Title	Perspectives on Preparedness for Chemical, Biological, Radiological, and Nuclear Threats in the Middle East and North Africa Region: Application of Artificial Intelligence Techniques
Type	Article
URL	https://clok.uclan.ac.uk/50798/
DOI	https://doi.org/10.1089/hs.2023.0093
Date	2024
Citation	Farhat, Hassan, Alinier, Guillaume, Helou, Mariana, Galatis, Ionnais, Bajow, Nidaa, Jose, Denis, Jouini, Sarra, Sezigen, Sermet, Hafi, Samia et al (2024) Perspectives on Preparedness for Chemical, Biological, Radiological, and Nuclear Threats in the Middle East and North Africa Region: Application of Artificial Intelligence Techniques. Health Security, 22 (3). pp. 190-202. ISSN 2326-5094
Creators	Farhat, Hassan, Alinier, Guillaume, Helou, Mariana, Galatis, Ionnais, Bajow, Nidaa, Jose, Denis, Jouini, Sarra, Sezigen, Sermet, Hafi, Samia, Mccabe, Sheena, Somrani, Naoufel, Aifa, Kawther El, Chebbi, Henda, Amor, Asma Ben, Kerkeni, Yosra, Al-Wathinani, Ahmed M., Abdulla, Nassem Mohammed, Jairoun, Ammar Abdulrahman, Morris, Brendon, Castle, Nicholas, Al-Sheikh, Loua, Abougalala, Walid, Dhiab, Mohamed Ben and Laughton, James

Cite as: Farhat H, Alinier G, Helou M, Galatis I, Bajow N, Jose D, Jouini S, Sezigen S, Hafi S, Mccabe S, Somrani N. Perspectives on Preparedness for Chemical, Biological, Radiological, and Nuclear Threats in the Middle East and North Africa Region: Application of Artificial Intelligence Techniques. Health security. 2024 Feb 9. Ahead of print

Puiblished article: https://www.liebertpub.com/doi/abs/10.1089/hs.2023.0093

https://doi.org/10.1089/hs.2023.0093

Perspectives on preparedness for chemical, biological, radiological, and nuclear threats in the Middle East and North Africa region: Application of artificial intelligence techniques

Short title: MENA CBRN Preparedness

Hassan Farhat[*,1,2,3], Guillaume Alinier[1,4,5,6], Mariana Helou[7,8], Ionnais Galatis[9], Nidaa Bajow[10], Denis Jose[11], Sarra Jouini[12,14], Sermet Sezigen15, Samia Hafi[16], Sheena Mccabe[17], Naoufel Somrani[18], Kawther El Aifa[1], Henda Chebbi[18], Asma Ben Amor[2], Yosra Kerkeni[18], Ahmed M. Al-Wathinani[19], Ahmed Al-Hajeri[20], Nassem Mohammed Abdulla[21], Ammar Abdulrahman Jairoun[12], Brendon Morris[1], Nicholas Castle[1], Loua Al-Sheikh[1], Walid Abougalala[22], Mohamed Ben Dhiab[1], James Laughton[1]

Affiliations:

1 Ambulance Service, Hamad Medical Corporation, PO Box 3050, Doha, Qatar.

2 Faculty of Medicine "Ibn El Jazzar", University of Sousse, 4000, Sousse, Tunisia.

3 Faculty of Sciences, University of Sfax, 3000, Sfax, Tunisia.

4 University of Hertfordshire, Hatfield, UK.

5 Weill Cornell Medicine-Qatar.

6 Northumbria University, Newcastle upon Tyne, UK.

7 Lebanese American University, School of Medicine, Beirut, Lebanon.

8 Department of Internal Medicine, Lebanese American University-Rizk Hospital, Beirut, Lebanon.

9 CBRN Knowledge Center, International CBRNE Institute, Belgium.

10 Disaster Medicine Unit, Emergency Department, Security Forces Hospital Program, Medical Services, Ministry of Interior, Riyadh - Saudi Arabia.

11 Alpes-Maritimes (France) Fire & Rescue Services, Paris, France.

13 Emergency department, Charles Nicolle Hospital, Tunis, Tunisia.

14 EL Manar University, Faculty of Medicine of Tunis, Tunis, Tunisia.

15 Health Sciences University, Gulhane Health Sciences Institute, 06010, Ankara, Turkey.

16 Emergency Medical Service 05, Ministry of Health, Tunis, Tunisia.

17 University of Central Lancashire, Doha, Qatar.

18 Strategic Health Operations Center, Ministry of Health, Tunis, Tunisia.

19 Department of Emergency Medical Services, Prince Sultan Bin Abdulaziz College for Emergency Medical Services, King Saud University, Riyadh 1154, Saudi Arabia.

20 The National Ambulance, Abu Dhabi, United Arab Emirates.

21 Health and Safety Department, Dubai Municipality, Dubai, United Arab Emirates.

22 Hamad Medical Corporation.

Corresponding author:

*: Ambulance Service, Hamad Medical Corporation, PO Box 3050, Doha, Qatar;

Mobile: 30507396, e-mail: Hfarhat1@hamad.qa

Author contributions:

MBD, WA, GA, LAS, and JL: supervising, reviewing and editing

IG, MH, NB, DJ, SS, SH, SM, NC, NMA, NS, KA, HC, AA, YK and BM: review and editing.

HF: Conceptualization, writing the original draft, software, formal Analysis and data validation.

Project administration: WA, MD, BM.

Conflicts of Interest:

The author declares no conflict of interest

Acknowledgements:

The authors thank the reviewers for their constructive feedback that helped improve this article.

The authors would like to thank all the experts who participated in the study and agreed to be acknowledged, including Craig Campbell, Ayşe Dökmeci, Mohammed Heriza, Henda Chebbi, Amira Jaafar, Sami Souissi, Nelson Olim, Mike Lynch, Ahmed Abdulsaboor, Joelle Khadra, Mohamed Abdelaziz, Ayla Sayın Öztürk, Gamal Eldin Khalifa, Mohammed Heriza, Adham Abulnour, Michael Spiteri, Rabih Asmar, Nelson Olim, and Fadi El Ters.

The first author would like to thank Ms Cyrine Abidi for all the Machine Learning knowledge she has previously shared.

The first author would like to thank the University of Sfax in Tunisia, especially Prof.Imed Gargouri, for all the data science courses in their institution.

Conference meeting

Work presented as a poster entitled “Promoting machine learning-based approach in disaster medicine quality improvement” on 12-14 May 2023, 19th Emirates Critical Care Conference Event Centre, Dubai Festival City, UAE.

==

Abstract:

The diversification of ethnic, religious, and political backgrounds worldwide, particularly in the Middle East and North Africa (MENA) countries, has increased the inter-country conflicts and terrorist attacks in the last three decades, sometimes involving chemical and biological agents. These warrants reassessing the need for a collaborative preparedness perspective. Artificial Intelligence (AI) techniques have been increasingly utilised in disaster medicine, allowing a thorough analysis by revealing unseen patterns. This study aimed to process open-ended feedback from multidisciplinary experts in disaster medicine regarding the MENA region's preparedness for chemical, biological, radiological, and nuclear (CBRN) risks using text mining (TM) and machine learning (ML) techniques. Open-ended feedback from 29 international experts in disaster medicine, selected based on their organisational roles and contributions to the academic field, was collected using a modified interview method (MIM) between October and December 2022. ML clustering algorithms, natural language processing, and sentiment analysis were used to analyse the data gathered using R-language accessed through the R-studio environment. Negative and fearful sentiments about the lack of accessibility to preparedness information were identified, in addition to positive sentiments toward the CBRN preparedness concepts raised by the MIM. The AI analysis techniques revealed a common consensus of the experts on the importance of having accessible and effective plans and improved health sector preparedness in MENA, especially for potential chemical and biological incidents. This would help the policymakers in the region to converge their efforts to build collaborative initiatives to strengthen the CBRN preparedness capabilities of the healthcare sector.

Keywords: CBRN, Health preparedness, MENA, Machine Learning, Artificial Intelligence.

1 Introduction

The risk of chemical, biological, radiological, and nuclear (CBRN) incidents has increased worldwide over the last 30 years [1]. Previous studies have indicated the tentative role of terrorist groups in weaponising CBRN agents worldwide [2], [3]. For example, the Aum Shinrikyo group released sarin in a metro subway in Japan in 1995, resulting in more than 600 casualties suffering from acute neurogenic and respiratory symptoms [4, p. 1]. In the Middle East and North Africa (MENA) region, terrorist groups have attempted to weaponise biological agents such as anthrax, ricin, and cyanide for use in metropolitan cities [3], [5]. Furthermore, CBRN incidents pose significant risks to the health sector and can lead to widespread public health emergencies if not adequately managed. CBRN incident planning is crucial for governments and the healthcare sector. Once an incident occurs, it has devastating consequences, including injury, illness, death, and widespread panic. Hence, effective preparedness involves identifying potential risks and vulnerabilities, developing preparedness and response strategies and protocols, and coordinating emergency services with other stakeholders to ensure a swift and effective response. Therefore, due to cross-border risks, collaboration across the MENA health sector is crucial to ensure appropriate assessment readiness for managing overwhelming numbers of victims with acute clinical presentations. Similar to other global experiences, such as in Europe [6], [7], engaging MENA experts in expressing their opinions about MENA international cooperation through interviews seems to be a helpful measure to drive perspectives of building a collaborative approach for CBRN threat health sector readiness.

Artificial intelligence (AI) is a broad and evolving science widely used in medicine for educational purposes, clinical diagnoses, therapeutic decision-making, and innovative health research. It stimulates cognisable human functions and provides a detailed problem-solving view [8]. Text mining (TM) is an AI technique that utilises natural language processing (NLP) and sentiment analysis techniques to extract meaningful insights by processing unstructured text data into a numeric form, exploring in-depth relationships between variables, and uncovering valuable insights and patterns in data that might be missed with manual analysis alone [9]. Recently, at a biological arms control conference, researchers put AI, usually used to search for helpful drugs, into a "bad actor" mode to show how easily it could be abused [10]. It took less than six hours for drug-developing AI to invent 40,000 potentially lethal molecules [10]. All the researchers had to tweak their approach to seek out, rather than weed out, toxicity [10]. The AI created thousands of new substances, some similar to the venomous agent X (VX), the most potent nerve agent ever developed [11]. Studies in healthcare have used TM and machine learning (ML) to explore public health information from patients' free-text feedback, which has rarely been fully exploited whilst valuable information may be hidden [12], [13]. Other disaster preparedness and resilience studies have utilised TM for mental health research by analysing sentiments beyond social media posts during unusual emergencies [14]. Another study utilised TM to process newspaper articles about the coronavirus-19 (COVID-19) pandemic's emotional impact on the community[15]. Overall, TM uses different approaches to help mitigate risks by building verbal input-based textual dictionaries that facilitate sentiment polarity assessment and ML algorithm modelling, enabling automated sentiment identification [16].

Expert feedback is considered an invaluable source of information that contributes to determining robust service delivery improvements in system problem-solving measures [17], [18]. Hence, using TM techniques could help produce an in-depth, objective analysis of experts' free-text feedback regarding the perspectives of MENA countries' coordination and preparedness for CBRN threats, considering the variability of political and geographical challenges and health system readiness levels across the region.

The MENA countries account for 6% of the global population [19]. They are strategically located between three continents, Asia, Africa, and Europe, which increases their risk of exposure to disasters, including deliberate and accidental CBRN incidents. The MENA region, particularly the Gulf Cooperation Council (GCC), is a global economic power with an important petroleum product hub. Over the last 20 years, the MENA region has witnessed multiple conflicts involving chemical weapons [20], significantly increasing the risk of exposure to CBRN incidents. Few studies have explored MENA experts' opinions toward regions' disaster preparedness levels [21]–[23]. To our knowledge, no previous study has used AI algorithms to explore experts' opinions concerning the MENA healthcare system's readiness to respond to CBRN incidents.

Therefore, this study used TM, including NLP and sentiment analysis, and ML techniques to analyse open-ended multidisciplinary disaster medicine experts' opinions and emotions concerning MENA countries' joint efforts and preparedness for potential CBRN incidents.

2 Methods

2.1 Study design and setting

This was a qualitative cross-sectional study of open-ended responses to a modified interview method (MIM) for multidisciplinary experts through an online link generated by the Phonic® application from October to December 2022. Phonic® allowed the participants to answer open-ended questions through recorded audio responses or free text according to their preference. Audio or textual responses were saved automatically within the application and were accessible only to the first author. International experts, selected according to their organisational roles and contributions to the scholars, were sent an invitation email explaining the study objectives and procedures. A research consent form was uploaded to Phonic®, and all participants had to sign it to access the MIM questions. The questions were written in English and French (Supplemental files). Both languages are used most in MENA from high school until post-graduate education levels and in the region's national and international scientific events, making them the most suitable for communicating scientific evidence [24], [25]. Transcripts translated into English were provided by the Phonic® application and double-checked by bilingual co-authors to ensure accuracy. By the end of the study period, responses were downloaded in a "comma-separated values" (CSV) format file. The MIM included 37 questions: 12 general questions, including five demographic questions; seven about national practices and policies; six regarding international cooperation; seven on hospital preparedness; and five about mass gatherings as the MENA frequently organise many mass-gathering events, such as the Hajj in Mecca in the Kingdom of Saudi Arabia, religious events in Karbala and Arbaeen in Iraq, Christian pilgrimages in Jerusalem, Jewish pilgrimages in Djerba in Tunisia [26], sporting events such as the recent International Federation of Association Football World Cup 2022 and the approaching basketball World Cup in 2027 in Qatar, Arabic, Asian, and African championships' leagues, and other festivals [27]–[29].

Data analyses were performed using R language accessed through the R-studio environment.

2.2 Participants

Purposeful sampling was used to determine information-rich individuals corresponding to the study objective to be selected [30]. Based on these criteria, 92 CBRN experts (76 English-speaking and 16 French-speaking) from different MENA countries, including Qatar, Kuwait, Oman, Saudi Arabia, the UAE, Egypt, Iran, Turkey, Tunisia, and Morocco, who were interested in the MENA region, were sent an invitation email explaining the study objectives. The study included experts with at least one master's degree and expertise in CBRN, toxicology, emergency and disaster medicine, or health sciences. Their expertise was defined through their organisational roles and contributions to the academic field. To ensure a meticulous selection, candidates were identified from the authors' global professional networks, prioritising those who met the inclusion criteria and were affiliated with national and international governmental and non-governmental organisations such as the MENA ministries of Health and Interior, World Health Organization and United Nations.

The research consent form, refresher leaflet, and Phonic®user guidelines, prepared in French and English, were attached to the email.

2.3 Data Analysis

The code generated in R for data cleaning and analysis is presented in the supplemental files.

2.3.1 Data Cleaning

Data cleaning is essential for rendering raw data suitable for descriptive, TM, and ML analyses. Code was created in R-Studio through the following steps: first, special characters such as "/", "@", and "|" were replaced with spaces; second, all spaces between words were removed; third, all dataset words were converted into lowercase words; fourth, stop words such as "the", "a", "is", "at", and "on", which are frequently used in the spoken language but were insignificant for the analysis, were removed; finally, numbers and punctuation were removed.

2.3.2 NLP and sentiment analysis

First, a word count was performed, followed by creating a "word cloud". This enabled easy visualisation of the most common words in the text according to size. The larger the word size on the word cloud, the more frequently the word was mentioned. The word cloud oriented us to the key themes the experts showed interest in.

Subsequently, the correlation coefficient was calculated for the top six repeated words with the remaining words in the text using the function "findAssocs()". This helped provide insights into the context in which these words were mentioned. After this, sentiment analysis was performed by calculating the sentiment scores of the participants using the "get_sentiment()" and "get_nrc_sentiment ()" functions. Several packages in R calculate sentiment scores and determine their ranges differently. In this study, the "Syuzhet package" was used [31]. It works well with other popular text analysis packages in R, such as "tm" and "tidytext", making it easy to incorporate sentiment analysis into a broader TM workflow and allowing multiple sentiment extraction methods [31], [32]. Upon examining the first vector score, the sentiment scores ranged from -0.5 (representing the most negative) to 10.3 (representing the most positive). A summary of the descriptive statistics of the sentiment scores and the Shewhart control chart were used to observe score variation among the participants. A Pareto chart was used to observe the overall sentiment score according to the participants' nationalities.

2.3.3 Unsupervised ML

Supervised and unsupervised ML can be used to analyse open-ended responses. Supervised ML is primarily used for sentiment analysis because it is labelled and known. This study used unsupervised ML to explore unlabelled text data and identify any existing similarities provided by the participants. Cluster analysis was conducted to explore the textual data [33]. The fewer the clusters, the better, indicating a consensus in opinions. It is widely used to explore data through open-ended text responses [34]. Hierarchical and k-means algorithms were used in this study. A dendrogram plot was used to identify the clusters in both algorithms [33]. The silhouette coefficient, which ranges between -1 and 1, was used to assess model accuracy. A value between 0 and 1 indicates good clustering. The closer the value is to 1, the better. A model with a high silhouette coefficient is preferred [33]. A principal component analysis (PCA) was conducted to identify participants' clusters.

3 Ethical board approval

This study was approved by the ethical review board of the Faculty of Medicine, "Ibn Eljazzar" of Sousse in Tunisia (CEFMS 110/2022) and by the Hamad Medical Corporation's Medical Research Centre Institutional Review Board Committee in Qatar (MRC-01-22-258).

4 Results

4.1 Demographic data

Thirty-five participants agreed to participate in this study. Twenty-nine participants achieved data saturation; hence, further data collection was stopped. The mean age of the participants was 34.37 years (standard deviation: 8.01 years). Table 1 and Figure 1 present the demographic characteristics of the study participants.

Table 1: Participants' demographic information.

<table>
<tr><th>Experts' demographic information</th><th colspan="6"></th><th>Count</th></tr>
<tr><td rowspan="12">Background</td><td colspan="6">Ministry of Health</td><td>7</td></tr>
<tr><td colspan="6">Medical doctor</td><td>5</td></tr>
<tr><td colspan="6">Academician</td><td>4</td></tr>
<tr><td colspan="6">CBRN instructor</td><td>2</td></tr>
<tr><td colspan="6">Ministry of interior</td><td>2</td></tr>
<tr><td colspan="6">Pharmacist</td><td>2</td></tr>
<tr><td colspan="6">Head of an emergency department</td><td>2</td></tr>
<tr><td colspan="6">Military</td><td>1</td></tr>
<tr><td colspan="6">Disaster medicine planning</td><td>1</td></tr>
<tr><td colspan="6">International Federation of Red Cross</td><td>1</td></tr>
<tr><td colspan="6">United Nations</td><td>1</td></tr>
<tr><td colspan="6">World Health Organization</td><td>1</td></tr>
<tr><td rowspan="2">Gender</td><td colspan="6">Female</td><td>11</td></tr>
<tr><td colspan="6">Male</td><td>18</td></tr>
<tr><td rowspan="2">Prefered Spoken Language to respond to the MIM</td><td colspan="6">English</td><td>23</td></tr>
<tr><td colspan="6">French</td><td>6</td></tr>
<tr><td rowspan="2">Age (years)</td><td>Min.</td><td>1st Qu.</td><td>Median</td><td>Mean</td><td>3rd Qu.</td><td>Max.</td><td>SD</td></tr>
<tr><td>28</td><td>41</td><td>48</td><td>47.37</td><td>54</td><td>64</td><td>8.01</td></tr>
</table>

Abbreviations: MIM, Modified interview method; CBRN, chemical, biological, radiological, and nuclear; Min, minimum; Qu, quartile; Max, maximum; SD, standard deviation.

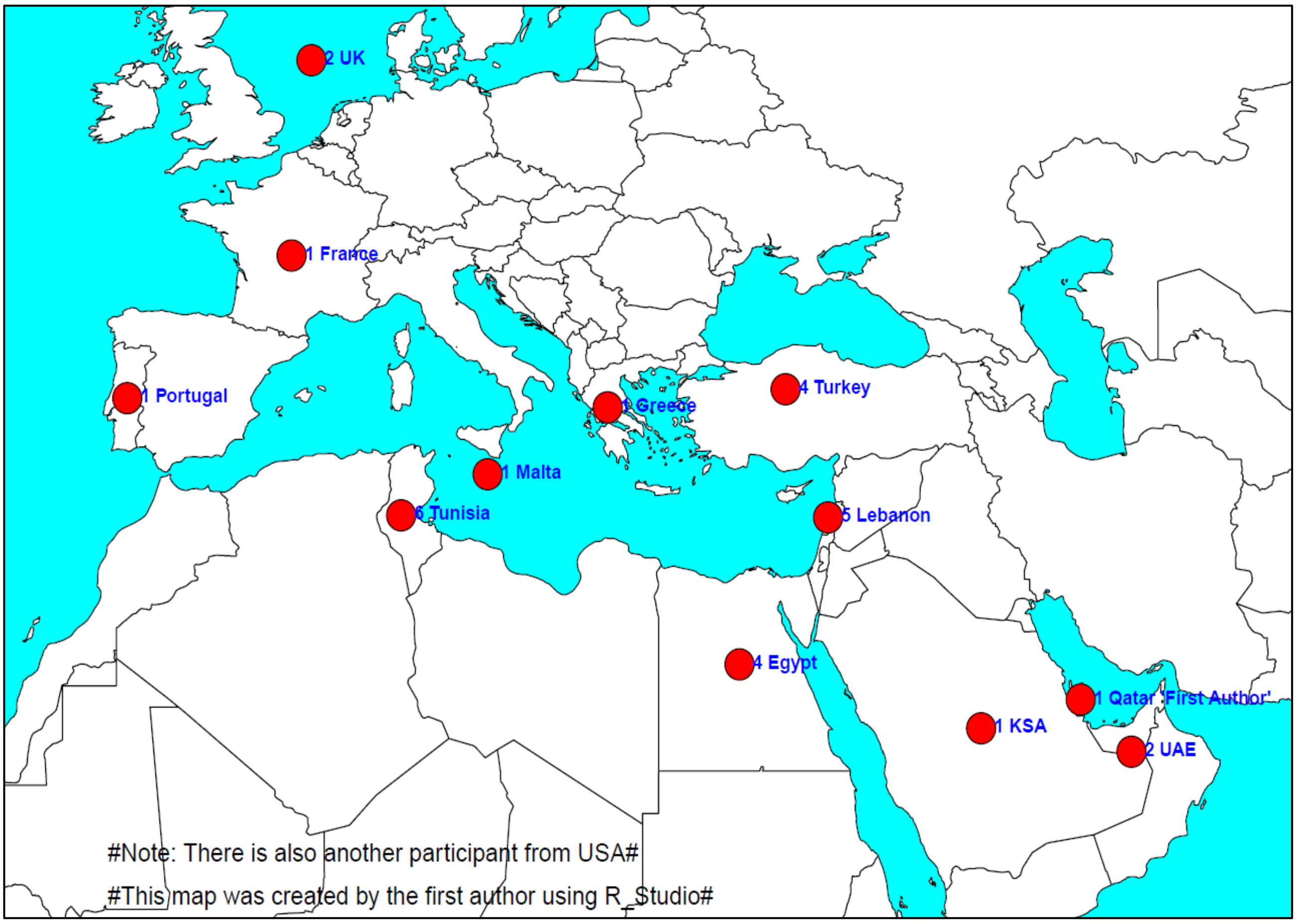

Figure 1: The geographic distribution of participants according to their country of residence.

4.1.1 Exploratory data with NLP and sentiment analysis

The word cloud in Figure 2 (A) demonstrates the most repeated words used by the participants represented by their size; the bigger they were, the more repeated they were. The scatter plot in Figure 2 (B) shows the top 50 words mentioned by the participants. The four primary prominent stem words were “plan” (n=54), “plans” (n=50), “training” (n=46), and “disaster” (n=38).

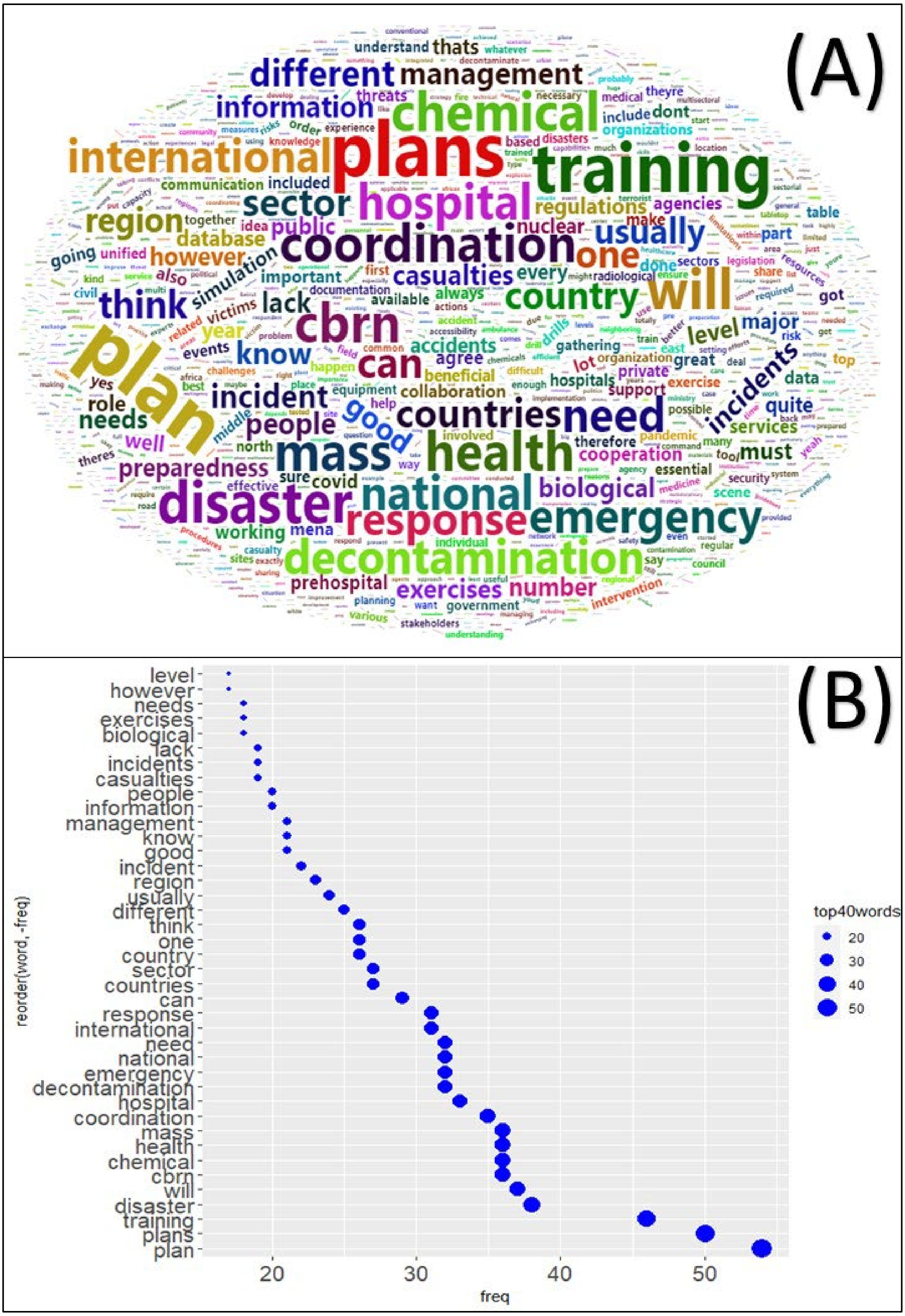

Figure 2: (A) Participants' word cloud based on their responses to the questions; (B) Scatter plot of the top 40 words mentioned by the participants.

Table 2 presents the correlation coefficients for the five primary words mentioned by the participants. First, “plan/plans” is highly correlated with words like “country”, “needed”, and “international”, among others. Participants mentioned a lack of available and accessible plans, which could cause challenges when managing CBRN emergencies. The word “training” was highly correlated with words like “incident”, “full”, and “preparedness”. Participants mentioned that frequent training sessions to explain the health outcomes of failure in managing a CBRN threat helped ensure adequate preparedness. The word “disaster” was highly correlated with “specialised”, “council”, and “activity”. The participants agreed that committees or councils, at the hospital and national level, must ensure readiness plans for each type of disaster according to a country’s identified risks. Furthermore, “CBRN” was unsurprisingly mainly correlated with “chemical” and “biological”. According to the participants, these were the most common threats out of the wider category of CBRN.

Table 2: The correlation coefficients of six prominent words in the top 10.

Plans		Plan		Training		Disaster		Will		CBRN	
Words	**Coeff**	**Word**	**Coeff**	**Word**	**Coeff**	**Word**	**Coeff**	**Word**	**Coeff**	**Word**	**Coeff**
Needed	0.85	Country	0.91	Much	0.83	Specialised	0.83	Lots	0.84	Chemicals biological	0.84
Just	0.84	Thing	0.9	Incident	0.72	Council	0.79	Politics	0.84	City there	0.84
Happens	0.83	Maybe	0.88	Use	0.72	Related	0.76	Chemicals	0.83	Complex problematic	0.84
Mass	0.83	International	0.86	Full	0.7	Activity	0.74	Happens	0.83	Define	0.84
Something	0.83	Extra	0.86	Sorry	0.68	Affected	0.74	Levels	0.83	Employees	0.84
Levels	0.82	Pre	0.85	Preparedness	0.66	Afflicted	0.74	Help	0.82	Equipped	0.84
Different	0.81	Issue	0.85	Course	0.65	Archive	0.74	Ambulance	0.81	Expert	0.84
Issue	0.8	Meet	0.84	Service	0.65	Assembled	0.74	Difficult	0.79	Hot	0.84
Extra	0.79	Try	0.84	Question	0.64	Bridged	0.74	Plans	0.78	Employees	0.84
Will	0.78	Whole	0.84	Trauma	0.64	Archive	0.74	Common	0.77	Equipped	0.84

Abbreviations: CBRN, chemical, biological, radiological, and nuclear; Coeff, coefficient.

The participants’ sentiment scores were calculated using “get_nrc_sentiment ()” of the Syzhut library. They were then plotted in the Shewhart chart in Figure 3 (a) using “qcc” library. The scores varied between -4, 4 and 15, and 25 with a mean of 3.11 (standard deviation: 5.22). The higher the score, the better. The Shewhart chart in Figure 3 (a) indicates that the variation in the sentiment scores between the participants was within the control limits with no significant difference, except for three participants (Figure 3 (b)) from Greece, the UK, and Portugal. They had high scores, likely reflecting positive thinking due to their experience in the field. For example, a participant from Greece with the highest sentiment score had over 35 years of experience as a military physician, had experience in hospitals' emergency departments' CBRN preparedness and response since 2001, had been involved in the 2004 Olympic Games CBRN preparedness, and was an international CBRN instructor working together with the Organisation for the Prohibition of Chemical Weapons to train the Olympic CBRN Response Unit of Army General Hospital of Athens. This is reflected in the Pareto chart in Figure 3 (c), which shows sentiment score distribution by nationality. The data was categorised as follows: positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Figure 3 (d) shows the sentiment scores classified according to the sentiment type and distribution.

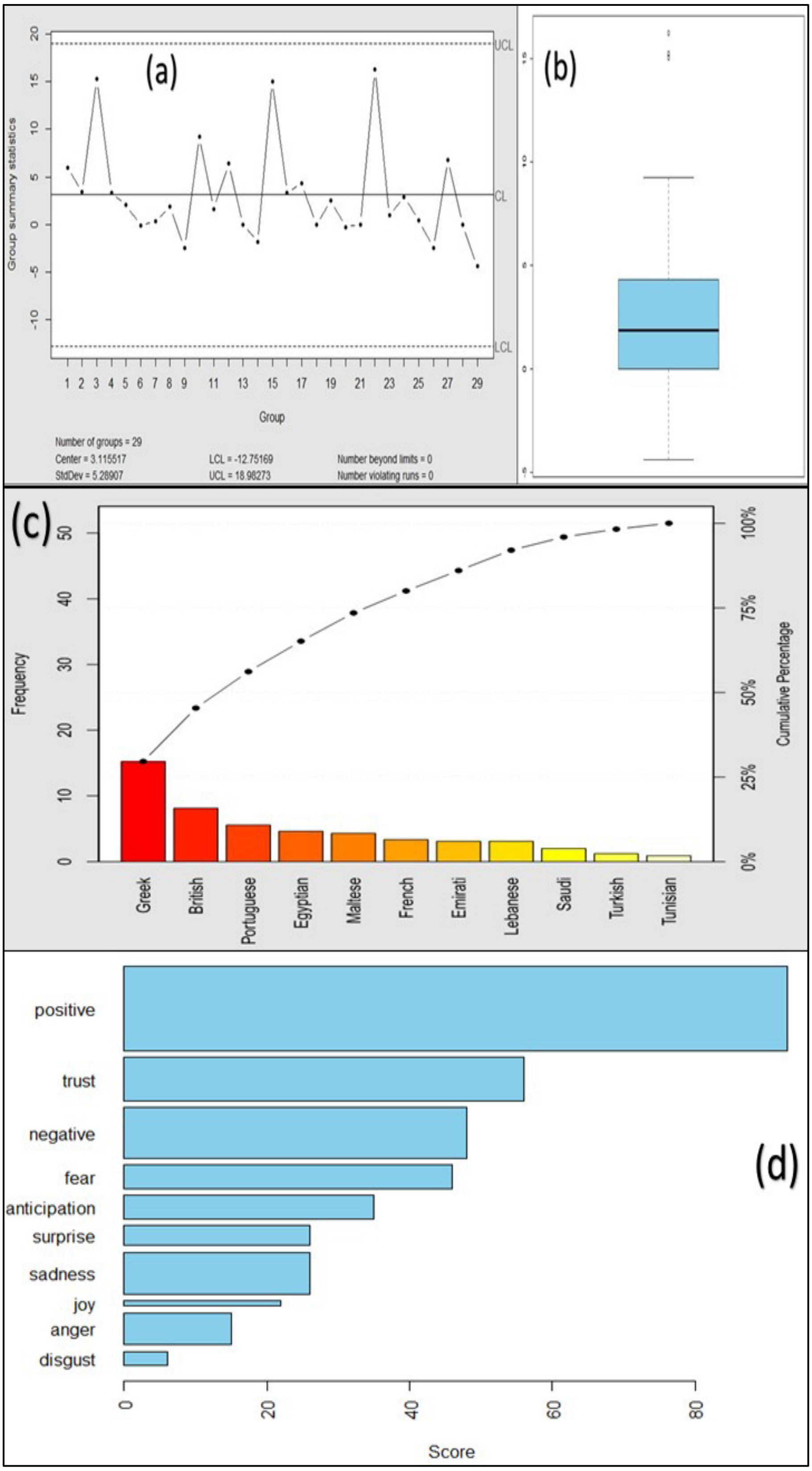

Figure 3: Shewhart (a) and boxplot chart; (b) of participants' sentiments scores; (c): Pareto chart of the participants' sentiment scores by country of residence; (d): Participants' sentiments identified during the modified interview.
StdDev: Standard deviation/ UCL=Upper control limit/ LCL=Lower control limit

In Figure 4, the coordinated parallel plot allows for the visualisation of multivariate data in a multidimensional manner. The correlation between emotions and text was calculated using the "COR ()" function. The resulting correlation matrix was used to create a heatmap using "ggplot2" and "reshape2" packages. Finally, the parallel coordinate plot and heat map displayed the combined chart using the "plot_grid()" function of the "cowplot" package. The parallel coordinate plot shows the distribution and relationships between different emotional variables.

The combined heatmap and parallel plot show the relationship between emotions and words or phrases mentioned in the open-ended text data related to CBRN preparedness in the MENA region. The parallel coordinate plot shows the relationships between each emotion within the text, whereas the heatmap shows the correlation between these variables and the text data of the words. The combination of these two plots assessed the relationship between the emotions and language used in the dataset. In Figure 4 (A), the emotions of "sadness", "negative", and "fear" have higher medians, as shown by the boxplots, meaning they were the emotions with higher values in the text data.

Furthermore, these emotions, except "disgust" and "sadness", converge in the word "accessibility", indicating that they were mainly associated with it. In contrast, the emotions of "positive", "trust", and "joy" diverge from the word "accessibility", which suggests that there was no association. This is also confirmed by the heatmap in Figure 4 (B), which shows the correlation between emotions and words. The closer the coefficient is to 1, the stronger the correlation.

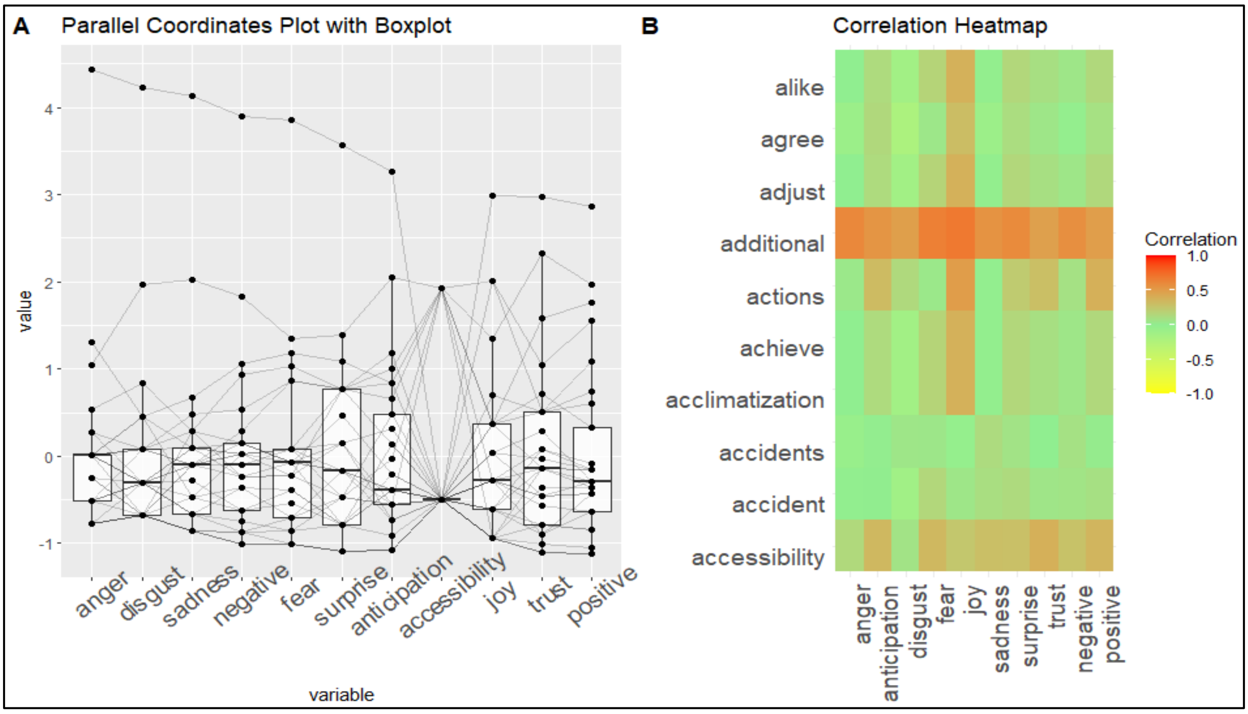

Figure 4: Combined heatmap and parallel plot with boxplot for participants' emotions. (A) Parallel coordinates plot with boxplots of participants' emotions. (B) Correlation heatmap between emotions and words.

In the heatmap (Figure 4 B), each row represents a different emotion, and each column represents a different word or phrase. The colour of each cell represents the frequency of the corresponding word or phrase in the text data associated with the corresponding emotion; darker colours indicate a higher frequency.

The parallel plot (Figure 4 A) shows how different emotions are related and the frequencies of words or phrases. Each line in the parallel plot represents a different emotion, and the vertical lines connecting the lines represent the values of each variable, that is, the frequency of the corresponding word or phrase.

4.1.2 Clustering analysis

First, the two-cluster method was utilised for the k-means algorithm as determined by the silhouette coefficient (s=0.57), which ranges between -1 and 1. A value close to 1 indicates good clustering. A silhouette plot was used to determine the distances between the resulting clusters.

This helps assess whether a cluster has a high degree of compatibility within its designated cluster and demonstrates low compatibility with neighbouring clusters. The dendrogram in Figure 5 (a) and (b) show the clusters according to colour. Figure 5 (c) and (d) show that the silhouette coefficient is 0.34, indicating the efficiency of the selected clustering in providing well-structured data.

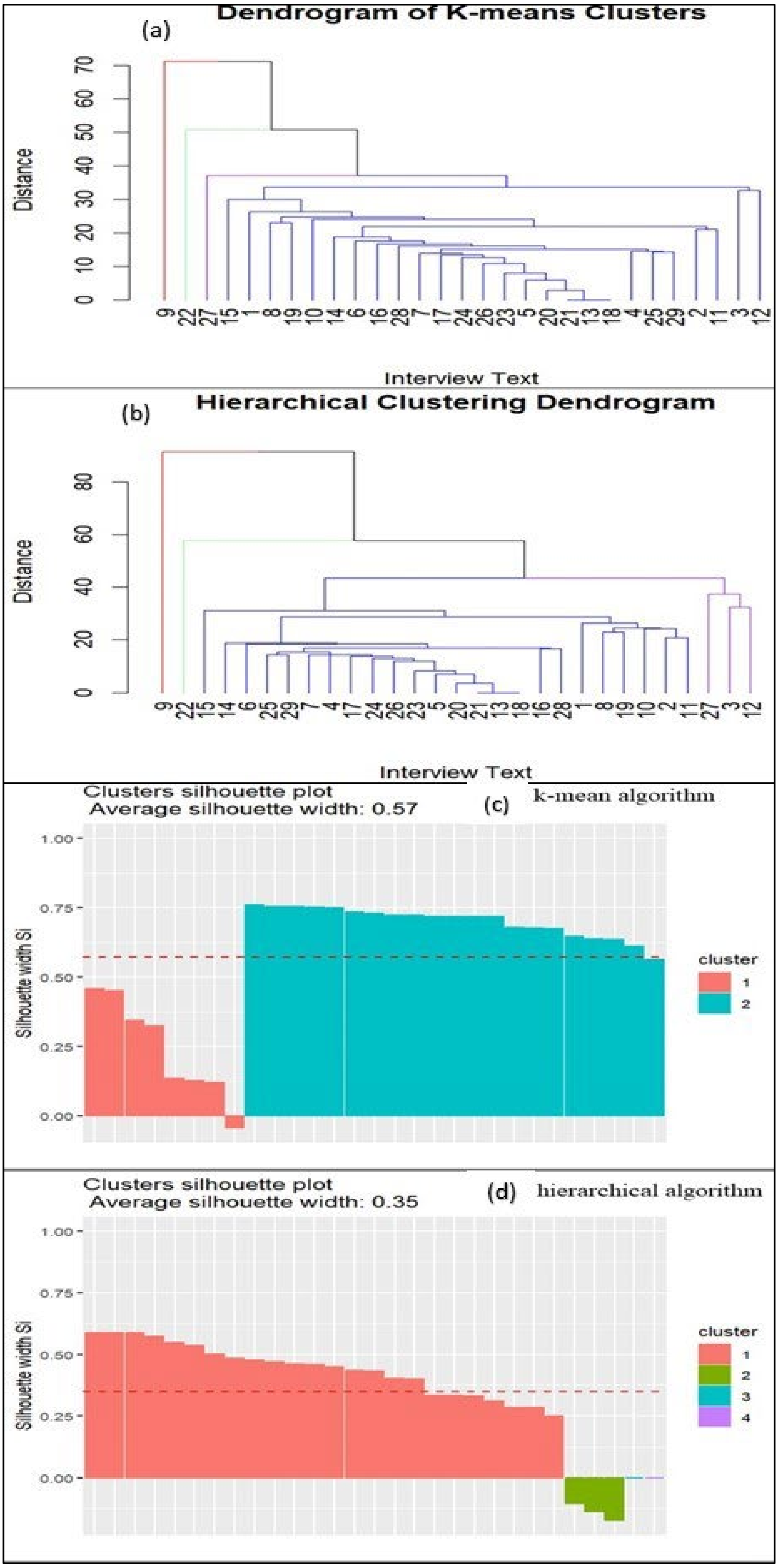
Dendrogram of K-means Clusters
(a)
Distance
Interview Text
Hierarchical Clustering Dendrogram
(b)
Distance
Interview Text
Clusters silhouette plot
Average silhouette width: 0.57
(c)
k-mean algorithm
Silhouette width Si
cluster
Clusters silhouette plot
Average silhouette width: 0.35
(d)
hierarchical algorithm
Silhouette width Si
cluster

Figure 5: (a) Dendrograms of K-mean clustering algorithm; (b): Dendogram of Hierarchical clustering algorithm; (c): Cluster silhouette plot for K-mean algorithms; (d) Cluster silhouette plot for Hierarchical algorithm

PCA was conducted to determine principal components 1 (PC1) and 2 (PC2), as shown in Figure 6, which depicts the distribution of data points depending on their values. Figure 8 shows that the texts in each cluster tend to cluster together, establishing different groupings.

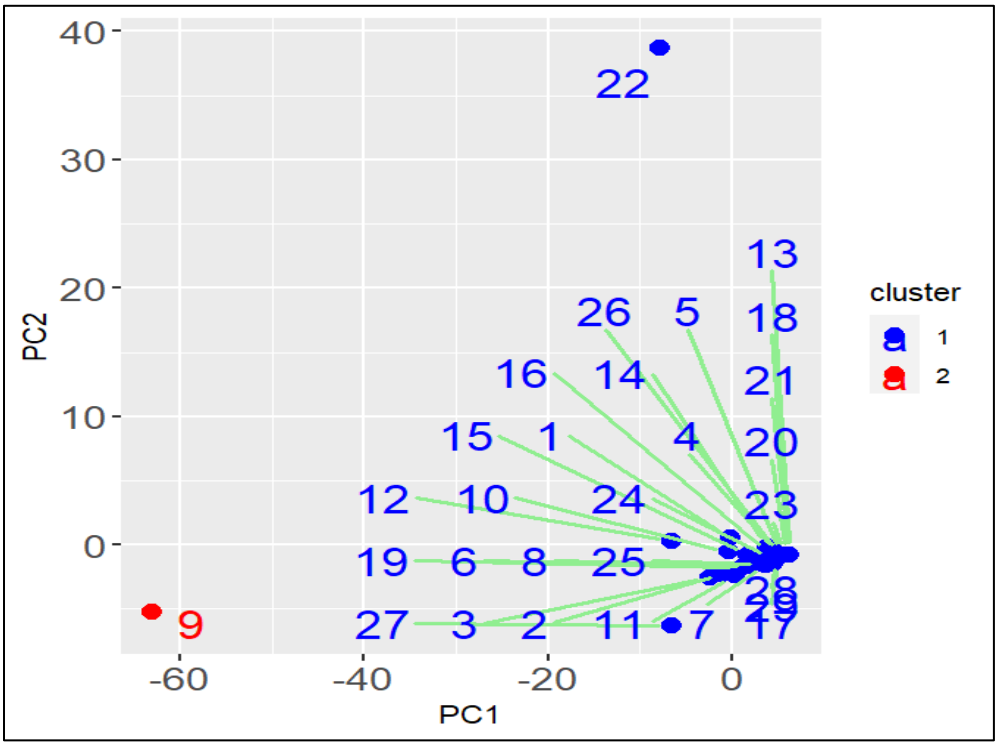

Figure 6: Participants' clustering according to principal compoenents 1 and 2.

In Figure 6, each point represents a participant's response, and the colours of the points determine the clusters.

5 Discussion

Undoubtedly, learning from the opinions of academically trained experts in disaster management, particularly CBRN, is crucial for decision-making and may help ensure the continuous improvement of preparedness and response plans. While using open-ended questionnaires and interviews may be time-consuming, they allowed us to explore various profound aspects of the study objective. Furthermore, the nature of the unstructured responses allowed the participants to express their opinions freely. However, determining whether experts have a consensus is a complex assessment that can be subjective through the researcher's interpretation and own views. Hence, in this study, TM and ML converted the textual input into numbers using a logical sequence of learning steps of a clustering algorithm and built an explanatory model through a dictionary-based approach to express their sentiments. They performed an in-depth analysis of open-ended feedback from the modified interview (MI) by transforming the text input data into numeric values and providing an overview of the most frequently used words in the form of a word cloud. They also identified the participants' sentiments, understood their correlation with the textual input, clustered the respondents' opinions and emotions according to their responses, and helped analyse the correlation between the words and sentiments (Figure 4 and Table 2).

Furthermore, sentiment analysis is considered a powerful tool in the medical field for identifying linguistic quirks of opinions and the emotions behind them [35]. In healthcare, for example, a substantial amount of feedback data is being gathered on the patient experience. However, this has not been exploited fully or adequately because of the difficulty in determining a pattern of free-text feedback. Sentiment analysis enabled an in-depth exploration of free-text data and an understanding of the emotions behind it, which was similarly utilised in a recent study analysing patient experience feedback [36].

In this study, although the experts had positive sentiments toward improving the MENA intercountry cooperation, fear and sadness were also identified regarding the accessibility of preparedness plans by healthcare staff at the hospital and national level (Figures 3 and 4). Access to a documented contingency plan was also identified as a primary challenge in effectively ensuring readiness to manage biological threats such as pandemics [36]. Furthermore, a few MENA studies have emphasised that a contingency plan must remain available to everyone, allowing multiple testing and evaluation of these plans by all interventions and guiding their improvements [37]. Experiences during the Severe Acute Respiratory Syndrome pandemic in

2002, the Middle East Respiratory Syndrome in 2012, and the latest worldwide COVID-19 pandemic have highlighted that designing response plans and disseminating them without testing may frustrate healthcare professionals if no prior training is conducted [38]. Testing, adapting, and updating plans is a never-ending process, in agreement with the disaster management cycle [39]. Another study showed that participants' confidence in responding to CBRN disasters increased with practice [40].Additionally, the clustering determined by unsupervised ML in Figures 5 and 6 helped to identify that most participants agreed with the elements raised in the MIM questions (MIMQ). These questions explored their opinions on a few themes, including CBRN preparedness, international cooperation perspectives within the MENA region, the value of tabletop exercises as a training modality for CBRN, and the role of the World Health Organization in leading this coordination within the region. In the same context, studies have identified that appropriate preparedness can be ensured based on adequate planning and training for all healthcare professionals [41]–[43]. Research in Saudi Arabia has also recommended implementing disaster management training in medical school undergrad programs to improve disaster preparedness metrics [44]. Despite identifying negative and fearful feelings toward some points, such as accessibility (Figure 4), the experts expressed similar positive sentiments, indicating a consensus toward the overall perspectives raised in the MIMQ.

Further, the first triad within the top 10 words, the experts mentioned "planning, training, and coordination" (Figure 2 (b)). A few studies have identified effective coordination and appropriate coordinated training as crucial to ensure the appropriate readiness of healthcare professionals for CBRN incidents in hospitals, thereby promoting the development of this concept at the regional level in MENA [43], [45]. However, the MENA experiences during previous re-emergent pandemics over the last two decades have identified a few challenges at different levels. These include challenges in the academic and research infrastructure that varied significantly between MENA countries, limiting initiatives and promoting a culture of not sharing information

[19]. This led to the isolation of the MENA academic research sectors of worldwide scientific advancement and weakened the healthcare readiness level for all CBRN threats, including re-emerging pandemics, by following other countries' newly identified therapeutic guidelines and not leading to any emerging initiatives [19]. A recent study in the MENA region identified that despite the availability of healthcare experts in the MENA region and the GCC as financial and industrial powers, no published clinical trials were conducted in the region during the Swine Flu, the Severe Acute Respiratory Syndrome, Ebola, Middle East Respiratory Syndrome, except a few related to the COVID-19 pandemic [19], [46].

The second triad within the top 10 words the experts mentioned was "health, mass, and chemical". Previous worldwide experiences in Japan, the UK, and a few MENA countries have taught us that chemical and biological agents are easier to weaponise and utilise in attacks, exhausting health-sector countermeasures [47]. They also identified the most accessible elements to be weaponised and utilised by terrorist groups in MENA [48]. Their immediate effects lead to the rapid onset of acute respiratory syndrome, which is sometimes fatal [49]. Furthermore, previous studies have identified that healthcare professionals, if not well-trained, are easily exposed to the secondary contamination of chemical and biological agents, including viruses, when delivering life-saving interventions [50], [51], explaining the concerns of the interviewed experts on the preparedness of healthcare staff for chemical and biological agents. In addition, the frequent mention of the word "mass" indicates that these incidents aim to create the highest number of casualties possible, resulting in mass exposures. A WHO report generated in Beirut in 2019 identified that most MENA countries have not conducted a national risk assessment for hosting large crowds or any focused risk assessment for specific large crowds, despite the fact that the MENA region annually hosts a number of the world's largest mass-gathering events. For example, the Muslim pilgrimage in Mecca and the gatherings in Arbaeen in Iraq are attended annually by worshipers from hundreds of neighbouring countries [52]. These events can be targeted by terrorist attacks that use chemical and biological agents and radiological dispersal or emitting devices. Therefore, the WHO encouraged the MENA countries to meet frequently, share their best practices, and enhance collaboration for mass-gathering incident preparedness.

The third triad within the top 10 words the experts mentioned was "disaster, CBRN, and will", which was the MIMQ's main interest.

While sentiment analysis inherently captures subjective perceptions, our integration of ML and TM aimed to systematically reduce interpretative biases [53]. This approach strived for a balanced, structured analysis of expert opinions, fostering a reproducible framework for understanding nuanced sentiments in the complex realm of disaster medicine.

Further, in light of our findings, outlining potential pathways forward for CBRN preparedness in the MENA region is imperative. First, a concerted effort is needed to make preparedness information more accessible to healthcare professionals. This could be achieved by establishing regional knowledge hubs or online platforms dedicated to CBRN preparedness. Such platforms could host webinars, training modules, and best practice guidelines, ensuring professionals have a one-stop shop for all their information needs. Secondly, fostering inter-country collaborations can streamline resource-sharing and knowledge transfer. Countries can collectively enhance their preparedness levels by leveraging strengths and sharing challenges. Lastly, policymakers must initiate public-private partnerships, bringing in expertise and resources from the private sector and ensuring that preparedness measures are holistic and well-rounded. This can be similar to the "CBRNE Research & Innovation Conference", which epitomises collaborative engagement with an array of professionals ranging from Emergency Medical Services to esteemed academics from institutions like the University of Strasbourg. It offers an interdisciplinary platform and emphasis on cutting-edge topics, from CBRN detection and protection to medical countermeasures and forensic sciences, which resonates with our study's aspirations [54]. Such events underscore the essence of face-to-face dialogues, workshops, and experiential learning, which could be indispensable for replicating and enhancing the outcomes of our research in wider contexts. With these steps, the MENA region can look forward to a more robust, informed, and collaborative approach to CBRN preparedness.

While the MENA region's significance in disaster medicine is paramount, our study sourced insights from leading experts despite a 31% (n=29) participation rate. This cohort sufficed for data saturation, underscoring that quality often prevails over quantity in qualitative research [55]. Although broader participation might have enriched perspectives, the depth of expertise captured was robust. AI analysis techniques enhanced our data's richness, spotlighting core themes and ensuring the findings' relevance and rigor.

6 Limitation

This study has several limitations. First, experts from some MENA countries, such as Qatar, Jordan, Iran, Oman, Morocco, and Kuwait, did not accept to participate. To foster a collaborative approach to CBRN readiness, it is imperative to comprehensively understand each country's unique challenges and strengths within the MENA region. While our study provides crucial insights from the participating countries, we acknowledge the limitation of non-responses from certain nations. Engaging with these non-responding countries in future research is vital. Their perspectives will enrich the dialogue and facilitate a more unified and holistic regional strategy, ensuring that collaborative CBRN preparedness efforts are both inclusive and effective.

Second, despite using the Phonic® application, which helped ensure a smooth process, participation in open-ended questionnaires and interviews was time-consuming for experts with responsibilities in their respective states. Third, TM and ML algorithms' performance could improve with the input text's increase. Hence, finding more disaster medicine experts who fulfil the inclusion criteria and are willing to participate and help ensure the information's diversity was challenging.

7 Conclusion

The results of the sentiment analysis showed that the overall sentiment towards preparedness and response capacity of the health sector in the MENA region was generally positive. However, some experts expressed concerns about various challenges, including infrastructure and training. These insights can be used to identify areas that require improvement and inform policies to enhance the preparedness and response capacity of the health sector in the MENA region. Moreover, this study demonstrated the potential of NLP techniques for analysing experts' opinions on complex issues related to public health and safety. Such methodologies may be a linchpin, offering policymakers and stakeholders invaluable insights, thereby facilitating judicious decision-making in enhancing the MENA region's readiness for CBRN incidents.

References

[1] E. Karmon, “Are We Nearer to a Major CBRN Terrorism Threat?,” in *Non-Traditional Security Challenges in Asia*, Routledge, 2015.

[2] S. Salama and L. Hansell, “Does Intent equal Capability? Al-Qaeda and Weapons of Mass Destruction,” *Nonproliferation Rev.*, vol. 12, no. 3, pp. 615–653, Nov. 2005, doi: 10.1080/10736700600601236.

[3] R. Ali. Shahzeb, “Is the Threat of ISIS Using CBRN Real?,” vol. 8, no. 2, pp. 4–10, 2016.

[4] R. Pangi, “Consequence Management in the 1995 Sarin Attacks on the Japanese Subway System,” *Stud. Confl. Terror.*, vol. 25, no. 6, pp. 421–448, Nov. 2002, doi: 10.1080/10576100290101296.

[5] T. Okumura *et al.*, “Report on 640 Victims of the Tokyo Subway Sarin Attack,” *Ann. Emerg. Med.*, vol. 28, no. 2, pp. 129–135, Aug. 1996, doi: 10.1016/S0196-0644(96)70052-5.

[6] European Network of CBRN Training Centers, “The eNotice Project,” 2023. https://www.h2020-enotice.eu/static/project.html (accessed Mar. 22, 2023).

[7] A. Baku *et al.*, “(231) European Front-Line Health Professionals and the New Public Health Threats: Assessment of Training Needs,” *Prehospital Disaster Med.*, vol. 22, no. S1, pp. S135–S136, Apr. 2007, doi: 10.1017/S1049023X0006369X.

[8] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, “Causability and explainability of artificial intelligence in medicine,” *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 4, p. e1312, 2019, doi: 10.1002/widm.1312.

[9] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, and E. Mayer, “Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review,” *BMJ Health Care Inform.*, vol. 28, no. 1, p. e100262, Mar. 2021, doi: 10.1136/bmjhci-2020-100262.

[10] F. Urbina, F. Lentzos, C. Invernizzi, and S. Ekins, “Dual use of artificial-intelligence-powered drug discovery,” *Nat. Mach. Intell.*, vol. 4, no. 3, Art. no. 3, Mar. 2022, doi: 10.1038/s42256-022-00465-9.

[11] J. E. Chambers and E. C. Meek, “Novel centrally active oxime reactivators of acetylcholinesterase inhibited by surrogates of sarin and VX,” *Neurobiol. Dis.*, vol. 133, p. 104487, Jan. 2020, doi: 10.1016/j.nbd.2019.104487.

[12] L. Ngamassi, H. Shahriari, T. Ramakrishnan, and S. Rahman, “Text mining hurricane harvey tweet data: Lessons learned and policy recommendations,” *Int. J. Disaster Risk Reduct.*, vol. 70, p. 102753, Feb. 2022, doi: 10.1016/j.ijdrr.2021.102753.

[13] N. Chintalapudi, G. Battineni, M. D. Canio, G. G. Sagaro, and F. Amenta, “Text mining with sentiment analysis on seafarers’ medical documents,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100005, Apr. 2021, doi: 10.1016/j.jjimei.2020.100005.

[14] L. Zou, N. S. N. Lam, H. Cai, and Y. Qiang, “Mining Twitter Data for Improved Understanding of Disaster Resilience,” *Ann. Am. Assoc. Geogr.*, vol. 108, no. 5, pp. 1422–1441, Sep. 2018, doi: 10.1080/24694452.2017.1421897.

[15] A. Meyer, W. Walter, and S. Seuring, “The Impact of the Coronavirus Pandemic on Supply Chains and Their Sustainability: A Text Mining Approach,” *Front. Sustain.*, vol. 2, 2021, Accessed: Mar. 31, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frsus.2021.631182

[16] M. Pejić Bach, Ž. Krstić, S. Seljan, and L. Turulja, “Text Mining for Big Data Analysis in Financial Sector: A Literature Review,” *Sustainability*, vol. 11, no. 5, Art. no. 5, Jan. 2019, doi: 10.3390/su11051277.

[17] T. Krueger, T. Page, K. Hubacek, L. Smith, and K. Hiscock, “The role of expert opinion in environmental modelling,” *Environ. Model. Softw.*, vol. 36, pp. 4–18, Oct. 2012, doi: 10.1016/j.envsoft.2012.01.011.

[18] Y. Duan, J. S. Edwards, and M. X. Xu, “Web-based expert systems: benefits and challenges,” *Inf. Manage.*, vol. 42, no. 6, pp. 799–811, Sep. 2005, doi: 10.1016/j.im.2004.08.005.

[19] H. Ibrahim, A. M. Kamour, T. Harhara, W. H. Gaba, and S. C. Nair, “Covid-19 pandemic research opportunity: Is the Middle East & North Africa (MENA) missing out?,” *Contemp. Clin. Trials*, vol. 96, p. 106106, Sep. 2020, doi: 10.1016/j.cct.2020.106106.

[20] G. D. Koblentz, “Chemical-weapon use in Syria: atrocities, attribution, and accountability,” *Nonproliferation Rev.*, vol. 26, no. 5–6, pp. 575–598, Sep. 2019, doi: 10.1080/10736700.2019.1718336.

[21] O. F. Al Kurdi, “A critical comparative review of emergency and disaster management in the Arab world,” *J. Bus. Socio-Econ. Dev.*, vol. 1, no. 1, pp. 24–46, Jan. 2021, doi: 10.1108/JBSED-02-2021-0021.

[22] N. M. Momani, “Academic role in disasters and crises management: opportunities and challenges,” *Int. J. Emerg. Manag.*, vol. 13, no. 2, pp. 150–159, Jan. 2017, doi: 10.1504/IJEM.2017.083113.

[23] E. Mastroianni, J. Lancaster, B. Korkmann, A. Opdyke, and W. Beitelmal, “Mitigating infrastructure disaster losses through asset management practices in the Middle East and North Africa region,” *Int. J. Disaster Risk Reduct.*, vol. 53, p. 102011, Feb. 2021, doi: 10.1016/j.ijdrr.2020.102011.

[24] A. Akkari, “Education in the Middle East and North Africa: The Current Situation and Future Challenges,” *Int. Educ. J.*, vol. 5, no. 2, pp. 144–153, 2004.

[25] D. S. Bouhlila, “The Quality of Secondary Education in the Middle East and North Africa: What Can We Learn from TIMSS’ Results?,” *Comp. J. Comp. Int. Educ.*, vol. 41, no. 3, pp. 327–352, 2011, doi: 10.1080/03057925.2010.539887.

[26] The Arab Weekly, “Jewish pilgrimage to Djerba goes ahead on smaller scale amid pandemic |,” *The Arab Weekly*, 2021. Accessed: Apr. 28, 2023. [Online]. Available:

https://thearabweekly.com/jewish-pilgrimage-djerba-goes-ahead-smaller-scale-amid-pandemic

[27] G. Alinier *et al.*, "Medical resources deployed for the 2019 World Athletics Championships in Doha, Qatar," *J. Emerg. Med. Trauma Acute Care*, vol. 2022, no. 1-Qatar Health 2022 Conference abstracts, p. 3, Jan. 2022, doi: 10.5339/jemtac.2022.qhc.3.

[28] M. Peyravi, M. Ahmadi Marzaleh, and H. Najafi, "An Overview of Health-Related Challenges in a Mass Gathering," *Trauma Mon.*, vol. 25, no. 2, pp. 78–82, Mar. 2020, doi: 10.30491/tm.2020.213574.1022.

[29] A. Alyami, C. L. Dulong, M. Z. Younis, and S. Mansoor, "Disaster Preparedness in the Kingdom of Saudi Arabia: Exploring and Evaluating the Policy, Legislative Organisational Arrangements Particularly During the Hajj Period," *Eur. J. Environ. Public Health*, vol. 5, no. 1, p. em0053, Jul. 2020, doi: 10.29333/ejeph/8424.

[30] L. A. Palinkas, S. M. Horwitz, C. A. Green, J. P. Wisdom, N. Duan, and K. Hoagwood, "Purposeful sampling for qualitative data collection and analysis in mixed method implementation research," *Adm. Policy Ment. Health*, vol. 42, no. 5, pp. 533–544, Sep. 2015, doi: 10.1007/s10488-013-0528-y.

[31] J. Silge and D. Robinson, *Text Mining with R: A Tidy Approach*. O'Reilly Media, Inc., 2017.

[32] P. Verma, A. Dumka, A. Ashok, A. Bhardwaj, and A. Dumka, *COVID-19: Origin, Detection and Impact Analysis Using Artificial Intelligence Computational Techniques*. CRC Press, 2021.

[33] E. A. L. M. Molenaar, D.-J. J. Barten, A. M. S. de Hoop, N. Bleijenberg, N. J. de Wit, and C. Veenhof, "Cluster analysis of functional independence in community-dwelling older people," *BMC Geriatr.*, vol. 22, no. 1, p. 996, Dec. 2022, doi: 10.1186/s12877-022-03684-2.

[34] V. Riccio, G. Jahangirova, A. Stocco, N. Humbatova, M. Weiss, and P. Tonella, “Testing machine learning based systems: a systematic mapping,” *Empir. Softw. Eng.*, vol. 25, no. 6, pp. 5193–5254, Nov. 2020, doi: 10.1007/s10664-020-09881-0.

[35] L. Abualigah, H. E. Alfar, M. Shehab, and A. M. A. Hussein, “Sentiment Analysis in Healthcare: A Brief Review,” in *Recent Advances in NLP: The Case of Arabic Language*, M. Abd Elaziz, M. A. A. Al-qaness, A. A. Ewees, and A. Dahou, Eds., in Studies in Computational Intelligence. Cham: Springer International Publishing, 2020, pp. 129–141. doi: 10.1007/978-3-030-34614-0_7.

[36] M. T. Khan and S. Khalid, “Sentiment Analysis for Health Care,” *Big Data: Concepts, Methodologies, Tools, and Applications*, 2016. https://www.igi-global.com/chapter/sentiment-analysis-for-health-care/www.igi-global.com/chapter/sentiment-analysis-for-health-care/150188 (accessed Apr. 13, 2023).

[37] F. M. Guangul, A. H. Suhail, M. I. Khalit, and B. A. Khidhir, “Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College,” *Educ. Assess. Eval. Account.*, vol. 32, no. 4, pp. 519–535, Nov. 2020, doi: 10.1007/s11092-020-09340-w.

[38] N. Al-Ali, “Covid-19 and feminism in the Global South: Challenges, initiatives and dilemmas,” *Eur. J. Womens Stud.*, vol. 27, no. 4, pp. 333–347, Nov. 2020, doi: 10.1177/1350506820943617.

[39] R. S. Oktari, K. Munadi, R. Idroes, and H. Sofyan, “Knowledge management practices in disaster management: Systematic review,” *Int. J. Disaster Risk Reduct.*, vol. 51, p. 101881, Dec. 2020, doi: 10.1016/j.ijdrr.2020.101881.

[40] H. Farhat, J. Laughton, A. Joseph, W. Abougalala, M. B. Dhiab, and G. Alinier, “The educational outcomes of an online pilot workshop in CBRNe emergencies,” *J. Emerg. Med. Trauma Acute Care*, vol. 2022, no. 5, p. 38, Dec. 2022, doi: 10.5339/jemtac.2022.38.

[41] S. Lennquist, “Education and Training in Disaster Medicine,” *Scand. J. Surg.*, vol. 94, no. 4, pp. 300–310, Dec. 2005, doi: 10.1177/145749690509400409.

[42] G. Alinier, G. Dacey, M. A. Segni, S. McCabe, and D. R. Dungan, “Planning a large-scale tabletop exercise to test Qatar’s healthcare system readiness to respond to a major incident during the 2022 FIFA World Cup,” *J. Contingencies Crisis Manag.*, vol. n/a, no. n/a, doi: 10.1111/1468-5973.12485.

[43] H. Farhat *et al.*, “Hazardous material and chemical, biological, radiological, and nuclear incident readiness among prehospital care professionals in the State of Qatar,” *Glob. Secur. Health Sci. Policy*, vol. 7, no. 1, pp. 24–36, Dec. 2022, doi: 10.1080/23779497.2022.2069142.

[44] N. Bajow, A. Djalali, P. L. Ingrassia, H. Ageely, I. Bani, and F. D. Corte, “Disaster medicine curricula in Saudi Arabian medical schools,” *J. Emerg. Med. Trauma Acute Care*, vol. 2015, no. 1, p. 8, Feb. 2015, doi: 10.5339/jemtac.2015.8.

[45] G. Stevens *et al.*, “Determinants of Paramedic Response Readiness for CBRNE Threats,” *Biosecurity Bioterrorism Biodefense Strategy Pract. Sci.*, vol. 8, no. 2, pp. 193–202, Jun. 2010, doi: 10.1089/bsp.2009.0061.

[46] H. Chemaitelly *et al.*, “mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar,” *Nat. Med.*, vol. 27, no. 9, Art. no. 9, Sep. 2021, doi: 10.1038/s41591-021-01446-y.

[47] V. Aroniadou-Anderjaska, J. P. Apland, T. H. Figueiredo, M. De Araujo Furtado, and M. F. Braga, “Acetylcholinesterase inhibitors (nerve agents) as weapons of mass destruction: History, mechanisms of action, and medical countermeasures,” *Neuropharmacology*, vol. 181, p. 108298, Dec. 2020, doi: 10.1016/j.neuropharm.2020.108298.

[48] D. Tin, S. Fares, M. A. Mulhim, and G. R. Ciottone, "Terrorist Attacks in the Middle East: A Counter-Terrorism Medicine Analysis," *Prehospital Disaster Med.*, vol. 37, no. 2, pp. 212–216, Apr. 2022, doi: 10.1017/S1049023X22000358.

[49] C. G. Raffagnato, T. A. de O. Cardoso, F. de V. Fontes, M. M. Carpes, S. C. Cohen, and L. A. Calçada, "Chemical terrorism: risk modeling proposal for attacks involving ricin in mass gatherings in Brazil," *Saúde Em Debate*, vol. 43, pp. 152–164, Jan. 2020, doi: 10.1590/0103-11042019S311.

[50] K. L. Koenig, "Preparedness for terrorism: managing nuclear, biological and chemical threats.," *Ann. Acad. Med. Singapore*, vol. 38, no. 12, pp. 1026–1030, Dec. 2009.

[51] S. M. Lim, W. C. Cha, M. K. Chae, and I. J. Jo, "Contamination during doffing of personal protective equipment by healthcare providers," *Clin. Exp. Emerg. Med.*, vol. 2, no. 3, pp. 162–167, Sep. 2015, doi: 10.15441/ceem.15.019.

[52] WHO Regional Office for the Eastern Mediterranean, "Mass gathering preparedness and response: cross-border collaboration and coordination between Iraq and neighbouring countries," *East. Mediterr. Health J.*, vol. 25, no. 7, pp. 521–522, Jul. 2019, doi: 10.26719/2019.25.7.521.

[53] T. H. Kung *et al.*, "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," *PLOS Digit. Health*, vol. 2, no. 2, p. e0000198, Feb. 2023, doi: 10.1371/journal.pdig.0000198.

[54] CBRNE Research & Innovation Conference, "Organizing committee CBRNE - CBRNE Conference," 2024. https://cbrneconference.fr/index.php?pagendx=287 (accessed Aug. 27, 2023).

[55] M. Q. Patton, "Enhancing the quality and credibility of qualitative analysis.," *Health Serv. Res.*, vol. 34, no. 5 Pt 2, pp. 1189–1208, Dec. 1999.

Table legends:

Table 1: Participants' demographic information.

Table 2: The correlation coefficients of six prominent words in the top 10.

Figure legends:

Figure 1: The geographic distribution of participants according to their country of residence.

Figure 2: (A) Participants' word cloud based on their responses to the questions; (B) Scatter plot of the top 40 words mentioned by the participants.

Figure 3: Shewhart (a) and boxplot chart; (b) of participants' sentiments scores; (c): Pareto chart of the participants' sentiment scores by nationality; (d): Participants' sentiments identified during the modified interview.

Figure 4: Combined heatmap and parallel plot with boxplot for participants' emotions. (A) Parallel coordinates plot with boxplots of participants' emotions. (B) Correlation heatmap between emotions and words.

Figure 5: (a) Dendrograms of K-mean clustering algorithm; (b): Dendogram of Hierarchical clustering algorithm; (c): Cluster silhouette plot for K-mean algorithms; (d) Cluster silhouette plot for Hierarchical algorithm.

Figure 6: Participants' clustering.

Supplemental Files:

Supplemental File 1: Modified interview method questions.

Supplemental File 2: R-code generated for the analysis.

Supplemental Files:

Supplemental File 1: Modified interview method questions

Answer's type	Items
Welcome	
Research consent form	
Display	
Selection	Please select the language: English French
English	
Display	I. General Questions
Dropdown list	Age
Dropdown list	Gender
Dropdown list	Nationality
Text	1. Which organisations are you working for?
Dropdown list	2. What's your country of residence?
Audio	3. What is your highest qualification?
Audio	4. What's your role within your organisation
Selection	5. Would you like to be recognised in a future publication?
Audio	6. What's your area of expertise and for how long?
Audio	7. What forms of mass casualty incidents have you encountered in your career?
Audio	8. What forms of Chemical, Biological, Radiological, or Nuclear (CBRN) threats do you think might occur in your country?
Audio	9. Some believe that the tabletop exercise is the most efficient education tool in disaster medicine (including CBRN). How far do you agree with that? What is the best educational tool in disaster medicine, according to you?
Display	II. National Practice and Policy
Audio	10. How do you evaluate the actual multi-sectorial coordination in your country's disaster preparedness plan for: a) Mass casualty incidents. b) Chemical, Biological, Radiological and nuclear (CBRN) incidents?
Audio	11. Can you give examples of major and minor challenges to multi-sectorial coordination for major incidents in your country?
Audio	12. How far do the regulations in your country help ensure appropriate multi-sectorial coordination to manage mass casualties? Can you expand on that? 12.1. What about CBRN-specific threats? 12.2. If not, how can you suggest improving it?
Audio	13. What do you think about the accessibility and documentation of disaster response planning in your country? 13.1. What do you suggest to make it easier to access?
Audio	14. What do you think about including the private healthcare sector in the national preparedness plan?
Audio	15. What do you think about having a national database about chemical and biological plants in the country and neighbouring countries? 15.1. What information do you suggest including in this database?
Audio	16. For the Middle East and North Africa (MENA) Region, how beneficial would it be to create the position of "National Liaison officer" who will ensure the national multi-sectorial coordination between the "inter-national" and the MENA countries?
Display	III. International cooperation
Audio	17. What do you think about collaborating between the Middle East and North Africa (MENA) region Experts to formulate unified contingency plans and guidelines for any potential CBRN threats?
Audio	18. What do you think about collaborating between the Middle East and North African countries for periodic unified training about health sector management of Chemical, Biological, Radiological or Radiological (CBRN) emergencies?
Audio	19. How frequently should this training be held?
Audio	20. What do you think about the actual perspectives of â€œinternational cooperationâ€□ concerning Worldwide Health Emergency Preparedness? Please explain.
Audio	21. What are the limitations of international cooperation between the Middle East and North African countries' health sectors?
Audio	22. What is your opinion about the role of the World Health Organization (WHO) in the health sector's international cooperation?
Display	IV. Mass Gathering
Audio	23. What is your expert opinion regarding decontamination when managing a mass gathering of CBRN victims? 23.1. Should it be performed on the scene, in a pre-hospital or in-hospital setting? Can you elaborate on that?
Audio	24. In relation to major incident management, what plans are you familiar with?

Audio	25. Do you think that the Middle East and North African countries should develop their own plans and follow the existing international plans? Please explain.
Audio	26. How should the health sector take extra measures to manage mass gathering incident victims?
Audio	27. Following a CBRN incident in a mass gathering event, is there a need for special consideration of Very Important Persons (VIP) during decontamination? Why?
Display	V. Hospital preparedness
Audio	28. When preparing the Hospital Emergency Plan, what departments do you think should be included?
Audio	29. Which disciplines should be included in the hospital's major incident response planning?
Audio Text	30. What do you think about including Forensic medicine in the hospital's major incident response planning?
Audio	31. How important it is to equip Emergency Departments with decontamination (showering) units?
Audio	32. To your knowledge, concerning chemical, biological, radiological, and nuclear (CBRN) exposure, what antidotes do you think should be available in the healthcare Emergency systems?
Audio	33. How frequently do you think healthcare facilities should perform a risk assessment about potential chemical, biological, radiological and nuclear (CBRN) threats?
Audio	34. What elements should the health sector focus on in disaster preparedness?
French	
Display	I. Questions générales
Dropdown	Age
Dropdown	Nationalité
Dropdown	Genre
Text	1. Pour quelles organisation travaillez-vous ?
Dropdown	2. Quel est votre pays de résidence ?
Text	3. Quelle est votre qualification académique ?
Text	4. Quelle est votre fonction au sein de votre organisation?
Selection	7. Souhaitez-vous d'etre remercié dans les futurs publications ?
Audio	8. Quel est votre domaine d'expertise ? Depuis combien du temps ?
Audio	9. Au cours de votre carrière, à quels types d'incidents comportant un grand nombre de blessés avez-vous été confronté ?
Audio	10. Quels types de menaces Chimiques, Biologiques, Radiologiques ou nuclèaires (CBRN) pourraient toucher votre pays ?
Audio	11. Certains pensent que le « Tabletop » est l'outil de formation le plus efficace en médecine de catastrophe (y compris CBRN)? Qu'en pensez-vous ? D'aprés vous, quel est le meilleur outil de formation en médecine de catastrophe ?
Display	II. Pratiques et politiques nationales
Audio	12. Comment pouvez-vous évaluer la coordination multisectorielle dans le cadre de préparation aux catastrophes dans votre pays: a) Une situation d'urgence de masse ? b) Incident Chimique, Biologiques, Radiologiques ou nucléaires (CBRN) ?
Audio	13. Pouvez-vous donner des exemples de défis majeurs et mineurs dans la coordination multisectorielle face à des incidents de masse dans votre pays ?
Audio	14. Dans quelle mesure la réglementation de votre pays permet-elle de s'assurer d'une bonne coordination multisectorielle pour gérer: 14.1. Un incident de masse. 14.2. Un cas de menace spécifique d'une urgence Chimique, Biologiques, Radiologiques ou nucléaires (CBRN) ? 14.3. Sinon, que proposeriez-vous pour l'améliorer ?
Audio	15. Que pensez-vous de l'accessibilité, dans votre pays, aux documentations liées à la planification de la gestion d'une situation de catastrophe ? 15.1. Que proposeriez-vous pour les rendre plus accessibles ?
Audio	16. Que pensez-vous d'inclure le secteur privé de la santé dans le plan national de la gestion des catastrophes ?
Audio	17. Que pensez-vous de disposer d'une base de données nationale recensant tous les sites chimiques et biologiques du pays et des pays limitrophes ? 17.1. Quelles informations proposez-vous d'inclure dans cette base de données ?
Audio	18. Pour la région Moyen-Orient et Afrique du Nord (MOAN), dans quelle mesure serait-il bènèfique de créer un poste « Agent national de liaison » qui pourrait s'assurer de la bonne coordination nationale multisectorielle entre les pays MOAN et « inter-nationaux» ?
Display	III. Coopération internationale
Audio	19. Que pensez-vous d'une collaboration entre experts de la région Moyen-Orient et Afrique du Nord (MOAN) pour formuler des plans dâ€™urgence et des directives unifiés afin de faire face à des éventuelles menaces Chimique, Biologiques, Radiologiques ou nucléaires (CBRN) ?

Audio	20. Que pensez-vous d'une collaboration entre les pays de la région Moyen-Orient et Afrique du Nord (MOAN) pour des formations réguliéres unifiées en matiére de gestion par le secteur de la santé de cas d'urgence CBRN ?
Audio	21. Ã€ quelle fréquence doivent avoir lieu ces formations ?
Audio	22. Que pensez-vous des perspectives réelles de la « coopération internationale » concernant le niveau de préparation mondiale à une urgence sanitaire ? Veuillez expliquer.
Audio	23. Quelles sont les limitations de la coopération internationale entre les secteurs de la santé des pays du Moyen-Orient et d'Afrique du Nord ?
Audio	24. Quel est votre avis à propos du role de l'Organisation Mondiale de la Santé (OMS) dans la coopération internationale du secteur de la santé ?
Display	VI. Rassemblement de masse
Audio	25. Quel est votre avis d'expert sur la décontamination lorsqu'il s'agit de gérer un grand nombre de victimes d'une urgence Chimique, Biologiques, Radiologiques ou nucléaires (CBRN) lors d'un rassemblement de masse ?
Audio	26. Faut-il la réaliser sur place, dans un cadre pré-hospitalier ou dans un environnement hospitalier ? Pouvez-vous développer votre argumentaire ?
Audio	27. En matière de gestion d'incident de catastrophe, quels plans connaissez-vous ?
Audio	28. Pensez-vous que les pays du Moyen-Orient et d'Afrique du Nord devraient développer leurs propres plans ou suivre les plans internationaux existants ? Veuillez expliquer.
Audio	29. Quelles mesures supplémentaires pourrait mettre en place le secteur de la santé pour gérer des victimes d'incident lors de rassemblements de masse ?
Audio	30. Suite à une situation d'urgence Chimique, Biologiques, Radiologiques ou Nucléaires (CBRN) lors dun évènement rassemblant un grand nombre de personnes, faut-il établir une prise en charge particulière des VIP (personnalités de marque) pendant la décontamination ? Pourquoi ?
Display	V. Niveau de préparation des hopitaux
Audio	31. Lors de la préparation du plan d'urgence d'un hopital, quels services devraient etre impliqués ?
Audio	32. Quelles disciplines devraient etre sollicitées dans la planification de la réponse de l'hopital à un incident majeur ?
Audio	33. Que pensez-vous d'impliquer la médecine légale dans la planification de la réponse de l'hopital à un incident majeur ?
Audio	34. Dans quelle mesure est-ce important d'équiper les services d'urgence avec des unités de décontamination (douches) ?
Audio	35. Ã€ votre connaissance, en cas d'exposition Chimique, Biologiques, Radiologiques ou nucléaires (CBRN), quels antidotes devraient etre disponibles dans les systèmes d'urgence sanitaires ?
Audio	36. A quelle fréquence les autorités sanitaires devraient-elle conduire une évaluation des risques des menaces Chimiques, Biologiques, Radiologiques ou nuclèaires (CBRN) potentielles ?
Audio	37. Sur quels èlements devraient se concentrer le secteur de la santé dans le cadre de la préparation de la gestion d'une situation de catastrophe ?

Supplemental File 2: R-code generated for the analysis

```
----------------------------------------------------------------------------------------------------------------------------------------------------
# Section 1: Loading all the packages we need
----------------------------------------------------------------------------------------------------------------------------------------------------
library(readxl)
library(ggplot2)
library(maps)
library(mapdata)
library(NLP)
library(readxl)
library(tm)
library(SnowballC)
library(wordcloud2)
library(shapes)
library(ggplot2)
library(caTools)
library(syuzhet)
library(dplyr)
library(tidyr)
library(caret)
library(cowplot)
library(ggparcoord)
library(reshape2)
library(qcc)
----------------------------------------------------------------------------------------------------------------------------------------------------
#Section 2: Demographic information
----------------------------------------------------------------------------------------------------------------------------------------------------
Demo <- read_excel("C:/Users/hasse/OneDrive/Application Data/Desktop/Demo.xlsx",col_types = c("date", "text", "text", "numeric", "text", "text"))
#Age
summary(Demo$Age)
sd(Demo$Age)
ggplot(Demo, aes(x=Age)) +geom_histogram(fill="skyblue", colour="black", binwidth=5)+theme(axis.text=element_text(size=15))+geom_vline(aes(xintercept=mean(Age)),colour="blue", linetype="dashed", size=2)
#Gender, nationalities and spoken language:
ggplot(Demo, aes(x = Gender))+geom_bar(fill = "skyblue", colour = "black")+theme(axis.text=element_text(size=15))+coord_flip()
ggplot(Demo, aes(x = Language))+ geom_bar(fill = "skyblue", colour = "black")+theme(axis.text=element_text(size=15))+coord_flip()
countries<- data.frame(Nationality=c("Tunisia", "Lebanon", "Egypt", "Turkey", "UAE", "Greece", "United Kingdom", "France", "Qatar"," Saudi Arabia", "Malta", "USA", "Portugal"), count=c(6,5,4,4,2,1,1,1,1,1,1,1,1))
countries
ggplot(data=countries, aes(x =reorder(Nationality,-count),y=count))+theme(axis.text = element_text(angle = 0, hjust = 1, size=15))+geom_bar(fill = "skyblue", colour = "black")+coord_flip()
backg<- data.frame(Background=c("Medical Docotor", "CBRN instructor", "Ministry of interior", "Pharmacist", "Head of emergency departement", "Academician", "International Federation of Red Cross", "United Nations", "Disaster medicine planning", "World Health Organization", "Military", "Ministry of Healtth"), number=c(5,2,2,2,2,4,1,1,1,1,1,7))
backg
countries
#Creating a map of nationalities
country_counts <- data.frame(country = c("Tunisia", "UAE", "KSA", "Egypt", "Lebanon", "Portugal", "France", "Turkey", "Greece", "Malta", "UK", "USA", "Qatar 'First Author'"),
  count = c(6, 2, 1, 4, 5, 1, 1,4,1,1,2,1,1), lat = c(34.0, 22.9, 24.0, 27.0, 33.9, 39.5, 46.2,39.92, 39.07, 35.93,55.37,38,25.35),lon = c(9.0, 54.33, 45.0, 30.0, 35.5, -8.0, 2.2,32.86,21.82,14.37,3.43,97,51.18))
# set up color palette for country names
color_palette <- c("blue", "blue", "blue", "blue", "blue", "blue", "blue", "blue","darkblue","blue", "blue","slategray2","navy" )
# plot the map of the included countries
map("world", col = "white", fill = TRUE, bg = "cyan", mar = c(0, 0, 0, 0), ylim = c(14, 58), xlim = c(-15, 64))
# add circles to the map at the latitude/longitude coordinates of each country
points(country_counts$lon, country_counts$lat, pch = 21,
       cex = 3, bg = "red")
for (i in 1:nrow(country_counts)) {
  text(country_counts$lon[i], country_counts$lat[i],
       paste(country_counts$count[i], country_counts$country[i]),
       cex = 0.8, col = color_palette[i], pos = 4, offset = 0.5, font = 2)
}
text(-10, 18, "#Note: There is also another participant from USA#", pos = 4, cex = 1.1, col = "black")
text(-10, 16, "#This map was created by the first author using R_Studio#", pos = 4, cex = 1.1, col = "black")
----------------------------------------------------------------------------------------------------------------------------------------------------
# Secion 3: Natural Langage Processing
```

```
----------------------------------------------------------------------------------------------------------------------------------------
Round1 <- read_excel("C:/Users/hasse/OneDrive/Application Data/Desktop/Round1.xlsx")
# Remove unnecessary columns
data <-Round1[, -(0:14)]
interview_data<-unite(data, col= 'interview_text', c(`16`,`17`, `18`, `19`, `20`,`21`, `22`, `23`, `24`, `25`, `26`,`27`,`28`, `29`,
`30`,`31`,`32`, `33`, `34`), sep= " ")
# Clean and preprocess the text data
corpus <- Corpus(VectorSource(interview_data$interview_text))
corpus <- tm_map(corpus, content_transformer(tolower))
corpus <- tm_map(corpus, removeNumbers)
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, removeWords, c(stopwords("english"), "amp"))
dtm <- DocumentTermMatrix(corpus)
# Perform exploratory data analysis by creating the word cloud and words' count
freq <- colSums(as.matrix(dtm))
cld <- data.frame(word = names(freq), freq = freq)
wordcloud2(cld, size = 0.4, shape = "cercle", rotateRatio = 0.2,  color = "random-dark", backgroundColor = "white")
# Create a bar chart of the top 50 most frequent words
top_words <- head(sort(colSums(as.matrix(dtm)), decreasing = TRUE), 50)
ggplot(data = data.frame(word = names(top_words), freq = top_words),
       aes(x = word, y = freq))+
geom_point(aes(size=top_words),colour="blue")+theme(axis.text=element_text(size=15))+theme(axis.text.y=element_text(size=13,
angle=0))+ geom_rug()+coord_flip()
# Examining associations for the top 6 words with a lower correlation limit is equal to 0.25
findAssocs(dtm, terms = c("plans"), corlimit = 0.25)
findAssocs(dtm, terms = c("plan"), corlimit = 0.25)
findAssocs(dtm, terms = c("training"), corlimit = 0.25)
findAssocs(dtm, terms = c("disaster"), corlimit = 0.25)
findAssocs(dtm, terms = c("will"), corlimit = 0.25)
findAssocs(dtm, terms = c("cbrn"), corlimit = 0.25)
----------------------------------------------------------------------------------------------------------------------------------------
# Section 4: sentiment analysis
----------------------------------------------------------------------------------------------------------------------------------------
#get emotions from the text
emotions<-get_nrc_sentiment(interview_data$interview_text)
head(emotions, 10)
#barplot of emotions according to the scores and distributions
par(mar=c(5, 10, 4, 2) + 0.1, xpd=TRUE)
barssentiment<-get_nrc_sentiment(test_data$interview_text)
barplot(sort(colSums(barssentiment)), ylab = " ", xlab = "Score", las = 2, size = 16, col = "skyblue", border = "black", horiz = TRUE,width =
colSums(barssentiment))
# Creating parallel coordinates plot combined with a heatmap
p <- ggparcoord(data = cbind(emotions[, 1:10], dtm_mat), columns = 1:11, alphaLines = 0.2, showPoints = TRUE, boxplot = TRUE, order =
"skewness") +
  theme(axis.text.x = element_text(size = 12, angle = 0)) +
  theme(axis.text.x = element_text(size = 14, angle = 40)) +
  ggtitle("Parallel Coordinates Plot with Boxplot")
p
# Calculating correlation between emotion variable and text in dtm
dtm_mat <- as.matrix(dtm)[, 1:10]
dtm_text_cor <- cor(emotions[, 1:10], dtm_mat)
# Creating heatmap
library(ggplot2)
heatmap <- ggplot(data = reshape2::melt(dtm_text_cor), aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "yellow", high = "red", mid = "lightgreen", midpoint = 0, limit = c(-1, 1), space = "Lab", name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 1, size = 14, hjust = 1)) +
 theme(axis.text.y = element_text(vjust = 1, size = 14, hjust = 1)) +
  labs(x = "", y = "") +
  ggtitle("Correlation Heatmap")
heatmap
# Combining plots into a single chart
combined_plots <- plot_grid(p, heatmap, ncol = 2, align = "h", axis = "tb", labels = c("A", "B"), label_size = 12)
combined_plots
# Getting the sentiments scores
score<-get_sentiment(interview_data$interview_text, method = "syuzhet")
score
summary(score)
```

```
sd(score)
boxplot(score, horizontal = FALSE, col="skyblue")
---------------------------------------------------------------------------------------------------------------------------
#Section 5:  Creating the quality control charts for sentiments scores
---------------------------------------------------------------------------------------------------------------------------
# Creating the control chart
qcc(score, type="xbar.one")
# Create a data frame for the input data for the Pareto chart of sentiment scores per nationality
my_data <- data.frame(category = c("Greek", "British", "Portuguese", "Egyptian", "Maltese", "French", "Emirati", "Lebanese", "Saudi",
"Turkish", "Tunisian"), frequency = c(15.25, 8.12, 5.55, 4.63, 4.3, 3.35, 3.1, 3.1, 2, 1.2, 0.9))
# Create a named vector from the data
category_freq <- setNames(my_data$frequency, my_data$category)
# Create the Pareto chart
pdf("Hassan.pdf") # This is only to export the chart as PDF to get the full-size chart without a trimmed plot
pareto.chart(category_freq, ylab = "Frequency", col = heat.colors(length(category_freq)), horiz = FALSE)
# Rotate x-axis labels
par(las=2)
dev.off()  # Chart exported to PDF
---------------------------------------------------------------------------------------------------------------------------
#Secttion 6: Unsupervised machine learning
---------------------------------------------------------------------------------------------------------------------------
#kmean clustering
kmeans(dtm, centers = 2, nstart = 20)
clustering <- kmeans(dtm, centers = 2, nstart = 20)
clustering
#creating a dendrogram for k-mean
kmeans_model <- kmeans(dtm, centers = 4)
dist_matrix <- dist(dtm)
hclust_model <- hclust(dist_matrix)
cluster_labels <- kmeans_model$cluster
#Plotting dendrogram for kmeans
cluster_dendkmean <- as.dendrogram(hclust_model)
cluster_colorskmean <- c("red","lightgreen","purple","blue")
cluster_col_dend <- color_branches(cluster_dendkmean, k = length(unique(cluster_labels)),col = cluster_colorskmean, lwd = 2)
plot(cluster_col_dend, main = "Dendrogram of K-means Clusters", xlab = "Interview Text", ylab = "Distance")
#Silhoutte plot
sil <- silhouette(clustering$cluster, dist(dtm))
fviz_silhouette(sil)
# Performing hierarchical clustering
distance_matrix <- dist(dtm1)
cluster_model <- hclust(distance_matrix, method = "ward.D2")
# Ploting dendrogram
suppressPackageStartupMessages(library(dendextend))
cluster_dend <- as.dendrogram(cluster_model)
cluster_colors <- c("red","lightgreen","blue","purple")
cluster_col_dend <- color_branches(cluster_dend, k = length(unique(cluster_labels)),col = cluster_colors, lwd = 2)
plot(cluster_col_dend, main = "Hierarchical Clustering Dendrogram", xlab = "Interview Text", ylab = "Distance")
# Extracting cluster labels to perform the analysis of the content in each cluster
cluster_labels <- cutree(cluster_model, k = 4)
# Adding cluster labels to original data
interview_data1$cluster <- as.factor(cluster_labels)
# Grouping data by cluster
cluster_groups <- split(interview_data1, interview_data1$cluster)
# Viewing first few rows of each cluster
lapply(cluster_groups, head)
# Calculating silhouette coefficient
sil2 <- silhouette(cluster_labels, distance_matrix)
# Plotting the second silhouette coefficient
fviz_silhouette(sil2)
#Performing the PCA
pca <- prcomp(dtm)
pca_df <- data.frame(pca$x[,1:2])
pca_df$cluster <- as.factor(clustering$cluster)
ggplot(pca_df, aes(x = PC1, y = PC2, color = cluster, label = rownames(pca_df))) +geom_point(size = 3) +labs(x = "PC1", y = "PC2")
+scale_color_manual(values = c("blue", "red")) +geom_text_repel(size = 6, box.padding = 0.5, point.padding = 0.5,segment.color =
"lightgreen", segment.size = 0.9, force = 2, max.overlaps = Inf) +theme(axis.text = element_text(size = 14))
```