

MASHA: A Multi-Agent System for Healthcare Sentiment Analysis Using AI for Migraine Detection in Arabic Tweets

¹Shadia Yahya Baroud

¹Department of Computer and Information Sciences,
Faculty of Computers and Information Technology, Al-Aqsa University, Gaza, Palestine, 4051

¹shadia.baroud@gmail.com

*Corresponding author

Abstract

Migraine detection and sentiment analysis in healthcare have become increasingly important, particularly with the rise of social media platforms like Twitter, where users often share their personal health experiences. This study presents MASHA (Multi-Agent System for Healthcare Sentiment Analysis), an artificial intelligence (AI)-driven framework that integrates multiple machine learning (ML) models for sentiment analysis of Arabic tweets related to migraines. The system leverages a multi-agent architecture to handle tasks such as data acquisition, pre-processing, model training and real-time decision-making. Key ML models, including Support Vector Machines (SVM), Naive Bayes (NB) and Logistic Regression (LR), are integrated using ensemble techniques, leading to improved classification performance.

Experiments conducted on a dataset of Arabic tweets demonstrate that MASHA outperforms traditional methods, achieving an accuracy of 90.0% and an F1-score of 89.46%. Moreover, the system's scalability and flexibility make it suitable for real-time public health monitoring, offering valuable insights into patient experiences and public sentiment regarding healthcare services.

MASHA's adaptability suggests its potential application for analysing other healthcare-related conditions, reinforcing the system's scalability and broader relevance. Future work will focus on incorporating deep learning (DL) models and expanding the dataset with content from additional social media platforms.

Keywords: Artificial Intelligence (AI), Multi-Agent Systems (MAS), Sentiment Analysis (SA), Natural Language Processing (NLP), Machine Learning (ML), Healthcare, Migraine Detection, Arabic Tweets, Social Media Analysis

1. Introduction

The integration of machine learning (ML) models is crucial in improving the performance of complex applications, particularly in the field of Natural Language Processing (NLP) (1). NLP involves the interaction between computers and human language, enabling machines to understand, interpret and generate human language (2–4). The effective integration of ML models in NLP is vital for managing and interpreting vast amounts of unstructured text data. However, traditional methods often face interoperability, scalability and efficiency challenges, limiting their effectiveness in real-world scenarios (5).

Healthcare is a critical global concern and headaches, particularly migraines, are recognized as one of the most common and disabling conditions worldwide (6,7). The World Health Organization (WHO) reports that nearly half of the global population experiences some form of headache, including tension-type headaches, cluster headaches and migraines (8). Migraines, often episodic, significantly impact individuals' quality of life and create a substantial economic burden on society (9).

With the rise of social media platforms such as Twitter, individuals increasingly share their personal experiences, including their health struggles (10,11). This presents a unique opportunity to analyze real-time data on conditions like migraines, providing insights into behavioural patterns and emotional responses. Textual data from Twitter can capture individuals' perceptions of migraine attacks, offering valuable information that can supplement existing medical knowledge and potentially improve predictive capabilities in public health (12).

While existing studies have explored various aspects of migraine epidemiology in the Arab world, there is a need to synthesize these findings, mainly using modern technologies such as SA and ML (13–16). Twitter data, especially in Saudi Arabia, where there is a significant number of active users, provides an opportunity to explore how individuals discuss their migraine experiences (17). However, extracting useful information from unregulated platforms like Twitter poses significant challenges, particularly in filtering relevant tweets and pre-processing Arabic language data (5,18,19). Moreover, current ML models often face difficulties integrating diverse data sources and handling large-scale datasets, limiting their scalability and flexibility (20–22).

Existing SA studies in healthcare often face challenges in effectively analyzing social media data due to the diverse nature of Arabic dialects, the volume of unstructured text and the

dynamic nature of social discussions. While approaches such as MARBERT and deep neural networks (DNNs) have shown success in structured datasets, the scalability and interoperability of these models remain limited when applied to real-world, large-scale, unstructured data environments like Twitter (23).

Most frameworks focus solely on individual models, overlooking the potential benefits of multi-agent-based frameworks that coordinate and integrate diverse ML models (24–27). This research bridges the gap by proposing MASHA (Multi-Agent System for Healthcare Sentiment Analysis), a novel multi-agent-based SA system that addresses the volume and complexity of Arabic healthcare-related tweets, specifically targeting migraine-related discussions.

This study presents the following contributions:

1. **Framework Development:** Introduction of **MASHA**, a novel Multi-Agent System (MAS) designed to manage the entire sentiment analysis workflow, from data collection and pre-processing to model training and decision-making.
2. **Scalability and Flexibility:** Integrating multiple ML models (support vector machines (SVM), Naive Bayes (NB) and Logistic Regression (LR)) through ensemble learning within a multi-agent framework to improve sentiment classification in resource-limited healthcare environments.
3. **Case Study Validation:** The proposed framework was applied to real-world Arabic tweets related to migraines, demonstrating improved accuracy, precision, recall and F1-score compared to traditional single-model approaches.
4. **Addressing Dialectal Complexity:** Implement customized pre-processing techniques to handle Arabic dialect variations, enhancing text normalization and feature extraction processes.

MASs involve multiple autonomous agents collaborating to achieve specific goals. In the context of ML, MAS facilitates the integration of diverse models by allowing agents to manage different aspects of the process (28). Prior research has demonstrated the potential of MAS in improving the efficiency and scalability of ML model integration (29–31).

Ensemble learning and hybrid models have traditionally been used to combine the strengths of different ML models, improving overall performance (32–34). However, these methods often struggle with scalability and flexibility in dynamic environments. In NLP applications, ensemble methods face challenges in coordination and efficiency during integration (32).

97 Despite advancements in ML model integration, existing methods struggle with handling
98 large-scale and dynamic data environments, particularly in NLP applications (5). This study
99 addresses these gaps by proposing a flexible and scalable multi-agent-based framework for
100 integrating diverse ML models.

101 Recent advancements in SA and ML for healthcare applications have demonstrated the
102 potential of analyzing patient experiences through social media data (35). AlMuhaideb et al.
103 (2023), introduced a novel HoPE-SA dataset that gathers patient experience sentiments from
104 Saudi Arabia's healthcare services using Arabic Twitter data. The study fine-tuned
105 transformer-based models, such as MARBERT, achieving high accuracy in sentiment
106 classification. This research underscores the growing role of social media SA in
107 understanding patient experiences, particularly in healthcare contexts (36).

108 Similarly, Khan et al. (2024) applied various ML models, including SVM, decision trees (DT)
109 and DNNs, to classify different types of migraines using patient data. The study employed
110 data augmentation techniques to expand the dataset, significantly improving the model's
111 accuracy. Notably, the DNN model achieved an accuracy of 99.66%, emphasizing the
112 potential of ML in diagnosing and predicting conditions such as migraines (15). These studies
113 highlight the importance of leveraging advanced ML models and SA to improve the accuracy
114 and scalability of healthcare-related NLP tasks. This aligns with the focus of this research on
115 developing a multi-agent-based framework for SA of Arabic tweets related to migraine
116 detection.

117 In summary, integrating ML models in NLP is a complex task that requires robust
118 frameworks to manage diverse data and model types effectively (37). This study proposes
119 **MASHA**, a multi-agent-based framework that addresses the challenges of ML models
120 integration. A case study on sentiment classification for migraine detection using Arabic
121 tweets is used to demonstrate the effectiveness of this framework, which enhances both
122 performance and efficiency in integrating diverse models.

123 The remainder of this study is structured as follows: Section 2 reviews the related work,
124 focusing on previous studies involving MAS, ML integration in NLP and SA, particularly in
125 the context of Arabic language and healthcare data. Section 3 outlines the methodology to
126 develop the multi-agent-based framework for sentiment classification of migraine-related
127 tweets, including data collection, filtering and integrating diverse ML models. Section 4
128 presents the results of the experiments and discusses the proposed framework's performance

using key metrics such as accuracy, precision, recall and F1-score. Finally, Section 5 concludes the study, summarizing the key contributions and suggesting potential avenues for future research.

2. Related Work

The integration of ML models into various applications has been extensively researched, particularly in the domain of NLP. However, there is room for improvement in healthcare-related tasks like SA of social media data for public health monitoring. This section explores the existing MAS, SA and ML model integration work, specifically in NLP and healthcare domains.

2.1 Multi-Agent Systems (MAS)

MASs are composed of multiple autonomous agents interacting to achieve common or individual goals (38). In ML contexts, MAS has shown great potential in managing complex processes by distributing tasks among agents. In their study (39), they laid the foundation for MAS by defining key concepts and applications, which have since expanded to various domains such as robotics, distributed computing and decision-making. MAS is particularly advantageous for integrating multiple ML models, as agents can manage different aspects of the integration process, improving system efficiency and scalability.

The key characteristics of agents in MAS, autonomy, reactivity, pro-activeness and social ability, enable them to operate independently, respond to changes, take initiative and collaborate effectively (28). These traits are essential for efficiently integrating ML models in the MASHA framework for SA and healthcare applications. Figure 1 visualizes these core characteristics, illustrating how agents' attributes support effective collaboration and task execution within the system.

Figure 1 Core Characteristics of Agents in MAS

In recent years, MAS has been employed to address challenges in data processing and integration of models in real-time systems (29). For example, MAS has been used in medical applications to diagnose and process patient data, by allowing different agents to handle diverse tasks such as data pre-processing, model development and model integration, MAS can significantly reduce bottlenecks and improve performance, especially when managing large datasets or complex data streams (41). Given these strengths, applying MAS in NLP tasks such as SA can significantly enhance system flexibility and scalability.

2.2 Sentiment Analysis (SA)

Sentiment Analysis (SA) is a popular approach in NLP that classifies text based on individuals' emotions, opinions or sentiments. SA has been widely applied across industries to gauge consumer feedback, track public opinion and monitor trends. In healthcare, SA can be particularly valuable for monitoring patient experiences and disease-related discussions on social media platforms (42).

Arabic sentiment analysis (ASA), in particular, presents unique challenges due to the language's rich morphological structure and the prevalence of dialects in social media conversations (43). Early research on SA in Arabic faced significant difficulties due to the lack of resources such as sentiment lexicons and annotated corpora (44). Also (45), provided a comprehensive overview of SA techniques while (46), discussed the importance of combining ML with NLP to enhance the accuracy of Arabic text classification.

Recent studies have explored deep learning (DL) techniques for ASA (47). The authors (48) surveyed the latest advances in this field, noting an increasing trend toward DNN for feature extraction and sentiment classification. However, these studies often focus on general-purpose applications rather than domain-specific tasks like healthcare. This gap highlights the need for further research on applying SA techniques to specific problems like migraine detection.

2.3 ML Integration in NLP

Extensive research has been done on integrating multiple ML models, especially in ensemble learning techniques like bagging, boosting and stacking (32,49). These methods combine different models' strengths to enhance prediction accuracy and robustness. However, despite their success, traditional ensemble methods face challenges when applied to large-scale and dynamic environments, such as social media data in healthcare contexts (50).

In NLP, ensemble learning methods have been used to combine the predictive power of different algorithms, particularly for tasks like SA, text classification and machine translation (51,52). However, these approaches often have inefficiencies in coordinating multiple models, particularly when handling heterogeneous data types and formats (53). The complexity of coordinating ML models increases when dealing with healthcare-related data, as the accuracy and timeliness of insights are crucial for improving patient outcomes (35).

Several studies have proposed hybrid models that combine ML with rule-based systems to handle complex NLP tasks better (54). However, these models often lack the flexibility and

scalability to process real-time data or handle dynamic input like social media streams. As a result, there is a growing interest in using MASs to manage the integration of ML models in such tasks. By distributing the workload across different agents, MAS can help address coordination, scalability and real-time processing issues.

2.4 Gaps in Current Research

Despite advances in SA, ML integration and MAS, current research faces several limitations when applied to healthcare-related NLP tasks like migraine detection. Traditional ML model integration methods often struggle with the sheer volume and variety of social media data, especially when processing real-time inputs from platforms like Twitter (42). Moreover, there is a shortage of high-quality Arabic datasets for NLP tasks, further complicating the development of effective sentiment classification models for Arabic-speaking populations (46).

Current SA models, particularly in Arabic, still face challenges in handling dialect variations, the lack of standardized lexicons and the complexity of Arabic morphology (44). Additionally, existing multi-agent frameworks, while promising, have yet to be fully exploited for healthcare-related SA. Specifically, their application in NLP tasks like migraine detection using Arabic tweets remains largely unexplored.

These gaps underscore the need for more flexible, scalable and context-aware frameworks that integrate diverse ML models to handle real-time healthcare data in dynamic environments. This research proposes a multi-agent-based framework to address these challenges, leveraging the strengths of MAS for efficient ML model integration in NLP tasks related to public health monitoring.

3. Methodology

This section describes the proposed **MASHA**, a multi-agent-based framework for integrating ML models in the context of SA for migraine detection using Arabic tweets. **MASHA** is designed to be scalable, flexible and efficient, leveraging multiple agents to handle different aspects of data acquisition, pre-processing, model development and decision-making. The architecture consists of three key layers: the Data Pre-processing Agents Layer, the Model Development Agents Layer, and the Decision-Making Agents Layer. Each agent in the **MASHA** framework is responsible for specific tasks to ensure the overall efficiency and performance of the system, enabling seamless integration of ML models and effective SA in real-time.

3.1 Overview of the Multi-Agent Framework

The proposed multi-agent framework consists of several autonomous agents, each responsible for specific tasks in the sentiment classification process. Figure 2 illustrates the overview of the **MASHA** framework architecture.

Figure 2 Overview of the MASHA Framework Architecture

3.1.1 Data Acquisition and Pre-processing Agents Layer

The data pre-processing layer handles all tasks related to collecting and preparing data for further analysis. This includes the following components:

- **Data Acquisition Agent:** This agent is responsible for gathering relevant tweets from Twitter. It tracks specific hashtags related to migraine and healthcare, ensuring the dataset contains real-time and historical tweets from Arabic-speaking users. The agent interacts with the Twitter API to continuously update the dataset. Figure 3 displays a word cloud generated from the dataset, highlighting the most frequently mentioned terms related to migraines.
- **Data Storage and Analytics:** After data is acquired, it is stored in a cloud-based system that supports real-time processing and ensures data security. Cloud storage enables the system to scale dynamically as more data is collected. The data analytics module prepares the raw data for the pre-processing pipeline.
- **Data Pre-processing Agent:** This agent performs essential tasks such as:
 - **Data Cleaning:** Removing irrelevant information such as URLs, mentions, non-Arabic characters and special symbols.
 - **Data Transformation:** Converting data into a consistent format that subsequent modules can process.
 - **Data Normalization:** Standardizing the text by handling variations in spelling, repeated letters and dialect-specific expressions in Arabic.
 - **Data Validation and Integration:** Ensuring the collected data is relevant and accurate and then integrating it into the system for further processing.
 - **Data Annotation:** Human annotators label the data, classifying tweets as positive or negative. This creates a gold standard dataset for model development.

- **Feature Extraction:** The system applies Term Frequency-Inverse Document Frequency (TF-IDF) vectorization and word embeddings to convert raw text into feature vectors suitable for ML models. Specifically, we applied AraVec, a pre-trained Arabic word embedding model, to generate dense vector representations of words. These embeddings capture semantic similarities and contextual relationships between words, enhancing the sentiment classification accuracy for Arabic tweets.

Figure 3 Word Cloud of Migraine-Related Terms in Arabic Tweets. The word cloud displays Arabic terms related to migraines. The most prominent terms include: الشقيقة (migraine), الصداع النصفي (migraine headache), ألم الشقيقة (migraine pain), داء الشقيقة (migraine disease), صداع نصفي (half-sided headache) and مرض الشقيقة (migraine illness).

3.1.2 Model Development Agents Layer

ML models are selected, trained, validated and deployed in the model development layer. The components of this layer include:

- **Model Selection Agent:** This agent chooses appropriate ML algorithms for the sentiment classification task. The framework employs several ML models, including SVM, NB and LR. The selection is based on the nature of the data and the classification task. The algorithms used in the framework are as follows:
 - **SVM:** obtains an optimal hyperplane that separates classes with maximum margin. The main steps in SVM are:
 - Calculate the decision boundary by solving the optimization problem.
 - Apply the kernel trick (linear, polynomial or RBF) for non-linear data if required.
 - Classify the sentiment based on the hyperplane's side.
 - **NB:** is based on Bayes' theorem assuming independence among features. It calculates the probability of each sentiment class C given the features X :

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

The class with the highest probability is chosen.

- **LR:** is used to estimate probabilities using the sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

Where $z = w^T X + b$. The model predicts positive or negative sentiment based on whether $\sigma(z)$ is above a threshold (usually 0.5).

- **Model Development Agent:** After selecting the models, the training agent uses the annotated and pre-processed data to train each model. The training process includes splitting the dataset into training and testing sets to evaluate the model's performance and avoid overfitting.
- **Model Validation Agent:** The trained models are validated against a separate validation set. The validation agent assesses each model's accuracy, precision, recall and F1-score to determine its effectiveness in classifying sentiment from migraine-related tweets. The formulas used for these metrics are as follows:

- **Accuracy:**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

where TP is True Positives, TN is True Negatives, FP is False Positives and FN is False Negatives.

- **Precision:**

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

- **Recall:**

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

- **F1-score:**

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

- **Model Deployment and Integration Agent:** Once models are validated, they are deployed into the system. This agent handles the integration of multiple models through ensemble techniques such as voting or stacking, improving overall classification accuracy by combining the strengths of each model.

3.1.3 Decision-Making Agents Layer

The decision-making layer coordinates the predictions and outputs generated by the model development layer, allowing the system to deliver actionable insights based on the SA of tweets. It consists of the following components:

- **Model Evaluation Agent:** This agent is responsible for continuously evaluating the integrated model's performance. It monitors the ensemble model's predictions and

continuously calculates performance metrics such as accuracy, precision, recall and F1-score to ensure the system operates optimally.

- **Decision-Making Agent:** This agent oversees the overall decision-making process based on the evaluated predictions. It coordinates the activities of all other agents and generates final predictions visualized on a control panel for stakeholders. The system also allows for error reporting, making it possible to identify and address issues in real-time.
- **Data Visualization and Reporting:** This module visualizes the predictions and trends identified by the ML models. It enables stakeholders to gain insights into the general sentiment about migraines and provides a control panel for monitoring system performance. The reporting module offers transparency into decisions and provides error reporting to alert administrators of any issues.

3.2 Dataset

3.2.1 Data Collection

The dataset used in this study was collected from Twitter, focusing on Arabic tweets related to migraines. The Twitter API was utilized to retrieve tweets in real-time and historical data based on predefined hashtags and keywords. The primary hashtags used for data collection included #صداع نصفي (#migraine), #صداع (#headache), #صداع الشقيقة (#migraine) and other health-related terms frequently associated with migraine discussions. The tweets were collected from March to May 2021, and only tweets written in Arabic were included in the dataset.

To ensure relevance, tweets were filtered based on geo-tags and user profile information to select content related to Saudi Arabia. The ease of use, availability and accuracy of the Twitter API made it ideal choice for data collection. We applied for a developer account to access the API and created an application to obtain the necessary OAuth parameters: “consumer key”, “consumer secret”, “access token” and “access secret”.

Python programming languages, related software packages and Microsoft Excel were used to collect tweets, preprocess them and build the annotated dataset. Verified users and accounts with significant activity were prioritized to filter irrelevant content, improving the quality of the collected data.

3.2.2 Data Characteristics

The dataset consisted of approximately 14,000 Arabic tweets. Each tweet was accompanied by metadata, including the tweet's timestamp, user information, location and the specific hashtag or keyword used. The collected data included real-time and historical tweets, providing a comprehensive view of how users discuss migraines over time. The geographic focus was on users from Saudi Arabia, where Twitter usage is exceptionally high and healthcare discussions are active. Figure 4 provides a snapshot of this study's sentiment classification dataset. The dataset consists of Arabic tweets related to migraines, annotated for sentiment as either positive or negative. This sample highlights the variety of textual data processed, showcasing the different structures and expressions found in the collected tweets, including the dialectical variations that pose a challenge for classification models.

Figure 4 Sample of the dataset

The dataset sample in Figure 4 shows Arabic tweets with their sentiment labels. English translations are provided below:

- Row 3: "RT @phSarah9A: 🧠 Types of headaches: 🧠 cluster headache - cluster 🧠 half-sided headache - migraine 🧠 frontal sinus headache - sinus 🧠 tension headache caused by stress - tension ht..." - (Positive)
- Row 4: "Half-sided headache is beautiful 🧠" - (Negative)
- Row 5: "I have half-sided migraine headache @XXIIIXVII" - (Positive)
- Row 6: "I'm cured, thank God, of half-sided headache" - (Positive)
- Row 7: "I'm really suffering from half-sided headache but alone and alone and alone lol 🧠🧠🧠 @amjad_awad99 @glewo_01" - (Negative)
- Row 8: "@CelebsArabicII Half-sided headache from envy <https://t.co/IVEYqOlKK6>" - (Negative)
- Row 9: "@AsfourSeham Expect pockets or half-sided headache even if he knows you 🧠🧠🧠" - (Negative)
- Row 10: "@InozukeT 🧠🧠♀️🧠? You or anyone else have chronic half-sided headache" - (Negative)
- Row 11: "@iKhaledM1 @Gladd Half-sided headache, what's its name, oh my God you're on my mind! I wish you all the best, may your eyes see the light of a perfect day" - (Negative)
- Row 12: "I sat with my brothers with half-sided headache" - (Positive)

□ Row 13: “It's a headache, not half-sided” - (Negative)

3.2.3 Pre-processing

Given the unstructured nature of Twitter data and the linguistic complexities of Arabic, extensive preprocessing was performed on the collected dataset. The following preprocessing steps were applied:

- **Text Normalization:** The normalization process involved removing diacritics, handling repeated characters and standardizing the spelling of common words related to migraines. Variants such as “الشقيقة” (migraine) and “شقيقة” (migraine - with repeated letters) were normalized to a single standard form to enhance consistency and model accuracy. Furthermore, mentions, unrelated hashtag, punctuation, URLs and non-Arabic words were removed to reduce noise. However, relevant hashtags (e.g., #صداع_نصفي (#migraine_headache)) were retained to capture crucial context related to migraines.
- **Data Validation and Integration:** Involves ensuring that the collected data is relevant, accurate and free from duplicates. Validation checks include confirming that tweets contain relevant migraine-related content by verifying specific keywords and hashtags. Duplicate tweets, bot-generated spam and unrelated texts are filtered out during this process. Integration involves merging real-time and historical data streams into a unified dataset, ensuring consistency in formats such as timestamps, text encoding and metadata fields. This step ensures that the final dataset is cohesive and ready for subsequent pre-processing and analysis.
- **Tokenization:** Each tweet was broken down into individual tokens (words) to allow for feature extraction.
- **Stemming and Lemmatization:** The “Arabic Stanford Lemmatizer” was used to standardize the text by reducing words to their base forms. This helped address the morphological complexity of the Arabic language, which is essential for handling various forms of the same word. Moreover, custom rules were applied to handle common dialectal variations and spelling inconsistencies. This confirmed that feature extraction was more consistent, leading to improved classification accuracy.
- **Stop-Word Removal:** Common Arabic stop words that do not contribute to the sentiment (e.g., prepositions, conjunctions) were removed to focus on the meaningful content of the tweets.

Below is an example of how a tweet from the dataset undergoes these steps:

Original tweet:

عندي صداع نصفي شديد جداً والله #الشقيقة <https://migraineinfo.com>

English translation: “I have a very severe migraine, I swear to God #migraine <https://migraineinfo.com>”.

1. **Data Cleaning:** The tweet is cleaned by removing unwanted elements, such as URLs, hashtags and emojis: “عندي صداع نصفي شديد جداً والله”. English translation: “I have very severe migraine I swear to God”.

2. **Text Normalization:** Spelling variations and repeated letters are standardized: “عندي صداع نصفي شديد جداً والله”. English translation: “I have very severe migraine I swear to God”.

3. **Feature Extraction:** The cleaned and normalized text is transformed into feature vectors using TF-IDF or word embeddings. For example: [0.14, 0.08, 0.30, ...]

This example demonstrates how the preprocessing steps convert unstructured text into a structured input format that the ML models can process efficiently.

3.2.4 Class Distribution and Labelling

A group of annotators manually annotated each tweet’s sentiment. The tweets were classified into two sentiment categories: “Positive” and “Negative”. The manual annotation process ensured that each tweet was accurately labelled based on its sentiment toward migraine experiences. The final distribution of the dataset is shown in Figure 5:

Figure 5 Sentiment Distribution in Arabic Tweets

The classification was based on specific keywords, context, and sentiment-related expressions. Table 1 presents the features and symptoms that were used to distinguish between “Positive” and “Negative” tweets:

Table 1 Classification Features for Positive and Negative Tweets

Positive	Negative
Contains the word “صداع نصفي” (migraine) indicating the user is experiencing or has experienced a migraine.	Contains the word “صداع نصفي” (migraine) but used in a general context without indicating a personal experience (e.g., informational tweets).
Includes words or phrases indicating symptoms	Includes words that describe unrelated contexts

such as “الم” (pain) or “صداع قوي” (strong headache).	(e.g., “My friend has migraine” or “تعرفوا على أعراض الصداع النصفي” – (Learn about migraine symptoms)).
Emojis or hashtags that imply distress (e.g., 😞, 🚑).	Neutral or unrelated hashtags (e.g., #صحة (#health), #معلومات (information)).

The classification process aimed to accurately reflect user sentiment and ensure the dataset contained relevant examples for training and testing the SA models.

3.3 Challenges and Limitations

Several challenges were encountered during the data collection and preprocessing stages. The primary challenge was the linguistic complexity of Arabic, including dialectal variations, which required extensive normalization and stemming. Additionally, Twitter’s unregulated nature made it difficult to filter irrelevant or off-topic tweets. However, the MAS’s preprocessing agent efficiently handled these issues, ensuring the final dataset was clean and relevant for SA.

4. Results and Discussion

This section presents the results of the proposed multi-agent-based framework for SA on migraine-related Arabic tweets. It evaluates the framework’s performance in key metrics such as accuracy, precision, recall and F1 score. The results are analyzed in comparison with traditional methods, emphasizing the improvements brought by the MAS. Additionally, this section discusses the implications of the findings, particularly in the context of healthcare and public health.

4.1 Sentiment Classification Results

The dataset was divided into 70% for training, 15% for validation and 15% for testing sets to ensure balanced evaluation. The training set was used to fit the models, while the validation set was used for hyperparameter tuning and preventing overfitting. The testing set was reserved for final model performance evaluation.

The training process involved splitting the data into batches, with a batch size of 32 for SVM, NB and LR models. The models were trained for (50) epochs with early stopping applied to prevent overfitting. The learning rate was set at (0.001) and the Adam optimizer was used for gradient updates. Furthermore, 5-fold cross-validation was used to further assess model robustness.

The sentiment classification results obtained from the models and their integration through ensemble learning techniques are summarized in Table 2. As shown in Figure 6, the ensemble model demonstrated the highest performance across all metrics, achieving an accuracy of 90.0% and an F1-score of 89.46%. The ensemble approach improved overall classification performance by effectively balancing precision and recall, as reflected by the increase in both metrics across all models.

Table 2 Performance of Individual Models and the Ensemble Model

Model	Accuracy	Precision	Recall	F1-score
SVM	88.0%	88.75%	86.90%	87.82%
NB	86.25%	84.00%	88.70%	86.29%
LR	87.55%	87.67%	85.85%	86.75%
Ensemble	90.0%	90.45%	88.50%	89.46%

The ensemble model, which combines the predictions of SVM, NB and LR, achieved the highest performance across all metrics, confirming the effectiveness of model integration through the **MASHA** framework.

Figure 6 Performance of SVM, NB, LR and Ensemble Models

In addition to the key performance metrics discussed in Table 1, the confusion matrix provides a more detailed breakdown of the model's performance in sentiment classification for migraine-related Arabic tweets. As shown in Figure 7, the matrix displays the correct and incorrect classifications made by the model, with 160 correctly classified positive tweets and 110 correctly classified negative tweets. However, some misclassifications occurred, with 21 negative tweets misclassified as positive and 24 positive tweets misclassified as negative.

Figure 7 Confusion Matrix for Sentiment Classification on Test Set (300 tweets)

4.2 Comparison with Traditional Approaches

Compared to traditional sentiment classification methods that typically rely on a single machine learning model, the multi-agent-based framework, **MASHA**, offers several key advantages:

- **Scalability:** The proposed system can handle large volumes of tweets in real-time thanks to the division of tasks among multiple agents. Each agent specializes in a specific task, allowing efficient processing without overwhelming a single system component.
- **Flexibility:** By integrating multiple ML models, the system is more adaptable to changes in data and can accommodate the diverse linguistic variations in Arabic, including dialects commonly used in Saudi Arabia.
- **Improved Accuracy:** The ensemble learning technique outperformed traditional methods, improving classification accuracy by approximately 2.5%. This improvement is significant in healthcare applications, where accuracy is crucial for making reliable predictions.

4.3 Error Analysis

Despite the MASHA framework's overall solid performance, particular challenges impacted classification accuracy. As depicted in Figure 8, the primary sources of error included handling Arabic dialect variations (40% of errors) and the short length of tweets (30% of errors), making it difficult to accurately assess sentiment.

- **Handling Arabic Dialects:** Dialect variation is one of the primary challenges in SA for Arabic tweets. Although the pre-processing agent performed text normalization and stemming, some dialect-specific expressions were complex to standardize, which impacted the accuracy of sentiment classification in some instances.
- **Short Text Length:** Tweets often contain very short messages, which limits the amount of context available for sentiment classification. As a result, some tweets were misclassified, mainly when they contained ambiguous language or lacked clear sentiment indicators.

During the analysis, several common error types were observed:

- **Ambiguity:** Tweets such as “الحمد لله بعد الصداع ارتحت” (“Thank God I feel better after the headache”) could be interpreted as both positive (relief) and negative (due to mentioning a headache).
- **Sarcasm:** Sentences like “ما في أحلى من صداع نصفي للصباح” (“Nothing better than a migraine in the morning”) were often misclassified due to the sarcastic tone.

- **Incomplete Context:** Some tweets lacked sufficient context for accurate classification, such as “أوه لا... الصداع بدأ!” (“Oh no... the headache started!”).

Although the dataset primarily contained tweets from “Saudi Arabia”, dialect variations such as “Najdi”, “Hejazi” and “Gulf Arabic” were present. For example, the word “راس” (head) can appear as “رأس” (head) in certain dialects due to phonetic spelling. Preprocessing included normalizing these variations to a standard form and the ML models were fine-tuned to recognize common synonyms and spelling differences to minimize misclassification due to dialect-specific terms.

Figure 8 Error Distribution in Sentiment Classification

4.4 Implications for Healthcare and Public Health

The findings from this study have significant implications for the healthcare sector, particularly in the context of public health monitoring. By analyzing real-time social media data, healthcare providers and policymakers can gain valuable insights into patient experiences with migraines, track public sentiment regarding healthcare services and detect emerging health trends.

- **Improved Patient Monitoring:** SA of social media data allows healthcare providers to monitor how patients perceive their migraine symptoms, treatments and overall healthcare experiences. This can help identify common challenges faced by patients and lead to improvements in treatment strategies.
- **Real-Time Public Health Insights:** The system’s real-time processing of tweets enables public health agencies to identify trends and patterns related to migraine episodes and their impact on society. This can inform public health campaigns and interventions to mitigate migraines’ impact on quality of life.
- **Scalability to Other Healthcare Conditions:** While this study focuses on migraines, the multi-agent-based framework can be adapted to analyze other health conditions, such as influenza, diabetes or mental health issues. The system’s flexibility and scalability make it a valuable tool for broader public health applications.

4.5 Limitations and Future Work

Despite the positive results, this study has several limitations that should be addressed in future research:

- **Dataset Size:** Although the dataset used in this study is significant, expanding the dataset to include more diverse sources of Arabic social media data would improve the generalizability of the results. Including data from other platforms like Facebook or Instagram could provide a more comprehensive understanding of public sentiment.
- **Further Model Optimization:** Future work should optimize the ML models used in the framework, particularly in handling complex Arabic language structures and dialects. Exploring DL techniques such as recurrent neural networks (RNNs) or transformers may improve classification performance.
- **Sentiment Lexicon Development:** Given the challenges associated with Arabic dialects, developing a more comprehensive sentiment lexicon specific to Arabic social media would improve sentiment classification accuracy.

MASHA's proposed multi-agent-based framework demonstrated its effectiveness in integrating multiple ML models for SA in healthcare-related tasks. The framework improved classification accuracy and provided valuable insights into patient experiences and public health trends. The system's scalability and flexibility suggest that it can be expanded for other health-related applications.

4.6 Discussion

This study's results demonstrate the **MASHA** framework's effectiveness in integrating multiple ML models for SA in healthcare-related tasks. By leveraging ensemble learning techniques, the system achieved higher accuracy, precision and F1-scores than individual models. The key takeaway from the research is the importance of integrating multiple models to balance different performance metrics to achieve an overall better sentiment classification outcome.

- **Healthcare Insights:** The system provides significant insights for the healthcare sector, especially in real-time tracking of patient experiences and public health trends. Analyzing sizeable social media data enables healthcare providers to stay updated on patient sentiment and emerging health conditions.
- **Adaptability:** The modular, adaptable structure of the **MASHA** framework ensures that it can be extended to various healthcare conditions, making it a valuable tool for public health campaigns and interventions. Furthermore, the flexibility in handling multiple Arabic dialects expand the system's application to different Arabic-speaking regions, further improving its utility for healthcare and public policy.

Addressing the current limitations, such as dataset expansion, further model optimization and developing an Arabic-specific sentiment lexicon, would improve the framework's performance even further. Additionally, exploring the potential of DL techniques could yield significant gains in handling the complexity of Arabic texts, ensuring better results for SA tasks.

5. Conclusion

This paper introduced **MASHA** (Multi-Agent System for Healthcare Sentiment Analysis), a scalable and flexible multi-agent-based framework that integrates multiple ML models for SA in healthcare. The proposed framework was validated through a case study focusing on sentiment classification for migraine detection using Arabic tweets. By leveraging the strengths of multiple ML models, including SVM, NB and LR and employing ensemble learning techniques, **MASHA** demonstrated significant improvements in classification performance compared to traditional methods.

The results showed that the ensemble approach improved the overall accuracy, precision, recall and F1-score, with the ensemble model achieving an F1-score of 89.46 %. This underscores the importance of integrating diverse ML models in handling complex healthcare data, especially when analyzing unstructured text from social media platforms like Twitter. Moreover, using a multi-agent architecture allowed the system to efficiently manage data acquisition, pre-processing, model development and decision-making tasks, ensuring the system's scalability for large-scale, real-time SA.

The implications of this study extend beyond migraine detection. **MASHA's** adaptability makes it suitable for analyzing sentiment in various healthcare conditions, providing valuable insights for healthcare providers and policymakers. By analyzing real-time social media data, **MASHA** offers a powerful tool for tracking public sentiment, improving patient monitoring and informing public health decisions.

However, this research also highlighted several challenges, particularly in handling the linguistic complexity of Arabic and the dialect variations present in tweets. Future work should expand the dataset to include other social media platforms and optimise the ML models, especially for Arabic text. Exploring advanced DL techniques such as transformer neural networks (TNNs) or RNNs could improve sentiment classification performance.

In conclusion, **MASHA** has demonstrated its potential as an effective solution for integrating ML models in SA for healthcare applications. Its modular design, scalability and ability to

handle large-scale data make it a valuable framework for real-time SA in healthcare and beyond. Future enhancements, such as incorporating state-of-the-art ML models, including transformer-based architectures and ensemble DL approaches, can further enhance the framework's predictive accuracy and robustness in public health monitoring. Additionally, expanding its application to other health-related domains, such as chronic disease management and mental health analysis, will demonstrate its versatility in addressing diverse healthcare challenges. Beyond healthcare, the framework's modular and adaptable design makes it well-suited for adoption in other sectors, such as finance and cybersecurity, where real-time SA and data integration are pivotal for timely and data-driven decision-making.

Acknowledgments section

The author would like to acknowledge the annotators who contributed to the manual labelling of the Arabic tweet dataset used in this study. The author also extends gratitude to colleagues and reviewers whose constructive feedback helped improve this work. This research was conducted independently, without institutional or external financial support.

Author contributions

Author contributions: S.Y.B.: Conceptualisation, Methodology, Software, Formal Analysis, Data Curation, Writing (Original Draft), Writing (Review & Editing).

Declaration of generative AI and AI-assisted technologies in the writing process:

During the preparation of this work, the author(s) used **Grammarly** in order to make grammar and spelling corrections and improve the sentence structure. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

1. Lauriola I, Lavelli A, Aiolfi F. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*. 2022;470:443–56. doi:10.1016/j.neucom.2021.05.103
2. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: A review. *Trends in Anaesthesia and Critical Care*. 2021;38:4–9. doi:10.1016/j.tacc.2021.02.007
3. Johri P, Khatri SK, Al-Taani AT, Sabharwal M, Suvanov S, Kumar A. Natural Language Processing: History, Evolution, Application, and Future Work. *Lecture Notes in Networks and Systems* [Internet]. Springer Singapore; 2021. 365–375 p. Available from: http://dx.doi.org/10.1007/978-981-15-9712-1_31 doi:10.1007/978-981-15-9712-1_31

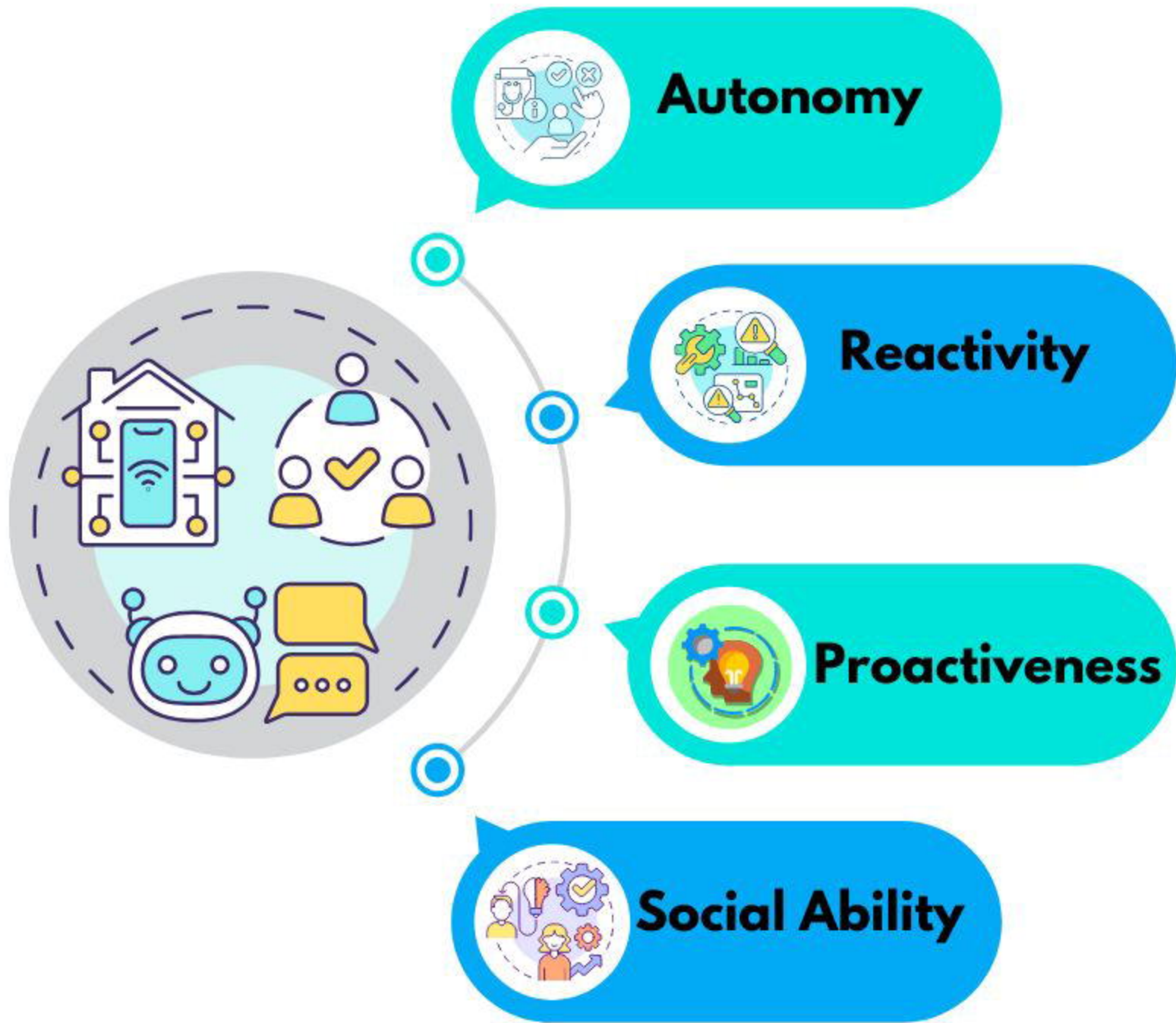
- 645 4. González TR, Rocha MA, Baltazar R, Casillas MÁ, del Valle J. Voice Control System Through NLP
646 (Natural Language Processing) as an Interactive Model for Scalable ERP Platforms in Industry
647 4.0. Smart Innovation, Systems and Technologies. 2022;306:207–17. doi:10.1007/978-981-19-
648 3359-2_18
- 649 5. Jim JR, Talukder MAR, Malakar P, Kabir MM, Nur K, Mridha MF. Recent advancements and
650 challenges of NLP-based sentiment analysis: A state-of-the-art review. Natural Language
651 Processing Journal. 2024;6(November 2023):100059. doi:10.1016/j.nlp.2024.100059
- 652 6. Mangrum R, Bryant AL, Gerstein MT, McCarrier KP, Houts CR, McGinley JS, et al. The impacts
653 of migraine on functioning: Results from two qualitative studies of people living with
654 migraine. Headache. 2024;64(2):156–71. doi:10.1111/head.14664 PubMed PMID: 38235605.
- 655 7. Buse DC, Cady R, Starling AJ, Buzby M, Spinale C, Steinberg K, et al. Headache/migraine-
656 related stigma, quality of life, disability, and most bothersome symptom in adults with current
657 versus previous high-frequency headache/migraine and medication overuse: results of the
658 Migraine Report Card survey. BMC Neurol. 2024;24(1):1–17. doi:10.1186/s12883-024-03732-
659 x PubMed PMID: 38965567.
- 660 8. El-Metwally A, Toivola P, Alahmary K, Bahkali S, Alkhathaami A, Al Ammar SA, et al. The
661 Epidemiology of Migraine Headache in Arab Countries: A Systematic Review. Scientific World
662 Journal. 2020;2020. doi:10.1155/2020/4790254 PubMed PMID: 32607079.
- 663 9. Seddik AH, Branner JC, Ostwald DA, Schramm SH, Bierbaum M, Katsarava Z. The
664 socioeconomic burden of migraine: An evaluation of productivity losses due to migraine
665 headaches based on a population study in Germany. Cephalalgia. 2020;40(14):1551–60.
666 doi:10.1177/0333102420944842 PubMed PMID: 32762249.
- 667 10. Chen J, Wang Y. Social media use for health purposes: Systematic review. J Med Internet Res.
668 2021;23(5):1–16. doi:10.2196/17917 PubMed PMID: 33978589.
- 669 11. Afful-Dadzie E, Afful-Dadzie A, Egala SB. Social media in health communication: A literature
670 review of information quality. Health Information Management Journal. 2023;52(1):3–17.
671 doi:10.1177/1833358321992683 PubMed PMID: 33818176.
- 672 12. Deng H, Wang Q, Turner DP, Sexton KE, Burns SM, Eikermann M, et al. Sentiment analysis of
673 real-world migraine tweets for population research. Cephalalgia Rep. 2020;3:1–9.
674 doi:10.1177/2515816319898867
- 675 13. Niiberg-Pikksööt T, Laas K, Aluoja A, Braschinsky M. Implementing a digital solution for
676 patients with migraine—Developing a methodology for comparing digitally delivered
677 treatment with conventional treatment: A study protocol. PLOS Digital Health. 2024;3(2):1–
678 15. doi:10.1371/journal.pdig.0000295
- 679 14. Amiri P, Kazeminasab S, Nejadghaderi SA, Mohammadinasab R, Pourfathi H, Araj-Khodaei M,
680 et al. Migraine: A Review on Its History, Global Epidemiology, Risk Factors, and Comorbidities.
681 Front Neurol. 2022;12(February):1–15. doi:10.3389/fneur.2021.800605
- 682 15. Khan L, Shahreen M, Qazi A, Jamil Ahmed Shah S, Hussain S, Chang HT. Migraine headache
683 (MH) classification using machine learning methods with data augmentation. Sci Rep.
684 2024;14(1):1–11. doi:10.1038/s41598-024-55874-0 PubMed PMID: 38431729.

- 685 16. Abdalrdha ZK, Al-Bakry AM, Farhan AK. Improving the CNN Model for Arabic Crime Tweet
686 Detection Based on an Intelligent Dictionary. In: 2023 16th International Conference on
687 Developments in eSystems Engineering (DeSE). 2023. p. 748–53.
688 doi:10.1109/DeSE60595.2023.10469560
- 689 17. Alzahrani F, Alahmadi YM, Thagfan SS Al, Alolayan S, Elbadawy HM. Migraine Management in
690 Community Pharmacies: Knowledge, Attitude and Practice Patterns of Pharmacists in Saudi
691 Arabia. *Pharmacy*. 2023;11(5):155. doi:10.3390/pharmacy11050155
- 692 18. Alrashidi B, Jamal A, Alkhathlan A. Abusive Content Detection in Arabic Tweets Using Multi-
693 Task Learning and Transformer-Based Models. *Applied Sciences (Switzerland)*. 2023;13(10).
694 doi:10.3390/app13105825
- 695 19. Baroud SY, El- Halees AM, Baraka RS, Yahaya NA. AI - Driven CBR and Ontological Integration
696 Framework for Enhancing Breast Cancer Care: Case Study. 2024 4th International Conference
697 on Robotics, Automation, and Artificial Intelligence, RAAI 2024. 2024;284–9.
698 doi:10.1109/RAAI64504.2024.10949518
- 699 20. Lwakatare LE, Raj A, Crnkovic I, Bosch J, Olsson HH. Large-scale machine learning systems in
700 real-world industrial settings: A review of challenges and solutions. *Inf Softw Technol*.
701 2020;127(December 2019). doi:10.1016/j.infsof.2020.106368
- 702 21. Munappy AR, Bosch J, Olsson HH, Arpteg A, Brinne B. Data management for production
703 quality deep learning models: Challenges and solutions. *Journal of Systems and Software*.
704 2022;191:111359. doi:10.1016/j.jss.2022.111359
- 705 22. Chai C, Wang J, Luo Y, Niu Z, Li G. Data Management for Machine Learning: A Survey. *IEEE*
706 *Trans Knowl Data Eng*. 2023;35(5):4646–67. doi:10.1109/TKDE.2022.3148237
- 707 23. Chaudhary L, Girdhar N, Sharma D, Andreu-Perez J, Doucet A, Renz M. A Review of Deep
708 Learning Models for Twitter Sentiment Analysis: Challenges and Opportunities. *IEEE Trans*
709 *Comput Soc Syst*. 2024;11(3):3550–79. doi:10.1109/TCSS.2023.3322002
- 710 24. Aguado G, Julian V, Garcia-Fornes A, Espinosa A. A Multi-Agent System for guiding users in on-
711 line social environments. *Eng Appl Artif Intell*. 2020;94(September 2019):103740.
712 doi:10.1016/j.engappai.2020.103740
- 713 25. Qian C, Xie Z, Wang Y, Liu W, Dang Y, Du Z, et al. Scaling Large-Language-Model-based Multi-
714 Agent Collaboration [Internet]. 2024. Available from: <http://arxiv.org/abs/2406.07155>
- 715 26. Alomari KM, Elsherif HM, Shaalan K. Arabic tweets sentimental analysis using machine
716 learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*
717 *Intelligence and Lecture Notes in Bioinformatics)*. 2017;10350 LNCS(October):602–10.
718 doi:10.1007/978-3-319-60042-0_66
- 719 27. Baroud SY. AI-Powered Multi-Agent System for Type 2 Diabetes Care: IoT , Federated
720 Learning and Mobile Health Integration. 2025 International Conference on Metaverse and
721 Current Trends in Computing (ICMCTC). 2025;1–5. doi:10.1109/ICMCTC62214.2025.11196336
- 722 28. Baroud SY, Yahaya NA, Elzamy AM. Cutting-Edge AI Approaches with MAS for PdM in Industry
723 4.0: Challenges and Future Directions. *Journal of Applied Data Sciences*. 2024;5(2):455–73.
724 doi:10.47738/jads.v5i2.196

29. Luzolo PH, Elrawashdeh Z, Tchappi I, Galland S, Outay F. Combining Multi-Agent Systems and Artificial Intelligence of Things: Technical challenges and gains. *Internet of Things (Netherlands)*. 2024;28(May):101364. doi:10.1016/j.iot.2024.101364
30. Alkazzi JM, Okumura K. A Comprehensive Review on Leveraging Machine Learning for Multi-Agent Path Finding. *IEEE Access*. 2024;12(April):57390–409. doi:10.1109/ACCESS.2024.3392305
31. Baroud SY, Yahaya NA. ML2MAS: A Multi-Agent Reinforcement Learning and BNNs-GAN Integration Framework for Smart Manufacturing Optimization. *Sustainable Operations and Computers*. 2025;6(July):217–28. doi:10.1016/j.susoc.2025.07.003
32. Rane N, Choudhary SP, Rane J. Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. *Studies in Medical and Health Sciences*. 2024;1(2):18–41. doi:10.48185/smhs.v1i2.1225
33. Zounemat-Kermani M, Batelaan O, Fadaee M, Hinkelmann R. Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology [Internet]*. Elsevier B.V.; 2021. p. 126266. Available from: <https://doi.org/10.1016/j.jhydrol.2021.126266> doi:10.1016/j.jhydrol.2021.126266
34. Mienye ID, Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*. IEEE; 2022. p. 99129–49. doi:10.1109/ACCESS.2022.3207287
35. Abualigah L, Alfar HE, Shehab M, Hussein AMA. Sentiment Analysis in Healthcare: A Brief Review. *Studies in Computational Intelligence*. 2020;874(January):129–41. doi:10.1007/978-3-030-34614-0_7
36. AlMuhaideb S, AlNegheimish Y, AlOmar T, AlSabti R, AlKathery M, AlOlyyan G. Analyzing Arabic Twitter-Based Patient Experience Sentiments Using Multi-Dialect Arabic Bidirectional Encoder Representations from Transformers. *Computers, Materials and Continua*. 2023;76(1):195–220. doi:10.32604/cmc.2023.038368
37. Ihara K, Dumkrieger G, Zhang P, Takizawa T, Schwedt TJ, Chiang CC. Application of Artificial Intelligence in the Headache Field. *Curr Pain Headache Rep*. 2024;28(10):1049–57. doi:10.1007/s11916-024-01297-5
38. Leman M. Multi-Agent Systems. In: *Objective Coordination in Multi-Agent System Engineering. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg; 2001. p. 9–32. doi:10.1007/3-540-44933-7_2
39. Wooldridge M, Jennings NR. Intelligent agents: Theory and practice. *Knowl Eng Rev*. 1995;10(2):115–52. doi:10.1017/S0269888900008122
40. Pulikottil T, Estrada-Jimenez LA, Ur Rehman H, Mo F, Nikghadam-Hojjati S, Barata J. Agent-based manufacturing — review and expert evaluation. *International Journal of Advanced Manufacturing Technology*. 2023. p. 2151–80. doi:10.1007/s00170-023-11517-8
41. Pretel E, Moya A, Navarro E, López-Jaquero V, González P. Analysing the synergies between Multi-agent Systems and Digital Twins: A systematic literature review. *Information and Software Technology*. 2024. doi:10.1016/j.infsof.2024.107503

- 764 42. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis:
765 Approaches, challenges and trends. Knowl Based Syst. 2021;226:107134.
766 doi:10.1016/j.knosys.2021.107134
- 767 43. Alruily M. Classification of arabic tweets: A review. Electronics (Switzerland). 2021;10(10).
768 doi:10.3390/electronics10101143
- 769 44. Sherif SM, Alamoodi AH, Albahri OS, Garfan S, Albahri AS, Deveci M, et al. Lexicon annotation
770 in sentiment analysis for dialectal Arabic: Systematic review of current trends and future
771 directions. Inf Process Manag. 2023;60(5):103449. doi:10.1016/j.ipm.2023.103449
- 772 45. Oueslati O, Cambria E, HajHmida M Ben, Ounelli H. A review of sentiment analysis research in
773 Arabic language. Future Generation Computer Systems. 2020;112:408–30.
774 doi:10.1016/j.future.2020.05.034
- 775 46. Elayeb B. Arabic Text Classification: A Literature Review. Proceedings of IEEE/ACS International
776 Conference on Computer Systems and Applications, AICCSA. 2021;2021-Decem:1–8.
777 doi:10.1109/AICCSA53542.2021.9686874
- 778 47. Nassif AB, Elnagar A, Shahin I, Henno S. Deep learning for Arabic subjective sentiment
779 analysis: Challenges and research opportunities. Appl Soft Comput. 2021;98:106836.
780 doi:10.1016/j.asoc.2020.106836
- 781 48. Khan J, Ahmad N, Khalid S, Ali F, Lee Y. Sentiment and Context-Aware Hybrid DNN With
782 Attention for Text Sentiment Classification. IEEE Access. 2023;11(February):28162–79.
783 doi:10.1109/ACCESS.2023.3259107
- 784 49. Abdalrdha ZK, Al-Bakry AM, Farhan AK. A Hybrid CNN-LSTM and XGBoost Approach for Crime
785 Detection in Tweets Using an Intelligent Dictionary. Revue d'Intelligence Artificielle.
786 2023;37(6):1–3. doi:10.18280/RIA.370630
- 787 50. Mohammed A, Kora R. A comprehensive review on ensemble deep learning: Opportunities
788 and challenges. Journal of King Saud University - Computer and Information Sciences.
789 2023;35(2):757–74. doi:10.1016/j.jksuci.2023.01.014
- 790 51. Jia J, Liang W, Liang Y. A Review of Hybrid and Ensemble in Deep Learning for Natural
791 Language Processing. arXiv preprint [Internet]. 2023;1–23. Available from:
792 <http://arxiv.org/abs/2312.05589>
- 793 52. Abdalrdha ZK, Mohsin Al-Bakry ProfDrA, K. Farhan ProfDrA. Arabic Crime Tweet Filtering and
794 Prediction Using Machine Learning. Iraqi Journal for Computers and Informatics.
795 2024;50(1):73–85. doi:10.25195/ijci.v50i1.479
- 796 53. Abro AA, Talpur MSH, Jumani AK. Natural Language Processing Challenges and Issues: A
797 Literature Review. Gazi University Journal of Science. 2023;36(4):1522–36.
798 doi:10.35378/gujs.1032517
- 799 54. Aftab F, Bazai SU, Marjan S, Baloch L, Aslam S, Amphawan A, et al. A Comprehensive Survey
800 on Sentiment Analysis Techniques. International Journal of Technology. 2023;14(6):1288–98.
801 doi:10.14716/ijtech.v14i6.6632

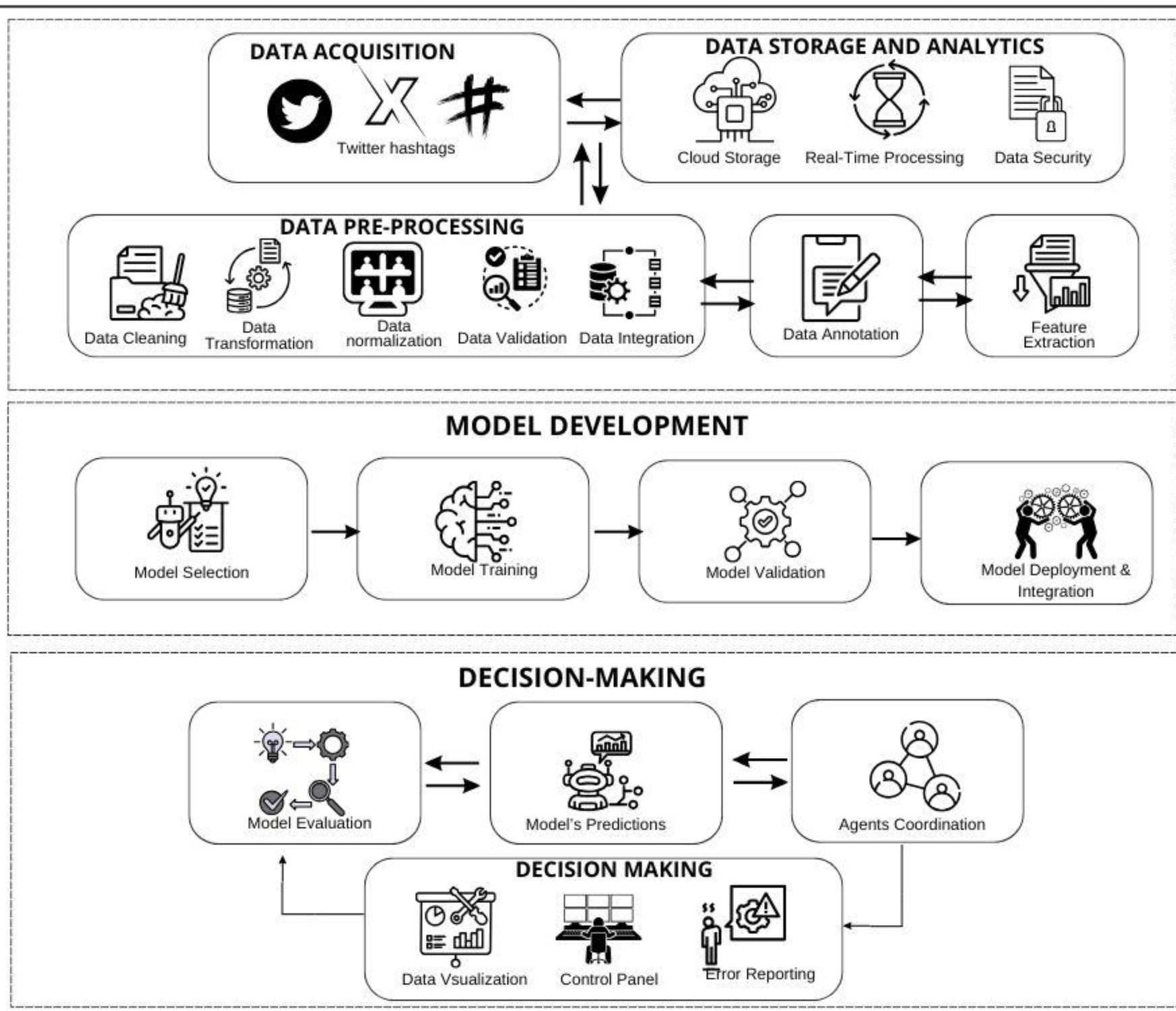
802



**Data Acquisition Pre-
processing Agents
Layer**

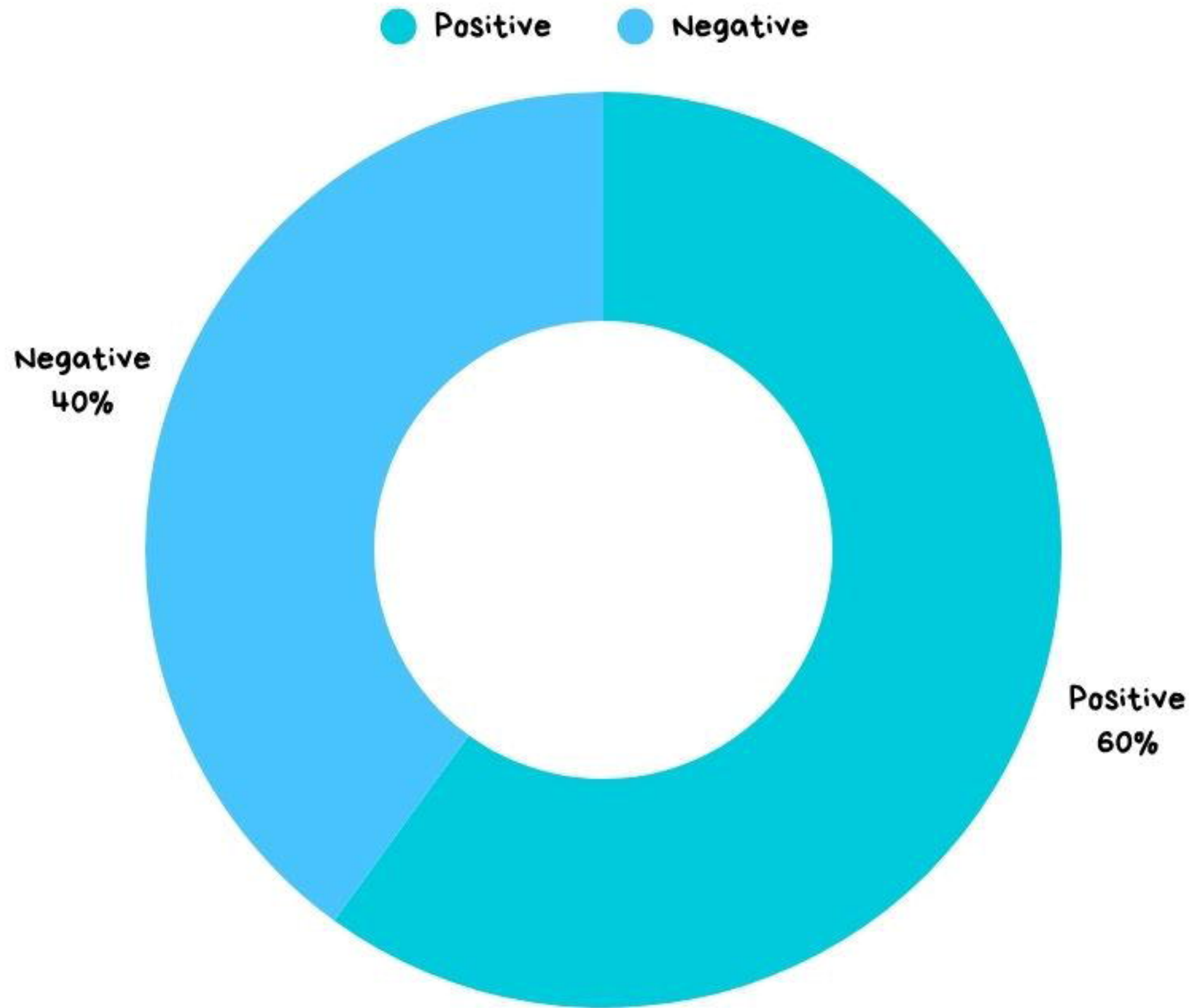
**Model
Development
Agents Layer**

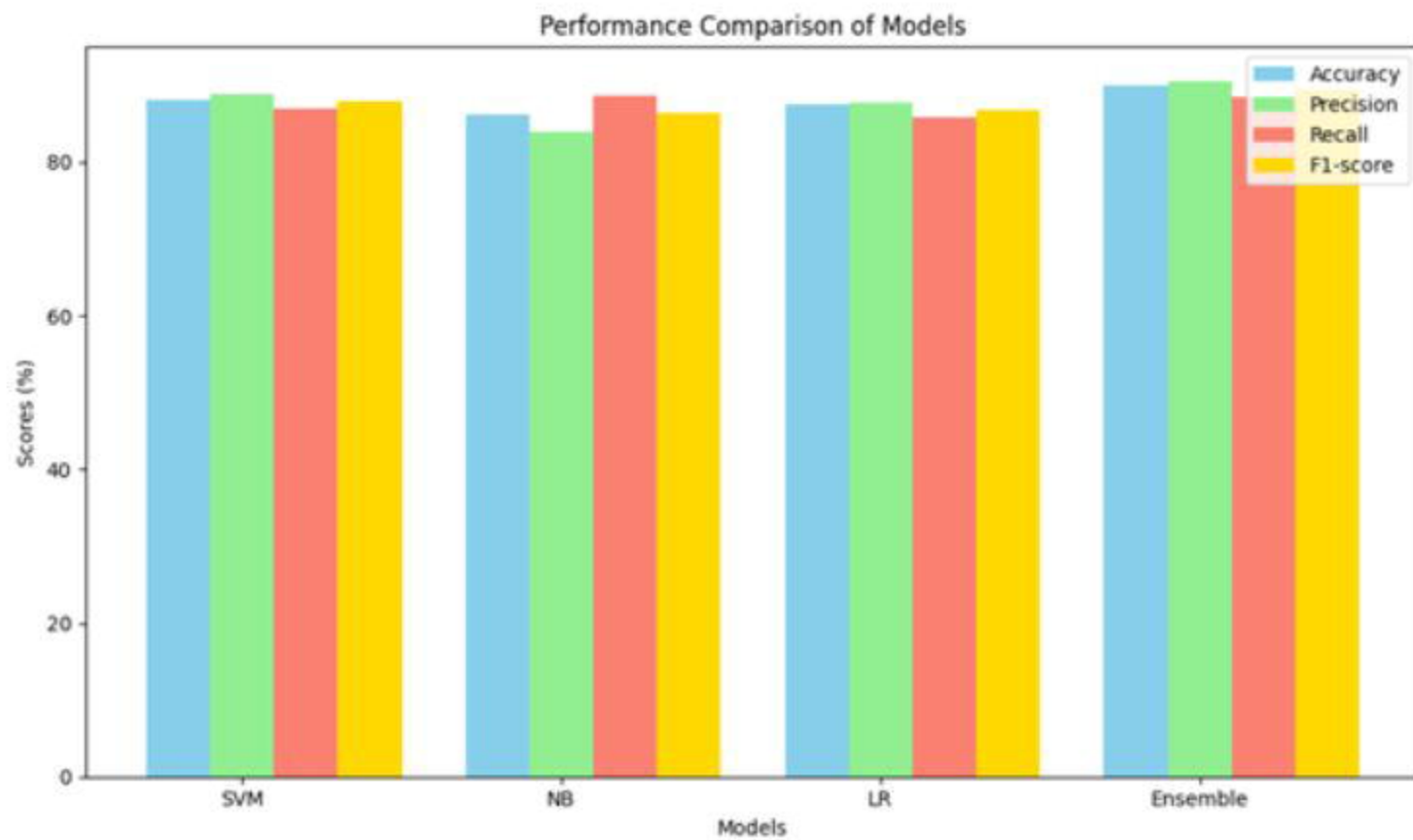
**Decision-Making
Agents Layer**

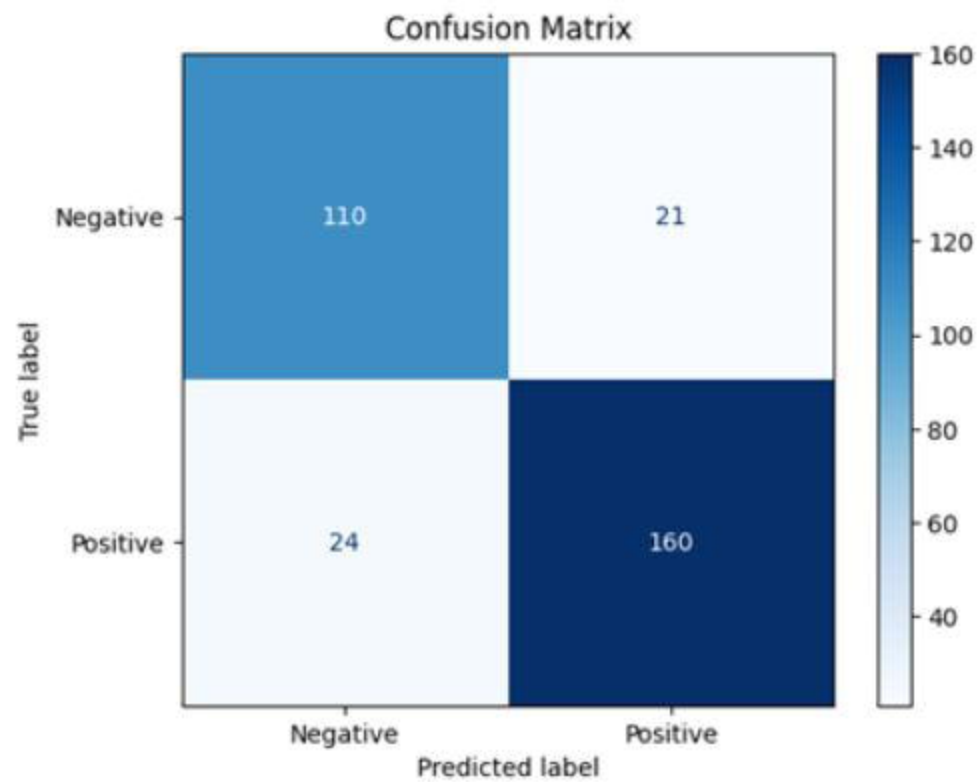


[illegible]

	A	B
	RT @phSarah9A: ♦ انواع الصداع:	
	♦ cluster - صداع حول العين	
	♦ migraine - صداع نصفي	
	♦ sinus - صداع الجيوب الانفية	
3	♦ tension ht... - صداع بسبب التوتر	Positive
4	😊 صداع نصفي جميل	Negativ
5	اخ حاشني صداع نصفي @XXIIIXVII	Positive
6	هلا جاني صداع نصفي	Positive
7	@amjad_awad99 @glewo_01 انا ركبت صداع لكن نصفي وحده ووحده لا 😊😊😊😊	Negativ
8	https://t.co/IVEYqOIKK6 صداع نصفي من الفخر @CelebsArabicII	Negativ
9	😊😊😊 اتوقع جيوب او صداع نصفي مش عارفه جد @AsfourSeham	Negativ
10	هيا لو عندك صداع نصفي مزمن ♀ @InozukeT 🇵🇸	Negativ
	! ودك تساله ،، مر عليك شيء اسمه صداع نصفي @Gladd @iKhaledM1	
11	! مر عليك النور يضرب عينك و تغيم ماعاد تشوف يوم كامل	Negativ
12	جلست مع اخواني جاني صداع نصفي	Positive
13	صداع ليس نصفي	Negativ







Other Misclassifications
10%

Ambiguous Language
20%

Dialect Variations
40%

Short Text Length
30%

