicon*.

Abdul-Mageed, M., Kübler, S., & Diab, M. (2014). SAMAR: a system for subjectivity and sentiment analysis of Arabic social media. *12 Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, *July*, 19–28.

Abdulkareem, M., & Tiun, S. (2017). Comparative analysis of ML POS on Arabic tweets. *Journal of Theoretical and Applied Information Technology*, *95*(2), 403–411.

Abdulla, N. A., Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., Al-Kabi, M. N., & Al-rifai, S. (2014). Towards improving the lexicon-based approach for arabic sentiment analysis. *International Journal of Information Technology and Web Engineering*, *9*(3), 55–71.

Abdulla, N. A., & Al-ayyoub, M. (2014). An extended analytical study of Arabic sentiments Mohammed Naji Al-Kabi. *International Journal of Big Data Intelligence*, *1*, 103–113.

Abu Kwaik, K., Saad, M., Chatzikyriakidis, S., & Dobnik, S. (2019). LSTM-CNN Deep Learning Model for Sentiment Analysis of Dialectal Arabic. *Communications in Computer and Information Science*, *1108*(October), 108–121.

Adouane, W., Semmar, N., Johansson, R., & Bobicev, V. (2016). Automatic Detection of {A}rabicized {B}erber and {A}rabic Varieties. *Proceedings of the Third Workshop on {NLP} for Similar Languages, Varieties and Dialects ({V}ar{D}ial3)*, 63–72.

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). *Sentiment Analysis Using Deep Learning Techniques: A Review*. *8*(6).

Al-Ayyoub, M., Khamaiseh, A. A., Jararweh, Y., & Al-Kabi, M. N. (2019). A comprehensive survey of arabic sentiment analysis. *Information Processing and Management*, *56*(2), 320–342.

Al-Ayyoub, M., Nuseir, A., Kanaan, G., & Al-Shalabi, R. (2016). Hierarchical Classifiers for Multi-Way Sentiment Analysis of Arabic Reviews. *International Journal of Advanced Computer Science and Applications*, *7*(2), 531–539.

Al-Azani, S., & El-Alfy, E.-S. M. (2017a). *Hybrid Deep Learning for Sentiment Polarity Determination of Arabic Microblogs* (pp. 491–500).

Al-Azani, S., & El-Alfy, E. S. M. (2017b). Hybrid Deep Learning for Sentiment Polarity Determination of Arabic Microblogs. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10635 LNCS*, 491–500. https://doi.org/10.1007/978-3-319-70096-0_51

Al-Dabet, S., & Tedmori, S. (2019). Sentiment analysis for Arabic language using attention-based simple recurrent unit. *2019 2nd International Conference on New Trends in Computing Sciences, ICTCS 2019 - Proceedings, November*.

Al-Kabi, M. ., Abdulla, N. A., & Al-Ayyoub, M. (2013). An analytical study of Arabic sentiments: Maktoob case study. *2013 8th International Conference for Internet Technology and Secured Transactions, ICITST 2013*, 89–94.

Al-Kabi, M. ., Al-Qwaqenah, A. A., Gigieh, A. H., Alsmearat, K., Al-Ayyoub, M., & Alsmadi, I. M. (2016). Building a standard dataset for Arabie sentiment analysis: Identifying potential annotation pitfalls. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 0*.

Al-Kabi, M., Al-Ayyoub, M., Alsmadi, I., & Wahsheh, H. (2016). A prototype for a standard Arabic sentiment analysis corpus. *International Arab Journal of Information Technology*, *13*(1A), 163–170.

Al-kabi, M., Gigieh, A., Alsmadi, I., Wahsheh, H., & Haidar, M. (2013). An_Opinion_Analysis_Tool_for_Colloquial. *The Fourth International Conference on Information and Communication Systems (ICICS 2013)*, 23–25.

Al-Subaihin, A. A., Al-Khalifa, H. S., & Al-Salman, A. S. (2011). A proposed sentiment analysis tool for modern Arabic using human-based computing. *ACM International Conference Proceeding Series*, 543–546.

Al-Twairesh, N., Al-Khalifa, H., & Al-Salman, A. (2014). Subjectivity and sentiment analysis of Arabic: Trends and challenges. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, *2014*(June), 148–155.

Al-Walaie, M. A., & Khan, M. B. (2017). Arabic dialects classification using text mining techniques. *2017 International Conference on Computer and Applications, ICCA 2017*, 325–329.

Al Omari, M., Al-Hajj, M., Sabra, A., & Hammami, N. (2019). Hybrid CNNs-LSTM Deep Analyzer for Arabic Opinion Mining. *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*, 364–368.

Al Shboul, B., Al-Ayyouby, M., & Jararwehy, Y. (2015). Multi-way sentiment classification of Arabic reviews. *2015 6th International Conference on Information and Communication Systems, ICICS 2015*, 206–211.

Alabbas, W., Al-Khateeb, H. M., Mansour, A., Epiphaniou, G., & Frommholz, I. (2017). Classification of colloquial Arabic tweets in real-time to detect high-risk floods. *2017 International Conference On Social Media, Wearable And Web Analytics, Social Media 2017*, *2017-June*, 1–8.

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2017a). *Arabic language sentiment analysis on health services*. 114–118. https://doi.org/10.1109/asar.2017.8067771

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2017b). *Arabic language sentiment analysis on health services*. 114–118. https://doi.org/10.1109/asar.2017.8067771

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2018a). A combined CNN and LSTM model for Arabic sentiment analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11015 LNCS*(July), 179–191.

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2018b). Improving sentiment analysis in Arabic using word representation. *ArXiv*.

Albraheem, L., & Al-Khalifa, H. S. (2012). Exploring the problems of sentiment analysis in informal Arabic. *ACM International Conference Proceeding Series*, 415–418.

Alharbi, A., Taileb, M., & Kalkatawi, M. (2019). Deep learning in Arabic sentiment analysis: An overview. *Journal of Information Science*, *47*(1), 129–140.

Alotaibi, S., & Anderson, C. (2017). Word Clustering as a Feature for Arabic Sentiment Classification. *International Journal of Education and Management Engineering*, *7*(1), 1–13.

Altowayan, A. A., & Tao, L. (2016). Word embeddings for Arabic sentiment analysis. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 3820–3825.

Alwehaibi, A., & Roy, K. (2019). Comparison of Pre-Trained Word Vectors for Arabic Text Classification Using Deep Learning Approach. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1471–1474.

Aly, M., & Atiya, A. (2013). LABR: A large scale arabic book reviews dataset. *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, *2*, 494–498.

Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based Model for Arabic Language Understanding. *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, *May*, 9–15.

Assiri, A., Emam, A., & Aldossari, H. (2015). Arabic Sentiment Analysis: A Survey. *International Journal of Advanced Computer Science and Applications*, *6*(12), 75–85.

Baly, R., Badaro, G., El-Khoury, G., Moukalled, R., Aoun, R., Hajj, H., El-Hajj, W., Habash, N., & Shaban, K. (2017). A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models. *Proceedings of the Third Arabic Natural Language Processing Workshop*, 110–118.

Baly, R., Badaro, G., Hamdi, A., Moukalled, R., Aoun, R., El-Khoury, G., Al Sallab, A., Hajj, H., Habash, N., Shaban, K., & El-Hajj, W. (2017). OMAM at SemEval-2017 Task 4: Evaluation of English State-of-the-Art Sentiment Analysis Models for Arabic and a New Topic-based Model. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 603–610.

Baniata, L. H., & Park, S.-B. (2017). Sentence Representation Network for Arabic Sentiment Analysis. *Korean Society of Information Sciences and Technology, January*, 470–472.

Baziotis, C., Pelekis, N., & Doulkeridis, C. (2017). DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, *1*, 747–754. https://doi.org/10.18653/v1/S17-2126

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P., Bottou, L., Collobert, R., Karlen, M., & Kavukcuoglu, K. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, *12*, 2493–2537.

Dahou, A., Xiong, S., Zhou, J., Haddoud, M. H., & Duan, P. (2016). Word embeddings and convolutional neural network for Arabic sentiment classification. *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, 2418–2427.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North*, *1*(Mlm), 4171–4186. https://doi.org/10.18653/v1/N19-1423

Duwairi, R. (2015). Sentiment analysis for dialectical Arabic. *2015 6th International Conference on Information and Communication Systems, ICICS 2015*, 166–170.

Duwairi, R., & El-Orfali, M. (2014). A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. *Journal of Information Science*, *40*(4), 501–513.

El-Beltagy, S. R., El kalamawy, M., & Soliman, A. B. (2017). NileTMRG at SemEval-2017 Task 4: Arabic Sentiment Analysis. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 790–795.

El-halees, A. (2011). *Arabic Opinion Mining Using Combined*.

El-makky, N., Nagi, K., El-ebshihy, A., Apady, E., Hafez, O., Mostafa, S., & Ibrahim, S. (2014). Sentiment Analysis of Colloquial Arabic Tweets. *2014 ASE BigData/SocialInformatics/PASSAT/BioMedCom 2014 Conference, Harvard University, December 14-16*, 1–9.

El-Masri, M., Altrabsheh, N., & Mansour, H. (2017). Successes and challenges of Arabic sentiment analysis research: a literature review. *Social Network Analysis and Mining*, *7*(1).

Eldesouki, M., Dalvi, F., Sajjad, H., & Darwish, K. (2016). QCRI @ DSL 2016: Spoken Arabic Dialect Identification Using Textual Features. *Proceedings of the Third Workshop on {NLP} for Similar Languages, Varieties and Dialects ({V}ar{D}ial3)*, 221–226.

Elhawary, M., & Elfeky, M. (2010). Mining Arabic business reviews. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1108–1113.

Elnagar, A., Al-Debsi, R., & Einea, O. (2020). Arabic text classification using deep learning models. *Information Processing and Management*, *57*(1), 102121.

Elnagar, A., & Einea, O. (2016). BRAD 1.0: Book reviews in Arabic dataset. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, *0*.

Elnagar, A., Khalifa, Y. S., & Einea, A. (2017). Hotel arabic-reviews dataset construction for sentiment analysis applications. *Studies in Computational Intelligence*, *740*, 35–52.

Elsahar, H., & El-beltagy, S. R. (2015). *Building Large Arabic Multi-domain Resources for Sentiment Analysis*. *3406*, 23–34.

Elsahar, H., & El-Beltagy, S. R. (2014). A fully automated approach for Arabic slang lexicon extraction from microblogs. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8403 LNCS*(PART 1), 79–91.

Fahrni, A., & Klenner, M. (2008). *Old wine or warm beer : target-specific sentiment analysis of adjectives Old Wine or Warm Beer : Target-Specific Sentiment Analysis of Adjectives*. *April*, 60–63.

Farra, N., Challita, E., Assi, R. A., & Hajj, H. (2010). *Sentence-Level and Document-Level Sentiment Mining for Arabic Texts Sentence-level and Document-level Sentiment Mining for Arabic Texts*. *December*.

Gao, S., Young, M. T., Qiu, J. X., Yoon, H. J., Christian, J. B., Fearn, P. A., Tourassi, G. D., & Ramanthan, A. (2018). Hierarchical attention networks for information extraction from cancer pathology reports. *Journal of the American Medical Informatics Association*, *25*(3), 321–330.

Ghoul, D., & Lejeune, G. (2019). MICHAEL: Mining character-level patterns for arabic dialect identification (madar challenge). *ACL 2019 - 4th Arabic Natural Language Processing Workshop, WANLP 2019 - Proceedings of the Workshop*, 229–233.

González, J.-À., Pla, F., & Hurtado, L.-F. (2017). ELiRF-UPV at SemEval-2017 Task 4: Sentiment Analysis using Deep Learning. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 723–727.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning Word Vectors for 157 Languages. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 1–5.

Gridach, M., Haddad, H., & Mulki, H. (2018). Empirical evaluation of word representations on arabic sentiment analysis. *Communications in Computer and Information Science*, *782*, 147–158.

Hamouda, A. E. A., & El-taher, F. E. (2013). Sentiment Analyzer for Arabic Comments System. *International Journal of Advanced Computer Science and Applications(IJACSA)*, *4*(3), 99–103.

Heikal, M., Torki, M., & El-Makky, N. (2018). Sentiment Analysis of Arabic Tweets using Deep Learning. *Procedia Computer Science*, *142*, 114–122.

Hessel, M., Soyer, H., Espeholt, L., Czarnecki, W., Schmitt, S., & Van Hasselt, H. (2019). Multi-Task Deep Reinforcement Learning with PopArt. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 3796–3803.

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. *Proceedings of the National Conference on Artificial Intelligence*, 755–760.

Ibrahim, H. S., Abdou, S. M., & Gheith, M. (2015). MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis. *2015 IEEE 2nd International Conference on Recent Trends in Information Systems, ReTIS 2015 - Proceedings*, *4*(2), 353–358.

Jia, L., Yu, C., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. *International Conference on Information and Knowledge Management, Proceedings*, *c*, 1827–1830.

Jin, N., Wu, J., Ma, X., Yan, K., & Mo, Y. (2020). Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification. *IEEE Access*, *8*, 77060–77072.

Johnson, R., & Zhang, T. (2015). Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, *2011*, 103–112.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 655–665.

Kim, S.-M., & Hovy, E. (2004). *Determining the sentiment of opinions*. 1367-es.

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *ArXiv Preprint ArXiv:1408.5882*.

Larochelle, H., & Hinton, G. (2010). Learning to combine foveal glimpses with a third-order Boltzmann machine. *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, 1–9.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion Observer : Analyzing and Comparing Opinions on the Web. *WWW '05: Proceedings of the 14th International Conference on World Wide Web*, 342–351.

Lu, G., Zhao, X., Yin, J., Yang, W., & Li, B. (2020). Multi-task learning using variational auto-encoder for sentiment classification. *Pattern Recognition Letters*, *132*, 115–122.

Lulu, L., & Elnagar, A. (2018). Automatic Arabic Dialect Classification Using Deep Learning Models. *Procedia Computer Science*, *142*, 262–269.

Mahmoud, E.-H., Paul, R., & Mariam, A. (2018). Arabic Dialect Identification in the Context of Bivalency and Code-Switching. *Eleventh International Conference on Language Resources and Evaluation. Paris, France: European Language Resources Association*, 3622–3627.

Maks, I., & Vossen, P. (2012). A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems*, *53*(4), 680–688.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.

Mehmood, F., Chen, E., Akbar, M. A., & Alsanad, A. A. (2021). Human Action Recognition of Spatiotemporal Parameters for Skeleton Sequences Using MTLN Feature Learning Framework. *Electronics*, *10*(21), 2708.

Miculicich, L., Ram, D., Pappas, N., & Henderson, J. (2019). Document-level neural machine translation with hierarchical attention networks. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, *i*, 2947–2954.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations ofWords and Phrases and their Compositionality. *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*, *2*, 3111–3119.

Mohammad, S. M., Salameh, M., & Kiritchenko, S. (2016). How translation alters sentiment. *Journal of Artificial Intelligence Research*, *55*(June), 95–130.

Mohcine, M., & Elouardighi, A. (2018). *Standard and Dialectal Arabic Text* (Vol. 1). Springer International Publishing.

Mulki, H., Haddad, H., Gridach, M., & Babaoğlu, I. (2017). Tw-StAR at SemEval-2017 Task 4: Sentiment Classification of Arabic Tweets. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, *July*, 664–669.

Mullen, T., & Collier, N. (2004). *Incorporating topic information into sentiment analysis models*. 25-es.

Nabil, M., Aly, M., & Atiya, A. (2014). *LABR: A Large Scale Arabic Sentiment Analysis Benchmark*. 1–9.

Nabil, M., Aly, M., & Atiya, A. (2015). ASTD: Arabic Sentiment Tweets Dataset. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, *September*, 2515–2519.

Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on Machine Learning*, *3*, 807–814.

Nassif, A. B., Elnagar, A., Shahin, I., & Henno, S. (2020). Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. *Applied Soft Computing*, *98*, 106836.

Nuseir, A., Al-Ayyoub, M., Al-Kabi, M., Kanaan, G., & Al-Shalabi, R. (2017). Improved hierarchical classifiers for multi-way sentiment analysis. *International Arab Journal of Information Technology*, *14*(4A Special Issue), 654–661.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2016). Sentiment analysis in Facebook and its application to e-learning. *COMPUTERS IN HUMAN BEHAVIOR*.

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *In Proceedings of the Seventh Conference on International Language Resources and Evaluation*, 1320–1326.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Ragab, A., Seelawi, H., Samir, M., Mattar, A., Al-Bataineh, H., Zaghloul, M., Mustafa, A., Talafha, B., Freihat, A. A., & Al-Natsheh, H. T. (2019). Mawdoo3 ai at madar shared task: Arabic fine-grained dialect identification with ensemble learning. *ACL 2019 - 4th Arabic Natural Language Processing Workshop, WANLP 2019 - Proceedings of the Workshop*, *2018*, 244–248. https://doi.org/10.18653/v1/w19-4630

Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 Task 4: Sentiment Analysis in Twitter. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 502–518.

Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press.

Safaya, A., Abdullatif, M., & Yuret, D. (2020). *KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media*. *Ml*, 2054–2059.

Shoukry, A., & Rafea, A. (2012). Sentence-level Arabic sentiment analysis. *Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012*, 546–550.

Siniscalchi, S. M., & Salerno, V. M. (2017). Adaptation to new microphones using artificial neural networks with trainable activation functions. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(8), 1959–1965.

Soliman, A., Eissa, K., & El-Beltagy, S. R. (2017). AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP. *Procedia Computer Science*, *117*, 256–265.

Soliman, T. H., Elmasry, M. A., Hedar, A., & Doss, M. M. (2014). Sentiment Analysis of Arabic Slang Comments on Facebook. *International Journal of Computers & Technology*, *12*(5), 3470–3478.

Srivastava, N., Geoffrey, H., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). *Dropout : A Simple Way to Prevent Neural Networks from Overfitting*. *15*, 1929–1958.

Tachicart, R., Bouzoubaa, K., Aouragh, S. L., & Jaafa, H. (2017). Automatic identification of moroccan colloquial arabic. *Communications in Computer and Information Science*, *782*(June 2021), 201–214.

Talafha, B., Fadel, A., Al-Ayyoub, M., Jararweh, Y., AL-Smadi, M., & Juola, P. (2019). Team JUST at the madar shared task on arabic fine-grained dialect identification. *ACL 2019 - 4th Arabic Natural Language Processing Workshop, WANLP 2019 - Proceedings of the Workshop*, 285–289.

Thrun, S. (1997). Multitask learning. *Machine Learning*, *28*(1), 41–75.

Touahri, I., & Mazroui, A. (2019). Studying the effect of characteristic vector alteration on Arabic sentiment classification. *Journal of King Saud University - Computer and Information Sciences*, *xxxx*.

Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn: Machine Learning in Python. *GetMobile: Mobile Computing and Communications*, *19*(1), 29–33.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999–6009.

Wray, S. (2018). Classification of closely related sub-dialects of Arabic using support-vector machines. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 3671–3674.

Xu, K., Ba, J. L., Kiros, R., & Courville, A. (2015). Show , Attend and Tell : Neural Image Caption Generation with Visual Attention. *Proceedings of the 32nd International Conference on Machine Learning*, *37*, 2048–2057.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 1480–1489.

Zhang, Y., & Wallace, B. (2015). *A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification*. 253–263.