University of Kerbala
College of Computer Science & Information Technology
Computer Science Department

# Emotion Classification in Dialectal Arabic Social Media Texts Using Deep Learning and Natural Language Processing Techniques

A Thesis

Submitted to the Council of the College of Computer Science & Information Technology / University of Kerbala in Partial Fulfillment of the Requirements for the Master Degree in Computer Science

**Written by**
Zakariya Hussein Ali Salih Al-Safar

**Supervised by**
Assist. Prof. Dr. Hiba Jabbar Abdul wahid Aleqabie

2025 A.D. 1446 A.H.

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

(إِنَّمَا يُؤْمِنُ بِآيَاتِنَا الَّذِينَ إِذَا ذُكِّرُوا بِهَا خَرُّوا سُجَّدًا وَسَبَّحُوا بِحَمْدِ
رَبِّهِمْ وَهُمْ لَا يَسْتَكْبِرُونَ ۩ تَتَجَافَى جُنُوبُهُمْ عَنِ الْمَضَاجِعِ
يَدْعُونَ رَبَّهُمْ خَوْفًا وَطَمَعًا وَمِمَّا رَزَقْنَاهُمْ يُنْفِقُونَ)

صَدَقَ اللهُ العَلِيُّ العَظِيم

سورة السجدة

الآية ١٥-١٦

## Supervisor Certification

I certify that the thesis entitled (**Emotion Classification in Dialectal Arabic Social Media Texts Using Deep Learning and Natural Language Processing Techniques**) was prepared under my supervision at the department of Computer Science / College of Computer Science & Information Technology / University of Kerbala as partial fulfillment of the requirements of the degree of Master in Computer Science.

Signature:
Supervisor Name: Assist. Prof. Dr. Hiba J. Aleqabie
Date: / /2025

**The Head of the Department Certification**

In view of the available recommendations, I forward the thesis entitled "Emotion Classification in Dialectal Arabic Social Media Texts Using Deep Learning and Natural Language Processing Techniques" for debate by the examination committee.

Signature:
Assist. Prof. Dr. Muhannad Kamil Abdulhameed
Head of Computer Science Department
Date: / / 2025

## Certification of the Examination Committee

We hereby certify that we have studied the dissertation entitled (**Emotion Classification in Dialectal Arabic Social Media Texts Using Deep Learning and Natural Language Processing Techniques**) presented by the student (**Zakariya Hussein Ali**) and examined him/her in its content and what is related to it, and that, in our opinion, it is adequate with (**Excellent**) standing as a thesis for the Master degree in Computer Science.

Signature:
Name: Ayad Hameed Mousa
Title: Assist. Prof. Dr.
Date: / / 2025
**(Chairman)**

Signature:
Name: Ahmad Shaker Abdalrada
Title: Assist. Prof. Dr.
Date: / / 2025
**(Member)**

Signature:
Name: Mohsin Hasan Hussein
Title: Assist. Prof. Dr.
Date: / / 2025
**(Member)**

Signature:
Name: Hiba Jabbar Abdul wahid
Title: Assist. Prof. Dr.
Date: / / 2025
**(Member and Supervisor)**

Approved by the Dean of the College of Computer Science & Information Technology, University of Kerbala.

Signature:
Assist. Prof. Dr. Mowafak Khadom Mohsen
Date: 2 / 4 / 2025
(**Dean of College of Computer Science & Information Technology**)

# Dedication

الى مقام صاحب العصر والزمان الامام المهدي المنتظر (عجل الله تعالى فرجه)، الى مقام السيد علي الحسيني السيستاني (دام ظله).

إلى من غرست فيّ بذور الطموح، وسهرت من أجلي الليالي، وتحملت عناء الحياة لأصل إلى ما أنا عليه اليوم... إلى والديّ الحبيبين، لكم مني كل الحب والتقدير، فأنتم النور الذي أضاء طريقي، والدعاء الذي رافقني في كل خطوة.

الى زوجتي العزيزة وعائلتي الثانية الذين طالما كانوا بجنبي ولم يتركوني لحظة.

إلى إخوتي وأخواتي، سندي في الحياة، الذين كانوا دائمًا مصدر قوتي ودعمي، أشارككم فرحتي بهذا الإنجاز الذي لم يكن ليتحقق دونكم.

إلى أساتذتي الكرام، الذين منحوني العلم والمعرفة، وفتحوا لي أبواب الفهم والإبداع، جزاكم الله عني خير الجزاء.

إلى كل من ساندني ووقف بجانبي خلال مسيرتي العلمية، أقدم لكم هذا العمل تعبيرًا عن امتناني العميق.

**إلى كل من يؤمن بأن العلم هو مفتاح النجاح والتقدم، أهديكم هذه الرسالة سائلا الباري عز وجل ان يجعله عملا صالحا ممتدا الى ابد الحدود.**

## Acknowledgement

To my esteemed supervisor, Assist. Prof. Dr. Hiba Jabar Aleqabie, words cannot fully express my deep gratitude for your invaluable support, guidance, and encouragement throughout my research journey. Your insightful suggestions, constructive feedback, and thorough evaluation have been instrumental in shaping my work. I sincerely ask Allah to bless you with continued success and advancement.

To Assist. Prof. Dr. Mowafak Khadom Mohsen, the dean of the college, and Assist. Prof. Dr. Mohsin Hasan Hussein, the scientific assistant of the college, I extend my heartfelt appreciation for their unwavering support in helping me complete my postgraduate studies. I would also like to express my gratitude to all my professors who have taught me and enriched me with their knowledge and wisdom.

# Abstract

Due to the popularity and consensus on social media, emotion classification in textual data has received considerable attention and has been used in various applications such as sentiment analysis, public opinion mining, and mental health monitoring.

Arabic, and specifically dialectal Arabic, such as Iraqi Arabic, has its own set of unique challenges when it comes to language processing due to morphological complexity, the availability of labeled datasets, and the informal nature of many Arabic-language written forms on social media.

BERT's strength is extracting context from the text by contextual representation of words, while LSTM capturing long-term dependencies of text sequences which helps in classification, this work proposes a hybrid deep learning approach based on BERT and LSTM to improve emotion classification in Arabic social media texts, specifically in Iraqi dialect.

Three pre-classified datasets were used for emotions, and the datasets used differed in dialects. The Iraqi dialect dataset was chosen, which is Iraqi Arabic Emotion Corpus (IAEC), the Saudi dialect dataset is Arabic Pan-Arab Emotion Dataset (ArPanEmo), and the Egyptian dialect dataset is Arabic Emotions Twitter Dataset (AETD). To remedy dialectal differences and data scarcity, various text preprocessing techniques such as normalization, emoji translation, and synonym-based data augmentation are implemented.

Based on the results we obtained when using BERT in classification, the accuracy rate was 81%, and when using the proposed hybrid model, the accuracy became 84%.

The results show the contribution of these summarized methods of how sequential and contextual learning could be combined to achieve the highest accuracy among the previous Arabic sentiment analysis models. the study presents credible contribution to Arabic natural language processing (NLP) as it proposes a scalable and efficient method for the detection of emotion in dialectal text.

## Declaration Associated with this Thesis

Some of the works presented in this thesis have been published or accepted as listed below.

* Z. H. Ali and H. J. Aleqabie, "Emotion Detection in Arabic Text in Social Media: A Brief Survey," Al-Furat Journal of Innovations in Electronics and Computer Engineering (FJIECE), vol. 3, no. 2, pp. 1-10, Jun. 2024. doi: 10.46649/fjiece.v3.2.27a.6.6.2024.

* H. J. Aleqabie and Z. Hussein Ali, "Hybrid BERT-LSTM Model for Emotion Detection in Iraqi Arabic Text," 2025 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Iraq, 2025, pp. 1-6, doi: 10.1109/CSASE63707.2025.11054005.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| Abbreviation | Description |
|---|---|
| ACMID | Annotated Corpus of Mesopotamian-Iraqi Dialect |
| AETD | Arabic Emotions Twitter Dataset |
| AI | Artificial Intelligence |
| AM4 | AMD Microarchitecture Version 4 |
| AMD | Advanced Micro Devices |
| API | Application Programming Interface |
| AraBERT | Arabic BERT |
| ArPanEmo | Arabic Pan-Arab Emotion Dataset |
| BERT | Bidirectional Encoder Representations from Transformers |
| BiLSTM | Bidirectional Long Short-Term Memory |
| CLS | Cumulative Layout Shift (CLS) |
| CNN | Convolutional Neural Network |
| CPU | Central Processing Unit |
| CUDA | Compute Unified Device Architecture |
| DA | Dialectal Arabic |
| Doc2 Vec | Document to Vector |
| DT | Decision Tree |
| EFL | English as a Foreign Language (EFL) |
| FFN | Feed-Forward Network |
| GB | Gigabyte |
| GloVe | Global Vectors for Word Representation |
| GPT | Generative Pre-trained Transformer |

| | |
|---|---|
| GPU | Graphics Processing Unit |
| GRU | Gated Recurrent Unit |
| IAD | Iraqi Arabic Dialect |
| IAEC | Iraqi Arabic Emotion Corpus |
| LLMs | Large Language Models |
| LSTM | Long Short-Term Memory |
| MARBERT | Multidialectal Arabic BERT |
| MB | Megabyte |
| mBERT | Multilingual BERT |
| ML | Machine Learning |
| MLM | Masked Language Modeling |
| MSA | Modern Standard Arabic |
| NB | Naïve Bayes |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| NSP | Next Sentence Prediction |
| QA | Question Answering |
| RAM | Random Access Memory |
| RNN | Recurrent Neural Network |
| SA | Sentiment Analysis |
| SVM | Support Vector Machine |
| TF-IDF | Term Frequency - Inverse Document Frequency |
| VM | Virtual Machine |

# CHAPTER ONE

## INTRODUCTION

## 1.1 Overview

The advent of social media has transformed the way people communicate, share opinions, and express emotions. Platforms like Twitter, Facebook, and Instagram have become rich sources of user-generated content, encapsulating diverse emotional expressions across languages and cultures. Arabic, spoken by over 400 million people, is among of the most prominent languages in the world. The linguistic complexity of Arabic, which includes Modern Standard Arabic and a variety of dialects, poses unique challenges for Natural Language Processing (NLP) tasks, especially emotion classification [1], [2].

Emotion classification in text involves identifying textual data into emotional categories such as happiness, sadness, anger, or surprise. In the context of Arabic social media, this task is particularly challenging due to the informal and often dialectal nature of the language used online. Social media texts frequently mix formal Arabic with colloquialisms, slang, and even foreign words, complicating text normalization and understanding [3].

Machine learning and NLP have emerged as powerful tools for tackling such challenges. Traditional methods of emotion classification have often relied on manually crafted lexicons or rule-based systems. While effective to some extent, these methods have struggled to capture the dynamic and context-dependent nature of emotions expressed in modern communication [4]. Advances in machine learning, particularly in deep learning, have revolutionized this field. Techniques like recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformers, used in models such as BERT and GPT, enable automatic feature extraction and context-aware processing of text [5], [6].

Applying those techniques to Arabic texts has shown promising results. Pre-trained models and embeddings such as AraBERT have been fine-tuned to handle the syntactic and semantic nuances of Arabic. Moreover, the integration of sentiment analysis and emotion classification models in Arabic social media analysis holds immense potential for applications in public opinion mining, marketing, and mental health monitoring [7], [8].

The importance of emotion classification extends beyond academia. Businesses can use these insights to understand customer sentiment and improve services. Policymakers can analyze public sentiment to gauge societal issues, while mental health professionals can use such tools to identify individuals at risk of emotional distress [9]. However, ethical considerations, such as privacy concerns and potential misuse of data, warrant careful scrutiny [10].

## 1.2 Social Media Network Microblogs

Social media microblogs, such as tweets in twitter, have evolved into influential platforms for sharing opinions and engaging in discussions on a wide range of topics. These platforms enable users to express emotions in real-time, often in brief, unstructured formats, making them a rich data source for emotion classification and sentiment analysis research. Recent advancements highlight the role of deep learning models, such as BERT and LSTM, in efficiently analyzing this data, especially for tasks like rumor detection and fake news identification [11], [12]. The use of these technologies in analyzing Arabic texts adds another layer of complexity due to the linguistic diversity and informal nature of Arabic on social media [13].

## 1.3 Sentiment Analysis of Texts

Sentiment analysis is a vital aspect of Natural Language Processing (NLP), focusing on extracting subjective information and classifying textual data based on emotions or opinions expressed. This field has gained significant importance due to its applications in various domains, including e-commerce, public health, and social behavior studies [14]. Techniques for sentiment analysis range from lexicon-based approaches to advanced machine learning methods, including deep learning models such as CNNs and LSTMs. These models have demonstrated exceptional performance in handling large datasets, particularly in detecting nuanced sentiments in mixed text formats [15], [16].

Recent advancements highlight the use of hybrid models combining contextual word embeddings like BERT with lexicon features to enhance sentiment classification accuracy. Such approaches are particularly effective in addressing challenges associated with complex text structures, including those in social media posts [17]. Furthermore, the integration of sentiment analysis in health monitoring, such as during the COVID-19 pandemic, has revealed valuable insights into public emotions and societal trends[18]. These advancements underscore the potential of sentiment analysis to transform text data into actionable intelligence.

## 1.4 Uses Sentiment Analysis in Texts

Sentiment analysis has diverse applications across multiple domains. In marketing, it is used to evaluate customer feedback, enabling businesses to tailor their services and products [14]. In healthcare, sentiment analysis of patient reviews helps identify issues in service delivery [18]. Governments use sentiment analysis to understand public opinion on policies or events [15]. Furthermore, during crises, such as the COVID-19 pandemic, sentiment analysis supports the monitoring of public sentiment to guide response strategies [16]. Social media platforms leverage it to identify and filter harmful content like hate speech or misinformation, contributing to safer online spaces [17].

## 1.5 Problem Statement

Despite significant progress in the field of natural language processing (NLP) and sentiment analysis, Arabic especially local dialects such as Iraqi still faces poor representation in this domain compared to English. The problem of this research lies in three interrelated challenges:

1. The structural and linguistic complexity of Arabic texts:
Arabic is characterized by a rich morphological and syntactic structure, containing diacritics that affect meaning, and multiple forms of the same letter (e.g., أ، إ، آ، ا), which complicates text preprocessing. Furthermore, the use of affixes and associated pronouns increases the complexity of segmentation and processing [19]. Most processing tools are primarily designed for English and do not support these features, requiring custom configuration for Arabic texts [20].

2. The difficulty of handling Iraqi dialect compared to Standard Arabic:
Arabic dialects, such as Iraqi dialect, differ significantly from Standard Arabic in terms of vocabulary and style, making it difficult for MSA models to accurately understand the semantic and emotional context of dialects [21]. Studies have shown that language models not specifically trained for dialect fail to accurately extract emotions or meanings [22]. Furthermore, the scarcity of resources for dialects, including the Iraqi dialect, represents a significant obstacle to building effective models [23].

3. The shortcomings of traditional models in accurately representing emotions:
Most sentiment analysis studies focus on binary or triadic classifications: positive, negative, and neutral [24]. This classification does not capture the diversity of emotions in real-world texts. Psychologically, Ekman's classification of basic emotions (happiness, sadness, anger, fear, surprise) is considered more accurate in representing basic emotional states [25]. Using this classification in texts in the Iraqi dialect requires advanced models capable of understanding subtle emotional expression within informal social contexts.

## 1.6 Aim and Objectives

- To develop an accurate, dialect aware emotion classification system for Iraqi text by combining contextual and sequential DL approaches, overcoming the unique challenges of Arabic morphological complex dialectal variation and limited labeled data.
- To classify emotions found in the Arabic text and the Iraqi dialect into five class (happiness, sadness, anger, fear, surprise).
- To conduct system performance evaluation using standard evaluation metrics and verify that system results are highly accurate.
- To compare the performance of the proposed system with existing methods for detecting emotions in Arabic text and the Iraqi dialect.

## 1.7 Related Work

The field of sentiment analysis in Arabic texts has witnessed significant development in recent years, driven by the proliferation of social media as a primary medium for expressing opinions and attitudes. However, the Arabic language and its diverse dialects especially the Iraqi dialect remain among the most significant challenges facing this field.

**First: Studies based on binary classification (positive/negative)**

**In a study** [26] used a Twitter dataset containing more than 41,000 Arabic tweets classified as positive or negative. The XGBoost model achieved an accuracy of 89.1%, outperforming traditional models such as SVM, KNN, and Decision Tree. However, the study faced challenges related to the diversity of dialects and the shortness of texts. The research gap lies in the scarcity of advanced models such as XGBoost for analyzing Arabic texts on social media platforms.

**In an Iraqi study** [27] movie comments in the Iraqi dialect were analyzed using hybrid models. The study achieved an improvement in accuracy from 65% to 77.5% after introducing improvements to the preprocessing stage. However, the limited effective processing of the Iraqi dialect remains one of the most significant gaps in previous research.

**A distinguished study** [28] focused on classifying texts in the Iraqi dialect into five main emotions (happiness, sadness, anger, fear, and surprise), using a hybrid model combining CNN and GRU. The model achieved accuracy ranging between 91.1% and 92.5% on three Iraqi datasets from Facebook (IAEDS, ACMID, and IAD). This study is one of the few to address the Iraqi dialect in such depth in sentiment analysis.

**Second: Studies based on multi-emotion classification (five or six emotions)**

The study, [29] relied on classifying only four emotions using CNN, SVM, NB, and MLP models, achieving an F1-score of 75.7% for the CNN model [4]. The study used data from the SemEval (Affect in Tweets) task in Arabic.

**Also,** [30] used a small hand-crafted dataset (1,353 Facebook posts in the Iraqi dialect) to classify six emotions (the full Ekman model). The PPM model achieved an accuracy of 84.6%, outperforming other traditional models such as Naïve Bayes and SMO.

**In a recent study** [31] approximately 11,000 manually classified tweets were collected from the ArECTD dataset. Researchers adopted a self-learning approach, integrating MARBERT and AraBERT models, achieving an accuracy of approximately 87% and an F1-score of 0.84. Despite these good results, the models struggled to transfer their performance to dialects like Iraqi without dedicated retraining.

*Table 1.1: Related Works conclusion*

| references | Emotions number | Accuracy | Technologies | Data sets use | limitation |
|---|---|---|---|---|---|
| [26] | (positive, negative) | Accuracy = 89.1% (XGBoost) | XGBoost ، Decision Tree ، KNN ،SVM | 41,125 tweets from Kaggle Unspecified | Many dialects and short tweets affect accuracy. |
| [27] | (positive, negative) | From 65% to 77.5% | Naive Bayes ، SVM ،Logistic Regression | Iraqi dialect movie comments | Lack of standardization of dialect, lack of studies |
| [28] | (positive, negative) | 91.1% to 92.5% | CNN + GRU | IAEDS, ACMID, IAD (Facebook posts in Iraqi dialect) | Difficulty understanding the Iraqi dialect, lack of resources |

| | | | | | |
|---|---|---|---|---|---|
| [29] | 4 (happiness, anger, sadness, fear) | F1-score = 75.7% (CNN) | CNN ،SVM ، NB ،MLP | SemEval Affect in Tweets, Classical Arabic | Lack of studies on subtle emotions |
| [30] | 6 (Ekman) | Accuracy = 84.6% (PPM) | PPM ،NB ، SMO ،J48 | Facebook posts in the Iraqi dialect (1,353 posts) | Not using modern techniques such as deep learning, small data size |
| [31] | 10 Emotion | Accuracy ≈ 87% F1 = 0.84 | MARBERT ، Self-Learning | ArECTD (~11k Manual + Auto-Extended) | Models not trained on dialects |

As clearly indicated from table 1.1 emotional classification still a challenge motivated to investigate in this area.

Studies show that most Arabic emotion classification models still focus on classical Arabic or binary classification. However, there is a clear lack of studies that have addressed the Iraqi dialect and adopted accurate classification of the five or six emotions according to the Ekman model. Furthermore, modern models (such as MARBERT), despite their superiority, do not achieve sufficient accuracy for dialects due to the scarcity of data dedicated to them.

**Research Gap:**

There are no current solutions that combine the complexity of the Arabic language, understanding the Iraqi dialect, and classifying emotions into five categories while achieving accuracy comparable to English models. Most studies either focus on Modern Standard Arabic or use general emotional classifications. This highlights the need to develop a specialized model that accurately classifies sentiment in Iraqi dialect texts, based on modern deep learning techniques and specialized language models.

# CHAPTER TWO

# THEORETICAL BACKGROUND.

## 2.1 Overview

In this part, we present the concepts and technologies that are essential for the emotion classification research in this study, focused on the dialect of the Iraqi people. The paper makes use of machine learning algorithms such as BERT and LSTM models that have been down streamed to predict emotions based on text input. For these strategies, where applicable, we cover the topics, such as AI, ML, NLP, transfer learning, deep learning.

## 2.2 Social Media Platforms

Communication, education and sharing information contemporary and social media platforms have become a very important part of all of them. Being highly favored by users, platforms such as Facebook and Twitter also provide distinct measurements of user activity and preferences. Facebook and YouTube dominate the social media landscape, according to Pew Research Center. Looking specifically at Facebook, it appears its usage maintains a steady stream across age groups, whereas some social networks (like Instagram, Snapchat), and most recently TikTok, are generally preferred by younger users [32].

Arab countries are witnessing a significant expansion in the use of social media. Facebook is the most popular platform, with usage rates ranging between 40% and 70%. Twitter (X) records lower rates, often below 35%, and is used by specific groups such as activists and influencers [33]. For example, the number of Facebook users in Iraq reached approximately 30.93 million (66.3%), compared to 2.65 million for Twitter (5.7%) [34]. In Saudi Arabia, the number of Twitter users reached 12.15 million (34.4%), and Facebook accounts for 44.7% of the population [35]. In Egypt, the number of Facebook users reached 50.75 million (44.6%), compared to 3.45 million for Twitter (3.1%) [36].

These figures highlight the differences in digital usage patterns and emphasize the need to focus sentiment analysis studies on the most popular platforms, especially in local dialects such as Iraqi.

In the realm of education, Facebook is now often seen as having the power to strengthen learning outcomes. According to research, its interactional characteristics like groups, captions and comments can engage learners in collaborative learning. And specifically, Facebook contributes to the promotion of writing skills by enabling students to produce and publish their own posts and content, obtain more collegial reviews, and create discussion, not only does this activities will enhance her vocabulary and grammar but they also promote positive thinking and communications [37]. The accessibility of the site and also its informal setting gives the members a relaxing surrounding for them to practice the language, which makes it a beneficial corollary for the students learning English as a Foreign Language (EFL)[37];

The incorporation of social media platforms into education approaches shows their wider influence on information sharing and social connections. With the expansion of social media, the impact of it on education, communication, and trends in society, still remains one of the prominent fields of study.

Moreover, understanding emotions from text has become more relevant, especially in social media, such as Facebook and Twitter. Users on social platforms often share their feelings and opinions in the form of posts, comments and tweets, and sentiment analysis can be used to help us better understand trends in public opinion, mental health, and user engagement. Identifying emotion in text is crucial for different applications including, disaster management, marketing tactics[38], and evaluating social wellness.

On Facebook, emotion classification assists in filtering out and monitoring harmful content, predicting customer feedback, and enhancing personalized user experiences. Intelligent NLP models can categorize sentiments as joy versus anger, happy versus sad, surprised versus not surprised, etc. [39] which help business and organizations respond to emotions of their users appropriately. If Facebook is a personal album of perceptions about users, then Twitter is a living newspaper of sentiment, where the public can react instantly to happenings, products, or policies around the world. They observed that Twitter emotion classification aids social studies, policymaking, and may even help identify mental health problems based on mood changes on user posts[40].

Emotion classification is on the verge of transformation; with the advancement of AI and deep learning methodologies, it has shown remarkable precision in detecting emotions in textual data available on social media. Such advancements are aiding in bridging digital communication gaps and creating an environment in the online space that is empathetic and responsive[41].

## 2.3 Machine Learning

Machine Learning (ML) is a core subfield of Artificial Intelligence (AI) that empowers systems to learn from data and make predictions or decisions without being explicitly programmed. ML focuses on developing algorithms that identify patterns and insights from data, enabling models to improve their performance over time. These algorithms form the foundation of numerous AI applications across various domains, including healthcare, finance, autonomous vehicles, and sentiment analysis.[42], [43] ML techniques are broadly categorized into four main types:

- Supervised Learning: Learning from labeled data to make predictions.
- Unsupervised Learning: Identifying patterns and relationships in unlabeled data.
- Semi-Supervised Learning: A hybrid approach utilizing both labeled and unlabeled data.
- Reinforcement Learning: Learning through rewards and penalties based on actions in an environment.

### 2.3.1 Supervised Learning

Supervised learning is a prominent branch of ML where models are trained using labeled datasets. Each data instance consists of input-output pairs, enabling the model to learn the mapping between the two. Once trained, the model generalizes this learning to predict outputs for unseen data.[42]

In this study, supervised learning is pivotal for teaching the model to process text and detect emotions. By using annotated data with specific emotions (e.g., anger, fear, happiness, sadness, surprise), the system learns to classify new, unseen texts into the appropriate emotional categories. This approach ensures robust and accurate emotion classification in texts, even in complex contexts like those found in Arabic and Iraqi dialects.

In summary, supervised learning encompasses a variety of algorithms, including SVMs, decision trees, and deep neural networks, each with unique strengths and applications. These algorithms play a crucial role in developing models that can effectively interpret and predict outcomes based on labeled data, facilitating advancements in fields such as sentiment analysis and emotion classification.

### 2.3.2 Unsupervised Learning

Is a branch of machine learning in which models are trained on unlabeled data, meaning they have no predefined outputs. This type of learning relies on uncovering hidden patterns and structures within the data, such as clustering and dimensionality reduction [44].

In the context of sentiment analysis and emotion classification, unsupervised learning is used to identify clusters or common features among texts without the need for labeled data. This makes it particularly useful in languages or dialects that lack labeled datasets, such as the Iraqi dialect. For example, algorithms such as K-means or DBSCAN can be used to cluster texts with similar emotional contexts [45]. While this approach does not achieve classification accuracy comparable to supervised learning, it paves the way for discovering new patterns and can be used in the preliminary stages of building datasets or to improve model performance in situations where sufficient data is not available.

### 2.3.3 Semi-Supervised Learning

Is a combination of supervised and unsupervised learning, relying on a small amount of labeled data and a large amount of unlabeled data. This type of learning is particularly effective when labeling data is expensive or difficult, as is the case in Arabic and its various dialects[46].

In sentiment analysis, researchers use this approach to improve model performance. The process begins by training the model on a small portion of labeled data, then the model is used to gradually classify unlabeled data, indirectly increasing the amount of labeled data [47].

Techniques such as self-training and co-training are used, as well as neural network-based algorithms that learn from both labeled and unlabeled data. This approach is particularly useful in environments with limited or imbalanced data, such as Arabic written content on social media.

### 2.3.4 Reinforcement Learning

Is based on the principle of interaction between an agent and its environment, where the agent learns to make decisions through a series of rewards or punishments based on its actions [48]. The system aims to maximize long-term rewards through an optimized behavior policy.

In the field of natural language processing, although the use of reinforcement learning is still relatively limited in sentiment analysis, it is increasingly being used to train intelligent conversational and text-generating models such as ChatGPT, where the quality of responses is enhanced through human feedback[49].

Theoretically, reinforcement learning could be combined with emotion classification in interactive applications (such as automated emotional responses or intelligent learning systems), where the system gradually learns how to respond based on the user's emotions, opening up new future prospects for emotional text analysis.

## 2.4 Natural Language Processing (NLP)

Natural Language Processing (NLP) has become one of the most basic parts of AI which helps to read human language. It includes techniques and methodologies used to process textual and spoken data and obtain valuable information and insights. So, the NLP comes in, with the aim of connecting human communication with machines by combining action from computational linguistics and machine learning [50] Human language, which is inherently ambiguous, context-sensitive, and varies raw syntactically, presents challenges in this area, especially for morphologically rich languages such as Arabic[43].

The development of NLP has been driven by advances in machine learning, and more recently deep learning models like Transformers. They [51] includes learning the internal relationships and syntactic structures between the tokens by training these models on input text (BERT and its variants). NLP has many applications, such as sentiment analysis, machine translation, text summarization, and emotion classification, all of which exploit linguistic modeling and contextual representations [52].

NLP is also essential in sentiment analysis and emotion classification, which helps to decode the emotional tone of texts and offers valuable information to the beneficiaries, customer reviews, and social patterns.[53]

## 2.4.1 Text Preprocessing

The preprocessing step is a set of treatments and modifications that we perform on data sets before starting the training process to maintain a general context in learning the model on the data. When dealing with texts in general and Arabic texts in particular, we need preprocessing with more accuracy and focus because the texts may contain the same words but with different meanings. Since the Arabic language and the Iraqi dialect have special characteristics such as vowels (ُ – ِ – َ... etc.) and a number of letters have different shapes such as (ا – إ – آ – ا or ى – ء – ئ)، this requires taking preprocessing seriously when working and unifying the letters in one way or another.[54], [55]

Natural Language Processing (NLP) model performance is highly dependent on preprocessing Arabic text and its dialects. Challenges of processing the Arabic language in general, and the Iraqi dialect in particular arise from its morphological richness, lack of standardization, and social media text that is very informal. It is important to note that Arabic is considered a highly inflected language, has complex morphology, and as such performs tokenization, stemming and lemmatization differently than Latin based languages. The difference between Modern Standard Arabic (MSA) and Iraqi dialect in terms of phonetics, spelling and grammatical constructions makes this extra-hard to preprocess. Moreover, informal text in social media involved with emojis, slang and code-switching, which need to be sick with caution in text normalization[27].

Preprocessing techniques might be familiar with could be normalization or diacritic removal that standardize Arabic letters (such as variations of Alef " ا ") and unnecessary diacritics that improve consistency in your text[30]. Such as, stop-word removal is applied to remove common function words that do not carry much meaning and stemming or lemmatization techniques, Khoja Stemmer, Farasa

Lemmatize are then used to reduce words to stems or lemmas to improve accuracy of the model. Tokenization is also very crucial and difficult in the agglutinative Arabic language, however, if performed satisfactorily it improves the bisecting of words and phrases. Furthermore, preprocessing for sentiment analysis or any other Arabic Natural Language Processing (NLP) applications is extremely important where emojis and slang words need to be mapped to their Arabic counterparts [56] These recent advancements in deep learning have led to further improvements in Arabic text preprocessing, especially with the introduction of transformer-based models such as AraBERT and MARBERT that capture the import of word embeddings trained on large scale Arabic corpora including dialects variants [57] These approaches have demonstrated substantial success in tasks like sentiment analysis and emotion recognition. Arabic Language preprocessing, especially in the Iraqi dialect, is still one of the most challenging yet vital part in short- and long-term applications. These techniques include normalization, stemming, and tokenization; or even custom changes tailored specifically for dialects that assimilate the text better. More tailored preprocessing methods for Arabic dialects leading to ensuring their compatibility in applications of deep learning need to be developed in the future[58].

## 2.5 Deep Learning

Deep learning has emerged as a pivotal technology in modern artificial intelligence, particularly within Natural Language Processing (NLP). It uses a hierarchy of many neurons model layers that perform nonlinear information processing, and techniques for supervised/unsupervised feature representation at deeper, more abstract levels [59]. In spite of its successes, it faces many challenges that are especially pronounced in the case of Arabic sentiment analysis.[60]

One of the fundamental limitations of deep learning is the quality and quantity of training data. Machine learning models generally need a lot of high-quality data to be accurate and generalizable [61]. That said, finding the right data size for continuous training isn't as simple as it seems; the size depends on the task's complexity and the input data quality. In the case of Arabic sentiment analysis, this problem is magnified due to the complexity of the language and its dialects. Moreover, noisy or mislabeled data can lead to poor performance, especially when the training data is systematically dissimilar than the real-world test cases [62].

Another major challenge is the interpretability of deep learning models. A models like BERT or LSTM takes inputs, processes them, and spits out outputs without telling us what specific features or interactions drove the decision. This nature of "black box" methods stands in opposition to simple linear models from which we derive feature importance by calculating weights and allowing us to interpret results more easily[63] .

However, deep learning still provides a strong framework for detecting emotion in Arabic text and dialects. Fine-tuning of pre-trained models such as BERT and combining them with sequential architectures such as LSTM could help capture the contextual and temporal nuances of language [51]. Although interpretability and data quality issues remain, there are continuous improvements in model performance and generalizability.

### 2.5.1 Transfer Learning

Transformance is based on a research paper published in 2017 entitled attention is all you need [64]. It seems that human learners naturally transfer knowledge from one job to another. In other words, when faced with new tasks, we identify and use pertinent information from prior learning experiences. We can master a new task more readily if it is connected to our prior knowledge. On the other hand, common machine learning algorithms typically deal with discrete jobs. By creating strategies to apply knowledge gained from one or more source activities to enhance learning in a related target task, transfer learning aims to change this. Knowledge transfer techniques are a step in the right direction towards making machine learning as effective as human learning.

There are three typical ways that transfer could enhance learning. First, it compares the first performance of an ignorant agent to the initial performance that can be achieved in the target task using only the transferred knowledge, prior to any additional learning. The second is how long it takes to learn the target task completely using the knowledge that has been transferred, as opposed to how long it takes to learn it from scratch. Thirdly, it is the difference between the final performance level without transfer and the final performance level that can be achieved in the target task. Negative transfer has taken place if a transfer method really reduces performance. Producing positive transfer between tasks that are adequately related while preventing negative transfer between tasks that are less related is one of the main issues in designing transfer strategies. It is frequently required to map the features of one task onto those of the other in order to specify correspondences when an agent applies information from one task to another. This mapping is done by a human in most transfer learning work, although there are ways to do it automatically as well. [65] Educational psychology may have been the original source of the idea of transfer learning. The generalization theory of transfer, put out by psychologist C.H. Judd, states that experience generalization leads to learning to transfer[66]

#### 2.5.1.1 Transformer vs Vactor

**Word2Vec** and **GloVe**, represented words as static vectors, lacking the ability to account for the contextual nuances that influence meaning. This limitation often resulted in ambiguous interpretations, as the same word could have different meanings depending on its usage in different contexts.[67]

Although all models before (transformance) were able to represent words as vectors, these vectors did not contain context, and as we know that the use of words changes based on context.[67]

For example:

- (حَتَّىٰ إِذَا بَلَغَ مَغْرِبَ الشَّمْسِ وَجَدَهَا تَغْرُبُ فِي عَيْنٍ حَمِئَةٍ وَوَجَدَ عِندَهَا قَوْمًا ۗ قُلْنَا يَا ذَا الْقَرْنَيْنِ إِمَّا أَن تُعَذِّبَ وَإِمَّا أَن تَتَّخِذَ فِيهِمْ حُسْنًا ﴿٨٦ الكهف﴾).
- (قَالَ أَلْقُوا ۖ فَلَمَّا أَلْقَوْا سَحَرُوا أَعْيُنَ النَّاسِ وَاسْتَرْهَبُوهُمْ وَجَاءُوا بِسِحْرٍ عَظِيمٍ) ﴿١١٦ الأعراف﴾
- (ثُمَّ لَتَرَوُنَّهَا عَيْنَ الْيَقِينِ) ﴿٧ التكاثر﴾.

In the above-mentioned Quranic verses, the word "عين" was mentioned with several meanings, each verse has a meaning, which reinforces that there is a meaning to the sentence that differs according to its context.

If we assume that vectors are used, they will have the same vector because they have the same shape (the letters " ن – ي – ع") ,And all this was before the emergence of attention mechanisms.

Transfer learning implies the use of experience that was received while solving one problem for a similar task. For example, models such as BERT that have access to and are trained on massive amounts of text data can be adapted to like those adapting to smaller datasets like the Iraqi Arabic Emotion Corpus. This approach improves model effectiveness in performing tasks which do not have large labelled datasets so that it is not necessary to train the models from beginning.

**Transformer**: It is an encoder and decoder model that uses attention mechanism. It can take advantage of summation and also process large amount of data at the same time due to its structure.[64]

The attention mechanism helps to improve the performance of machine translation applications.

Transformer models are built with attention mechanisms at the core, Transformer consists of encoder and decoder, where encoder encodes the input sequence and passes it to decoder, decoder decodes the representation for the relevant task, Encoder component is a set of encoders with the same number, The paper presented for transformers stacks six encoders on top of each other [64] Sach as Figure 2.1.

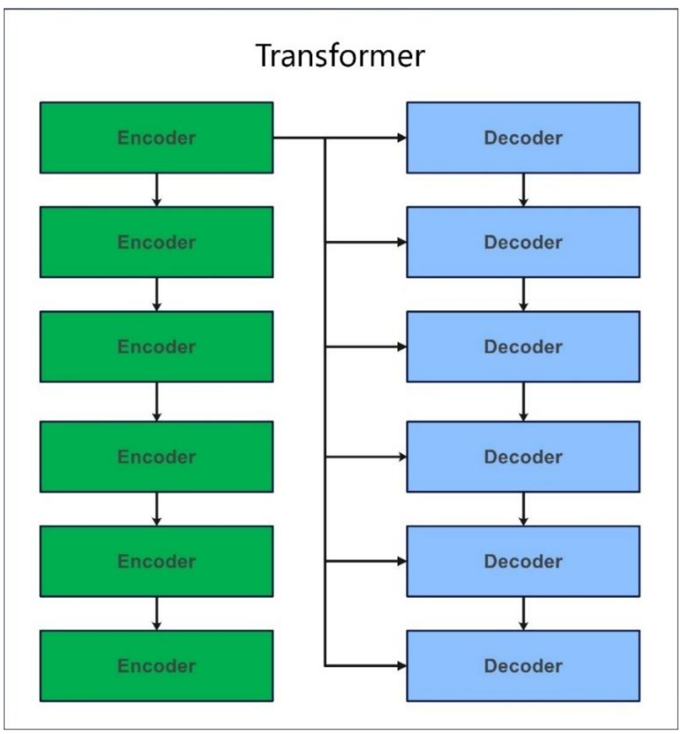


*Figure 2.1: Transformer[64]*

Six is not a magic number, it's just a hyperparameter, all encoders are identical in structure, but with different weights.

Each encoder can be divided into two sub-layers[64] as Figure 2.2:

- The first layer is called the self-attention layer.
  The encoder input first flows through the self-attention layer, which helps encode the relevant parts of words or look at them while encoding a central word in the input sentence.
- The second layer is called the feed-forward layer.
  The output of the self-attention layer is fed to the feed-forward neural network.
  The same feed-forward neural network is applied independently to each location.

The decoder contains both the self-attention and feed-forward layers, but in between is the encoder-decoder, the attention layer that helps the decoder focus on the relevant parts of the input sentence.

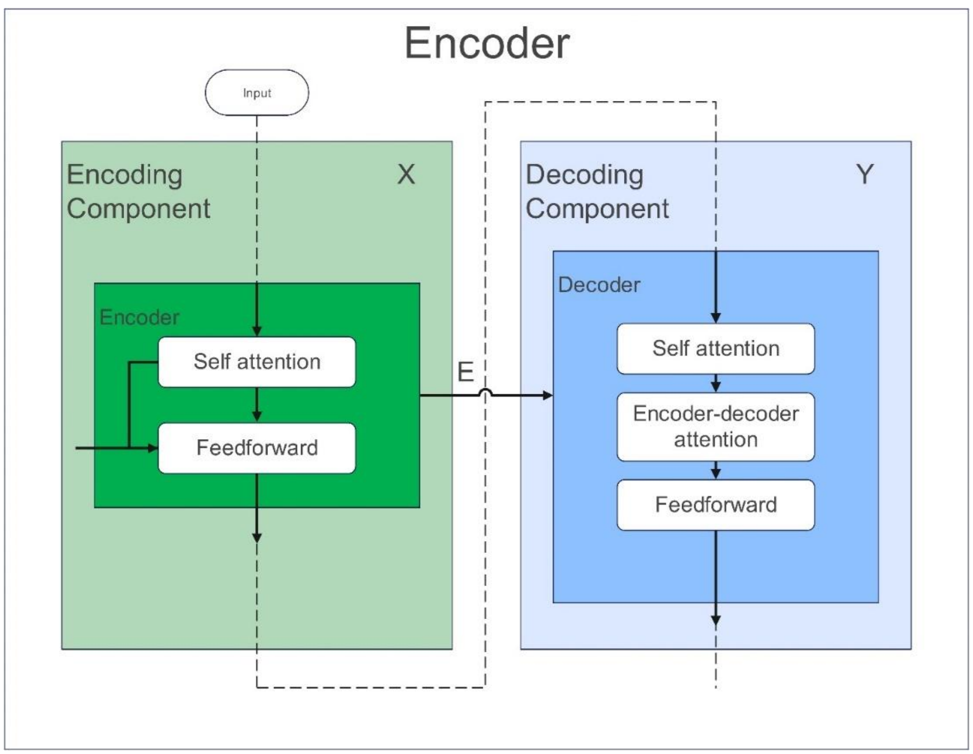


*Figure 2.2: Encoder[64]*

## 2.5.2 Bidirectional Encoder Representations from Transformers (BERT)

BERT, introduced by Devlin et al., is a cutting-edge Natural Language Processing model based on the Transformer architecture [51]. BERT is a revolutionary Natural Language Processing model that utilizes the Transformer architecture with a bidirectional training approach. Introduced initially by Devlin et al., its bidirectional nature enables BERT to consider the full context of a word by analyzing both preceding and succeeding tokens in a sentence. This is particularly critical for handling languages like Arabic, where context significantly influences word meaning. BERT’s training consists of two stages: pre-training and fine-tuning. During pre-training, it performs Masked Language Modeling (MLM) to predict masked words and Next Sentence Prediction (NSP) to model sentence relationships. Fine-tuning then adapts the model to specific tasks, such as emotion classification or text classification, by adding task-specific layers[52], [53].

### 2.5.2.1. Structure of BERT:

BERT's architecture is built upon the Transformer encoder introduced by Vaswani et al. [64]. The main components of BERT are:

1. Input Embeddings:

   BERT processes input as a combination of token embeddings, positional embeddings, and segment embeddings [64]:

$$E = T + P + S \ldots\ldots\ldots\ldots \text{(2.1)}$$

where T represents token embeddings, P positional embeddings, and S segment embeddings.

2. Transformer Encoders:

   BERT employs a stack of Transformer encoders, each consisting of:

- A multi-head self-attention mechanism to model relationships between all input tokens.
- A feed-forward neural network (FFN) applied independently to each token.
- Layer normalization and residual connections to stabilize training.

### 2.5.2.2. Equations in BERT

1. Self-Attention Mechanism:

   The self-attention mechanism computes dependencies between tokens[64]:

   $$\text{Attention}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \ldots\ldots\ldots\ldots \textit{(2.2)}$$

   where Q (query), K (key), and V (value) matrices are derived from the input embeddings, and d, k is the dimensionality of the key vector.

2. Multi-Head Attention:

   BERT uses multi-head attention to learn information from multiple representation subspaces[64]:

   $MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^o$ ………… *(2.3)*

   where each attention head is computed as:

   $$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V \ldots\ldots\ldots\ldots \textit{(2.4)}$$

3. Feed-Forward Network (FFN):

   Each encoder contains a position-wise FFN[64]:

   $$FFN(x) = ReLU(xW_1 + b_1)W_2 + b_2 \ldots\ldots\ldots\ldots \textit{(2.5)}$$

   where $W_1, W_2$ are weight matrices, and $b_1, b_2$ are bias terms.

4. Masked Language Modeling (MLM):

   BERT is pre-trained using MLM, where random tokens in the input are replaced with '[MASK]'. The model predicts the masked tokens[64]:

$$\mathcal{L}_{MLM} = -\sum_{t \in masked} \log P(x_t | x_{\backslash t}; \theta) \text{………… } (2.6)$$

where $x_t$is the masked token, and $x_{\backslash t}$represents the rest of the tokens.

5. Next Sentence Prediction (NSP):

   BERT is also pre-trained with NSP, where it predicts whether two sentences are consecutive[64]:

$$\mathcal{L}_{NSP} = -\sum_{(A,B)} [y \ \log P(A,B) + (1-y) \log(1 - P(A,B))] \text{ ………… } (2.7)$$

   where A, B are sentences and y is a binary label (1 for consecutive, 0 otherwise).

#### 2.5.2.3. Fine-Tuning BERT

For specific downstream tasks, BERT is fine-tuned by adding task-specific output layers. For multi-class classification, the final equation is[64]:

$$\hat{y} = \text{softmax}(W_h \cdot h + b \text{ ………… } (2.8)$$

where h is the pooled output from the '[CLS]' token, $W_h$ is the weight matrix, and b is the bias.

**AraBERT v0.2** is a BERT-based model specially trained on various Arabic Natural Language Processing (NLP) tasks, devised by Mind Lab 3 at the American University of Beirut. The Arabic language is morphologically rich and syntactically complex and poses its own challenges for NLP applications including Sentiment Analysis (SA), Named Entity Recognition (NER), and Question Answering (QA).[68] AraBERT uses large corpus of Modern Standard Arabic (MSA) and Dialectal Arabic (DA) to pre-train with a core architecture that bears resemblance to BERT, building upon Transformer-based architectures. It was trained with Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) through pre-training and fine-tuning processes for various downstream tasks. AraBERT outperformed even Google's multilingual BERT (mBERT), and was a substantial improvement over other Arabic-specific models, reaching state-of-the-art for a number of Arabic NLP tasks. The model adopts Farasa segmentation for tokenization (to appropriately separate words) and introduces [69] as a new subword algorithm that more accurately represents Arabic morphology. On GitHub and Hugging Face, the AraBERT models are publicly available, and they play a pivotal role in the Arab NLP research community[68].

### 2.5.3 LSTM (Long Short-Term Memory)

Long Short-Term Memory (LSTM) networks have significantly advanced Arabic Natural Language Processing (NLP) by effectively modeling the language's complex morphological and syntactic features. Arabic's rich morphology, including root-based word formations and extensive inflections, poses challenges for traditional NLP models. LSTMs, capable of capturing long-term dependencies in sequential data, have been instrumental in overcoming these challenges across various NLP tasks[70], [71].

Recent studies have demonstrated the efficacy of LSTM networks in Arabic sentiment analysis, where they capture contextual and semantic nuances unique to the language, leading to improved classification accuracy [72], [73] . In machine translation, LSTM-based models have enhanced Arabic-English translation systems by effectively handling syntactic complexities and word reordering inherent in Arabic [73]. Additionally, LSTMs have been applied to Arabic Named Entity Recognition (NER), utilizing character-level embeddings to manage morphological variations and achieve higher recognition rates [73], [74]. These advancements highlight the significant impact of LSTM networks in overcoming the unique challenges of Arabic NLP during this period.

#### 2.5.3.1 LSTM Architecture

Long Short-Term Memory (LSTM) An RNN variant that processes data serially and has a superior ability to connect with context is the LSTM. The "gate" is an intrinsic mechanism of LSTM. It has input, forgetting, and output gates that let it to store and remove data. The structure of LSTM is depicted in Figure 2.3 and is made up of many neuronal units.[75]

The preceding neuron's state is represented by $C_{t-1}$, its output by $h_{t-1}$, the current input by $X_t$, and the Sigmoid function by σ, which collectively determine the current neuron's output $h_t$.[75]

The Sigmoid function is used to modify the gate so that its output value falls between 0 and 1. When it is zero, the output gate disregards the recently computed states, and when it is zero, the forgetting gate disregards all prior memories.[75]

The core architecture of an LSTM unit consists of the following components [71]:

1. Cell State:

   The cell state ($C_t$) acts as the memory of the network, storing long-term dependencies. Information is added or removed from the cell state via carefully controlled gates.

2. Gates:

   LSTM uses three types of gates to regulate the flow of information:

- Forget Gate: Determines what portion of the cell state from the previous time step ($C_{t-1}$) should be retained.
- Input Gate: Controls how much new information is added to the cell state.
- Output Gate: Decides how much of the cell state contributes to the hidden state ($h_t$) at the current time step.

#### 2.5.3.2 LSTM Equations

To explain equations used in LSTM [76]

1. Forget Gate:
   The forget gate decides which parts of the previous cell state ($C_{t-1}$) to retain[71]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \text{ ………… } (2.9)$$

$f_t$: Forget gate activation.

$W_f$: Weight matrix for forget gate.

$b_f$: Bias term for forget gate.

$[h_{t-1}, x_t]$: Concatenated vector of the previous hidden state and current input.

$\sigma$: Sigmoid activation function [71].

2. Input Gate:
   The input gate determines how much new information is stored in the cell state[71]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i \text{ ………… } (2.10)$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C \text{ ………… } (2.11)$$

$i_t$: Input gate activation.

$\tilde{C}_t$: Candidate cell state (new information).

$W_i, W_C$: Weight matrices for input gate and candidate state.

$b_i, b_C$: Bias terms for input gate and candidate state.

$tanh$: Hyperbolic tangent activation function.

3. Cell State Update:
   The cell state is updated by combining the forget gate and input gate outputs[71]:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \text{ ………… } (2.12)$$

$\odot$: Element-wise multiplication.

4. Output Gate:
   The output gate determines the hidden state ($h_t$) for the current time step[71]:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o \text{ ………… } (2.13)$$

$$h_t = o_t \odot \tanh(C_t) \quad \ldots\ldots\ldots\ldots (2.14)$$

$o_t$: Output gate activation.

$W_o$: Weight matrix for output gate.

$b_o$: Bias term for output gate.

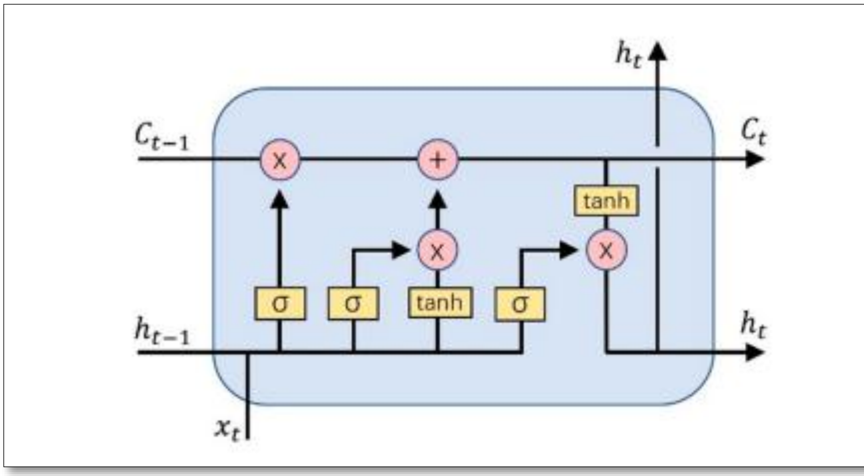


*Figure 2.3: Architecture of LSTM Gets.[77]*

## 2.6 Performance Evaluation

Refer have also considered two widely used evaluation metrics:

The **confusion matrix** is an essential tool for evaluating the performance of machine learning classification models. It enables researchers to analyze the model's effectiveness in correctly or incorrectly classifying samples across four main categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). This matrix helps calculate metrics such as precision, recall, F1-score, and overall accuracy, providing a balanced and comprehensive view of the model's quality [78].

**precision** and **recall**, as well focusing on the F1-score for both text-level emotion classification at word level. F1-score balances precision and recall, it is the harmonic mean of both[60]. Recall says the false negative rate is low, and precision means a lower false positive rate. F1-score is the best metric to use in presence of class imbalance. Those metrics are calculated using the formulas below: [70], [74]

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \quad [70],[74] \ldots (2.15)$$

$$\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \qquad [70],[74] \ldots (2.16)$$

F1-score is harmonic mean of Precision and Recall can be calculated by using the following formula:

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad [70],[74] \ldots (2.17)$$

**Accuracy** is one of the most common metrics used to evaluate classification models. It represents the percentage of correctly classified samples (either positive or negative) to the total number of samples. It is particularly useful in cases where class balance is high. The importance of using accuracy within a cross-validation framework has been highlighted to ensure more reliable results in machine learning and text classification tasks [78].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad [78] \ldots (2.18)$$

## 2.7 Summary

The chapter presented an introduction to detect emotions using advanced machine learning techniques like BERT (Bidirectional Encoder Representations from Transformers) & LSTM (Long Short-Term Memory) algorithms, where the whole chapter mainly focus on the Iraqi dialect. emotion classification is an important task in social media applications since it is the primary usage by people (Facebook, Twitter etc. found to be used for expressing their thoughts). Beyond mere communication, these technologies will enrichment education, promote collaborative learning and develop language skills, making them more relevant than ever for English as a Foreign Language (EFL) students. Artificial Intelligence (AI) and Machine Learning (ML) play a crucial role in determining sentiments and emotions thereby helping systems understand human expressions that are not easily translatable, making them applicable in marketing, customer support, and even in mental health assessment. Natural Language Processing (NLP) is a crucial avenue that works to narrow the gap between humans communicating with machines, especially for morphologically rich languages, such as Arabic, which pose a number of unique challenges owing to their non-linear and highly ambiguous nature, and the considerable lack of standardization in writing. Arabic text preprocessing methods, including normalization, stemming and tokenization, are crucial for dealing with Arabic text data, especially in the case of informal social media data. Using transfer learning techniques, as seen with models like BERT, we can leverage knowledge from large datasets and fine-tune the models on small task-specific datasets, making the process much more efficient in emotion classification. Pre-trained Embeddings Pre-trained embeddings allow the transfer of prior knowledge, ensuring that the deep learning models capture long-term dependencies and contextual nuances in the Arabic text during preprocessing, enhancing their suitability for tasks in sentiment analysis and machine translation.

Note: Close attention is also given to Performance evaluation metrics such as precision, recall and F1 to measure the performance of the model. In sum, these developments in AI, ML, and NLP are revolutionizing emotion classification and allowing for more accurate and empathetic communication as code and data in a digital context, especially for non-standard complex languages and dialects such as Iraqi Arabic.

# CHAPTER THREE

# PROPOSED METHOD

## 3.1 Overview

This chapter describe the architecture and implementation of the proposed system that design to detect emotions in Arabic text, the Iraqi dialect. The system makes use of recent developments in deep learning and Natural Language Processing (NLP), Long Short-Term Memory (LSTM) networks and the Bidirectional Encoder Representations from Transformers (BERT) model. the chapter is divided into multiple sections. The first section provides a thorough description of the dataset, and the subsequent sections discuss the preprocessing methods used to enhance and clean the data. Next, we go over how to extract features using BERT and create word embeddings. Lastly, we describe how the LSTM model is trained, using these embeddings as input to identify sequential patterns and forecast the emotion classes. Our suggested system's outcomes show how well it can identify emotions from Arabic text, with an emphasis on taking into account the subtleties of the Iraqi dialect

## 3.2 Proposed System

The proposed system is a sentiment analysis model designed specifically for Arabic text and the Iraqi dialect, combining a pre-trained BERT model with an LSTM network to classify text into five emotions: anger, fear, happiness, sadness, and surprise. The system incorporates advanced preprocessing techniques, including Arabic text normalization, emoji translation, and synonym replacement, to enhance text quality. It uses BERT embeddings to capture contextual information and a bidirectional LSTM layer to model sequential dependencies, followed by a classification layer for emotion prediction as show in figure 3.1. Trained on the IAEC dataset, showcasing its effectiveness in handling the complexity of Arabic language and dialectal variations for emotion classification.

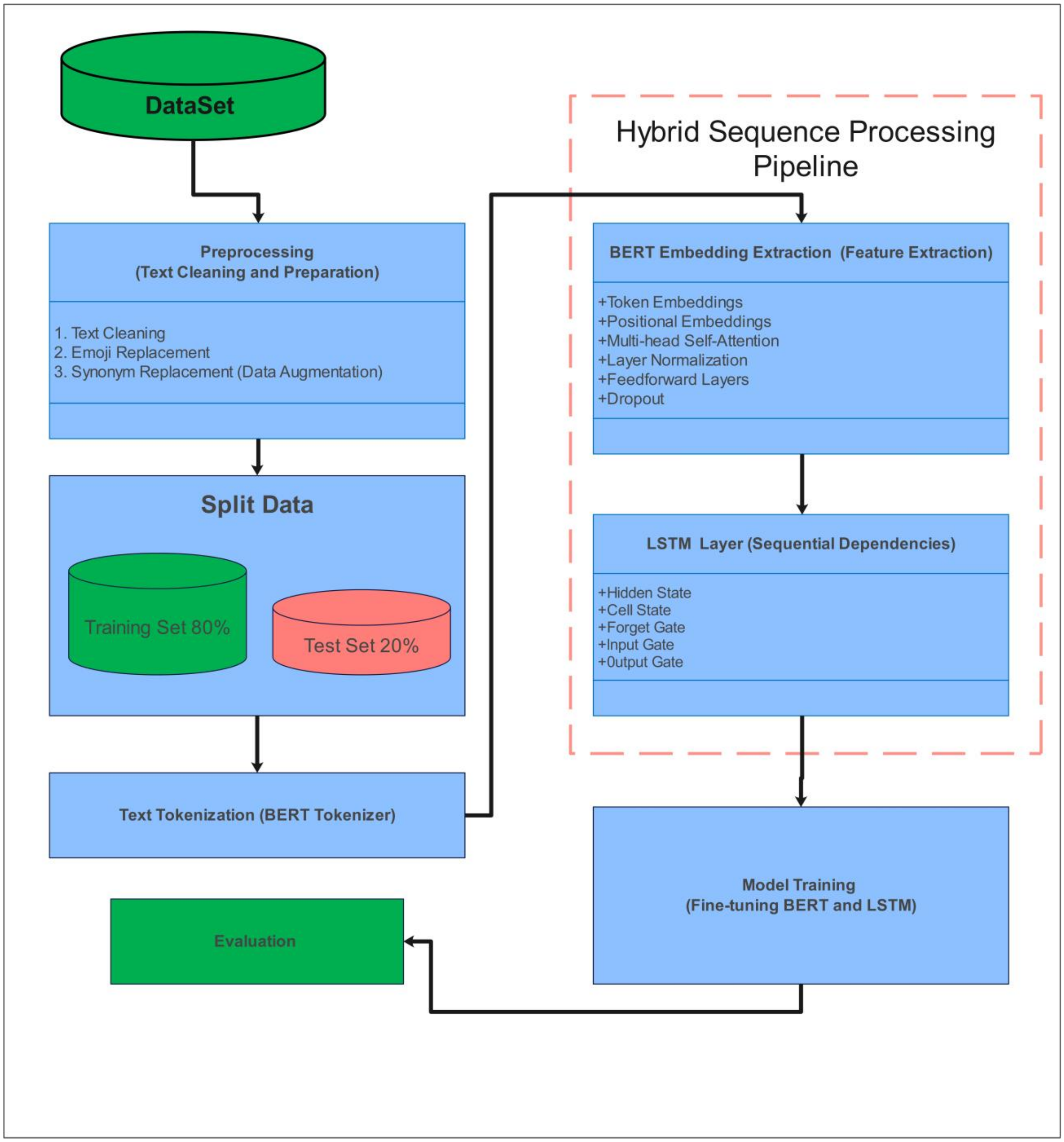


*Figure 3.1 the proposed system*

### 3.2.1 Dataset

The dataset we are utilizing for our study is known as the Iraqi Arabic Emotion Corpus (IAEC). The aim of this collection is to identify emotions in Arabic texts, with a focus on the Iraqi dialect. The dataset has been discussed in chapter four in details.

### 3.2.2 Data Preprocessing

Preprocessing is one of the most important pillars of our work, as we clean the data and remove words, symbols and letters while preserving the true meaning and context of the sentence written in Arabic and the Iraqi dialect. Preprocessing was carried out in multiple stages to clean and enhance the datasets.

1) **Text Cleaning**: the methodical and comprehensive process of eliminating extraneous characters, symbols, and punctuation, as well as words that are borrowed from foreign languages or are not local phrases. To further guarantee that only textual content is preserved and undisturbed by outside influences, this painstaking procedure also excludes numerical values.

2) **Emoji Translation**: Conversion of emojis to their corresponding Arabic terms using a specialized Python-based translator application that has been specifically developed to accurately interpret and translate emojis into their respective Arabic equivalents.

   **Emojis** are widely used in Arabic content and have been shown to have a strong impact on emotion analysis. Emotion scores were calculated for 222 common emojis in Arabic tweets (58,000 tweets), and it was confirmed that replacing emojis with textual phrases enhances the performance of emotion classification models [79]. for example:

متت قهر ☹ ☹
متت قهر حزين حزين

3) **Data Augmentation**: Expansion of the dataset by generating new sentences through synonym replacement. we resorted to expanding the data set for the purpose of training the new model on the largest amount of data. We chose the method of using synonyms to generate new sentences based on the **SynonymsDataset** is dataset comprises 500 synsets derived from the Arabic Wordnet. Every synset is supplemented with a compilation of potential synonyms. Therefore, we processed the thesaurus data set and created an Arabic language dictionary to benefit from it. A new sentence is generated from the same previous sentence, which is present in the original data set, by replacing the original word with the word that has a synonym in the thesaurus data set, and thus we obtain a new sentence[80].

   **Data augmentation** techniques are an effective strategy for addressing the shortage of Arabic datasets, especially due to the diversity of dialects and the complexity of morphology. There are several methods for expansion in Arabic text, such as paraphrasing, noising, and unbalanced techniques such as oversampling and undersampling[81]. for example:

يعني والله العظيم كلها خوف والمدينة مهجورة
يعني والله العظيم جميعها رعب والاماكن متروكة

   To display the preprocessing steps work and sequence, showing the Algorithm 3.1, it’s described the Preprocessing text to IAEC dataset.

**Algorithm (3.1): Preprocessing the IAEC Dataset**

**Input: Line Text IAEC Dataset.**

**Output: Processed Line IAEC Dataset.**

**Begin**

**Step 1: Text Cleaning**

- Remove all symbols, foreign letters, and punctuation marks from the text.
- Remove all numbers to retain only textual content.

**Step 2: Emoji Translation**

- Identify emojis in the dataset.
- Convert the emoji to its corresponding meaning in English using Python library.
- Translate the English meaning of the emoji into Arabic using the Python translator library.
- Replace the emoji in the text with its Arabic translation to maintain sentence integrity.

**Step 3: Data Augmentation**

- Use the **SynonymsDataset** to identify synonyms for words in the dataset.
- Replace original words with their synonyms from the dataset to generate new sentences.
- Validate and include the newly generated sentences alongside the original dataset to expand the data.

**End**

## 3.3. Feature Extraction Using BERT

Our emotion recognition algorithm relies heavily on feature extraction and representation, especially considering the dialectal differences and morphological complexity of Arabic. In order to overcome these obstacles, we utilized a variety of cutting-edge methods to make sure that the characteristics our model uses appropriately capture the text's emotional content.

We used the BERT model to generate contextual word embeddings, which provide a deep understanding of the meaning of words within their specific contexts. These embeddings work especially well at preserving the subtleties of the Iraqi dialect, where context can change the meaning of a word. By fine-tuning BERT on our specific dataset, we ensured that the generated embeddings are highly relevant to the task of emotion classification in Arabic.

Sentence-BERT was employed in conjunction to word-level embeddings to generate sentence-level representations. Sentence-BERT combines all of the word embeddings into a single vector that expresses the sentence's general mood. This approach is particularly useful for tasks where the emotion is conveyed at the sentence level rather than by individual words.

We also included conventional NLP methods like TF-IDF and N-gram models to supplement the deep contextual characteristics offered by BERT. With the help of these techniques, we were able to detect frequency-based patterns in the text, which are especially helpful for recognizing typical expressions and phrases that are exclusive to the Iraqi dialect. Our model was guaranteed to have a thorough comprehension of both the text's surface-level language patterns and its profound contextual meaning because to our hybrid method to feature extraction.

Furthermore, we explored the effectiveness of using semantic representation with BERT in emotion classification on Arabic text and Iraqi dialect, and the results were promising.

BERT used to extract features, and then we fed the features to an LSTM model to identify sequential patterns in the text data. When dealing with sequential data, such as text, where word order is essential to understanding meaning, LSTM is especially useful. With a learning rate of 3e-5, the model was trained over 10 epochs using the Adam optimizer and the categorical cross-entropy loss function.

The following is the LSTM model architecture[82] Sach as section (2.5.3):

- **Input Layer**: The LSTM receives its word embeddings from the BERT model as input. These embeddings, which are dense vectors, represent the text's semantic meaning in relation to its surroundings.
- **LSTM Layer**: To identify both short- and long-term dependencies in the text data, a multi-layer LSTM with 2 layers.
- **Dropout Layer**: To avoid overfitting, a dropout layer with a rate of 0.3 was added. During training, some of the units are randomly dropped.
- **Dense Layer**: We employed dense layers to convert the LSTM outputs into emotion classes after the LSTM layer.
- **Softmax Activation**: The probability distribution across the five emotion classes (Anger, Fear, Happiness, Sadness, and Surprise) was output in the last layer by applying the softmax activation function.

The LSTM model training revealed consistent gains in accuracy and loss. After 10 epochs, the model improved to a final validation accuracy of 84%, with a corresponding validation loss of 0.7068, from a starting point of 57.7% in epoch 1. Early stopping was used to track validation loss during training, and training was terminated if the model did not improve after a predetermined number of epochs. This ensured that the model generalized well on unobserved data and helped prevent overfitting.

### 3.3.1 Hyperparameters Used in Training

- Batch size: 16
- Optimizer: AdamW (lr=3e-5)
- Loss function: CrossEntropyLoss
- LSTM hidden size: 128
- LSTM layers: 2
- Dropout rate: 0.3
- Bidirectionality: Enabled (bidirectional LSTM)
- Epochs: 10
- Gradient Scaling: Applied using (*torch.cuda.amp.GradScaler*) for mixed-precision training

The BERT-LSTM hybrid model is obtained by following sequential and interconnected steps, and they have been confined in the algorithm (3.2):

**Algorithm (3.2): Hybrid BERT-LSTM Model for Arabic Text Emotion Classification**

**Input: Raw IAEC Dataset.**

**Output: Processed IAEC Dataset.**

---

**Begin**

**Step 1: Data preprocessing:**

- apply Preprocessing the IAEC Dataset algorithm (3.1) in section (3.2.2).

**Step 2: Tokenization and Input Preparation:**

- Use the BertTokenizer associated with the pretrained Arabic BERT model (*aubmindlab/bert-base-arabertv02*) to tokenize the preprocessed text.
- Generate *input_ids* and *attention_mask* tensors, ensuring truncation and padding for consistent input size.

**Step 3: Feature Extraction with BERT:**

- Pass the tokenized inputs through the pretrained BERT model to extract contextual embeddings for each token.
- Use the *last_hidden_state* output from BERT to capture rich, contextual features of the text.

**Step 4: Sequence Modelling with LSTM:**

- Feed the extracted BERT embeddings into a bidirectional LSTM layer to model sequential dependencies.
- Use stacked LSTM layers to capture deeper and more complex patterns in the text.

**Step 5: Dropout and Classification:**

- Apply a dropout layer to the output of the LSTM to prevent overfitting.
- Pass the final LSTM output (last timestep) through a fully connected (dense) layer with softmax activation to predict the emotion class.

---

**Step 6: Training the Model:**

- Train the hybrid BERT-LSTM model end-to-end using cross-entropy loss and the AdamW optimizer.
- Use a learning rate of 3e-5 for BERT layers and a slightly higher rate for LSTM and classification layers.
- Train for 10 epochs with batch size 16, monitoring training and validation loss and accuracy.

**Step 7: Evaluation and Fine-Tuning:**

- Evaluate the model using metrics like accuracy, precision, recall, F1-score, and confusion matrix on the validation set.
- If necessary, fine-tune the model by adjusting hyperparameters or training epochs.

**End**

---

## 3.4. Summary

The chapter looks at the proposed model and the way in which BERT and LSTM were linked together, in addition to a quick overview of the Iraqi dataset used in the work, and a reference to the basic steps in preprocessing and their importance, as we review the preprocessing algorithms and the algorithm for obtaining the hybrid model with its training.

# CHAPTER FOUR

# RESULTS AND DISCUSSIO

## 4.1 Overview

This chapter looks into the outcomes and analysis of the suggested hybrid model to recognize emotions in Arabic texts in the Iraqi dialect. The study uses several datasets to assess and confirm how well the model works. These datasets include the Iraqi Arabic Emotion Corpus (IAEC), ArPanEmo, and Arabic Emotions Twitter Dataset (AETD).

The Iraqi dialect's casual and tricky nature, with its unique slang and sayings, makes it hard to figure out emotions. This gets even tougher because there's not much stuff out there for this specific dialect it's need to think outside the box. The model we're talking about uses the BERT pre-trained model, which is tweaked for Arabic (Arabert). It also adds an LSTM layer to catch BERT vectors and dealing with strong two-way vectors is better than using LSTM alone which uses word transformation to one-way vector.

This chapter starts with displaying using the datasets with both the basic BERT model and the combined BERT-LSTM model. The obtained result was extensively discussed and how they compare to earlier studies. look at big challenges, like differences in dialects and not having enough resources, and suggest areas to explore in the future. To wrap up, we sum up what learned and how it helps improve emotion recognition in Arabic texts for the Iraqi dialect.

## 4.2 Computer Specifications

The hybrid model was developed using robust hardware:

- Processor: AMD Ryzen 7 3800X (8 cores, 16 threads, 3.81 GHz base, 4.16 GHz boost, 7nm, AM4 socket, 32 MB L3 cache).
- Motherboard: Gigabyte B450 AORUS ELITE with AMD Ryzen SOC chipset.
- Memory (RAM): 32 GB DDR4 (4x8 GB, 3200 MHz, CAS latency 16).
- Graphics Card: NVIDIA GeForce RTX 2070 SUPER (8 GB GDDR6, 256-bit, 375 MHz core clock, 12nm architecture).
- The GPU, leveraging thousands of cores, was used for optimized training and inference via Torch and CUDA modules. It enabled faster matrix computations, parallel processing for large datasets, and mixed-precision training using cuda.amp.GradScaler and torch.cuda.amp.autocast for efficient memory management and numerical stability.
- **Development Environment:** The development was conducted using **PyCharm 2024.1.3 (Community Edition)** on Windows 11, with OpenJDK 64-bit VM and 16 CPU cores. Key features included intelligent code editing, debugging, Git integration, and virtual environment compatibility. A machine learning plugin (Version 241.17890.15) enhanced support for NLP tasks, facilitating efficient development and testing of the BERT-LSTM hybrid model.

## 4.3 Dataset

Several datasets were applied as shown respectively

### 4.3.1 Iraqi Arabic Emotion Corpus (IAEC) Dataset

The dataset we are utilizing for our study is known as the Iraqi Arabic Emotion Corpus (IAEC). The aim of this collection is to identify emotions in Arabic texts, with a focus on the Iraqi dialect. The dataset has a number of features and characteristics [30]. 1,365 post that were carefully gathered from Facebook make up the dataset with 286,775 characters overall, there are 22,438 words in all. Users writing in the Iraqi dialect have rich emotional sentiments in these messages. Between December 2016 and August 2018, the posts were gathered[30]. Ekman identified six main emotions[25], which form the basis of the dataset: anger, disgust, fear, happiness, sadness, and surprise. Nevertheless, just five emotions aside from disgust are employed in our study, until the total number of posts has become 1180 post. Six classes one for each emotion are created from the dataset. One of the most important pre-processing steps mentioned in the algorithm () is the data expansion process, through which the number of posts for each category will change, and thus the results will be according to the following table 4.1.

*Table 4.1: classes post of IAEC dataset*

| **class** | **No. of posts before Prepressing** | **No. of posts after augmentation** |
|---|---|---|
| Anger | 309 | 447 |
| Disgust | 185 | Don't used |
| Fear | 148 | 215 |
| Happiness | 256 | 348 |
| Sadness | 238 | 329 |
| Surprise | 229 | 318 |

### 4.3.2 ArPanEmo dataset:

the ArPanEmo dataset were also used, which is a fine-grained dataset for emotion recognition in Arabic online contents during the COVID-19 Pandemic. Here are its features[83] that includes 11,128 online posts, which were obtained manually from three platforms: Twitter, YouTube comment sections, and online newspaper comments. The date range of the data collection was from March 2020 to March 2022, while the content covered the topics of COVID-19, touching on the medical aspect of COVID-19, as well as how it affects various aspects of life[83]. There are 11 emotion categories in the dataset, which are anger, anticipation, confusion, disgust, fear, happiness, neutral, optimism, pessimism, sadness, and surprise. Based on manual labeling, each post is annotated to one of these categories[83] .We use only five Emotion {anger, fear, happiness, sadness, surprise}, In line with the main objective of the work. This dataset includes a total of 223,008 tokens (words) captured in 11,128 posts. Of these, there are 8,751 posts which are concerned with issues related to healthcare and 2,377 posts related to the life of the pandemic.[83] Major hurdles cropped up with posts in Saudi dialect. This made normal emotion word lists useless. As a result, people had to sort through and group things again to check if they fit and matched specific feelings. Also, a Fleiss Kappa score of 0.71 showed that the people marking things agreed a lot. This made sure the labels were trustworthy.

### 4.3.3 Arabic Emotions Twitter Dataset (AETD)

The dataset used in this study is the Arabic Emotions Twitter Dataset (AETD dataset). Its goal is to enable emotion classification in Arabic social media text, focusing on the diversity of emotions expressed on Twitter. The dataset has several distinctive features and characteristics[84] , The AETD dataset consists of over 10,000 tweets manually annotated for emotions. These tweets were aggregated from multiple sources. A pre-existing dataset of 1167 polarity-labeled tweets re-annotated for emotions. Tweets collected using Twitter's API during the 2016 Olympics, specifically filtered for Egyptian tweets. A random subset of more than 500,000 tweets collected using the NileULex sentiment lexicon[84].

The dataset includes tweets distributed among eight emotion categories as follows table 4.2:

*Table 4.2: classes tweet of AETD dataset*

| **Class** | **No. of tweets** |
|---|---|
| Sadness | 1256 tweets |
| Anger | 1444 tweets |
| Joy | 1281 tweets |
| Surprise | 1045 tweets |
| Love | 1220 tweets |
| Sympathy | 1062 tweets |
| Fear | 1207 tweets |
| None | 1550 tweets |
| Total tweets are 10,065 | |

The dataset is carefully balanced to ensure each category is well-represented, making it suitable for machine learning tasks.[84] Many tweets contained multiple emotions, complicating the annotation process. Rare emotions, like disgust, were often confused with anger, necessitating category adjustments. Balancing underrepresented emotions required targeted queries and manual verification to avoid bias.[84].

### 4.3.4 Iraqi Arabic Dialect (IAD) dataset

This is a dataset of sentiment texts in Arabic and Iraqi dialect, collected from user comments on Facebook pages, and manually labeled for sentiment analysis tasks. This dataset is divided into 2000 comments, 1000 of which are positive and 1000 are negative, and it also contains approximately 250,000 unlabeled comments for training purposes.[85]

In this study, IAD was used this dataset for testing only in the Iraqi hybrid model trained on 5 types of emotions, where we considered {anger, fear, sadness} as negative emotions and {happiness and surprise} as positive emotions, as mentioned in the research paper.[28]

Note that there is a difference in considering surprise emotions, which may be biased if the surprise is negative or positive.

## 4.4 Result of Model

Firstly, the obtained results by applying traditional methods in classifying emotions with IAEC dataset are visualized in Table 4.3:

*Table 4.3: Model Performance Comparison*

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|
| **BERT + LSTM** | 84 | 82 | 81 | 81.5 |
| **SVM** | 81 | 82 | 78 | 80 |
| **Logistic Regression** | 79 | 81 | 76 | 77 |
| **Naive Bayes** | 78 | 80 | 74 | 76 |
| **Random Forest** | 76 | 77 | 75 | 75 |

Later, the trained different models and compared their results to see whether the model contributed to improving the basic model or not. Thus, the model "*Arabert*" was tested once alone and once after using the proposed hybrid model which is based on the "*Arabert*" model. The results were shown on two sets of Arabic sentiment texts dataset, but since there is no dataset other than IAEC dataset in the Iraqi dialect, we compared it to the ArPanEmo dataset in the Saudi dialect.
Later we will use the trained models to predict other datasets.

### 4.4.1 BERT Technique

In the stage of using the BERT model only, which deals with texts well, we test by training the "*Arabert*" pre-trained model to deal with the Arabic language on the data set and storing the new trained model to test it on other data sets.

#### 4.4.1.1 IAEC Dataset:

In the stage of using the BERT model alone, which handles texts well, we obtained an accuracy of 82% according to the results attached in the table 4.4 with the confusion matrix:

*Table 4.4: Result using BERT Model only with IAEC Dataset*

| | **precision** | **recall** | **f1-score** | **support** |
|---|---|---|---|---|
| anger | 0.75 | 0.89 | 0.81 | 89 |
| fear | 0.94 | 0.72 | 0.82 | 43 |
| happiness | 0.82 | 0.89 | 0.85 | 70 |
| sadness | 0.9 | 0.83 | 0.87 | 66 |
| surprise | 0.79 | 0.7 | 0.74 | 64 |
| | | | | |
| accuracy | | | 0.82 | 332 |

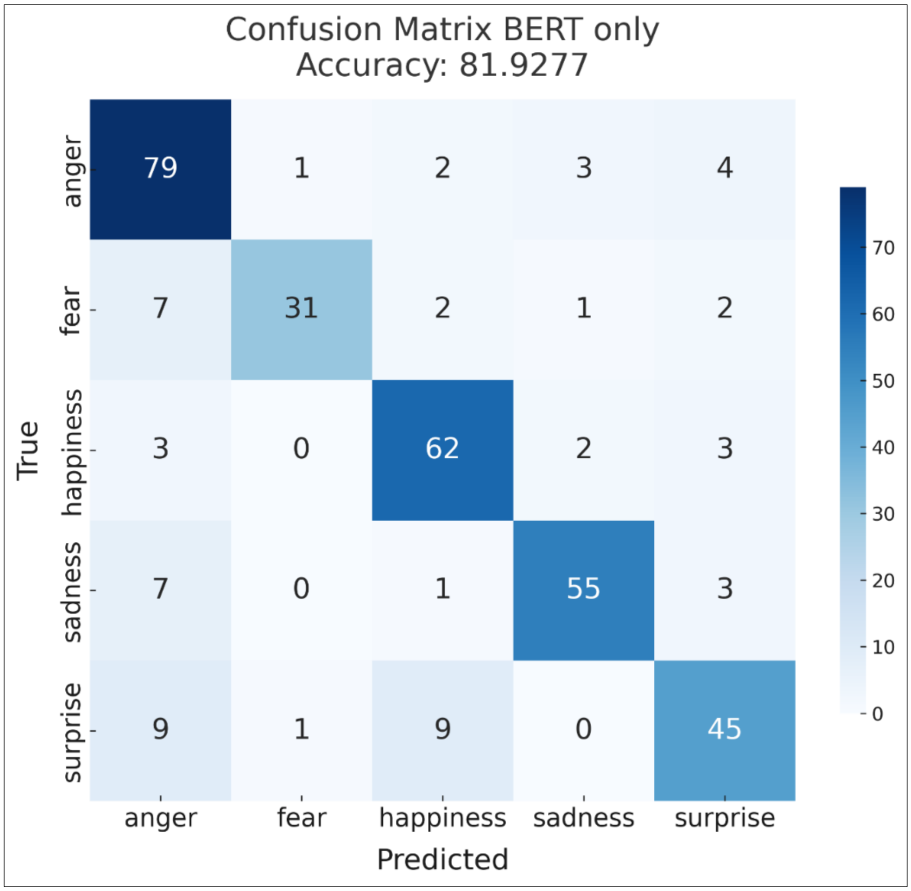


*Figure 4.1: Confusion matrix BERT Model only with IAEC Dataset*

The "*Arabert*" pre-trained BERT model when implemented by training and testing on the Iraqi dataset (IAEC Dataset) and the model getting an accuracy of 82% is a good step and this is due to the model being trained to deal with the Arabic language in general and since the dataset is primarily Arabic language data this gave the above result.

Note that the most errors were in predicting the text from the surprised category where it was wrongly classified 9 times as angry and 9 times as happy and this is due to the fact that basically the surprise can be a happy surprise or an angry surprise.

#### 4.4.1.2 ArPanEmo Dataset:

Using the other dataset ArPanEmo after training the model on it and using only the selected emotions to compare it with the results of the IAEC dataset, we obtained an accuracy of 87.6% according to the table 4.5.

*Table 4.5: BERT Model only with ArPanEmo Dataset*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.85 | 0.89 | 0.87 | 53 |
| **fear** | 0.9 | 0.81 | 0.85 | 58 |
| **happiness** | 0.95 | 0.92 | 0.93 | 59 |
| **sadness** | 0.79 | 0.87 | 0.83 | 47 |
| **surprise** | 0.88 | 0.92 | 0.9 | 25 |
| | | | | |
| **accuracy** | | | 0.88 | 242 |

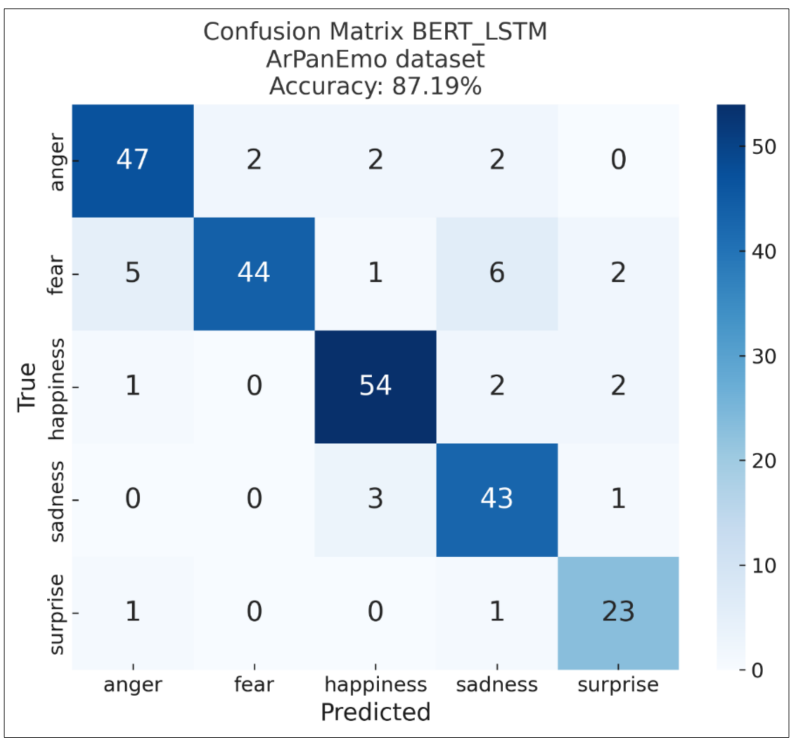


*Figure 4.2: Confusion matrix BERT Model only with ArPanEmo Dataset*

The first look at the accuracy of the model shows that it obtained a good score with the Saudi data set, and this may be due to the fact that the Saudi dialect is based on words expressing emotions on classical Arabic words in general, taking into account the nature of the data set used.

We note that the highest prediction errors were with feelings of fear, and this is natural since fear, sadness, and anger are close in terms of negativity, as previously indicated by researchers.[28]

#### 4.4.1.3 AETD Dataset:

now, when applying the hybrid model to the third dataset, AETD Dataset, the accuracy of the training and testing result was obtained as 77.2%, according to Table 4.6:

*Table 4.6: BERT Model with AETD Dataset*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.68 | 0.88 | 0.77 | 289 |
| **fear** | 0.96 | 0.92 | 0.94 | 242 |
| **happiness** | 0.81 | 0.78 | 0.79 | 256 |
| **sadness** | 0.72 | 0.69 | 0.7 | 251 |
| **surprise** | 0.74 | 0.56 | 0.64 | 209 |
| | | | | |
| **accuracy** | | | 0.77 | 1247 |

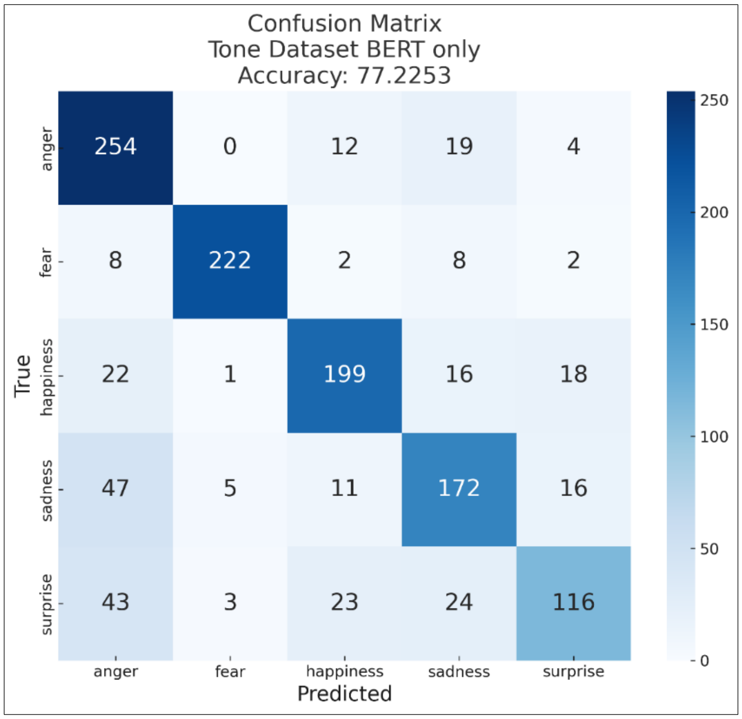


*Figure 4.3: Confusion matrix BERT Model with AETD Dataset*

### 4.4.2 Hybrid Model BERT and LSTM Layer

To increase the accuracy of the model, we added the last layer to be an LSTM model and tested whether it contributes to increasing the accuracy of the results or not. The results were as follows:

#### 4.4.2.1 IAEC Dataset:

With the IAEC Dataset the hybrid model achieved an accuracy of 84% which is a good change compared to using the BERT model alone, according to the table 4.7:

*Table 4.7: Hybrid Model with IAEC Dataset*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **Anger** | 0.85 | 0.81 | 0.83 | 101 |
| **Fear** | 0.84 | 0.75 | 0.79 | 51 |
| **Happiness** | 0.81 | 0.95 | 0.88 | 59 |
| **Sadness** | 0.94 | 0.89 | 0.91 | 70 |
| **Surprise** | 0.75 | 0.80 | 0.77 | 51 |
| | | | | |
| **Accuracy** | | | 0.84 | 332 |

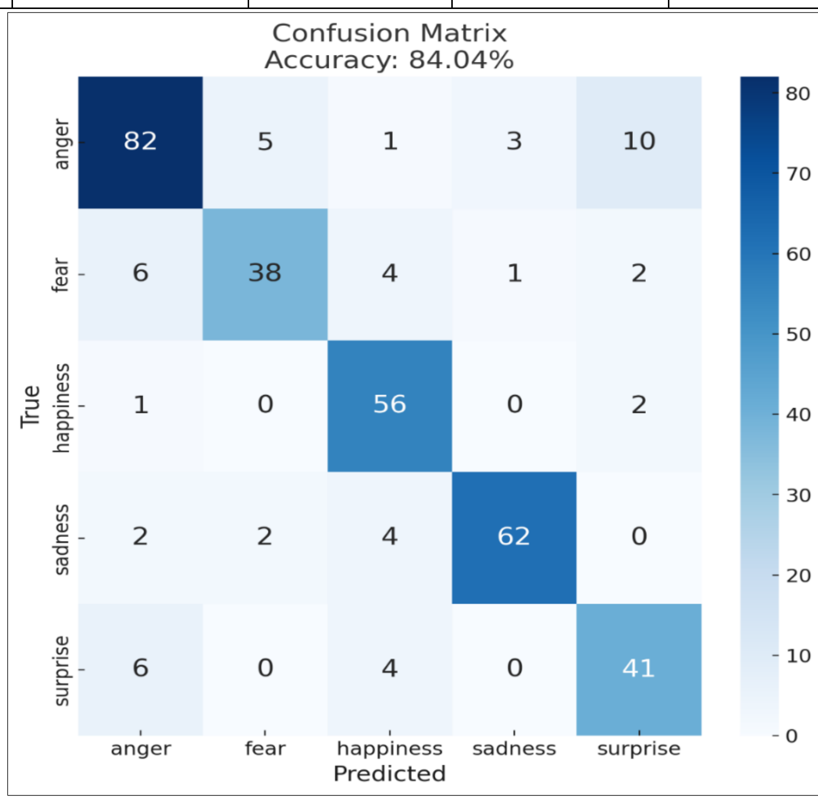


*Figure 4.4: Confusion matrix Hybrid Model with IAEC Dataset*

### 4.4.2.2 ArPanEmo Dataset:

By applying the hybrid model to the ArPanEmo dataset in the Saudi dialect, we notice that the accuracy of the model did not change much, as it obtained 87.1%. This is due to the nature of the Saudi dialect, as shown in the table 4.8:

*Table 4.8: Hybrid Model with ArPanEmo Dataset*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Anger | 0.87 | 0.89 | 0.88 | 53 |
| Fear | 0.96 | 0.76 | 0.85 | 58 |
| Happiness | 0.9 | 0.92 | 0.91 | 59 |
| Sadness | 0.8 | 0.91 | 0.85 | 47 |
| Surprise | 0.82 | 0.92 | 0.87 | 25 |
| | | | | |
| Accuracy | | | 0.87 | 242 |

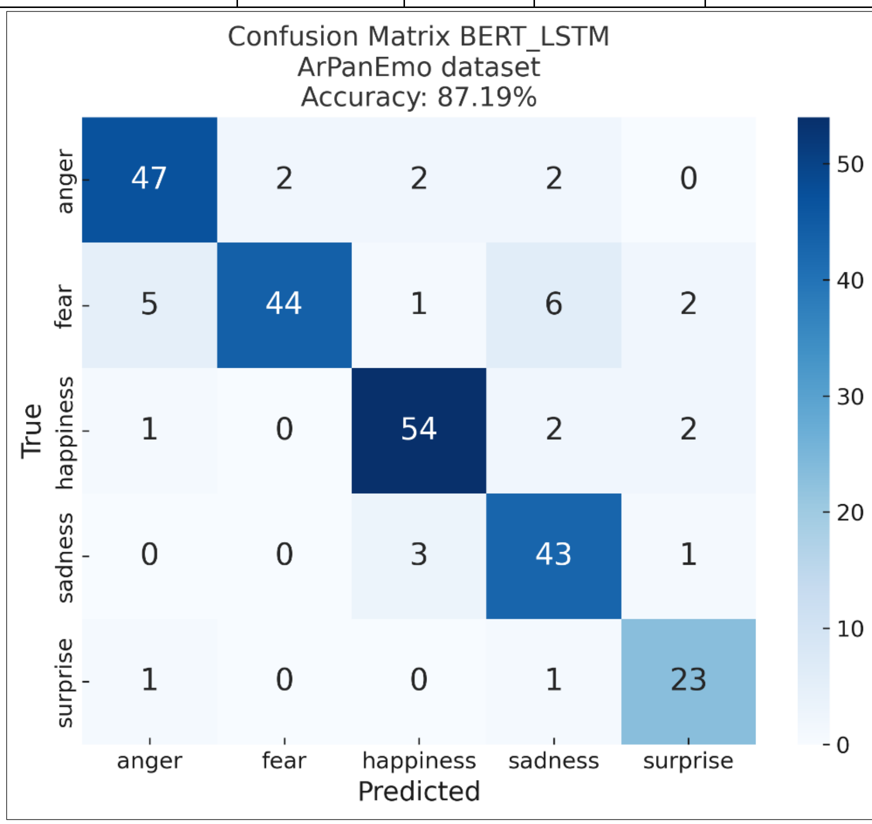


*Figure 4.5: Confusion matrix Hybrid Model with ArPanEmo Dataset*

### 4.4.2.3 AETD Dataset:

Finally, when applying the hybrid model to the third dataset, AETD Dataset, the accuracy of the training and testing result was obtained as 77.3%, according to Table 4.9.

*Table 4.9: Hybrid Model with AETD Dataset*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.79 | 0.77 | 0.78 | 310 |
| **fear** | 0.91 | 0.95 | 0.93 | 237 |
| **happiness** | 0.79 | 0.74 | 0.77 | 233 |
| **sadness** | 0.68 | 0.76 | 0.72 | 261 |
| **surprise** | 0.69 | 0.64 | 0.66 | 206 |
| | | | | |
| **accuracy** | | | 0.77 | 1247 |

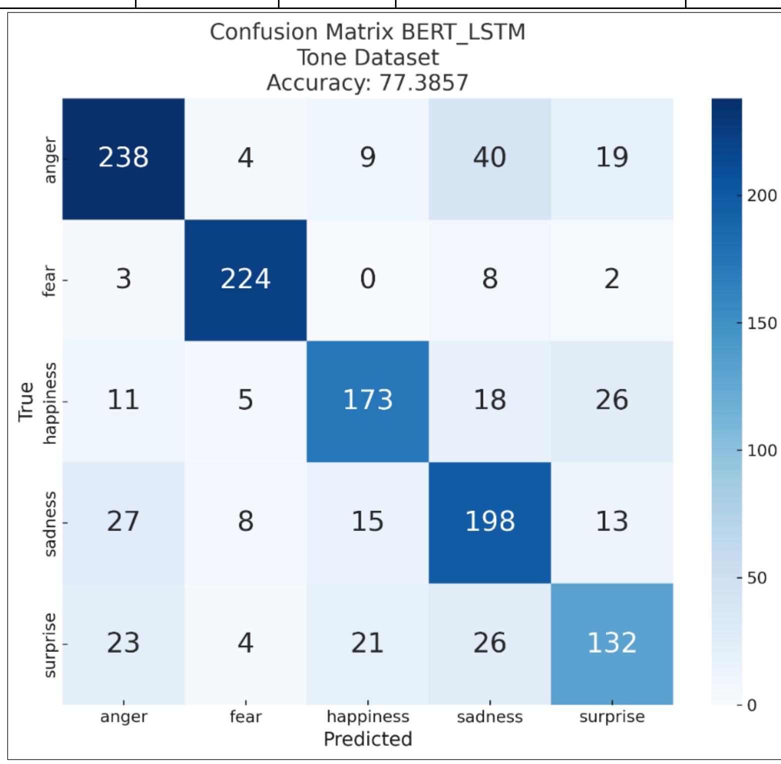


*Figure 4.6: Confusion matrix Hybrid Model with AETD Dataset*

## 4.4.3 Testing Proposed Models After Training

At this stage we will review the test results only after training the model and saving it for later use. We will also test the Iraqi model on the Saudi ArPanEmo dataset, and test the model trained on the Saudi ArPanEmo dataset with the IAEC Iraqi dataset.

### 4.4.3.1 Proposed Iraqi Hybrid Model:

After training the hybrid model on the Iraqi IAEC dataset and saving it for later use, we tested it on the following datasets:

#### 4.4.3.1.1 IAD dataset:

This is an Iraqi dataset divided into two classes, positive and negative. Considering that anger, fear and sadness are negative classes, and happiness and surprise are positive[28], the model was tested by prediction and the results were as follows table 4.11:

*Table 4.11: Iraqi Hybrid Model with IAD Dataset*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **NEG** | 0.82 | 0.76 | 0.79 | 1000 |
| **POS** | 0.78 | 0.83 | 0.81 | 1000 |

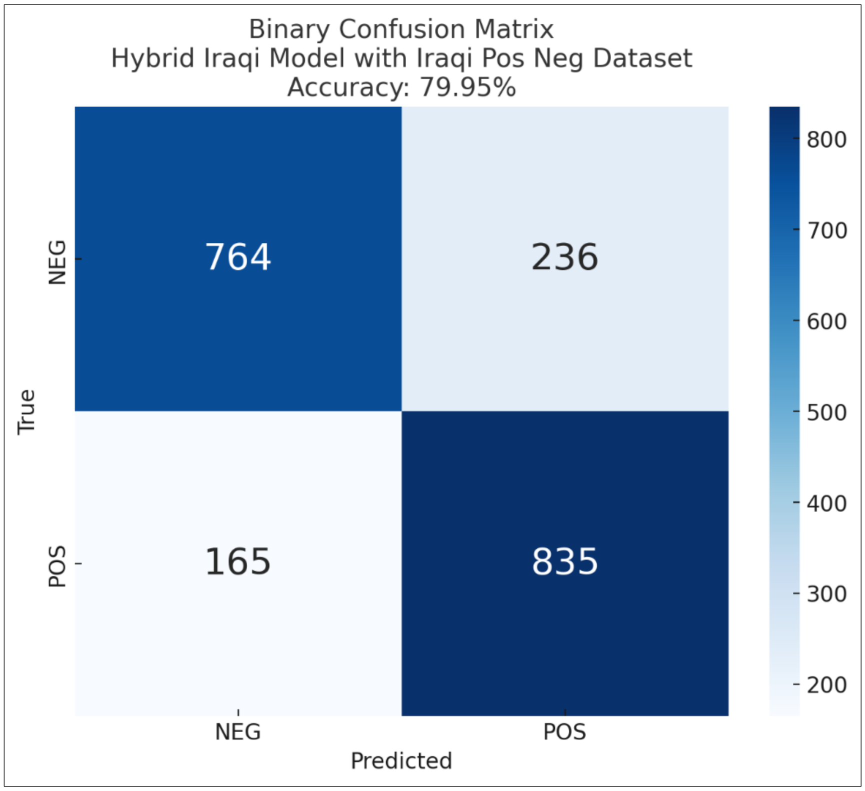


*Figure 4.8: Confusion matrix Iraqi Hybrid Model with IAD Dataset*

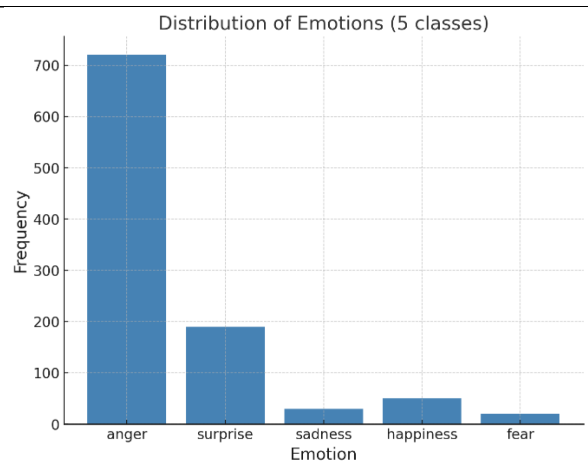


*Figure 4.9.A: Distributed negative dataset to 5 class*

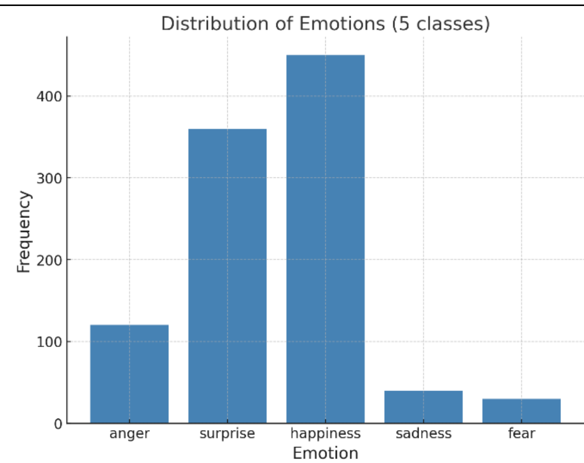


*Figure 4.9.B: Distributed positive dataset to 5 class*

### 4.4.3.1.2 Saudi ArPanEmo Dataset:

When the TEST model trained on the IAEC dataset was executed on the ArPanEmo Saudi Arabic dataset, the result was 45.5% and according to the table 4.12:

*Table 4.12: Iraqi Hybrid Model with ArPanEmo Dataset Test*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.49 | 0.67 | 0.57 | 262 |
| **fear** | 0.48 | 0.63 | 0.54 | 243 |
| **happiness** | 0.94 | 0.15 | 0.27 | 310 |
| **sadness** | 0.63 | 0.41 | 0.49 | 231 |
| **surprise** | 0.23 | 0.49 | 0.32 | 156 |
| | | | | |
| **accuracy** | | | 0.46 | 1202 |

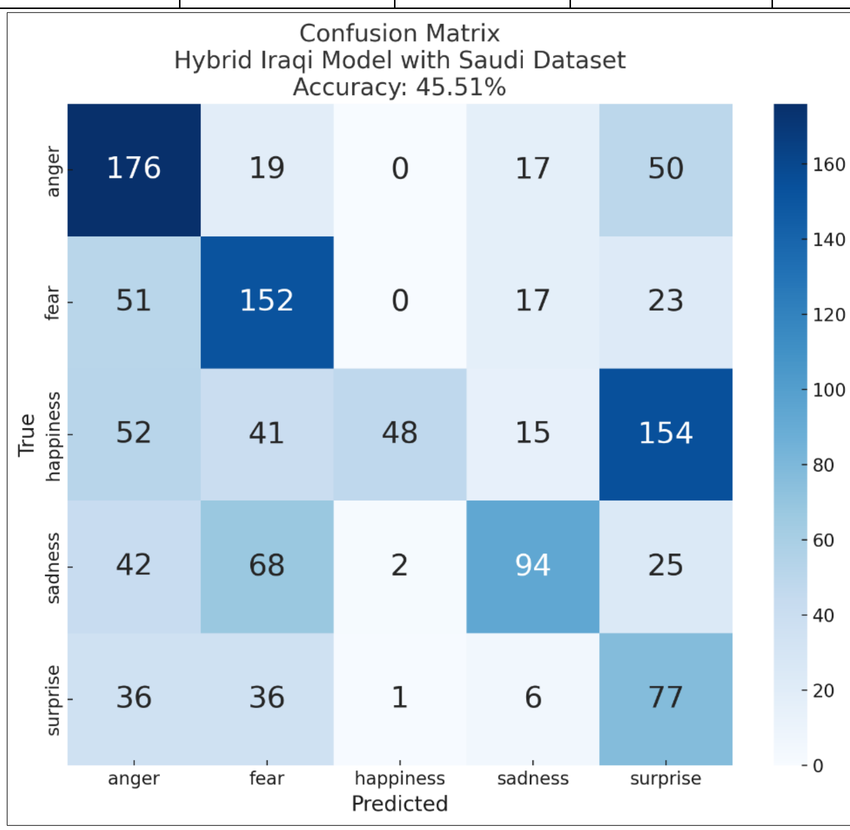


*Figure 4.10: Confusion matrix Iraqi Hybrid Model with ArPanEmo Dataset Test*

#### 4.4.3.1.3 AETD Dataset:

Using the hybrid model trained on Arabic texts in the Iraqi dialect IAEC Dataset When used for prediction on Arabic texts in the Egyptian dialect AETD Dataset the accuracy was 55.7% according to Table 4.13:

*Table 4.13: Iraqi Hybrid Model with AETD Dataset Test*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.57 | 0.7 | 0.63 | 1444 |
| **fear** | 0.82 | 0.85 | 0.84 | 1205 |
| **happiness** | 0.9 | 0.27 | 0.42 | 1280 |
| **sadness** | 0.63 | 0.31 | 0.42 | 1251 |
| **surprise** | 0.32 | 0.67 | 0.43 | 1045 |
| | | | | |
| **accuracy** | | | 0.56 | 6225 |

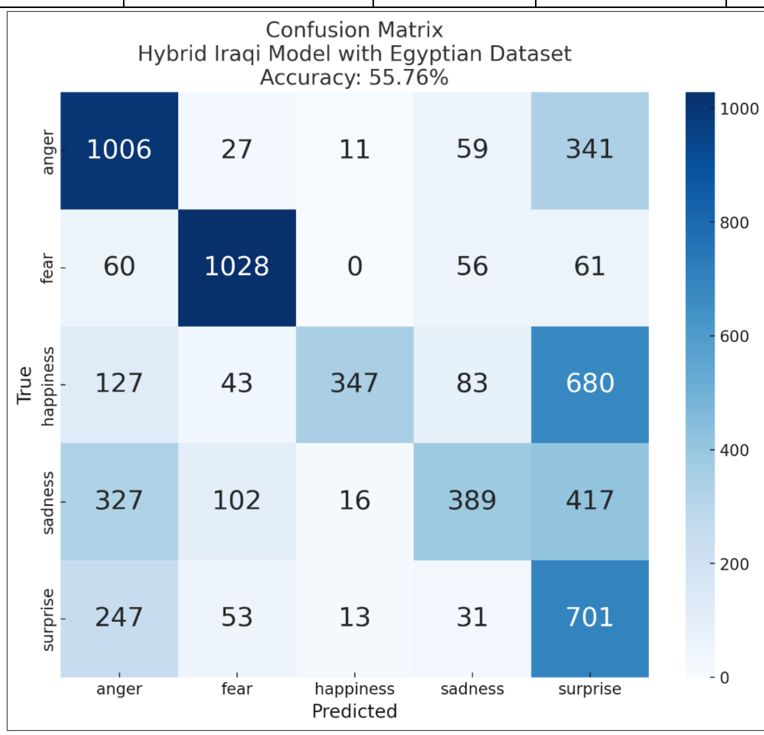


*Figure 4.11: Confusion matrix Iraqi Hybrid Model with AETD Dataset Test*

#### 4.4.3.2 Proposed Saudi Hybrid Model

The hybrid model trained on the ArPanEmo Dataset Saudi dialect was saved, and used in prediction on the following datasets:

##### 4.4.3.2.1 IAEC Dataset:

When the model predicted on the Iraqi IAEC dataset, the accuracy result was 57.1% according to Table 4.15:

*Table 4.15: Saudi Hybrid Model with IAEC Dataset Test*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.69 | 0.57 | 0.62 | 446 |
| **fear** | 0.52 | 0.41 | 0.46 | 215 |
| **happiness** | 0.53 | 0.84 | 0.65 | 348 |
| **sadness** | 0.57 | 0.77 | 0.66 | 329 |
| **surprise** | 0.47 | 0.18 | 0.26 | 315 |
| | | | | |
| **accuracy** | | | 0.57 | 1653 |

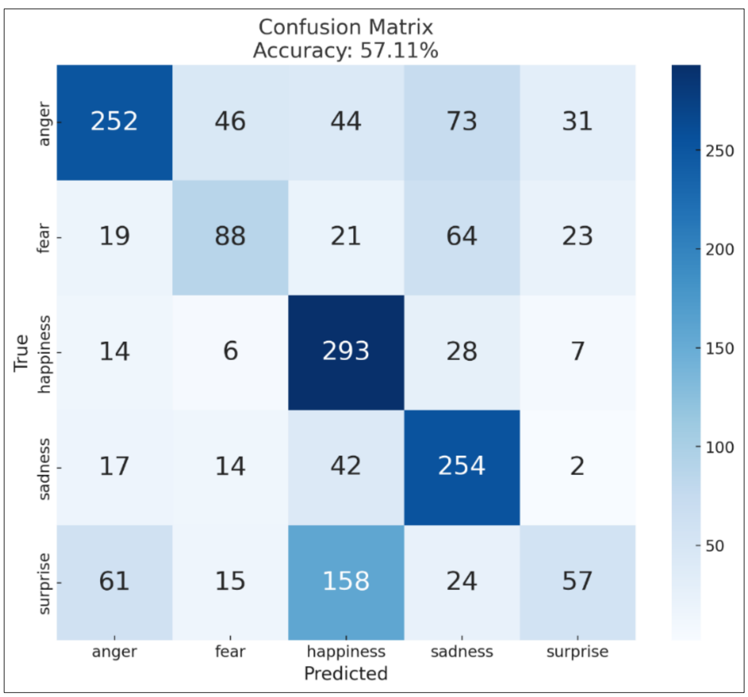


*Figure 4.13: Confusion matrix Saudi Hybrid Model with IAEC Dataset Test*

#### 4.4.3.2.2 AETD dataset:

When the model predicts on the Egyptian AETD dataset, the accuracy result was 62.99% according to Table 4.16:

*Table 4.16: Saudi Hybrid Model with AETD Dataset Test*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.55 | 0.69 | 0.61 | 1444 |
| **fear** | 0.73 | 0.91 | 0.81 | 1205 |
| **happiness** | 0.71 | 0.64 | 0.68 | 1280 |
| **sadness** | 0.64 | 0.5 | 0.56 | 1251 |
| **surprise** | 0.49 | 0.37 | 0.42 | 1045 |
| | | | | |
| **accuracy** | | | 0.63 | 6225 |

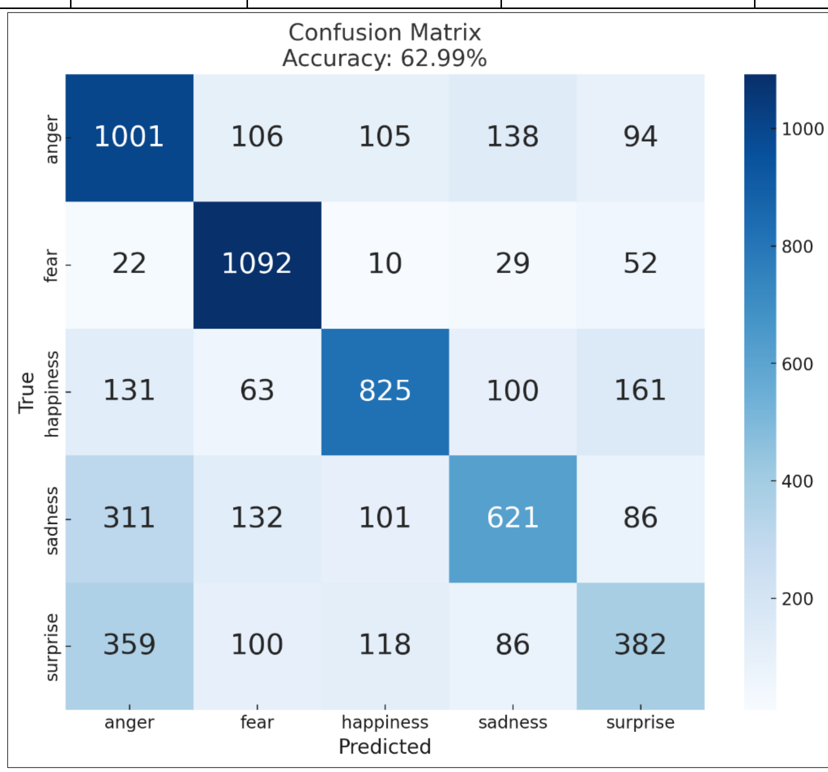


*Figure 4.14: Confusion matrix Saudi Hybrid Model with AETD Dataset Test*

#### 4.4.3.3 Proposed Egyptian Hybrid Model:

Here is the hybrid model after training it on the Egyptian Arabic dataset (AETD Dataset) and using it to predict on the rest of the datasets:

##### 4.4.3.3.1 IAEC Dataset:

When predicting the Iraqi dialect Arabic dataset (IAEC Dataset), the model accuracy was 54.4%, according to Table 4.18:

*Table 4.18: Egyptian Hybrid Model with IAEC Dataset Test*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.68 | 0.49 | 0.57 | 446 |
| **fear** | 0.66 | 0.41 | 0.5 | 215 |
| **happiness** | 0.54 | 0.87 | 0.66 | 348 |
| **sadness** | 0.46 | 0.8 | 0.59 | 329 |
| **surprise** | 0.42 | 0.09 | 0.15 | 315 |
| | | | | |
| **accuracy** | | | 0.54 | 1653 |

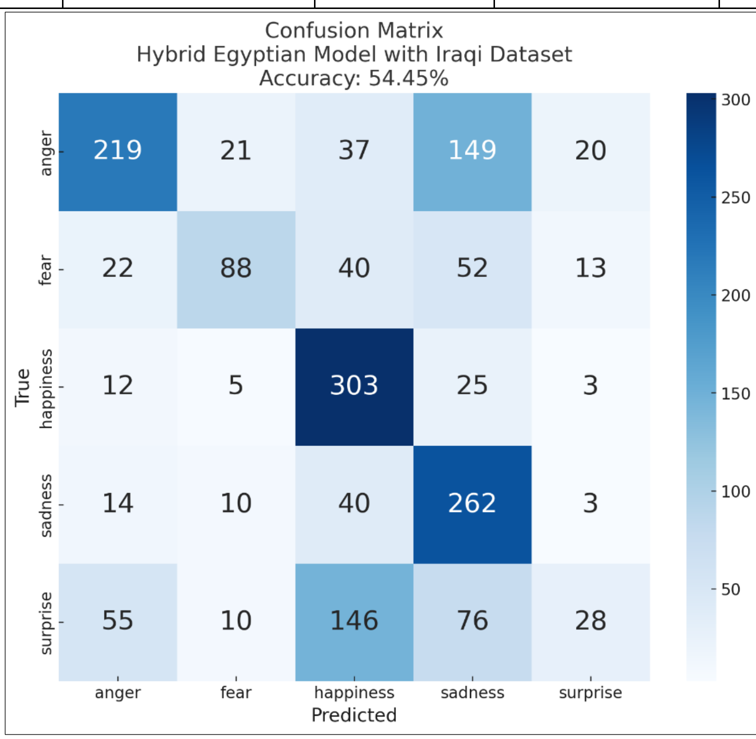


*Figure 4.16: Confusion matrix Egyptian Hybrid Model with IAEC Dataset Test*

#### 4.4.3.3.2 ArPanEmo Dataset:

When predicting the Saudi dialect Arabic dataset (ArPanEmo Dataset), the model accuracy was 57.8%, according to Table 4.19:

*Table 4.19: Egyptian Hybrid Model with ArPanEmo Dataset Test*

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **anger** | 0.67 | 0.41 | 0.51 | 262 |
| **fear** | 0.59 | 0.65 | 0.62 | 243 |
| **happiness** | 0.82 | 0.61 | 0.7 | 310 |
| **sadness** | 0.41 | 0.77 | 0.54 | 231 |
| **surprise** | 0.55 | 0.4 | 0.47 | 156 |
| | | | | |
| **accuracy** | | | 0.58 | 1202 |

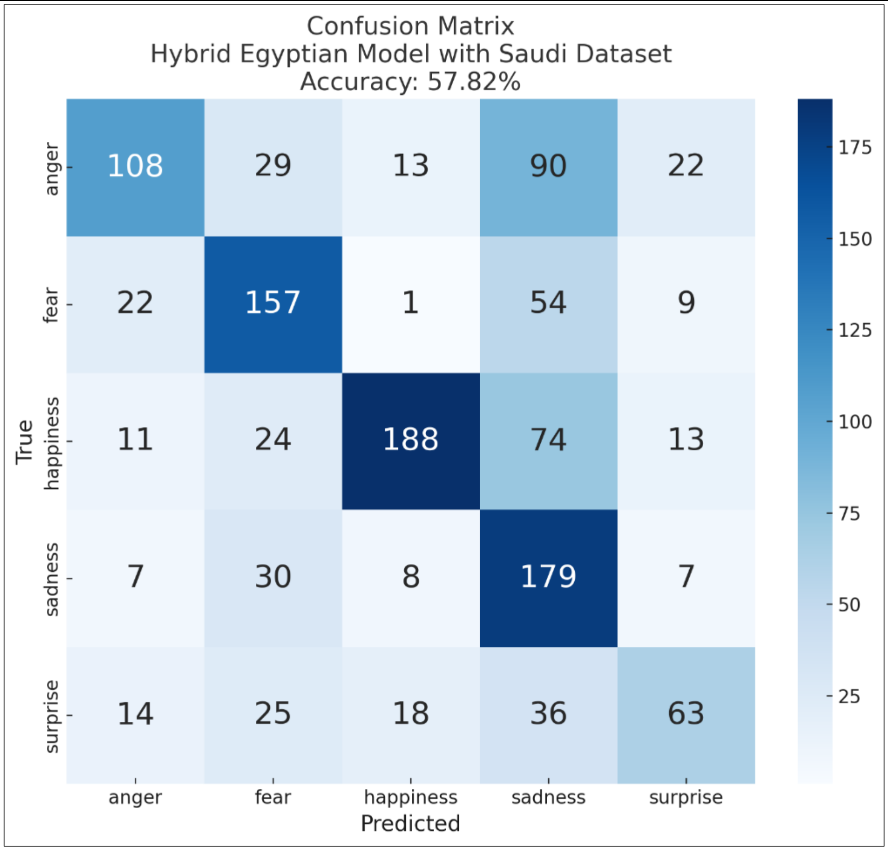


*Figure 4.17: Confusion matrix Egyptian Hybrid Model with ArPanEmo Dataset Test*

## Conclusion results:

In this section we will review the results more clearly to understand the work results more according to Table 4.20:

*Table 4.20: summary table for all result models*

| Action | Dataset | Model name | accuracy |
|---|---|---|---|
| Training and Test | IAEC Dataset | BERT model "arabertv02" | 81.9% |
| | | Proposed Hybrid Model | 84.0% |
| | ArPanEmo Dataset | BERT model "arabertv02" | 87.6 |
| | | Proposed Hybrid Model | 87.1 |
| | AETD Dataset | BERT model "arabertv02" | 77.2% |
| | | Proposed Hybrid Model | 77.3% |
| Test | IAEC Dataset | Saudi Hybrid Model | 57.1% |
| | | Egyptian Hybrid Model | 54.4% |
| | IAD Dataset | Iraqi Hybrid Model | 79.9% |
| | ArPanEmo Dataset | Iraqi Hybrid Model | 45.5% |
| | | Egyptian Hybrid Model | 57.8% |
| | AETD Dataset | Iraqi Hybrid Model | 55.7% |
| | | Saudi Hybrid Model | 62.9% |

Now, there are several observations.

- First, note that BERT alone was able to handle the Arabic dataset in the Iraqi dialect. However, when we added the LSTM layer, the model's accuracy increased. This is because BERT extracts context well, and LSTM can leverage the context vectors extracted by BERT for classification. LSTM converts words into vectors, which reduces understanding of the sentence context and, consequently, emotion classification in the text.
- Secondly, the results of using BERT model only are very close to the hybrid model for the Saudi dialect dataset. This indicates that the BERT model, pre-trained on Arabic, can recognize the Saudi dialect well. It also indicates that

the words used in the Saudi dialect contain a number of words from the classical Arabic language. This also applies to the Egyptian dialect dataset.

- Third, the results obtained from testing the proposed models reveal that dialect differences significantly affect the model's performance, even if everyone uses Arabic. This opens the door to research studies to find similarities between dialects.

## 4.5 Discussion

Social media is an integral part of people lives as much time is spent on social media sites sharing daily activities and stating their opinions directly. As a result, there is an increased demand for emotion classification in textual data to gain insights into user sentiment in contexts such as opinion mining, mental health, and marketing. Clearly the categorization has been done and a group of previous studies attempted to specify the nature of the text (pos/neg) (binary sentiment classification) but the goal here is to go deeper and tackle the task of fine-grained emotion classification in 5 categories: (angry, fear, happy, sad, surprised). Researching emotions to improve accuracy is thus significant considering emotions are subjective, context-dependent, and often quite ambiguous, making their prediction one of the key challenges researchers faced, and to do so researchers needed better models, greater linguistic resources, and more advanced computational methods.

The most important challenge faced by the researchers working in this area is the absence of high-quality emotion labelled datasets especially for Arabic dialects. The second challenge in extracting emotions from text is related to the nature of dialect Arabic itself, especially that the Iraqi dialect substantially deviates from MSA and includes unique expressions, informal hashes and vocabularies which are not present in standard Arabic dictionaries. In contrast to MSA's grammatical structure, Iraqi Arabic has phonetic spelling, borrowed words, and shortened phrases, complexities that make it more challenging for traditional NLP models to process. This requires a specific model that is built to understand the Iraqi dialect, its linguistic elements, and visible emotions.

Processing of Arabic texts and its dialects is a complex process and involves several challenges such as word normalization, diacritics removal, stopwords filtering, handling elongated words etc. Moreover, social media texts also include informal structures, abbreviated keywords, and multilingual (e.g. Arabic-English code-switching) data types that create additional challenges for the process of emotion classification. Emojis in the text is one important feature for emotion classification because they often directly indicate the emotion or enhance the context. Instead of excluding emojis from text (common in traditional NLP models), our method converts them to Arabic words, retaining their semantic meaning and augmenting the model's training corpus. As an example, it maps the " 😢 " emoji to "حزين" (sad) making sure that the emotion is preserved in text form.

In this paper, we propose an advanced deep learning model that achieves Arabic text emotion recognition over five emotion classes for Arabic text and the Iraqi dialect. We utilized a hybrid BERT-LSTM architecture to train our model over a carefully pre-processed data-set using a pre-trained BERT model, AraBERT, to tokenize the text and extract contextual embeddings. Model trained on it is able to encode into embeddings to ensure that the representation of textual data is dense and keeps the similar meaning words close to each other because such embeddings can take benefit of not only in; BERT serves as a static contextualized feature extractor but fails to capture sequential dependencies. To mitigate this, we add an LSTM layer, which is good at handling sequential data and allows the model to learn how words change during a sentence. This leads to effective encoding for temporal dependency and contextual meaning, and hence improves emotion classification performance.

Presented experiments on three datasets, representing various Arabic dialects (IAEC, ArPanEmo, AETD). It produced the best on IAEC (Iraqi Arabic) dataset, which is indicative of capturing dialectic emotions (7-20). However, the model accuracy on AETD data (Egyptian dialect) was low, which shows that the cross-dialect generalization is not so easy. It implies substantial domain adaptation on a multi-dialect Arabic corpus is needed to improve performance on other regional variants.

To improve training duration, we used parallel computing techniques to speed up model training through GPU acceleration where available. Doing so greatly decreased computation time, opening the door to faster experimentation and model tuning.

Conditioning the BERT model with an LSTM layer was significantly important to the enhancement of the model's performance. This is where the LSTM layer comes into play, as it allows the model to identify patterns of ramifications within a sentence by respecting the sequential dependencies in the embeddings created by BERT. This hybrid approach led to a substantial increase in accuracy compared to BERT-only. The accuracy achieved from our proposed hybrid model with Iraqi Arabic text proves that a sufficient degree of its dialectal expressions is processed properly by the model. This proves the concept and demonstrates it could be used in the wild in Arabic NLP applications.

In Tables, we evaluated the accuracy,  precision, recall, and F1-score of our model. These metrics gave a  broader evaluation of the model's strengths and weaknesses, indicating opportunities for future enhancements.

## 4.6 Summery

In the previous chapter, explained the data sets referred to in Chapter 3 and explained their nature. We also performed the preprocessing steps mentioned in Chapter 2 for each group separately, and we conducted training and testing procedures for each data set separately and with both models (pure BERT model - proposed hybrid model). Then we reviewed the results, and the models trained on a data set with a specific dialect were tested on a data set with another dialect. Finally, look forward to a detailed discussion of all aspects of the work accomplished in this thesis and the results that were achieved.

# CHAPTER FIVE

# CONCLUSION AND FUTURE WORK

## 5.1 Overview

In this last chapter, we summarize the work and findings of the thesis, and present suggestions for future researchers and work that should be highlighted and that our thesis will contribute to developing based on the work mentioned in this thesis.

## 5.2 Challenges and limitation

The Challenges of sentiment Analysis in Arabic has been underexplored for this area of research and development for various reasons, despite sentiment analysis being widely studied for the English language and other widely spoken languages:

- Morphological Complexity – Arabic is considered morphologically rich and inflectionally and derivationally varied, making both tokenization and feature extraction challenging [85].

- Dialectal Variations – Arabic has Modern Standard Arabic (MSA) and many dialects which vary drastically in terms of vocabulary, grammar and spelling, causing performance discrepancies in NLP models [27].

- Limited High-Quality Datasets: Compared to English, there are fewer annotated datasets in Arabic for sentiment and emotion analysis, especially for dialects such as Iraqi Arabic [28].

- Restricted Annotated Datasets – There are few publicly accessible, high-quality datasets for emotion recognition in Arabic dialect [28].

- Dialectal Variations Handling – Most Arabic emotion classifiers are MSA based, which does not work for dialects, including Iraqi Arabic [27].

- Multi-Language Influences – Code-switching (mixing Arabic with English or French) is frequently used by Arabic speakers, leading to intricate emotion classification models [86]

These challenges have led to various approaches and methods to characterize sentiment in social media Arabic texts [27].

## 5.3 Conclusion

1. The first objective was achieved by using the BERT model pre-trained on Arabic texts and combining it with the LSTM layer to increase accuracy, which produced a hybrid model trained on Arabic texts and the Iraqi dialect, and contributed to increasing accuracy to 84%.
2. The second objective was archived by classify emotions in the Arabic text and the Iraqi dialect into five class (happiness, sadness, anger, fear, surprise).
3. The third objective is achieved by using accuracy metrics to identify model performance.
4. The fourth objective was achieved by demonstrating the results of the dataset against traditional classification models such as SVM, Logistic Regression, Naive Bayes, and Random Forest.
5. The research indicates that leveraging natural language processing is a fundamental step in processing texts, especially when the texts contain dialects different from those of standard Arabic.
6. The research also takes a significant step toward processing emojis and converting them to their Arabic meanings using Python libraries.
7. The research suggests working with small amounts of data and attempting to increase it through data augmentation, using a new concept: using synonyms of words to generate new sentences with the same meaning.
8. The research also pointed to dealing with texts with different dialects and showed the difference in dialects in the Arabic language between Iraqi, Saudi and Egyptian.

## 5.4 Future Work

In this section, we review several proposed ideas directly related to our thesis, which may contribute to expanding the scope of the study and using its application in other fields:

1. Using the proposed model to detect similarity between dialects in the Arabic language
2. Creating various datasets from social media platforms in Arabic and the Iraqi dialect such as Twitter, Instagram, Thread, etc., to be available to scientific researchers.
3. Developing the proposed model and programming a writing application that analyzes the text and adds emojis based on the feelings of the person writing.
4. Using the proposed model in artificial intelligence chat tools and obtaining a tool that deals with Arabic texts written in the Iraqi dialect.

## 5.5 Summary

The last chapter of the thesis provides a summary of the most important points that we have addressed in the thesis and the main results in the structure of the model. It also outlines future work that is likely to be related to the field of the thesis and that we can adopt and work on at a later time.

# REFERENCES


[1] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," in *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Inc, Mar. 2021, pp. 610–623. doi: 10.1145/3442188.3445922.

[2] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed Tools Appl*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023, doi: 10.1007/s11042-022-13428-4.

[3] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in Social Media Research: Past, Present and Future," *Information Systems Frontiers*, vol. 20, no. 3, pp. 531–558, Jun. 2018, doi: 10.1007/s10796-017-9810-y.

[4] M. V Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," *Comput Sci Rev*, vol. 27, pp. 16–32, 2018, doi: https://doi.org/10.1016/j.cosrev.2017.10.002.

[5] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[6] S. F. Tsao, H. Chen, T. Tisseverasinghe, Y. Yang, L. Li, and Z. A. Butt, "What social media told us in the time of COVID-19: a scoping review," Mar. 01, 2021, *Elsevier Ltd*. doi: 10.1016/S2589-7500(20)30315-0.

[7] J. A. Tucker *et al.*, "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," 2018. [Online]. Available:

https://ssrn.com/abstract=3144139Electroniccopyavailableat:https://ssrn.com/abstract=3144139

[8] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, “Sentiment analysis based on deep learning: A comparative study,” *Electronics (Switzerland)*, vol. 9, no. 3, Mar. 2020, doi: 10.3390/electronics9030483.

[9] A. D. ’ Andrea, F. Ferri, and P. Grifoni, “Approaches, Tools and Applications for Sentiment Analysis Implementation,” 2015. [Online]. Available: http://messenger.yahoo.com/features/emoticons

[10] Q. T. Ain *et al.*, “Sentiment Analysis Using Deep Learning Techniques: A Review,” 2017. [Online]. Available: www.ijacsa.thesai.org

[11] W. H. Bangyal *et al.*, “Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches,” *Comput Math Methods Med*, vol. 2021, 2021, doi: 10.1155/2021/5514220.

[12] D. Rao, X. Miao, Z. Jiang, and R. Li, *STANKER: Stacking Network based on Level-grained Attention-masked BERT for Rumor Detection on Social Media*. 2021. doi: 10.18653/v1/2021.emnlp-main.269.

[13] J. Wang, H. Mao, and H. Li, “FMFN: Fine-Grained Multimodal Fusion Networks for Fake News Detection,” *Applied Sciences (Switzerland)*, vol. 12, no. 3, Feb. 2022, doi: 10.3390/app12031093.

[14] J. Mutinda, W. Mwangi, and G. Okeyo, “Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network,” *Applied Sciences (Switzerland)*, vol. 13, no. 3, Feb. 2023, doi: 10.3390/app13031445.

[15] L. Xiaoyan, R. C. Raga, and S. Xuemei, “GloVe-CNN-BiLSTM Model for Sentiment Analysis on Text Reviews,” *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/7212366.

[16] M. Ojeda-Hernández, D. López-Rodríguez, and Á. Mora, "Lexicon-based sentiment analysis in texts using Formal Concept Analysis," *International Journal of Approximate Reasoning*, vol. 155, pp. 104–112, 2023, doi: https://doi.org/10.1016/j.ijar.2023.02.001.

[17] T. Zhu, L. Li, J. Yang, S. Zhao, H. Liu, and J. Qian, "Multimodal Sentiment Analysis With Image-Text Interaction Network," *IEEE Trans Multimedia*, vol. 25, pp. 3375–3385, 2023, doi: 10.1109/TMM.2022.3160060.

[18] N. Thakur, "Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox," *Big Data and Cognitive Computing*, vol. 7, no. 2, Jun. 2023, doi: 10.3390/bdcc7020116.

[19] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, no. 4, pp. 1–22, 2009.

[20] N. Habash, *Introduction to Arabic Natural Language Processing*. Springer, 2010.

[21] W. Magdy, "ArASL: A large-scale Arabic slang lexicon for sentiment analysis," in *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2018.

[22] A. Abdelali, K. Darwish, K. Duh, and H. Mubarak, "Dialectal Arabic Processing: Challenges and Solutions," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 18, no. 3, pp. 1–21, 2019.

[23] Z. AlKhamissi, H. Alsudais, and I. Alabdulmohsin, "Dialect Identification for Arabic: A Comprehensive Survey," *Inf Process Manag*, vol. 60, no. 2, 2023.

[24] A. Elmadany, M. Abdul-Mageed, and M. Hashemi, "ArSentD-LEV: A Multi-Topic Corpus for Sentiment Analysis of Arabic Social Media," in

*Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 2020.

[25] P. Ekman, “An Argument for Basic Emotions,” *Cogn Emot*, vol. 6, no. 3–4, pp. 169–200, 1992.

[26] H. M. Alkhateeb and A. Alnahhas, “Emotion detection in Arabic texts extracted from twitter network by using machine learning techniques,” *Journal of Artificial Intelligence, Machine Learning and Neural Network*, no. 21, pp. 27–36, Jan. 2022, doi: 10.55529/jaimlnn.21.27.36.

[27] H. A. Nasrullah, M. A. Nasrullah, and W. N. Flayyih, “Sentiment Analysis in Arabic Language using Machine Learning: Iraqi Dialect Case Study.”

[28] A. A.-R. Alfarhany and N. A. Z. Abdullah, “Iraqi Sentiment and Emotion Analysis Using Deep Learning,” *Journal of Engineering*, vol. 29, no. 09, pp. 150–165, Sep. 2023, doi: 10.31026/j.eng.2023.09.11.

[29] M. Baali and N. Ghneim, “Emotion analysis of Arabic tweets using deep learning approach,” *J Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0252-x.

[30] A. J. Almahdawi and W. J. Teahan, “A New Arabic Dataset for Emotion Recognition,” in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2019, pp. 200–216. doi: 10.1007/978-3-030-22868-2_16.

[31] A. El-Sayed, M. Abougabal, and S. Lazem, “Benchmarking a large Twitter dataset for Arabic emotion analysis,” *SN Appl Sci*, vol. 5, no. 8, Aug. 2023, doi: 10.1007/s42452-023-05437-1.

[32] B. Brooke Auxier and M. Anderson, “Social Media Use in 2021 FOR MEDIA OR OTHER INQUIRIES,” 2021. [Online]. Available: www.pewresearch.org.

[33] S. Kemp, "Digital 2025: Global Overview Report," DataReportal – Kepios, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-global-overview-report

[34] S. Kemp, "Digital 2025: Iraq," DataReportal – Kepios, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-iraq

[35] S. Kemp, "Digital 2025: Saudi Arabia," DataReportal – Kepios, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-saudi-arabia

[36] S. Kemp, "Digital 2025: Egypt," DataReportal – Kepios, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-egypt

[37] N. Putri and D. Aminatun, "USING FACEBOOK TO PRACTICE WRITING SKILL: WHAT DO THE STUDENTS THINK?," *Journal of English Language Teaching and Learning*, vol. 2, pp. 45–50, Jun. 2021, doi: 10.33365/jeltl.v2i1.852.

[38] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective Computing and Sentiment Analysis," in *A Practical Guide to Sentiment Analysis*, E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, Eds., Cham: Springer International Publishing, 2017, pp. 1–10. doi: 10.1007/978-3-319-55394-8_1.

[39] S. Tiwari and A. Sinha, "Sentiment Analysis of Facebook Data using Machine Learning," *International Journal of Innovative Research in Applied Sciences and Engineering*, vol. 4, no. 4, pp. 735–742, Oct. 2020, doi: 10.29027/ijirase.v4.i4.2020.735-742.

[40] H. Alhuzali, T. Zhang, and S. Ananiadou, "Emotions and Topics Expressed on Twitter During the COVID-19 Pandemic in the United Kingdom: Comparative Geolocation and Text Mining Analysis," *J Med Internet Res*, vol. 24, no. 10, Oct. 2022, doi: 10.2196/40323.

[41] C. Li and F. Li, “Emotion recognition of social media users based on deep learning,” *PeerJ Comput Sci*, vol. 9, p. e1414, Jun. 2023, doi: 10.7717/peerj-cs.1414.

[42] H. Byeon, “Advances in Value-Based, Policy-Based, and Deep Learning-Based Reinforcement Learning,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023, doi: 10.14569/ijacsa.2023.0140838.

[43] Z. H. Ali and H. J. Aleqabie, “Emotion Detection in Arabic Text in Social Media: A Brief Survey,” *Al-Furat Journal of Innovations in Electronics and Computer Engineering*, vol. 3, no. 2, pp. 412–421, Jul. 2024, doi: 10.46649/fjiece.v3.2.27a.6.6.2024.

[44] S. Raschka, J. Patterson, and C. Nolet, “Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence,” Apr. 01, 2020, *MDPI AG*. doi: 10.3390/info11040193.

[45] C. Bishop, “Pattern Recognition and Machine Learning,” in *Journal of Electronic Imaging*, vol. 16, 2006, pp. 140–155. doi: 10.1117/1.2819119.

[46] L. T. Hardanto and L. A. Wulandhari, “Lithofacies Classification Using Supervised and Semi-Supervised Machine Learning Approach,” *Int J Adv Sci Eng Inf Technol*, vol. 11, no. 2, pp. 542–548, 2021, doi: 10.18517/ijaseit.11.2.11764.

[47] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow Google Brain, “Realistic Evaluation of Deep Semi-Supervised Learning Algorithms.” [Online]. Available: https://github.com/brain-research/realistic-ssl-evaluation

[48] R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction Second edition, in progress.”

[49] L. Ouyang *et al.*, “Training language models to follow instructions with human feedback”, doi: 10.48550/arXiv.2203.02155.

[50] D. Jurafsky and J. H. Martin, “Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models Third Edition draft Summary of Contents.”

[51] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. doi: 10.48550/arXiv.1810.04805.

[52] A. Rogers, O. Kovaleva, and A. Rumshisky, “A Primer in BERTology: What We Know About How BERT Works,” *Trans Assoc Comput Linguist*, vol. 8, pp. 842–866, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:211532403

[53] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” *ArXiv*, vol. abs/1907.11692, 2019, [Online]. Available: https://api.semanticscholar.org/CorpusID:198953378

[54] “The effects of Pre-Processing Techniques on Arabic Text Classification,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 1, pp. 41–48, Feb. 2021, doi: 10.30534/ijatcse/2021/061012021.

[55] H. Alangari and N. Algethami, “Exploring the Effects of Pre-Processing Techniques on Topic Modeling of an Arabic News Article Data Set,” *Applied Sciences (Switzerland)*, vol. 14, no. 23, Dec. 2024, doi: 10.3390/app142311350.

[56] I. Abu Farha and W. Magdy, “A comparative study of effective approaches for Arabic sentiment analysis,” *Inf Process Manag*, vol. 58, no. 2, Mar. 2021, doi: 10.1016/j.ipm.2020.102438.

[57] B. M. Sabbar, N. T. Yousir, and L. A. Habeeb, “SENTIMENT ANALYSIS FOR IRAQIS DIALECT IN SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS,” 2018. [Online]. Available: https://ijict.edu.iq

[58] A. Odeh, A. Abu-Errub, Q. Shambour, and N. Turab, “Arabic Text Categorization Algorithm Using Vector Evaluation Method,” *International Journal of Computer Science and Information Technology*, vol. 6, no. 6, pp. 83–92, Dec. 2014, doi: 10.5121/ijcsit.2014.6606.

[59] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[60] S. K. Bharti *et al.*, “Text-Based Emotion Recognition Using Deep Learning Approach,” *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/2645381.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[62] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Jan. 2013, [Online]. Available: http://arxiv.org/abs/1301.3781

[63] L. Zhang, S. Wang, and B. Liu, “Deep Learning for Sentiment Analysis : A Survey,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, Jan. 2018, doi: 10.1002/widm.1253.

[64] A. Vaswani *et al.*, “Attention Is All You Need,” 2023.

[65] R. Gupta, K. K. Bhardwaj, and D. K. Sharma, “Transfer Learning,” in *Machine Learning and Big Data*, 2020, pp. 335–360. doi: https://doi.org/10.1002/9781119654834.ch13.

[66] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021, doi: 10.1109/JPROC.2020.3004555.

[67] W. Ge *et al.*, "Identifying Patients With Delirium Based on Unstructured Clinical Notes: Observational Study," *JMIR Form Res*, vol. 6, no. 6, p. e33834, 2022, doi: 10.2196/33834.

[68] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based Model for Arabic Language Understanding," Feb. 2020, [Online]. Available: http://arxiv.org/abs/2003.00104

[69] T. Kudo and J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing," Aug. 2018, [Online]. Available: http://arxiv.org/abs/1808.06226

[70] B. Jang, M. Kim, G. Harerimana, S. U. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism," *Applied Sciences (Switzerland)*, vol. 10, no. 17, Sep. 2020, doi: 10.3390/app10175841.

[71] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

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

[73] L. Yuan, X. Zheng, Y. Zhou, C.-J. Hsieh, and K.-W. Chang, "On the Transferability of Adversarial Attacks against Neural Text Classifier."

[74] V. Yadav, P. Verma, and V. Katiyar, "Long short term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages," *International Journal of Information Technology (Singapore)*, vol. 15, no. 2, pp. 759–772, Feb. 2023, doi: 10.1007/s41870-022-01010-y.

[75] H. Wang and F. Li, "A text classification method based on LSTM and graph attention network," *Conn Sci*, vol. 34, no. 1, pp. 2466–2480, 2022, doi: 10.1080/09540091.2022.2128047.

[76] X. Zhou, Y. Q. Bao, T. Ji, and Q. Wang, "Two-Layer Optimization Scheduling Model of Integrated Electricity and Natural Gas Energy System Considering the Feasibility of Gas-Fired Units' Reserve," *IEEE Access*, vol. 8, pp. 40337–40346, 2020, doi: 10.1109/ACCESS.2020.2976835.

[77] M. Hu and B. Liu, "Mining and summarizing customer reviews," *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, [Online]. Available: https://api.semanticscholar.org/CorpusID:207155218

[78] M. Douglass, "Book Review: Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, 2nd edition by Aurélien Géron: O' Reilly Media, 2019, 600 pp., ISBN: 978-1-492-03264-9," *Phys Eng Sci Med*, vol. 43, Aug. 2020, doi: 10.1007/s13246-020-00913-z.

[79] A. Khan, D. Majumdar, and B. Mondal, "Sentiment analysis of emoji fused reviews using machine learning and Bert," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-92286-0.

[80] S. Ghanem, M. Jarrar, R. Jarrar, and I. Bounhas, "A Benchmark and Scoring Algorithm for Enriching Arabic Synonyms." [Online]. Available: https://portal.sina.birzeit.edu/synonyms

[81] A. A. ElSabagh, S. S. Azab, and H. A. Hefny, “A comprehensive survey on Arabic text augmentation: approaches, challenges, and applications,” *Neural Comput Appl*, vol. 37, no. 10, pp. 7015–7048, 2025, doi: 10.1007/s00521-025-11020-z.

[82] L. Khan, A. Amjad, K. M. Afaq, and H. T. Chang, “Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media,” *Applied Sciences (Switzerland)*, vol. 12, no. 5, Mar. 2022, doi: 10.3390/app12052694.

[83] M. J. Althobaiti, “ArPanEmo: An Open-Source Dataset for Fine-Grained Emotion Recognition in Arabic Online Content during COVID-19 Pandemic”, doi: 10.17632/d9yy8w52ns.2.

[84] A. Al-Khatib and S. R. El-Beltagy, “Emotional Tone Detection in Arabic Tweets,” 2017. [Online]. Available: https://www.researchgate.net/publication/320271778

[85] A. Alnawas and N. Arici, “Sentiment analysis of Iraqi Arabic dialect on Facebook based on distributed representations of documents,” *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 18, no. 3, Jan. 2019, doi: 10.1145/3278605.

[86] E. A. H. Khalil, E. M. F. El Houby, and H. K. Mohamed, “Deep learning for emotion analysis in Arabic tweets,” *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00523-w.

# الخلاصة

نظرًا للشعبية الواسعة والتأثير المتزايد لوسائل التواصل الاجتماعي، حظي تصنيف المشاعر في النصوص المكتوبة باهتمام كبير، وتم توظيفه في تطبيقات متعددة مثل تحليل الميول، وتنقيب الرأي العام، ومراقبة الصحة النفسية.

تُعَدّ اللغة العربية، وبالأخص اللهجات العربية مثل اللهجة العراقية، من أكثر اللغات تحديًا في مجال المعالجة الحاسوبية بسبب تعقيدها الصرفي، وقلة توافر مجموعات البيانات المعرّفة مسبقًا، والطابع غير الرسمي الذي يغلب على النصوص العربية المكتوبة في منصات التواصل الاجتماعي.

تتمثل قوة نموذج BERT في قدرته على استخراج السياق من النص عبر التمثيل السياقي للكلمات، في حين يتميز نموذج LSTM بفعاليته في التقاط الاعتماديات طويلة المدى في تسلسل النصوص مما يساعد في عملية التصنيف. وبناءً على ذلك، تقترح هذه الدراسة منهجًا هجينًا يجمع بين BERT وLSTM بهدف تحسين تصنيف المشاعر في النصوص العربية على وسائل التواصل الاجتماعي، مع التركيز بشكل خاص على اللهجة العراقية.

تم استخدام ثلاث مجموعات بيانات مصنفة مسبقًا للمشاعر، وهي: (IAEC) للهجة العراقية (ArPanEmo) للهجة السعودية، (AETD) للهجة المصرية.
ولمعالجة الفروقات اللهجية وندرة البيانات، طُبّقت تقنيات مختلفة للمعالجة المسبقة للنصوص مثل: التطبيع، وتحويل الرموز التعبيرية (الإيموجي) إلى نصوص، وزيادة البيانات بالاعتماد على المرادفات.

أظهرت النتائج أن استخدام BERT وحده حقق دقة بلغت ٨١%، بينما ارتفعت الدقة إلى ٨٤% عند اعتماد النموذج الهجين المقترح. وتبين هذه النتائج أن الجمع بين التعلم السياقي والتعلم التسلسلي يحقق أفضل أداء مقارنة بالنماذج السابقة في مجال تحليل المشاعر العربية. وبالتالي، تقدم هذه الدراسة إسهامًا موثوقًا في مجال معالجة اللغة الطبيعية العربية (Arabic NLP)، من خلال اقتراح منهج فعال وقابل للتوسع للكشف عن المشاعر في النصوص اللهجية.

جامعة كربلاء
كلية علوم الحاسوب وتكنولوجيا المعلومات
قسم علوم الحاسوب

# تصنيف المشاعر في نصوص وسائل التواصل الاجتماعي باللهجة العربية باستخدام تقنيات التعلم العميق ومعالجة اللغة الطبيعية

رسالة ماجستير
مقدمة الى مجلس كلية علوم الحاسوب وتكنولوجيا المعلومات / جامعة كربلاء وهي جزء من متطلبات نيل درجة الماجستير في علوم الحاسوب

**كتبت بواسطة**
زكريا حسين علي صالح الصفار

**بإشـراف**

أ.م.د هبة جبارعبدالواحد العقابي

١٤٤٦ هـ　　　　٢٠٢٥ م