Hypergraph Text Classification for Mental Health Misleading Advice

Lulwah Alkulaib* ‡, Abdulaziz Alhamadani*, Shailik Sarkar*, and Chang-Tien Lu*

* Department of Computer Science, Virginia Tech, Falls Church, VA 22043 USA

‡ Department of Computer Science, Kuwait University, Kuwait

{lalkulaib, hamdani, shailik, ctlu}@vt.edu

Abstract—This paper introduces HyperMAD, a novel Hypergraph Convolutional Network model designed for the multiclass classification of mental health advice in Arabic tweets. The model distinguishes between misleading and valid advice, further categorizing each tweet into specific classes of advice. HyperMAD leverages high-order relations between words in short texts, captured through the definition of four types of hyperedges that represent local and global contexts as well as semantic similarity. Extensive experiments demonstrate the effectiveness of HyperMAD, with results outperforming those from existing baselines. The study also includes an ablation study to investigate the significance and contribution of each hyperedge type. The paper presents a case study analyzing the accuracy and types of Arabic mental health advice on Twitter, revealing that about 9% of the advice in response to mental health expressions on Twitter was accurate in general. The paper concludes with the hope that the application of HyperMAD can be utilized in flagging misleading responses on social media, providing the correct resources for those who choose to share their mental health struggles online.

Index Terms—hypergraph learning, hypergraph convolutional network, text classification, short text classification

I. INTRODUCTION

The rise in mental health advice on social media platforms like Instagram, Twitter, and TikTok is a response to increased awareness and reduced stigma around mental health issues. These platforms host a variety of content, from self-care tips to professional insights, aimed at promoting mental well-being. However, the credibility of these sources varies, and not all advice is accurate or reliable. The abundance of information can make it difficult for individuals to distinguish between valid advice and misinformation, which can potentially worsen their mental health conditions. Therefore, it's essential for users to critically evaluate these sources, verify their credibility, and consult with qualified professionals to ensure they receive accurate, evidence-based guidance.

People turn to social media for mental health advice due to its accessibility, affordability, and anonymity. Some cannot afford traditional services, while others prefer online platforms to avoid judgment or stigma in their immediate social circles. For those living in societies where mental health is stigmatized, social media offers a discreet way to find support and connect with others facing similar challenges. Thus, the convenience and supportive community on these platforms make them an attractive alternative for those who find traditional mental health services inaccessible.

Our work examines mental health advice given on Twitter to Arabic users posting mental health concerns. Twitter's anonymity allows users to express feelings without fear of stigma. Several datasets have been published discussing the mental states of Arab Twitter users, whose potential audience spans 22 countries in the Middle East and North Africa. According to Statista[1], 54% of Arab youth reported difficulty in accessing quality mental health help in their region. Given the stigma associated with mental illness in the Arab world [7], we aim to investigate the nature of advice provided to those seeking help online.

Despite extensive research on medical misinformation on social media [30], [36], no studies specifically target mental health misinformation. Existing works focus on general health misinformation or fake news detection, employing various methods such as knowledge graphs [3], [23], transformer-based models [10], [11], and graph-based approaches [9], [20]. However, none concentrate solely on mental health misinformation. This gap presents challenges: **1) accurately labeling misleading mental health advice from Twitter while considering the advice type** and **2) capturing high-order word associations in tweets to enhance tweet class representation.** Our goal extends beyond binary classification of misleading advice; we aim for multiclass classification of advice types, requiring labeled data not readily available from Twitter. Furthermore, we aim to capture complex word relations in tweets, which hypergraphs can effectively represent, thereby improving classification.

To address those challenges, we propose **HyperMAD**, **M**isleading mental health **ad**vice classification for Arabic tweets using hypergraph convolutional networks. In particular, the model distinguishes between misleading advice tweets and valid advice and meanwhile matches each tweet to a specific

---

[1]https://www.statista.com/statistics/1013488/mena-arab-youth-on-degree-of-difficulty-to-get-quality-help-for-mental-health-issues-in-their-country/

class of advice (medical, alternative medicine, meditation, physical activity, religious advice, and other). We summarize our contributions as follows:

- We propose HyperMAD, a hypergraph convolutional network model for short text multiclass classification. Our proposed model captures high-order relations between words in short texts and utilizes those learned representations in text classification.
- We define four types of hyperedges: sentence hyperedges, DQE hyperedges, tweet hyperedges, and co-occurrence hyperedges, to connect the words extracted from our dataset to capture local and global contexts as well as semantic similarity. These hyperedge types help capture the complex relations between words in our dataset.
- We conduct comprehensive experimental evaluations to validate the efficacy of our proposed model. Our model outperforms seven baseline models in classifying short texts retrieved from Twitter. The experiment results demonstrate the superiority of the proposed approach.

II. RELATED WORK

A. Misinformation Detection

Misinformation research has employed various techniques, including NLP [1], [26], network-based analysis [27], and advanced methods like Graph Convolutional Networks [25] and gated GNN [33]. Studies have focused on rumor identification using user profiles [28], rumor veracity on social media [29], and fact-checking COVID-19 vaccine data [24]. Multi-modal approaches [4], [13] and heterogeneous graphs have been used for fake news detection [21], [22]. Despite these methods' success, they generally address health misinformation, not specifically mental health. Our work uniquely uses Twitter short texts for classification without additional features like likes, retweets, or user profiles, reducing data collection needs.

B. Hypergraph Learning

Hypergraphs, a generalized graph structure, allow for hyperedges connecting multiple nodes, capturing high-order relationships ideal for modeling real-world data like social networks [37]. Hypergraph learning, extending from graph learning, includes tasks like node classification [6], link prediction [17], and community detection [2]. In text classification, a hypergraph attention network was proposed for long documents [5], and hypergraph convolutional networks were used for short texts like reviews or titles [14]. However, [5]'s model was built for long documents and did not address how to handle short texts like those retrieved from social media posts, and [14]'s model represents the text association graph using graph convolutional networks, which overlooks the high-order representations that are captured by the hypergraph. Our work uniquely uses a hypergraph for short texts from Twitter, capturing high-order relations with four types of hyperedges.

C. Dynamic Query Expansion

Dynamic Query Expansion (DQE) is a widely used data mining technique applied in various domains. It is particularly useful in finding keywords in published tweets similar to a seed query word. By leveraging the tweets dataset, DQE expands initial seed queries and reformulates the queries to retrieve tweets containing similar keywords. It has been successfully used in multiple applications to track civil unrest events [34], detect metro emerging threats [38], monitor cyber-attack events [15], and track flu outbreaks [35] through social media data. Our work utilizes the DQE algorithm to capture any similar keywords in tweets that are then used in constructing a semantic DQE-hyperedge in our hypergraph.

III. PRELIMINARIES

In this section, we define the hypergraph since we use it to represent our data. A hypergraph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_i\}_{i=1}^M$ represents the set of nodes in the graph, and $\mathcal{E} = \{e_j\}_{j=1}^N$ represents the set of hyperedges. Unlike traditional graphs, in a hypergraph, a hyperedge e_j can connect two or more nodes from $\mathcal{V}$. The hypergraph $\mathcal{G}$ is represented by an incidence matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$, where the entries are defined as:

$$\mathbf{H}_{i,j} = \begin{cases} 1 & \text{if } v_i \in e_j \\ 0 & \text{if } v_i \notin e_j \end{cases} \quad (1)$$

In this representation, $\mathbf{H}_{i,j}$ is set to 1 if the node v_i is part of the hyperedge e_j, otherwise, it is set to 0. For our text-based hypergraph, nodes in $\mathcal{V}$ represent words that appear in each tweet $\mathcal{T}_{\parallel}$, where each word node has an attribute vector $\mathcal{X}$ of the pre-trained word embeddings denoted as $\mathcal{X} = [x_1, x_2, ..., x_m]^T \in \mathbb{R}^{m \times d}$.

IV. METHODOLOGY

In this section, we first summarize the notations used in this paper. Then, we define our problem and explain our proposed method, HyperMAD, with detailed information, including the hypergraph construction, hypergraph convolution, and text classification task. Figure 1 shows our proposed framework.

A. Problem Definition

Let $G = (H, X)$ be a hypergraph, where $H = (V, E)$ represents the hypergraph structure and X represents the node attributes. In this work, we propose a method for short text classification that leverages the hypergraph G to capture complex relationships between words in short texts. The goal is to construct H with four types of hyperedges and use a ReLU function to control the information flow from the node attributes X to the hypergraph convolutional network layers. The final classification is performed using an MLP, which takes as input the embeddings from the modified hypergraph convolutional network. The MLP is trained using a cross-entropy loss function for multi-class classification. The objective is to minimize this loss function to achieve accurate classification of the short text tweets.
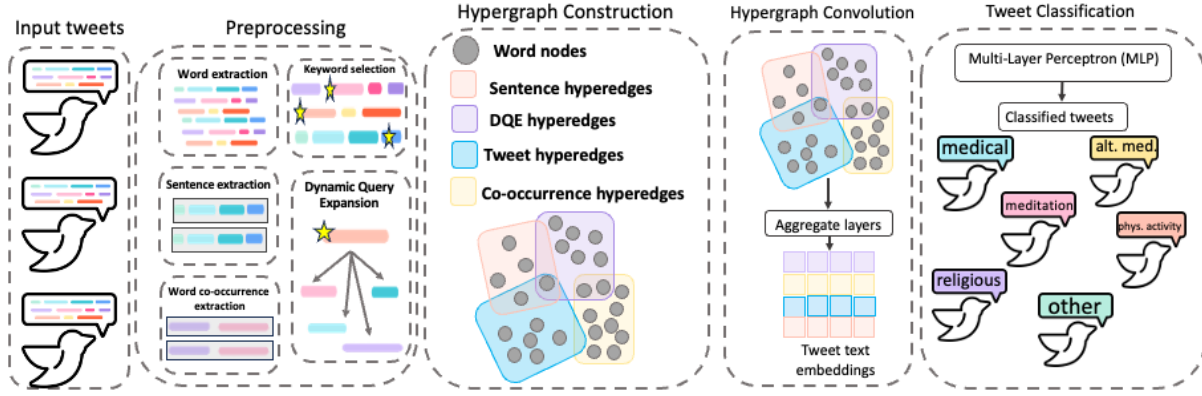
Fig. 1: The illustrative architecture of the proposed HyperMAD model.

TABLE I: Notations and their explanations

Notation	Explanation
G	The hypergraph
H	The hypergraph structure, consisting of nodes and hyperedges
X	The node attribute matrix, representing the term embeddings
V	The set of nodes in the hypergraph, representing words in the tweets
E	The set of hyperedges in the hypergraph, representing relationships between words
$X^{(0)}$	The initial node attribute matrix
$X^{(l)}$	The node attribute matrix after l layers of hypergraph convolution
ϵ	A hyperparameter controlling the contribution of the self-connections in the hypergraph convolution
W_1	The weight matrix in the hypergraph convolution and the MLP
b_1	The bias vector in the hypergraph convolution and the MLP
Z	The short text embedding matrix, obtained by aggregating the node attributes
y	The output of the MLP, representing a probability distribution over the classes
$\hat{y}$	The true label vector
$L(y, \hat{y})$	The cross-entropy loss function
T	A set of tweets
S	Seed query
K	Expanded query
Q_0	Initial query, set equal to the seed query S
F_0	Initial feature vector, set equal to S
$w_e(F_0)$	Initial weight of F_0, set to 1
p	Iteration index in the DQE algorithm
F_p	Feature vector at the p-th iteration
$w_e(F_p)$	Weight of F_p
T_p	Tweet at the p-th iteration
$w_e(T_p)$	Weight of T_p
σ	Difference between the minimum weight of T_p and the maximum weight of the remaining tweets
idf	Inverse document frequency

B. Hypergraph Construction

In our proposed method, a hypergraph $H = (V, E)$ is constructed. The set V denotes the collection of nodes, where each individual node represents a unique word present in the dataset. On the other hand, E signifies the set of hyperedges. Each hyperedge within the set E is a specific subset of V and symbolizes a relationship amongst the words. To capture the complexity of relationships between words in the dataset, four distinct types of hyperedges are defined in this context:

1) Sentence Hyperedges: For each sentence in the dataset, we create a hyperedge that connects all the words in the sentence. This captures the local context of each word. For a sentence s containing n words, we define a hyperedge $e_s = \{v_1, v_2, \ldots, v_n\}$, where each v_i is a node representing a word in the sentence.

2) DQE Hyperedges: For each selected word in the dataset that we call our seed query, we use an extended DQE al-

gorithm [34] for tweets that contain up to 280 characters to iteratively find similar keywords to our seed query. Algorithm 1 iteratively retrieves all tweets from our dataset that contain keywords related to the seed query and produces a list of those keywords after it converges. We create a hyperedge that connects the seed query with its similar keywords. This captures the semantic similarity between words in our dataset. So, for a keyword k and its set of similar keywords K, we define a hyperedge $e_k = \{v_k\} \cup \{v_i | v_i \in K\}$. The DQE hyperedges are meant to connect expanded words of similar meaning, which in turn would introduce connections between tweets that belong to the same topic.

Algorithm 1 Dynamic Query Expansion

Require: A set of Tweets $\langle T_0, T_1, \ldots, T_t \rangle$, Seed Query S
Ensure: Expanded Query K
1: Set $Q_0 = S = F_0$, $w_e(F_0) = 1$, $p = 0$
2: **repeat**
3:      $p = p + 1$
4:      $w_e(F_p) = \text{idf}(F_p) \cdot C \cdot w_e(T_{p-1})$ ▷ Use $w_e(T_{p-1})$ weight to compute $w_e(F_p)$
5:      $w_e(T_p) = \Phi \cdot C' \cdot w_e(F_p)$
6:      **repeat**
7:          swap $(\min(w_e(T_p)), \max(w_e(T - T_p)))$
8:          $\sigma = \min(w_e(T_p)) - \max(w_e(T - T_p))$
9:      **until** $\sigma \leq 0$
10: **until** $w_e(F_p) = w_e(F_{p-1})$
11: $K = F_p$

---

3) Tweet Hyperedges: For each tweet in the dataset, we create a hyperedge that connects all the words in the tweet. For a tweet t containing n words, we define a hyperedge $e_t = \{v_1, v_2, \ldots, v_n\}$, where each v_i is a node representing a word in the tweet. This type of hyperedge is meant to capture the high-order relations in the tweet as a short document.

4) Co-occurrence Hyperedges: For each set of words that co-occur in more than 15% of the dataset, we create a hyperedge that connects all the words in the set. This captures the global context of each word. So for a set of co-occurring

words C that appear in more than 15% of the dataset, we define a hyperedge $e_c = \{v_i | v_i \in C\}$.

By defining these four types of hyperedges, we construct a hypergraph that captures both local and global context, as well as semantic similarity.

C. Hypergraph Convolutional Network

With the objective of capturing high-order relationships among word terms within the tweets, we design short text hypergraph convolution networks to learn word term embeddings. The proposed model, HyprMAD, incorporates several convolution layers and utilizes $X^{(l)} \in R^{d \times N}$ to represent the d-dimensional word embeddings learned at each level. Recognizing that different perspectives may contribute differently to the final classification results, it is not advisable to indiscriminately distribute the initial term embedding $X(0)$ across all layers of the hypergraph convolutional network. Therefore, to manage the distribution of the initial term embedding across the layers of the hypergraph convolutional network, we apply the ReLU function to the term embedding at each layer, which is defined as follows:

$$X^{(0)} = ReLU(X^{(0)} + \epsilon X^{(0)} W_1 + b_1) \tag{2}$$

where $X^{(0)}$ is the initial term embedding, ϵ is a small constant, W_1 is the weight matrix, and b_1 is the bias vector. And for each layer l in the hypergraph convolutional network, the term embedding matrix $X^{(l)}$ is updated using the ReLU function as defined in Equation 3:

$$X^{(l)} = \text{ReLU}(X^{(l-1)} + \epsilon X^{(l-1)} W_1 + b_1) \tag{3}$$

The main challenge in applying convolutional operations on hypergraphs is determining an effective method for the propagation of lexical item embeddings due to the non-Euclidean nature of hypergraph data. We present the short textual hypergraph convolution following the evidence in [8], as follows:

$$X^{(l+1)} = \sum_{\epsilon=1}^{N} \sum_{j=1}^{M} H_{\epsilon j} P^{(l)} x_j^{(l)} \tag{4}$$

The hypergraph convolution operation, a key mechanism in our model, refines term embeddings within the hypergraph structure. Each term starts with an initial term embedding, $X^{(0)}$, typically derived from pre-trained models like Word2Vec or GloVe. The operation $Z^T X^{(l)}$ symbolizes the aggregation of information from individual terms to the short text, while the pre-multiplied Z signifies the reverse aggregation. This operation unfolds over L layers, each broadening the scope of information exchange across the hypergraph. The final term embeddings, which are the result of this operation, are then used in Equation 5 to compute the final short text embedding:

$$X^* = \frac{1}{L+1} \sum_{l=0}^{L} X^{(l)} \tag{5}$$

The embedding for a specific short text tweet, represented as $d_i = \{v_1, v_2, \ldots, v_s\}$, is derived by aggregating the representations of the words it contains, as given by:

$$d_i = \frac{1}{s} \sum_{t=1}^{s} x_t^* \tag{6}$$

Consequently, the short text embedding matrix for the hypergraph, denoted as Θ_h, is computed as the average of tweet text embeddings:

$$\Theta_h = \frac{1}{M} \sum_{i=1}^{M} d_i \tag{7}$$

D. Text Classification

In this work, we propose a short text classification method that leverages a hypergraph convolutional network to overcome the challenges posed by the high-order relations between words in tweet texts. Our approach involves the construction of a hypergraph that represents the complex relationships between words in the short texts, using four distinct types of hyperedges. The ReLU function is then employed to control the flow of information in the network. The final classification is carried out using an MLP, which is trained on the embeddings derived from the hypergraph convolutional network using a cross-entropy loss function for multi-class classification.

The hypergraph is constructed with four types of hyperedges, represented as:

$$E = E_1 \cup E_2 \cup E_3 \cup E_4 \tag{8}$$

where E_1, E_2, E_3, E_4 represent the sentence hyperedges, query expansion hyperedges, tweet hyperedges, and co-occurrence hyperedges, respectively.

In the final stage of our method, we perform classification using an MLP. The MLP takes as input the embeddings derived from the hypergraph convolutional network. Each layer of the MLP involves a linear transformation followed by the application of the ReLU activation function, which introduces non-linearity and allows the MLP to learn more complex patterns in the data. The MLP is represented as:

$$y = softmax(ReLU(W_1 x + b_1)) \tag{9}$$

where W_1 is the weight matrix, b_1 is the bias vector, x is the input vector, and y is the output vector. The softmax function is used for the output layer to produce a probability distribution over the classes.

The MLP is trained using the cross-entropy loss function, which is defined as:

$$L(y, \hat{y}) = -\sum_{i} y_i \log(\hat{y}_i) \tag{10}$$

where y is the true label vector, $\hat{y}$ is the predicted label vector, and the sum is over all classes. This loss function is

minimized during training to achieve accurate classification of the short text tweets.

V. EXPERIMENTS

To assess the performance of our proposed model, HyperMAD, we conduct extensive experiments to validate its effectiveness. These experiments aim to answer the following questions:

- **RQ1:** How does the performance of HyperMAD compare to the baseline methods with regard to short-text multiclass classification?
- **RQ2:** Does the imbalance in the data affect the classification results for each class?
- **RQ3:** How does each type of hyperedge contribute to the performance, and which combination is the best?
- **RQ4:** How does the complexity of the hypergraph and the number of hyperedges compare to the baseline methods?

Algorithm 2 Hypergraph Convolutional Network for Short Text Classification

---

Require: Tweet dataset $D = \{d_1, d_2, ..., d_M\}$, term set $V = \{v_1, v_2, ..., v_N\}$, hyperedge set $E = E_1 \cup E_2 \cup E_3 \cup E_4$, term embedding matrix $X^{(0)} \in R^{N \times d}$, number of layers L, learning rate η, batch size B, number of epochs T.

Ensure: The tweets classification results.

1: **for** $t = 1$ to T **do**
2:     **for** each mini-batch B **do**
3:         Construct the hypergraph $G = (V, E)$ with four types of hyperedges: sentence hyperedges, DQE hyperedges, tweet hyperedges, and co-occurrence hyperedges.
4:         Initialize the term embedding matrix $X^{(0)}$ using the ReLU function (Eq. 2).
5:         **for** $l = 1$ to L **do**
6:             Update the term embedding matrix $X^{(l)}$ using the ReLU function (Eq. 3):
7:         **end for**
8:         Compute the short text embedding matrix H by aggregating the term embeddings of each short text.
9:         Compute the output of the MLP using the short text embeddings (Eq. 9):
10:         Compute the cross-entropy loss (Eq. 10):
11:         Update the parameters of the model using backpropagation and the Adam optimizer with learning rate η.
12:     **end for**
13: **end for**
14: **return** The class with the highest probability in the output vector y for each short text.

A. Datasets

Since we were interested in classifying mental health advice on Twitter, we collected a dataset from responses to tweets about mental health states and moods. In order to do that, we found a Twitter dataset that collected Arabic tweets describing different users' moods, anxieties, sleep modules,

TABLE II: Dataset description

Tweet/Advice class	# of tweets	Type	Avg. length in words	# of words
Medical	23,000	v: 1,150 m: 21,850	8.26	37,227
Alternative medicine	15,298	v: 1,223 m: 14,075	9.73	35,091
Meditation	1,279	v: 38 m: 1,241	7.45	5,764
Physical activity	10,241	v: 1,228 m: 9,013	8.88	30,514
Religious	30,682	v: 4,295 m: 26,387	10.12	42,719
Other	11,500	v: 920 m: 10,580	9.67	32,120

and depression states [19]. Since we were interested in the response and advice given to those tweets, we had to retrieve the data from Twitter and collect all responses. The dataset had about 48.8K tweets, but at the time of retrieval, we incurred a 10% loss which still left us with 43.9K tweets to investigate responses. The resulting dataset of responses consisted of 131.7K tweets. After removing tweets that responded with emojis only or did not include advice, we had 92K tweets left, each corresponding to one of 6 classes. The dataset was labeled by three native Arabic-speaking psychology graduate students to confirm which tweet/advice was valid (v) or misleading (m) and to assign it to the corresponding class. A summary of the dataset statistics is shown in table II. In our experiments, we adopt the train/test data splits similar to previous works [5], [32]. For each run, we randomly select 90% of the training samples to train the model, while the remaining 10% of the data is used for validation. This consistent approach allows for a fair and comparable assessment of our results to previous works.

B. Baselines and Evaluation Metrics

In our experiments, we compare our model to word embedding methods such as: CNN [16], LSTM [18], BiLSTM [31], graph-based methods such as: Graph-CNN [32], TextGCN [32], TextGNN [12], and hypergraph-based methods such as: HyperGAT [5], and STHCN [14].

To evaluate the performance of our model in comparison to the baselines, we use accuracy and F1-score as our metrics, as shown in Table III.

C. Implementations Details

Our HyperMAD model, a Hypergraph Convolutional Network, is built with PyTorch and optimized using Adam. It's run on a system with an AMD Ryzen Threadripper 1920X, Windows 10 64-bit, and NVIDIA GeForce RTX 2080Ti GPU with 64 GB RAM. We use $d = 50$ for the word embedding dimension and $L = 4$ convolutional layers. All network parameters are initialized using a Normal distribution. Word embeddings are initialized with models like Word2Vec or GloVe. The results reported are the averages from 10 runs under optimal hyperparameter settings.

TABLE III: Overall performance of baseline methods in comparison to HyperMAD using Accuracy (Acc) and F1-Score (F1)

Models	Twitter AR-MHAD	
	ACC	F1
CNN	0.41	0.33
LSTM	0.55	0.43
BiLSTM	0.56	0.45
TextGCN	0.58	0.47
TextGNN	0.56	0.45
HyperGAT	0.71	0.61
STHCN	0.77	0.67
HyperMAD	**0.87**	**0.86**

TABLE IV: HyperMAD fine-grained misleading advice per class results

Class	HyperMAD	
	ACC	F1
Medical	0.895	0.875
Alternative medicine	0.872	0.865
Meditation	0.745	0.710
Physical activity	0.900	0.880
Religious	**0.905**	**0.890**
Other	0.815	0.780
Weighted avg.	0.883	0.865

VI. RESULTS AND ANALYSIS

A. Overall Performance

We conduct extensive experiments to evaluate the performance of our proposed model, HyperMAD, on short-text classification answering questions **RQ1 and RQ2** and show the results in Table III and TableIV, respectively. Based on our experiments, we are able to make the following observations:

• Overall, HyperMAD outperforms the seven baseline methods on the Twitter Arabic Mental Health Advice Dataset (Twitter AR-MHAD).

• Hypergraph-based methods outperform graph-based and word-embedding methods on our short-text dataset. This indicates the importance of the hypergraphs' ability to represent high-order relations between words in text datasets.

• While HyperMAD and STHCN are both hypergraph methods for short-text datasets, our proposed model only represents the data in hyperedges. Whereas STHCN incorporates the text association using a graph convolutional network. We attribute the improvement in performance to the critical role that hypergraphs play in capturing text associations that existing graph-based methods cannot capture.

• In the fine-grained classification task in Table IV, it is noticeable that the class with the most number of tweets, Religious advice, is the best performing one. The second best is the physical activity advice; we attribute that to the distinctive words associated with that class.

• The most difficult tweets to classify are those in the meditation class. Since it has the least amount of tweets, it was harder for the model to differentiate them. Also, the other advice class contains advice that holds a combination

TABLE V: Ablation study: Evaluation of our proposed model, HyperMAD, on different combinations of hyperedges types where S, D, T, and C correspond to the Sentence, DQE, Tweet, and Co-occurrence hyperedges, respectively

Model Variants					
S	D	T	C	ACC	F1
x				0.723	0.714
	x			**0.766**	**0.748**
		x		0.739	0.722
			x	0.706	0.689
x	x			0.811	0.797
x		x		0.797	0.782
x			x	0.778	0.764
	x	x		**0.820**	**0.801**
	x		x	0.772	0.754
		x	x	0.708	0.691
x	x	x		**0.854**	**0.837**
x	x		x	0.819	0.803
	x	x	x	0.849	0.831
x	x	x	x	**0.880**	**0.856**

TABLE VI: GPU memory consumption comparison of the best-performing method in each baseline type

Model	Twitter AR-MHAD
HyperMAD	**576.12 MB**
STHCN	1064.36 MB
TextGCN	13,427.21 MB
BiLSTM	9,825.37 MB

of two or more classes, which makes them more likely to be misclassified.

• Despite the variance in the number of tweets among classes, we find the fine-grained results significant in differentiating between advice classes.

B. Ablation Study

To answer **RQ3**, we have to investigate the significance and contribution of each hyperedge type to our model, HyperMAD. We conduct an ablation study where we test the model with various combinations of hyperedges, starting with one type at a time and ending with all four types. Table V shows the 15 variants that we tested. The table is color-coded based on the number of hyperedges that appear, and for each color group, we mark the best-performing variant in bold and the second best by underlining its results. We notice that all four types of hyperedges achieve the best performance, confirming that the hypergraph learns the most from the high-order relations in all four types. It is worth noting that the hyperedges generated from the DQE seem to be the most informative among the other single types of hyperedges. This outcome confirms that the hypergraph learns more information from the expanded queries than word co-occurrence, sentences, and tweets. Combining the DQE hyperedges and tweet hyperedges seems to improve the models' expressiveness better than other combinations. Moreover, adding the sentence hyperedges to the previous combination indicates that the context and seman-
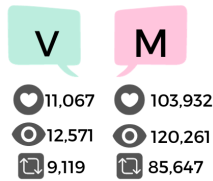
Fig. 2: Social Media Reach of Misleading and Valid Arabic Mental Health Advice on Twitter

Language	Initial seed queries	Expanded queries
Arabic	معالج نفسي، تشخيص، مشكلة نفسية، اضطرابات	متخصص، العلاج السيكولوجي، الاضطرابات السلوكية، فحص، العلاج المناسب، المعالج النفساني، أخصائي، الأمراض النفسية، مراكز الإرشاد النفسي المتخصصة، علاج، المشكلات النفسية
English	Psychotherapist, diagnosis, psychological problem, disorders	Specialist, psychological treatment, behavioral disorders, examination, appropriate treatment, psychotherapist, specialist, mental illness, specialized psychological counseling centers, treatment, psychological problems

Fig. 3: Examples of initial queries expanded by our DQE algorithm translated to English

tic information heterogeneity significantly enhances overall performance.

C. Computational Efficiency

In order to evaluate the computational efficiency of our model and answer question **RQ4**, we perform the comparison between HyperMAD and the best-performing method in each baseline type. The results in Table VI show that Hyper-MAD has an advantage pertaining to the amount of memory used over the baseline methods. We attribute this efficiency improvement to HyperMAD's learning which is conducted on the tweet level and the small batches of tweets being stored during training. STHCN and TextGCN both require constructing the tweet graphs for training and testing data which consumes additional memory. Due to our dataset's large unique vocabulary size, the BiLSTM model involves handling a larger number of word embeddings and conducting more calculations for each word, making it consume more memory.

D. Case Study

To show an example of how our DQE algorithm works, we translated the seed keywords we used in the medical category and displayed the initial seeds and expanded queries in Figure 3.

In this case study, we try to analyze the accuracy and types of Arabic mental health advice on Twitter. Seeking help for mental illnesses can be challenging, stigmatized, and expensive, thus, leading many Arabic speakers to vent on social media. The responses in the form of advice that they get to the description of their mental state could affect them negatively or positively depending on the advice. We try to answer the following question in our case study: **Could the Arabic Twitter community support each other with accurate advice?** We analyzed 92K tweets containing Arabic

mental health advice from Twitter. As seen in Figure 4, about 9% of the advice in response to mental health expressions was accurate in general. Meanwhile, over 90% of the advice belonging to the medical, alternative medicine, and meditation categories was misleading. It was also worrisome to see that the misleading advice had a higher reach in terms of likes, views, and retweets than advice and responses to tweets posted by verified medical doctors or institutions, as seen in Figure 2. We hope that these insights raise awareness in fighting misleading mental health advice on social media platforms and highlight the importance of requiring warning labels on advice that could cause more harm than good.
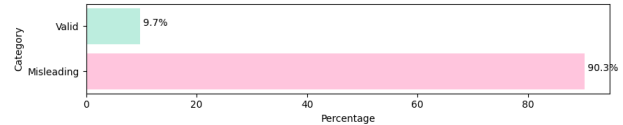
(a) Arabic Mental Health Advice on Twitter
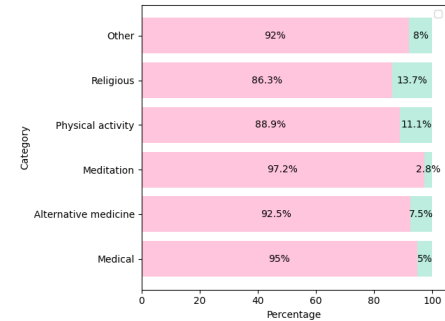
(b) Breakdown of Misleading and Valid Advice on Arabic Mental Health Advice on Twitter

Fig. 4: Arabic Mental Health Advice on Twitter Analysis

VII. CONCLUSION

We proposed HyperMAD, a novel hypergraph short text classification model that captures the high-order relations between words in short texts. Specifically, we define four types of hyperedges to capture the complexity of the relations between words in our Twitter dataset. Our constructed hypergraph incorporates several convolutional layers to learn the text embeddings. Additionally, utilizing the MLP allowed us to perform multiclass classification of the advice in response to tweets corresponding to the topic they fall under. Extensive experiments demonstrated the effectiveness of our proposed model. To the best of our knowledge, HyperMAD is the first model to focus on the classification of misleading mental health advice. We hope that this application is utilized in flagging such responses on social media with the correct resources for those who choose to share their mental health struggles online.

REFERENCES

[1] Sabur Butt, Shakshi Sharma, Rajesh Sharma, Grigori Sidorov, and Alexander Gelbukh. What goes on inside rumour and non-rumour tweets and their reactions: A psycholinguistic analyses. *Computers in Human Behavior*, 135:107345, 2022.

[2] I Chien, Chung-Yi Lin, and I-Hsiang Wang. Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 871–879. PMLR, 2018.

[3] Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 492–502, 2020.

[4] Mudit Dhawan, Shakshi Sharma, Aditya Kadam, Rajesh Sharma, and Ponnurangam Kumaraguru. Game-on: Graph attention network based multimodal fusion for fake news detection. *arXiv preprint arXiv:2202.12478*, 2022.

[5] Kaize Ding, Jianling Wang, Jundong Li, Dingcheng Li, and Huan Liu. Be more with less: Hypergraph attention networks for inductive text classification. *arXiv preprint arXiv:2011.00387*, 2020.

[6] Quan Fang, Jitao Sang, Changsheng Xu, and Yong Rui. Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning. *IEEE Transactions on Multimedia*, 16(3):796–812, 2014.

[7] Feten Fekih-Romdhane, Haitham Jahrami, Manel Stambouli, Amthal Alhuwailah, Mai Helmy, Hanaa Ahmed Mohamed Shuwiekh, Cheikh Mohamed fadel Mohamed Lemine, Eqbal Radwan, Juliann Saquib, Nazmus Saquib, et al. Cross-cultural comparison of mental illness stigma and help-seeking attitudes: a multinational population-based study from 16 arab countries and 10,036 individuals. *Social Psychiatry and Psychiatric Epidemiology*, 58(4):641–656, 2023.

[8] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019.

[9] Siva Charan Reddy Gangireddy, Deepak P, Cheng Long, and Tanmoy Chakraborty. Unsupervised fake news detection: A graph-based approach. In *Proceedings of the 31st ACM conference on hypertext and social media*, pages 75–83, 2020.

[10] Adeep Hande, Karthik Puranik, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. Evaluating pretrained transformer-based models for covid-19 fake news detection. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 766–772. IEEE, 2021.

[11] Maryam Heidari, Samira Zad, Parisa Hajibabaee, Masoud Malekzadeh, SeyyedPooya HekmatiAthar, Ozlem Uzuner, and James H Jones. Bert model for fake news detection based on social bot activities in the covid-19 pandemic. In *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0103–0109. IEEE, 2021.

[12] Lianzhe Huang, Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. Text level graph neural network for text classification. *arXiv preprint arXiv:1910.02356*, 2019.

[13] Raj Jagtap, Abhinav Kumar, Rahul Goel, Shakshi Sharma, Rajesh Sharma, and Clint P George. Misinformation detection on youtube using video captions. *arXiv preprint arXiv:2107.00941*, 2021.

[14] Liu Jin, Zhaochun Sun, and Huifang Ma. Short text classification method with dual channel hypergraph convolution networks. In *2022 8th International Conference on Systems and Informatics (ICSAI)*, pages 1–6. IEEE, 2022.

[15] Rupinder Paul Khandpur, Taoran Ji, Steve Jan, Gang Wang, Chang-Tien Lu, and Naren Ramakrishnan. Crowdsourcing cybersecurity. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, nov 2017.

[16] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[17] Dong Li, Zhiming Xu, Sheng Li, and Xin Sun. Link prediction in social networks based on hypergraph. In *Proceedings of the 22nd international conference on world wide web*, pages 41–42, 2013.

[18] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605.05101*, 2016.

[19] Ashwag Maghraby and Hosnia Ali. Modern standard arabic mood changing and depression dataset. *Data in Brief*, 41:107999, 2022.

[20] Nikhil Mehta, Maria Leonor Pacheco, and Dan Goldwasser. Tackling fake news detection by continually improving social context representations using graph neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1363–1380, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[21] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. Fang: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1165–1174, 2020.

[22] Yuxiang Ren and Jiawei Zhang. Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[23] Lanyu Shang, Yang Zhang, Zhenrui Yue, YeonJung Choi, Huimin Zeng, and Dong Wang. A knowledge-driven domain adaptive approach to early misinformation detection in an emergent health domain on social media. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 34–41. IEEE, 2022.

[24] Shakshi Sharma, Ekanshi Agrawal, Rajesh Sharma, and Anwitaman Datta. Facov: Covid-19 viral news and rumors fact-check articles dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1312–1321, 2022.

[25] Shakshi Sharma and Rajesh Sharma. Identifying possible rumor spreaders on twitter: A weak supervised learning approach. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[26] Shakshi Sharma, Rajesh Sharma, and Anwitaman Datta. (mis) leading the covid-19 vaccination discourse on twitter: An exploratory study of infodemic around the pandemic. *IEEE Transactions on Computational Social Systems*, 2022.

[27] Kai Shu, H Russell Bernard, and Huan Liu. Studying fake news via network analysis: detection and mitigation. *Emerging research challenges and opportunities in computational social network analysis and mining*, pages 43–65, 2019.

[28] Kai Shu, Xinyi Zhou, Suhang Wang, Reza Zafarani, and Huan Liu. The role of user profiles for fake news detection. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 436–439, 2019.

[29] P Suthanthira Devi, S Karthika, P Venugopal, and R Geetha. Veracity analysis and prediction in social big data. In *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018*, pages 289–298. Springer, 2020.

[30] Rishabh Upadhyay, Gabriella Pasi, and Marco Viviani. Vec4cred: a model for health misinformation detection in web pages. *Multimedia Tools and Applications*, 82(4):5271–5290, 2023.

[31] Shengxian Wan, Yanyan Lan, Jiafeng Guo, Jun Xu, Liang Pang, and Xueqi Cheng. A deep architecture for semantic matching with multiple positional sentence representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[32] Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7370–7377, 2019.

[33] Jiawei Zhang, Bowen Dong, and S Yu Philip. Fakedetector: Effective fake news detection with deep diffusive neural network. In *2020 IEEE 36th international conference on data engineering (ICDE)*, pages 1826–1829. IEEE, 2020.

[34] Liang Zhao, Feng Chen, Jing Dai, Ting Hua, Chang-Tien Lu, and Naren Ramakrishnan. Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling. *PLOS ONE*, 9(10):e110206, 2014.

[35] Liang Zhao, Jiangzhuo Chen, Feng Chen, Fang Jin, Wei Wang, Chang-Tien Lu, and Naren Ramakrishnan. Online flu epidemiological deep modeling on disease contact network. *GeoInformatica*, 24(2):443–475, Apr 2020.

[36] Yuehua Zhao, Jingwei Da, and Jiaqi Yan. Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. *Information Processing & Management*, 58(1):102390, 2021.

[37] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. *Advances in neural information processing systems*, 19, 2006.

[38] Omer Zulfiqar, Yi-Chun Chang, Po-Han Chen, Kaiqun Fu, Chang-Tien Lu, David Solnick, and Yanlin Li. Risecure: Metro incidents and threat detection using social media. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 531–535, 2020.