ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

A Language-Adaptive Ensemble Clustering Framework for Emotion Detection in Multilingual Social Media Text

Wafa Saadi[1], Fatima Zohra Laallam[2] and Messaoud Mezati[3]

ABSTRACT

Social media platforms generate vast streams of emotionally rich textual data, offering valuable opportunities for critical applications, including mental health assessment and the analysis of collective public sentiment. However, detecting emotions in noisy and multilingual content remains challenging, particularly for under-resourced varieties such as dialects. Moreover, supervised learning techniques strongly depend on the availability of manually annotated corpora, whose creation requires substantial human effort and domain expertise. In contrast, unsupervised methods, while avoiding the need for human intervention, often lack sufficient robustness when confronted with the variability and complexity of natural language across diverse linguistic and cultural contexts. We present an ensemble clustering framework that automatically generates emotion labels from Twitter data, without human intervention. Our approach incorporates three emoji-handling strategies in the preprocessing step, enabling diverse semantic representations of emojis. We applied the BERT embeddings combined with PCA for dimensionality reduction within the same experimental framework. An ensemble clustering strategy integrating K-Means, Agglomerative clustering, and Gaussian Mixture Models (GMM) is adopted using multiple ensemble configurations. Experimental evaluation conducted on 10017 English tweets and 4134 Arabic tweets demonstrates that the proposed method achieves a silhouette score of 0.808 on English data using K-Means with Agglomerative and K-Means with GMM ensemble configurations. For Arabic data, silhouette scores of 0.728 and 0.718 are obtained using English and Arabic keywords, respectively. Emoji semantic integration enhances ensemble clustering performance, suggesting its importance for contextual disambiguation. The proposed framework provides a scalable solution for emotion detection in low-resource languages, enabling language-aware applications in multilingual contexts, particularly within linguistically diverse and multilingual populations.

Article information:

Keywords: Emotion Detection, Artificial Intelligence, Ensemble Clustering, Social Media, Twitter

Article history:
Received: August 11, 2025
Revised: November 20, 2025
Accepted: December 4, 2025
Published: December 27, 2025 (Online)

DOI: 10.37936/ecti-cit.2026201.263493

1. INTRODUCTION

In recent years, social media platforms have become pivotal channels through which individuals communicate their thoughts, experiences, and express a wide range of emotions. These platforms generate vast and continuously evolving streams of textual data, characterized by substantial linguistic variability and rich emotional content. Leveraging this data through advanced artificial intelligence (**AI**) and machine learning (**ML**) techniques has enabled new developments in computational social science, digital health, and affective computing. Among the various applications, detecting and interpreting emotions remains particularly important for understanding user behavior and interaction patterns. Emotion detection from social media [1, 2] contributes to a deeper understanding of online discourse and supports the development of emotion-aware systems for personal-

[1]The author is with the Computer Science Department, Mohamed Khider University, Biskra, Algeria, Email: wafa.saadi@univ-biskra.dz
[2,3]The authors are with the Computer Science department, Kasdi Merbah University, Ouargla, Algeria, Email: laallam.fatima_zohra@univ-ouargla.dz and mezati.messaoud@univ-ouargla.dz

ized services, mental health monitoring, and real-time public sentiment analysis.

Emotion detection, also referred to as *emotion recognition*, constitutes a critical task within the broader field of affective computing [3] and natural language understanding. It involves the automatic identification and categorization of emotional expressions across multiple data modalities, including text, images, audio, and video, based on established emotion models [4], ranging from discrete emotions such as joy, anger, sadness, and fear to more complex affective dimensions. Within the textual modality, emotion recognition depends on the analysis of semantic, syntactic, and contextual cues that are often subtle and highly variable across users and domains.

Advances in natural language processing (**NLP**), alongside machine learning and deep neural architectures, have driven the development of advanced models for inferring emotional content from unstructured text with improved accuracy. These models leverage linguistic representations, including contextual embeddings and attention mechanisms, to extract affective signals from large-scale textual data. Text-based emotion detection plays an important role in various real-world applications, such as consumer behavior analysis, digital mental health assessment, crisis monitoring, and emotion-aware conversational agents. In particular, integrating emotion recognition into social media analytics [5] provides insights into public sentiment, user engagement patterns, and population-level affective trends.

Emotion detection from textual data [6] fundamentally relies on the availability of robust and well-trained machine learning models. However, the effectiveness of such models is highly dependent on the quality and scale of labeled datasets. Constructing such datasets is especially challenging for under-resourced languages and dialects, including the Algerian dialect, where manual annotation remains both time-consuming and labor-intensive. To address this limitation, this work focuses on the automatic labeling of emotion-related data, aiming to generate high-quality corpora while reducing the need for extensive human annotation. To achieve this objective, we relied on unsupervised learning techniques, particularly clustering methods, which provide a viable alternative for tasks where labeled data is scarce or unavailable and manual annotation is impractical or costly.

Clustering in isolation often proves insufficient to adequately capture the intrinsic complexity and variability of real-world data, particularly when selecting a single optimal algorithm. Individual clustering methods, including k-means and DBSCAN, exhibit algorithm-specific biases and limitations that may adversely affect result consistency and quality. Consequently, reliance on a single clustering technique is insufficient for achieving robust performance across diverse datasets. To address these shortcomings, we adopt an ensemble clustering approach. Ensemble clustering, also referred to as cluster ensemble or consensus clustering [7, 8, 9, 10], is a technique that combines multiple base clustering results into a single, unified partition. The fundamental principle consists of leveraging the diversity across multiple clustering algorithms, as well as repeated executions of the same algorithm with different parameter configurations, to obtain a consensus clustering that exhibits enhanced stability, accuracy, and robustness to noise and initialization effects. This aggregation can be achieved using various consensus functions, including co-association matrices, graph-based methods, and voting schemes. By integrating multiple perspectives on the data, ensemble clustering mitigates the weaknesses of individual algorithms and enhances the overall reliability and interpretability of the clustering outcome. In the context of text-based emotion detection, this strategy enables more coherent groupings of emotionally relevant content, even when emotional expressions are subtle or context-dependent.

This work contributes to the expanding body of research in multilingual contexts by illustrating the effectiveness of ensemble clustering as a bridging mechanism between unstructured textual data and supervised learning frameworks. By leveraging algorithmic diversity and semantic mapping, the proposed methodology enables the development of scalable, language-adaptive, emotion-aware systems without relying on resource-intensive manual annotation processes.

The resulting datasets are annotated based on the Ekman emotional model [4], which defines six fundamental emotion categories: happiness, sadness, fear, disgust, anger, and surprise. To construct these labeled corpora, we employ ensemble clustering techniques that integrate multiple clustering algorithms, thereby enabling the generation of emotion-consistent clusters suitable for downstream supervised learning tasks. This methodology is evaluated on both English and Arabic textual datasets, with particular attention to the Algerian dialect. The main characteristic of the Algerian dialect is its multilingual nature, namely the simultaneous use of several languages within a single sentence to convey an idea. This phenomenon is characterized by the integration of linguistic elements from Algerian Arabic, French, Spanish, and Turkish. To address this constraint, transformer-based large language model (LLM) architectures, including BERT, were employed during data preprocessing to facilitate a more precise and nuanced analysis of emotional content in text. BERT, a transformer-based model, is pre-trained on an extensive multilingual dataset via self-supervised learning. By learning directly from raw text without human labeling, it can effectively exploit large volumes of publicly available data.

Unsupervised learning techniques provide an effective alternative for labelling textual data while minimizing human intervention. Within this framework, our primary contribution lies in leveraging an ensemble clustering approach to automatically annotate Twitter data for emotion detection, using the six basic emotions defined by Ekman as labels. Within the employed ensemble clustering framework, our contribution lies in the generation mechanism across three distinct levels: (i) the object representation level, achieved through the use of diverse emoji representations; (ii) the clustering algorithm level, via the application of multiple algorithms in various combinations; and (iii) the parameter initialization level, involving the configuration of different parameters for each algorithm and algorithmic combination. The automatic annotation of tweets with emotional labels results in labeled datasets for both Arabic and English. These resulting datasets form a robust foundation for training supervised models designed to achieve accurate emotion detection.

The remainder of this paper is organized as follows. Section 2 presents the background and related works of the study. Section 3 details the proposed methodology. Section 4 reports and discusses the results. Finally, Section 5 concludes the paper and outlines future work.

2. RELATED WORKS

Emotion detection from textual data [6, 11, 12] has emerged as a core task in affective computing and natural language understanding, attracting increasing attention across disciplines such as computational linguistics, psychology, and human-computer interaction. Overall, it involves the application of natural language processing and machine learning techniques to identify and classify the emotional states conveyed in written language. In this context, several approaches have been explored in the literature [13, 14, 15, 16, 17]. Explicit emotions are commonly identified through keyword-based methods [18, 19, 20, 21, 22], which rely on the direct occurrence of emotionally salient terms. In contrast, the recognition of implicit emotions [19], which are not overtly expressed in text, has driven the adoption of more advanced strategies. These include rule-based methods [23, 24] based on predefined affective lexicons (e.g., the NRC Emotion Lexicon [25] and WordNet-Affect [26]) as well as supervised learning models that treat emotion detection as a multiclass classification task [1, 2, 27, 28, 29, 30, 31, 32] using labeled datasets [15, 33]. Recent advances have introduced deep learning architectures [1, 2, 15, 27, 28, 29], including convolutional neural networks (CNNs) [34], recurrent neural networks (RNNs) [35], and transformer-based models (e.g., BERT and RoBERTa) [36, 37, 38], enabling the capture of contextual and semantic nuances in textual data. Due to the limitations of individual approaches, hybrid methods [39, 40] that combine the previously discussed techniques are employed to enhance the accuracy and robustness of emotion detection.

Although successful, these methods typically require large annotated corpora, limiting their applicability in low-resource and dialect-rich environments, such as Arabic dialects. To address this challenge, researchers have employed lexicon-based approaches adapted to Arabic (e.g., ArSEL [41]), as well as translation-based projection methods, in which emotion labels from English are transferred to Arabic text [42]. However, these strategies often suffer from semantic drift, idiomatic mismatch, and limited coverage when applied to informal and noisy texts, particularly those found on social media. This challenge has prompted researchers to explore weakly supervised [43, 44] and semi-supervised [45, 46]. The authors in [47] proposed a dialect-aware framework for Arabic dialect and emotion classification that combines customized preprocessing, fast Text-DBSCAN clustering for building dialect-specific emotion lexicons, and AraBERT-based classification. The proposed approach achieved high accuracy in dialect identification and consistently improved emotion detection compared with general models. The results demonstrate that incorporating dialectal context significantly enhances performance, thereby establishing a new benchmark for Arabic NLP. Only a limited number of studies have explored the use of clustering techniques to uncover emotional structures in unlabeled text corpora, particularly in scenarios where annotation is unavailable. These approaches have been applied to short texts, such as tweets and user comments, to identify groups expressing similar emotional content. When combined with semantic features, such as vector embeddings or affective term frequencies, clustering methods have demonstrated potential for organizing emotional information and reducing the need for manual labeling.

In [48], an unsupervised ensemble clustering approach is proposed for emotion detection on multilingual Twitter data, integrating emoji semantics and emotion-related keywords. Using BERT embeddings and multiple clustering algorithms, the proposed method demonstrated improved performance, particularly when combining emoji maps and keywords, on both Arabic and English datasets. The results highlight the effectiveness of emoji-aware clustering in enhancing emotion detection in short, noisy texts without human intervention.

In [49], a lightweight clustering-based approach is introduced to identify nuanced emotional expressions in social message streams without relying on predefined emotion categories. In contrast to traditional supervised models, the approach employs unsupervised clustering techniques to capture subtle emotional variations.

In [50], the authors enhance emotion-based text classification by integrating fuzzy entropy into the Fuzzy C-Means (FCM) clustering process. This hybrid approach enables more flexible and accurate grouping of emotionally similar content, particularly in cases involving ambiguity or overlapping emotional states. The incorporation of fuzzy entropy improves cluster separability and supports a more nuanced interpretation of emotional expressions.

In [51], tweets related to Anies Baswedan's Indonesian presidential candidacy were analyzed using Agglomerative Hierarchical Clustering. After preprocessing and TF-IDF vectorization, ten clusters were identified, distinguishing informational content from opinions. The results showed predominantly positive emotions, particularly excitement and anticipation, with limited expressions of doubt, highlighting the effectiveness of hierarchical clustering in capturing topical and affective dimensions in political social media data.

While existing research has explored clustering and, more recently, ensemble clustering for emotion detection, most studies have focused on high-resource languages, with limited attention given to Arabic, particularly dialectal varieties such as Algerian Arabic. The authors in [52] introduced an unsupervised sentiment analysis approach that combines contextual enrichment via SentiWordNet with an ensemble clustering framework featuring improved k-means initialization.

Furthermore, the challenge of generating labeled datasets without manual annotation remains unresolved mainly in this context. Existing research has primarily focused on English or other high-resource languages, leaving a gap in scalable solutions for emotion detection in multilingual and under-resourced settings. Our work builds on this line of research by extending ensemble-based approaches to multilingual and dialect-rich Arabic social media data, addressing both the scarcity of labeled resources and the complexity of linguistic variation.

3. METHODOLOGY

Emotion detection from textual data constitutes a comprehensive and methodologically rigorous process that encompasses successive stages, including data acquisition, text preprocessing, feature extraction, and the application of advanced machine learning algorithms to achieve reliable emotion classification.

The figure presented below illustrates the sequential architecture of the proposed framework, detailing the flow of data from initial collection through preprocessing to the emotion inference phase.

This schematic representation illustrates the methodological coherence of the proposed approach and emphasizes the integration of diverse natural language processing techniques with affective computing strategies. The resulting pipeline ensures scalability

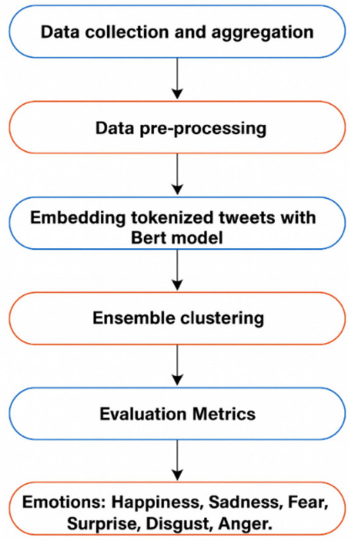

Fig.1: *Methodology Pipeline.*

and adaptability, facilitating the effective processing of heterogeneous textual inputs and the accurate extraction of emotional signals.

3.1 Data collection

The corpus used in this study comprises two complementary datasets. The first dataset was collected via the Twitter API and consists of approximately **4134** Arabic tweets derived from real-time, user-generated textual data. In parallel, the second dataset was sourced from Kaggle and consists of a publicly available corpus containing **10017** English tweets. The characteristics of textual data play a critical role in determining the accuracy and effectiveness of emotion detection, particularly given the linguistic and structural differences between the English and Arabic datasets. The following table summarizes the key characteristics of the datasets used in this study.

Table 1: *Key Features of the corpus.*

Futures	Arabic Dataset	English Dataset
Total number of Tweets	4134	10017
Number of Tweets with Emojis	919	2257
Number of Tweets without Emojis	3215	7760
Total number of Words	24964	131265
Total number of Emojis	3023	4806
Percentage of Emojis compared to words	12.11%	3.66%

This dual-sourcing strategy establishes a robust basis for the construction of a heterogeneous and multilingual corpus, ensuring both representativeness and diversity, which in turn improves the interpretation of emotions derived from short textual data.

While the English dataset obtained from Kaggle is partially pre-processed, the Arabic dataset collected via an API is raw and inherently noisier. Despite this disparity, both datasets were processed using an identical preprocessing and cleaning pipeline to ensure consistency. This initial difference may affect the distribution of linguistic patterns as well as the level of preprocessing required.

3.2 Data preprocessing

Before initiating the automatic labeling process, the raw textual data underwent a structured preprocessing workflow designed to clean, normalize, and prepare the input for reliable label assignment. This step is essential for ensuring semantic clarity and minimizing noise-related inconsistencies, which are particularly prevalent in informal, user-generated content, especially in short texts such as tweets. By standardizing the data, the preprocessing phase facilitates more accurate and context-aware emotion annotation in subsequent stages. This phase includes several sub-phases, which are described below.

3.2.1 Duplicate Letter Normalization

This step addresses the overuse of repeated characters by reducing sequences of identical letters to a maximum of two consecutive instances, thereby contributing to lexical normalization and reducing noise in the textual data.

3.2.2 Text Cleaning

This process involves the removal of special characters, punctuation marks, numerical digits, URLs, user mentions, stock tickers, and obsolete retweet indicators. For Arabic textual data, additional cleaning operations include the removal of diacritics, elongated characters, and Arabic-specific punctuation, such as question marks, to preserve the integrity and uniformity of the dataset.

Several further operations are applied to the textual data, notably:

Hashtag Processing: Hashtags, typically consisting of alphanumeric characters and, in some cases, emojis, and prefixed with the "♯" symbol, function as metadata that categorize content and enhance its discoverability on social media platforms. Considering that the textual components of hashtags frequently encapsulate significant semantic and emotional content, this step required the removal of the "♯" symbol, while retaining the associated terms.

Text Formatting: this step aims to standardize the structure of the textual input by eliminating newline characters, replacing underscores with whitespace, and removing redundant spaces. These operations contribute to a cleaner and more uniform text format, thereby facilitating reliable tokenization and subsequent processing stages.

Lowercasing: converting all text to lowercase ensures lexical uniformity and prevents the model from treating identical words with different capitalizations as distinct tokens. This normalization step helps reduce vocabulary sparsity and ensures more consistent feature representation.

3.3 Embedding with the BERT model

To process English textual data, we utilize BERT (Bidirectional Encoder Representations from Transformers), a deep contextual language model based on the Transformer encoder architecture. These tasks enable the model to learn bidirectional contextual representations, effectively capturing linguistic dependencies from both preceding and succeeding tokens.

Within the proposed methodology, BERT is employed to generate contextualized embeddings from raw textual inputs. These embeddings capture the contextual semantics of words as well as their interdependencies within a sentence. Token identifiers (token IDs) serve as model inputs, allowing each token to be mapped to its corresponding vector representation. The resulting outputs are contextualized embeddings that encode fine-grained linguistic information derived from the surrounding context.

The representations generated by BERT are general-purpose and can be effectively fine-tuned for a wide range of downstream tasks, including emotion detection. These dense representations encode rich syntactic and semantic information, which is crucial for accurately identifying emotions in short and informal texts, such as those commonly encountered on social media platforms.

The output shape of the BERT model is (**4134, 768**) for Arabic data and (**10017, 768**) for English data, where each row corresponds to a high-dimensional vector representation of a text instance. While rich in semantic information, these high-dimensional embeddings introduce computational and performance challenges in downstream clustering tasks due to the curse of dimensionality. To address this issue, we employed **P**rincipal **C**omponent **A**nalysis (PCA) as a dimensionality reduction technique. In our case, PCA projects the original feature space onto a two-dimensional subspace that preserves most of the data variance, thereby retaining the most informative structures. By reducing tweet embeddings to two principal components, this approach produces a more compact and tractable representation, facilitating more efficient and meaningful clustering

3.4 Ensemble Clustering

In this study, an ensemble clustering approach is employed to enhance the robustness and stability of the final clustering solution. Ensemble clustering, also referred to as consensus clustering, aggregates multiple base clusterings generated using different algorithms, random initializations, or data subspaces into a single consolidated partition. This strategy mitigates the inherent variability and sensitivity of individual clustering methods by leveraging the collective information captured across diverse clustering solutions. As illustrated in the figure below, the ensemble clustering process comprises two principal phases: (1) the generation of base clustering partitions and (2) the consensus phase, during which these partitions are integrated into a single unified solution.

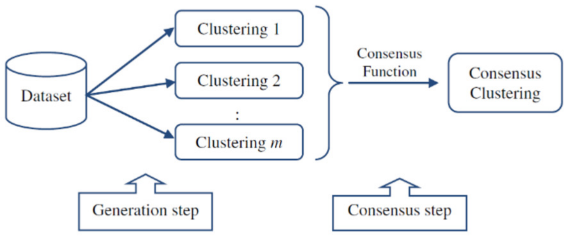

Fig.2: *Ensemble clustering Process [10].*

3.4.1 The generation mechanism

Generation is the first step in the clustering ensemble process; in this step, the set of clusterings to be combined is generated. This phase involves generating a diverse set of base partitions by applying one or more clustering algorithms to the same dataset. The different generation strategies are illustrated in the figure below. This research focuses on applying clustering to different object representations and using different clustering algorithms, as well as multiple runs of the same algorithm with varied initializations or parameter settings. Based on the foundational work presented in [48], the proposed model is designed to contribute across multiple stages of the clustering process. This includes enhancements in the generation of base clusterings, the selection of diverse representations, and the aggregation strategy used for consensus formation. By systematically addressing these distinct phases, the proposed approach improves the robustness and adaptability of the ensemble clustering framework to the complexities inherent in emotion detection tasks.

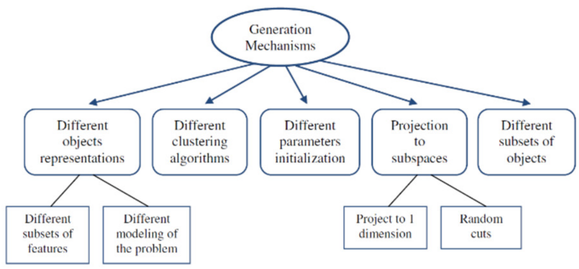

Fig.3: *Generation Mechanisms tasks [10].*

In different object representation: the removal of emojis is not considered an optimal strategy for emotion detection tasks. In the mechanism defined in [48], particular attention is given to the joint handling of emojis and keywords to enhance the emotional analysis of textual data. By simultaneously addressing these two semantic components, the method aims to achieve a more nuanced and comprehensive interpretation of the emotions conveyed in Arabic and English text. Accordingly, the proposed approach incorporates three techniques for processing emojis, aiming to maximize their contribution to the accurate identification of emotions embedded in textual content. The first technique consists of replacing emojis with their textual descriptions using a specialized package. The second technique consists of grouping emojis that convey the same emotion into a single representative emoji using an emoji map derived from multiple sources. Two representation strategies are then applied to model the six basic emotions by replacing the resulting emoji either with its corresponding meaning or with its textual description. The third technique consists of extracting and duplicating emojis while preserving their order, followed by a classification step that groups them into predefined categories based on the closest matching target emoji. The resulting data are stored in a structured dictionary for further analysis. In keyword processing, emotion-related word lists are condensed into single representative terms using a keyword map derived from the NRC Emotion Lexicon. It is important to note that, for Arabic data analysis, both English and Arabic keywords are utilized.

Using different clustering algorithms: in this mechanism, the methodology centers on the integration of multiple clustering algorithms, namely K-Means, Gaussian Mixture Models (GMM), and Agglomerative Clustering, to enhance the effectiveness of emotion detection from textual data. By integrating the complementary strengths and inductive biases of these algorithms, the proposed approach seeks to enhance both the accuracy and robustness of the clustering results. This ensemble strategy facilitates more nuanced modeling of emotional expressions by capturing the complex and subtle patterns often present in short, informal texts. Consequently, it enables a more comprehensive understanding of the underlying emotional content conveyed in user-generated data. Following comprehensive preprocessing and the integration of emoji-handling techniques across both the English and Arabic datasets, the methodology proceeds with the application of multiple clustering algorithms. Each algorithm is executed under distinct parameter configurations, enabling a systematic evaluation of their effectiveness in capturing emotion-relevant structures within the processed textual representations. The integration method employed during the consensus phase is based on a voting strategy, whereby the final partition is obtained by aggregating the outputs of individual base clusterings through majority agreement across data instances.

In different parameter initialization: In this mechanism, emphasis is placed on exploring diverse parameter initialization strategies for clustering algorithms. This approach is applied explicitly to K-Means and Gaussian Mixture Models (GMMs) to evaluate the effects of different initialization techniques on their performance and stability. By introducing controlled variation at the initialization phase, the methodology aims to enhance the consistency and expressiveness of the resulting emotion-based clusters.

3.5 Evaluation Metrics

As clustering is an unsupervised learning technique, there is no inherent mechanism for directly measuring the accuracy of the resulting models. To address this limitation, a range of evaluation approaches has been proposed, including internal, external, and manual evaluation methods. In this research, commonly used internal evaluation metrics are employed to assess the performance of clustering algorithms. These metrics allow us to evaluate the quality of the generated clusters independently of external class labels. The main internal metrics considered in this work include:

3.5.1 Silhouette Coefficient

The Silhouette Coefficient evaluates clustering quality by comparing the mean intra-cluster distance and the mean inter-cluster distance for each data point. It is computed using the following equation:

$$Silhouette\ Score = \frac{b-a}{\max(a, b}$$

Where **a** is the mean distance between the current data point and all other data points in the same cluster, and **b** is the mean distance between the current data point and all other data points in the next nearest cluster.

The silhouette coefficient ranges from -1 to 1, where a value close to -1 indicates that a data point may be incorrectly clustered, a value around 0 suggests that clusters are not clearly separable, and a value close to 1 reflects dense and well-separated clusters. In general, higher values approaching 1 indicate superior clustering quality.

3.5.2 Calinski-Harabasz Index

The Calinski-Harabasz (CH) Index, also referred to as the Variance Ratio Criterion, quantifies the similarity of an object to its own cluster (cohesion) relative to other clusters (separation). Cohesion is quantified by the distances between data points and their respective cluster centroids, whereas separation is determined by the distances between cluster centroids relative to the global centroid. It can be calculated as follows:

$$CH\ Index = \frac{Trace(Bc)}{Trace(Wc)} * \frac{n_E - c}{c-1}$$

Where, c denotes the number of clusters, n_E represents the size of the dataset E, and trace(Bc) refers to the trace of the between-cluster (inter-cluster) dispersion matrix, and trace(Wc) denotes the trace of the within-cluster (intra-cluster) dispersion matrix. Higher values indicate more well-defined clusters and, consequently, superior clustering performance, whereas lower values suggest poorer clustering quality.

3.5.3 Davies-Bouldin Index

The Davies-Bouldin (BD) Index evaluates clustering models by measuring the average similarity of each cluster to its most similar cluster. Similarity is defined as the ratio between intra-cluster distance and inter-cluster distance, where lower values indicate more compact clusters with better separation. It can be computed as follows:

$$BD\ Index = \frac{1}{c}\sum_{i=1}^{c} \max_{i\neq j} \frac{(\sigma_i + \sigma_j)}{d(c_i + c_j)}$$

Where $\boldsymbol{c}$ is the number of clusters, $\boldsymbol{\sigma_i}$ is the dispersion of cluster $\boldsymbol{i}$, $\boldsymbol{\sigma_j}$ is the dispersion of cluster $\boldsymbol{j}$, and $\boldsymbol{d(c_i, c_j)}$ is the distance between the centroids of clusters i and j. In practice, lower Davies–Bouldin (DB) index values indicate superior clustering performance, whereas higher values reflect poorer cluster quality.

3.5.4 Resulting Emotional Labelling

After applying the ensemble clustering algorithms, both the Arabic and English datasets were systematically annotated. The annotation was performed based on Ekman's basic emotion model, which comprises six basic emotions: happiness, sadness, fear, disgust, anger, and surprise. Ekman's model is the most commonly adopted emotional framework in text-based emotion detection, as it provides a concise, empirically validated, and cross-culturally robust theoretical basis for the systematic classification of emotional states. Its applicability across diverse linguistic and cultural datasets makes it particularly well suited for studies involving multilingual corpora.

4. RESULTS AND DISCUSSIONS

In this section, the performance of the proposed ensemble clustering framework for emotion detection is evaluated on both English and Arabic Twitter datasets. We systematically compare multiple base clustering mechanisms, including K-Means, Gaussian Mixture Models (GMM), and agglomerative clustering, under various initialization strategies. Preprocessing configurations are selected based on the best-performing results. During the consensus phase, a voting-based aggregation method is employed. The quality of the ensemble clustering is evaluated using three metrics: the Silhouette coefficient, the Calinski–Harabasz index, and the Davies–Bouldin index.

4.1 In different object representations

Table 2 shows the results of applying the K-means ensemble clustering algorithm to both English and Arabic data. It evaluates different object representation techniques described in Section 3.4.1 using the Silhouette score.

Interpretative notes on the Table 2 Legend:
K1: 4718 words,
K2: 9436 words,
K3: 14154 words,
D: Emojis description,
M: Emojis map and meaning,
MD: Emojis map and description.

Table 2: *The silhouette score of the K-Means ensemble clustering algorithm.*

	Technique	English Data	Arabic data with Arabic keyword	Arabic data with English keyword
Keywords	K1	0.63551307	0.44243714	0.67778635
	K2	0.8201926	0.095876314	0.7015065
	K3	0.6870811	0.2314722	0.31340364
Emojis description	D	0.50868887	0.4847758	0.48477587
	D+K1	0.7347591	0.3865655	0.44216302
	D+K2	0.10922659	0.2603351	0.3857908
	D+K3	0.73475957	0.123009235	0.10213358
Emojis map and meaning	M	0.68193233	0.713531	0.71353114
	M+K1	0.7980847	0.37672126	0.72585523
	M+K2	0.06491469	0.34749	0.69142765
	M+K3	0.79808414	0.54133844	0.399814
Emojis map and description	MD	0.5086886	0.4847758	0.484776
	MD+K1	0.734759	0.38656536	0.44216287
	MD+K2	0.10922663	0.26033512	0.38579088
	MD+K3	0.7347591	0.12300883	0.102133326

The results presented in Table 2 show that, for English data, mid-sized keywords (K2) achieve the highest clustering quality, as indicated by the highest Silhouette score (0.8201926), while combining emoji meanings with smaller keyword sets (M + K1) further improves performance.

For the Arabic data, keyword-based representations alone were less effective, yielding substantially lower silhouette scores, particularly when relying on Arabic keywords. Notably, the incorporation of emoji semantics alone produced the most favorable results (0.713531), highlighting the critical role of emojis in disambiguating emotional content within Arabic text.

Similarly, combining emoji mapping with K1 keywords in the Arabic data, supplemented by English keywords, resulted in notable performance improvements (0.72585523), confirming the advantages of leveraging cross-linguistic resources. Overall, the integration of selective keyword partitions with emoji representations proved essential for achieving robust

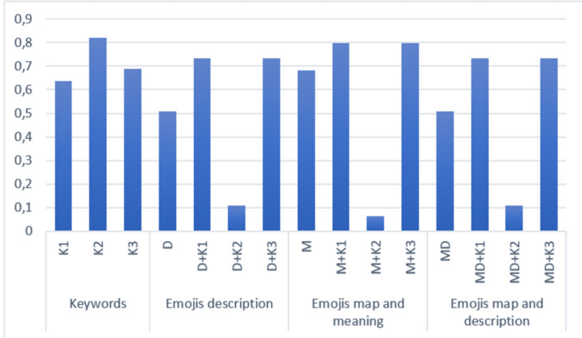

Fig.4: *Performance of techniques on the English Data.*

and discriminative emotion clusters, particularly in Arabic text.

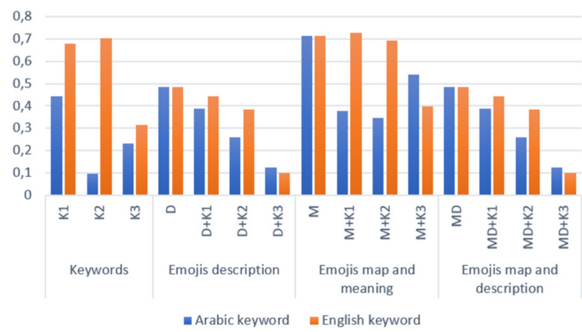

Fig.5: *Arabic Data Performance.*

4.2 Using different clustering algorithms

In an initial evaluation, the three clustering algorithms (K-means, GMM, and Agglomerative clustering) were applied to both the English and Arabic datasets using the best-performing configurations identified in Table 2. For the English dataset, the selected emoji representation technique is the keyword-based approach (K2), whereas for the Arabic dataset, using both Arabic and English keywords, the selected technique is the emoji map and meaning approach (denoted as M and M+K1, respectively). Clustering quality was evaluated using the Silhouette coefficient (S), the Calinski–Harabasz index (CH), and the Davies-Bouldin index (DB), with the results reported in Table 3. It presents the results obtained using different clustering algorithms with specified parameter settings to construct ensemble clusterings for both English and Arabic datasets. These algorithms were applied to the best-performing configurations identified across the various object representation mechanisms.

The initial parameters of the different algorithms are given by:
- *Kmeans clustering ensemble:* (n clusters=6, n splits=22) [KMeans(n clusters=6,int='k-means++', n init=10, random state=42)]

- Agglomerative clustering ensemble: (n clusters=6, n splits=22) [Agglomerative(n components=6,linkage = ‘ward’)].

Table 3: *Results of using different Clustering ensemble algorithms.*

Dataset	Measure	K-Means	Agglomerative	GMM
English Dataset	S	0.8201926	0.80711025	0.21689884
	CH	91745.62641078627	82095.06365364435	1293.232981595158
	DB	0.3754570366450628	0.4382128364154188	4.179389738257317
Arabic Dataset with Arabic Key words	S	0.713531	0.71850187	0.1593421
	CH	20809.897109448524	21486.400541401115	1649.8963204892182
	DB	0.45602117108059553	0.4750010235891053	2.3276551648394856
Arabic Dataset with English Key words		0.72585523	0.7281388	0.30612046
	CH	22248.16336927542	22183.87983856446	3000.356822271811
	DB	0.4537078290501955	0.4537509510534045	0.7349608995940897

- Gaussian Mixture Models clustering ensemble: (n clusters=6, n splits=22) GMM (n components=6, covariance type=‘diag’).

The comparative analysis of the clustering results presented in Table 3 shows that K-Means and Agglomerative clustering exhibit consistently strong performance across all datasets, with Silhouette scores ranging from approximately 0.72 to 0.82. In contrast, Gaussian Mixture Models display substantially weaker performance, suggesting their limited relevance for short-text emotion detection when applied as standalone methods. Consistent trends are observed for the Calinski-Harabasz and Davies-Bouldin indices, where K-Means and Agglomerative clustering yield comparable values indicative of well-separated clusters, whereas GMM records markedly inferior scores.

By using different clustering combinations, we identified the best-performing combination across different configurations and datasets.

The parameters for these ensembles are n clusters=6 and n splits=22. The following combinations are employed:

Ensemble K-Means, GMM, and Agglomerative clustering:
- GMM (n components=6, covariance type=‘diag’)
- KMeans(n clusters=6, init=‘k-means++’, random state=42)
- Agglomerative (n components=6, linkage=‘ward’)

Ensemble K-Means and Agglomerative clustering:
- KMeans (n clusters=6, init=‘kmeans++’, random state=42)
- Agglomerative (n components=6, linkage=‘ward’)

Ensemble K-Means and GMM clustering:
- GMM (n components=6, covariance type=‘diag’)
- KMeans (n clusters=6, init=‘k-means++’, random state=42)

Ensemble GMM and Agglomerative clustering:
- GMM (n components=6, covariance type=‘diag’)
- Agglomerative (n components=6, linkage=‘ward’)

Table 4: *Results of the evaluation of different Ensemble clustering combinations.*

Dataset	Measure	Ensemble of 3 Algorithmes	K-Means + Agglomerative	K-Means+ GMM	GMM+ Agglomerative
English Dataset	S	0.8082085	0.80820864	0.011694864	0.80820876
	CH	79735.94280219184	79736.03153231436	920.8410284863651	79736.50107745678
	DB	0.3684250721751218	0.36842434308936906	2.956858445295924	0.36842543818738943
Arabic Dataset with Arabic Key words	S	0.7000599	0.71850175	0.6988605	0.71850175
	CH	22634.541956694975	21486.342681856633	22778.816289394357	21486.346321423087
	DB	0.5526452547771405	0.47500135258218773	0.5543781892085458	0.47500127521381236
Arabic Dataset with English Key words	S	0.41997117	0.72813874	0.6819453	0.72813886
	CH	1307.139659600541	22183.864683043903	17757.426513964136	022183.87621494166
	DB	1.0342339813020565	0.4537510778891658	0.5445479014871696	0.45375091599260187

As shown in Table 4, the combination of K-means and Agglomerative clustering consistently demonstrates strong performance. Notably, the highest silhouette score observed (0.80820876) was achieved by the combination of Gaussian Mixture Models (GMM) and Agglomerative clustering. However, this improvement is only marginal compared to the result produced by the KMeans + Agglomerative ensemble, indicating that the overall performance of both methods is nearly identical across both English and Arabic datasets. However, ensembles combining GMM with K-Means or incorporating all three algorithms exhibited slightly lower scores, suggesting that GMM consistently introduces instability compared to purely partition-based or hierarchical approaches. Similar trends are observed for the Calinski-Harabasz index, where ensembles including K-Means and Agglomerative clustering yield high values across different datasets, particularly for English data, confirming strong cluster separation. In contrast, the K-Means + GMM combination produces very low values, especially for English data. Likewise, low Davies-Bouldin index values are maintained for ensembles involving K-Means + Agglomerative clustering, whereas K-Means + GMM exhibits high values, reflecting poor cluster compactness.

In summary, the K-Means + Agglomerative ensemble clustering outperforms alternative combinations across the majority of datasets and evaluation metrics, highlighting the strong reliability of this configuration for emotion annotation via ensemble clustering techniques.

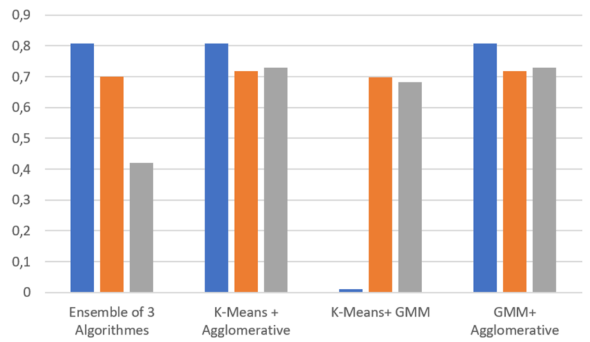

Fig.6: *Ensemble techniques Performance across different Datasets.*

Table 5: *Results of different Ensemble clustering algorithms' initialisations.*

Dataset	Measure	2 K-Means + GMM	2 Agglomerative + GMM
English Dataset	S	0.8202998	0.5521399
	CH	93012.11590289751	8362.180473101296
	DB	0.3782209000635028	0.8169632690347536
Arabic Dataset with Arabic Key words	S	0.7002942	0.7000601
	CH	22864.951720807945	22634.575392025527
	DB	0.5525221284687479	0.5526448834732988
Arabic Dataset with English Key words	S	0.6858503	0.42923573
	CH	18595.923876520785	1311.265848280646
	DB	0.5596238566122017	0.7645868761976568

4.3 In different parameter initialization

We performed a comprehensive examination of the results derived from applying diverse parameter initialization strategies within the ensemble clustering framework. These strategies were selectively implemented on the best-performing clustering approach identified in the evaluation. For all ensemble configurations, the experimental parameters were held constant, with six clusters and 22 splits. The following parameter settings were used:

Ensemble 2-KMeans and GMM

- KMeans (n clusters=6, init='k-means++', random state=42),
- GMM (n components=6, covariance type='diag'),
- KMeans (n clusters=6, init='center', random state=42).

Ensemble 2-Agglomeratives and GMM

- Agglomerative (n components=6, linkage='ward'),
- GMM (n components=6, covariance type='diag'),
- Agglomerative (n components=6, linkage='complete').

Based on Table 5, we observe that, for the English dataset, the 2K-Means + GMM ensemble achieves a substantially higher silhouette score (0.8202998) compared to the 2Agglomerative + GMM configuration, which yields a considerably lower value (0.5521399). This performance gap indicates that initializing the ensemble with K-Means partitions leads to more coherent and well-structured clusters in this setting. This observation is further confirmed by the Calinski-Harabasz index, which shows markedly higher values for the 2K-Means + GMM ensemble, reflecting superior cluster separation, whereas the 2Agglomerative + GMM configuration exhibits noticeably lower scores. In addition, the Davies-Bouldin index reinforces these findings: the 2K-Means + GMM ensemble maintains low values indicative of compact and well-separated clusters, while the 2Agglomerative + GMM ensemble produces significantly higher values, signaling inferior clustering quality.

In the Arabic dataset incorporating Arabic keywords, both ensemble techniques yielded nearly identical performance (0.7002942 vs. 0.7000601), indicating that under this condition, the choice of the initial clustering strategy has a limited impact on the outcome. The Calinski-Harabasz index exhibits slightly higher values for the 2 K-Means + GMM ensemble compared to the 2 Agglomerative + GMM configuration; however, the difference remains marginal. Similarly, the Davies-Bouldin index yields nearly identical values for both ensembles (0.553 vs. 0.553), indicating comparable levels of cluster compactness and separability.

Conversely, for the Arabic dataset incorporating English keywords, the 2 K-Means + GMM ensemble again outperforms the agglomerative-based configuration, achieving a substantially higher silhouette score (0.6858503 vs. 0.42923573). This result highlights the strong sensitivity of agglomerative initialization to keyword representation in this setting. Consistent with this observation, the Calinski-Harabasz index reveals a pronounced performance gap, with the 2 K-Means + GMM ensemble attaining markedly higher values, while the 2 Agglomerative + GMM dominated ensemble exhibits very poor performance. Similarly, the Davies-Bouldin index further supports these findings: the 2 K-Means + GMM configuration yields lower values, indicating more compact and well-separated clusters, whereas the 2 Agglomerative + GMM ensemble produces substantially higher values, reflecting inferior clustering quality. Overall, the results indicate that the 2 K-Means + GMM ensemble generally outperforms alternative configurations across evaluation metrics, especially for the English dataset and the Arabic dataset with English keyword mappings. The limited performance gap observed when using Arabic keywords suggests that semantic alignment between keyword representations and data distributions plays a role in determining ensemble clustering outcomes.

5. CONCLUSIONS

This study addresses the critical challenge of generating labelled datasets for emotion detection, with a particular focus on under-resourced languages and dialects such as Algerian Arabic, where annotation typically requires substantial time investment and considerable human effort. We propose and validate an unsupervised ensemble clustering approach that effectively generates high-quality, automatically labelled corpora for emotion analysis from Twitter text. Our findings demonstrate that combining multiple clustering algorithms, including K-Means, Gaussian Mixture Models, and Agglomerative Clustering, produces more robust and stable results than relying on individual methods. A key insight from this research is the critical importance of integrating emoji semantics with effective keyword handling strategies. Combinations of K-Means and Agglomerative clustering consistently demonstrated strong performance across datasets. In addition, the 2 K-Means + GMM ensemble exhibited robust performance, particularly on the English dataset (silhouette score of 0.8202998), highlighting the benefits of carefully designed initialization strategies. By leveraging algorithmic diversity alongside semantic mapping, the proposed methodology provides a scalable and language-adaptive solution that effectively bridges the gap between unstructured textual data and the requirements of supervised learning frameworks. This work makes a significant contribution to advancing emotion-aware systems by eliminating the reliance on resource-intensive manual annotation processes. Future research may extend this framework in several promising directions. First, investigating more advanced consensus functions and base-clustering generation mechanisms within the ensemble clustering framework could further enhance clustering quality and stability. Second, the integration of external knowledge resources, such as enriched emotion lexicons, may provide deeper semantic context and improve the detection of nuanced or implicit emotional expressions. Third, the use of automatically generated labeled datasets for training and evaluating supervised emotion detection models would facilitate a direct evaluation of the effect of the proposed unsupervised labelling strategy on downstream classification performance. Finally, extending the proposed approach to additional textual sources, such as user-generated comments on platforms like YouTube, represents a promising avenue for increasing its applicability across heterogeneous and large-scale social media environments.

AUTHOR CONTRIBUTIONS

Conceptualization, W.S.; methodology, W.S.; software, W.S., M.M.; validation, W.S., F.Z.L. and M.M.; formal analysis, W.S.; investigation, W.S.; data curation, W.S.; writing—original draft preparation, W.S.; writing—review and editing, W.S., F.Z.L. and M.M.; visualization, W.S. and M.M..; supervision, , F.Z.L.; All authors have read and agreed to the published version of the manuscript.

References

[1] V. Ahire and S. Borse, "Emotion Detection from Social Media Using Machine Learning Techniques: A Survey," in *Applied Information Processing Systems*, Springer Singapore, pp. 83–92, 2022.

[2] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu and V. Varadarajan, "AI Based Emotion Detection for Textual Big Data: Techniques and Contribution," *Big Data and Cognitive Computing*, vol. 5, no. 3, p. 43, Sep. 2021.

[3] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1–2, pp. 55–64, Jul. 2003.

[4] P. Ekman, "Are there basic emotions?," *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992.

[5] B. Batrinca and P. C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms," *AI & SOCIETY*, vol. 30, no. 1, pp. 89–116, Feb. 2015.

[6] A. Seyeditabari, N. Tabari, and W. Zadrozny, "Emotion detection in text: A review," *arXiv preprint* arXiv:1806.00674, 2018.

[7] T. Alqurashi and W. Wang, "Clustering ensemble method," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 6, pp. 1227–1246, Jun. 2019.

[8] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[9] A. Strehl and J. Ghosh, "Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions," *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2002.

[10] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 03, pp. 337–372, May 2011.

[11] A. Agrawal and A. An, "Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations," *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Macau, China, pp. 346-353, 2012.

[12] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora and I. Pappas, "A review on text-based emotion detection—Techniques, applications, datasets, and future directions," *arXiv preprint* arXiv:2205.03235, 2022.

[13] S. Al-Saqqa, H. Abdel-Nabi and A. Awajan, "A Survey of Textual Emotion Detection," *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, Amman, Jordan, pp. 136-142, 2018.

[14] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recogni-

tion in text," *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2937–2987, Aug. 2020.

[15] A. A. Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha and Z. Aung, "Challenges and Opportunities of Text-Based Emotion Detection: A Survey," in *IEEE Access*, vol. 12, pp. 18416-18450, 2024.

[16] R. B. Patel, "A survey of Emotion Detection", *TechRxiv Preprints* techrxiv.174285087.74645354, 2025.

[17] M. Zidan, I. Elhenawy, A. R. Abas and M. Othman, "Textual emotion detection approaches: A survey," *Future Computing and Informatics Journal*, vol. 7, no. 1, pp. 32–58, Jun. 2022.

[18] A. Bandhakavi, N. Wiratunga, S. Massie and D. Padmanabhan, "Lexicon Generation for Emotion Detection from Text," in *IEEE Intelligent Systems*, vol. 32, no. 1, pp. 102-108, Jan.-Feb. 2017.

[19] S. Y. M. Lee, "A Linguistic Analysis of Implicit Emotions," in *Lu, Q., Gao, H. (eds) Chinese Lexical Semantics. CLSW 2015. Lecture Notes in Computer Science*, vol 9332. Springer, Cham, 2015.

[20] S. N. Shivhare, S. Garg and A. Mishra, "EmotionFinder: Detecting emotion from blogs and textual documents," *International Conference on Computing, Communication & Automation*, Greater Noida, India, pp. 52-57, 2015.

[21] J. Tao, "Context based emotion detection from text input," in *INTERSPEECH 2004 - ICSLP, 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, pp. 1337–1340, Oct. 2004.

[22] Z. Dong and Q. Dong, "HowNet - a hybrid language and knowledge resource," *International Conference on Natural Language Processing and Knowledge Engineering*, 2003. Proceedings. 2003, Beijing, China, pp. 820-824, 2003.

[23] A. Álvarez Núñez, M. D. C. Santiago Díaz, A. C. Zenteno Vázquez, J. Pérez Marcial and G. T. Rubín Linares, "Emotion detection using natural language processing," *The international Journal of Combinatorial Optimization Problems and Informatics*, vol. 15, no. 5, pp. 108–114, Nov. 2024.

[24] M. Kumar, "Emotion Recognition in Natural Language Processing: Understanding How AI Interprets the Emotional Tone of Text," *Journal of Artificial Intelligence & Cloud Computing*, pp. 1–5, Dec. 2024.

[25] S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.

[26] C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.

[27] C. C. Aggarwal, *Machine Learning for Text*. Cham: Springer International Publishing, 2022.

[28] P. Hegde, S. B. Mukkamala, and A. P. Jyothi, "Intelligent Techniques for Emotion Recognition in Social Media and Business Applications," in *Advances in Marketing, Customer Relationship Management, and E-Services*, IGI Global, pp. 219–236, 2024.

[29] B. D. Jasvitha, K. Kanagaraj, K. Murali and M. Venugopalan, "Emotion Detection from Text Using ML Framework," in *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*, Shivamogga, India: IEEE, pp. 1–6, May 2024.

[30] O. Kaminska, C. Cornelis, and V. Hoste, "Nearest neighbour approaches for emotion detection in tweets," *arXiv preprint* arXiv:2107.05394, 2021

[31] K. Machová, M. Szabóova, J. Paralič and J. Mičko, "Detection of emotion by text analysis using machine learning," *Frontiers in Psychology*, vol. 14, p. 1190326, Sep. 2023.

[32] I. Perikos and I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," *Engineering Applications of Artificial Intelligence*, vol. 51, pp. 191–201, May 2016.

[33] V. Maheshkar and S. K. Sarin, "Review and Analysis of Emotion Detection from Tweets using Twitter Datasets," in *Proceedings of the Workshop on Applied Computing (WAC-2022)*Á, Chennai, India, Jan. 2022.

[34] K. Shrivastava, S. Kumar and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," *Multimedia Tools and Applications* , vol. 78, no. 20, pp. 29607–29639, Oct. 2019.

[35] N. S. Kumar, M. Amencherla and M. G. Vimal, "Emotion Recognition in Sentences - A Recurrent Neural Network Approach," *Computational Intelligence in Data Science*, vol. 578, pp. 3–15, 2020.

[36] F. A. Acheampong, H. Nunoo-Mensah and W. Chen, "Transformer models for text-based emotion detection: a review of BERT-based approaches," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5789–5829, Dec. 2021.

[37] A. Chaves-Villota, A. Jimenez and A. Bahillo, "UAH-UVA in EmoSpeech-IberLEF2024: A Transfer Learning Approach for Emotion Recognition in Spanish Texts based on a Pretrained DistilBERT Model," in *Proceedings of the Iberian Languages Evaluation Forum, co-located with the Conference of the Spanish Society for Natural Language Processing*, SEPLN 2024, 2024.

[38] N. A. P. Masaling, R. R. Siswanto and A. S. Girsang, "Indonesian Tweet Emotion Detection Using IndoBERT," *2024 International Conference on Information Management and Technol-*

ogy (ICIMTech), Bali, Indonesia, pp. 478-482, 2024.

[39] W. Amelia and N. U. Maulidevi, "Dominant emotion recognition in short story using keyword spotting technique and learning-based method," *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, Penang, Malaysia, pp. 1-6, 2016.

[40] M. H. F. F, V. R. T and S. Lakumarapu, "Integrating Supervised and Probabilistic Learning: A Hybrid SVM-Naive Bayes Framework for Enhanced Emotional Insight Extraction from Twitter Streams," *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, pp. 2116-2121, 2024.

[41] G. Badaro, H. Jundi, H. Hajj, W. El-Hajj and N. Habash, "ArSEL: A Large Scale Arabic Sentiment and Emotion Lexicon," in *Proceedings of the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT).* Miyazaki, Japan: European Language Resources Association (ELRA), Apr. 2018.

[42] S. Hassan, S. Shaar and K. Darwish, "Cross-lingual Emotion Detection," *presented at the International Conference on Language Resources and Evaluation*, arXiv, 2021. [Online]. Available: `https://api.semanticscholar.org/CorpusID:235417255`

[43] M. J. Althobaiti, "An open-source dataset for arabic fine-grained emotion recognition of online content amid COVID-19 pandemic," *Data in Brief*, vol. 51, p. 109745, Dec. 2023.

[44] R. AlYami and R. Al-Zaidy, "Weakly and Semi-Supervised Learning for Arabic Text Classification using Monodialectal Language Models," in *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, pp. 260–272, 2022.

[45] M. Abdullah, M. Hadzikadicy and S. Shaikhz, "SEDAT: Sentiment and Emotion Detection in Arabic Text Using CNN-LSTM Deep Learning," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, 2018, pp. 835-840,

[46] M. J. Althobaiti, "ArPanEmo: An open-source dataset for fine-grained emotion recognition in Arabic online content during COVID-19 pandemic," *arXiv preprint* arXiv:2305.17580, 2023.

[47] N. Alsadhan, "A Novel Dialect-Aware Framework for the Classification of Arabic Dialects and Emotions," *Journal of Computer Science*, vol. 21, no. 1, pp. 88–95, Jan. 2025.

[48] W. Saadi, F. Z. Laallam, M. Mezati, D. L. Youmbai and N. E. Messaoudi, "Enhancing emotion detection on Twitter: an ensemble clustering approach utilizing emojis and keywords across multilingual datasets," *Studies in Engineering and Exact Sciences*, vol. 5, no. 2, p. e10548, Nov. 2024.

[49] F. Di Martino, S. Senatore and S. Sessa, "A lightweight clustering–based approach to discover different emotional shades from social message streams," *International Journal of Intelligent Systems*, vol. 34, no. 7, pp. 1505–1523, Jul. 2019.

[50] B. Cardone, F. Di Martino and S. Senatore, "Improving the emotion-based classification by exploiting the fuzzy entropy in FCM clustering," *International Journal of Intelligent Systems*, vol. 36, no. 11, pp. 6944–6967, Nov. 2021.

[51] C. C. Sujadi, Y. Sibaroni and A. F. Ihsan, "Analysis Content Type and Emotion of the Presidential Election Users Tweets using Agglomerative Hierarchical Clustering," *Sinkron*, vol. 7, no. 3, pp. 1230–1237, Jul. 2023.

[52] M. T. AL-Sharuee, F. Liu and M. Pratama, "Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison," *Data & Knowledge Engineering*, vol. 115, pp. 194–213, May 2018.

Wafa Saadi received the Diplôme d'Étude Universitaire Appliquée (DEUA) in Computer Science from the University of Mohamed Khider Biskra in 2000, the State Engineer degree in 2003, the Magister degree in 2009 from the same institution, and currently pursuing her PhD in the same University. She is an Assistant Professor in the Department of Computer Science and Information Technology at the University of Kasdi Merbah, Ouargla, Algeria. Her research interests include Artificial Intelligence and Big Data analytics. She has contributed to the academic community through various publications and presentations, focusing on emotion recognition artificial intelligence, Natural Language Processing, and Social media analysis.

Fatima Zahra Laallam received her State Engineering degree in Computer Science from the University of Annaba in 1991, specializing in Management Information Systems. She is currently a Professor of Computer Science at the University of Kasdi Merbah, Ouargla, Algeria, with expertise in Artificial Intelligence. She is widely recognized as a pioneer in southern Algeria, being the first female computer engineer, the first woman in the region to earn a doctorate in the field, and the first to attain the rank of professor in Computer Science. Her professional journey began at the National Electricity and Gas Company as an IT study engineer, where she developed long-lasting management applications. Upon joining the University of Ouargla, she became one of the founding members of the Computer Science Department and later held several key administrative positions, including serving as the first Dean of the Faculty of New Information and Communication Technologies.

She is also the founder of a research laboratory dedicated to artificial intelligence and information technologies and has been continuously active in scientific committees and academic councils. She currently serves as Chair of the Scientific Committee of the Department of Computer Science, a member of the Scientific Council of the Faculty of NICT, and a member of the National Council for Scientific and Technological Research. Her scientific contributions are substantial, encompassing leadership in major research projects, participation as an expert evaluator at national and international levels, and the publication of numerous internationally recognized works. She has received numerous awards, most recently winning first place nationally for the best doctoral thesis in 2021. Her research spans databases, big data analytics, ontologies, IoT, energy optimization, multi-agent systems, and knowledge representation, with a current focus on applying artificial intelligence to enhance educational systems.

Messaoud Mezati received the Diplôme d'Étude Universitaire Appliquée (DEUA) in Computer Science from the University of Biskra in 2002, the State Engineer degree in 2005, the Magister degree in 2008, and the Ph.D. in Computer Science in 2017 from the same institution. Since 2017, he has been a Maître de Conferences at the Department of Computer Science at the University of Kasdi Merbah, Ouargla, Algeria. His research interests include behavioral simulation, image synthesis, virtual reality, artificial life, and machine learning. He has authored multiple scientific publications on topics such as machine learning, emotion detection, clustering algorithms, and semantic representation of virtual humans. Dr. Mezati serves as a Member of the Editorial Board for the International Journal of Artificial Intelligence & Applications (IJAIA) and is a Program Committee Member for Computer Science & Information Technology as CNSA 2018.