Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

A Domain Adaptive Graph Learning Framework to Early Detection of Emergent Healthcare Misinformation on Social Media

Lanyu Shang, Yang Zhang, Zhenrui Yue, YeonJung Choi, Huimin Zeng, Dong Wang

School of Information Sciences, University of Illinois Urbana-Champaign, Champaign, IL, USA
{lshang3, yzhangnd, zhenrui3, yc55, huiminz3, dwang24}@illinois.edu

Abstract

A fundamental issue in healthcare misinformation detection is the lack of timely resources (e.g., medical knowledge, annotated data), making it challenging to accurately detect emergent healthcare misinformation at an early stage. In this paper, we develop a crowdsourcing-based early healthcare misinformation detection framework that jointly exploits the medical expertise of expert crowd workers and adapts the medical knowledge from a source domain (e.g., COVID-19) to detect misleading posts in an emergent target domain (e.g., Mpox, Polio). Two important challenges exist in developing our solution: (i) How to leverage the complex and noisy knowledge from the source domain to facilitate the detection of misinformation in the target domain? (ii) How to effectively utilize the limited amount of expert workers to correct the inapplicable knowledge facts in the source domain and adapt the corrected facts to examine the truthfulness of the posts in the emergent target domain? To address these challenges, we develop CrowdAdapt, a crowdsourcing-based domain adaptive approach that effectively identifies and adapts relevant knowledge facts from the source domain to accurately detect misinformation in the target domain. Evaluation results from two real-world case studies demonstrate the superiority of CrowdAdapt over state-of-the-art baselines in accurately detecting emergent healthcare misinformation.

Introduction

With the ubiquity of digital content and the proliferation of social networks, the far-reaching spread of misinformation on social media has become a severe societal issue and raised wide public concerns (Zhou and Zafarani 2020). Among diverse domains of misinformation, healthcare misinformation is a critical category that has caused serious societal impacts by threatening the health and well-being of the general public and undermining the trustworthiness of mass media (Kou et al. 2022b). A fundamental issue in healthcare misinformation is the *early detection* of misinformation in an emergent healthcare domain, such as the recent outbreak of Mpox (or Monkeypox) and Polio, due to the lack of timely resources (e.g., up-to-date medical knowledge, annotated data). Specifically, an emergent healthcare domain refers to an emerging health event/topic (e.g., disease outbreak, food safety incident) that requires prompt responses and immediate actions (Shang et al. 2022b). In addition, the discrepancy between different healthcare domains can lead to degraded model performance when a misinformation detector is directly applied to an emergent domain different from the source domain it was trained in (Zhang et al. 2020). In this paper, we study the problem of domain adaptive early healthcare misinformation detection, where the goal is to explore the domain discrepancy between different healthcare domains and detect emergent misinformation. The outcome of our study can be adopted by healthcare agencies (e.g., Department of Public Health) and social media platforms to take timely response actions (e.g., providing precautionary information to the public) to help healthcare stakeholders and online users make well-informed healthcare decisions.

With the advanced information processing ability in machine learning and deep learning, existing healthcare misinformation detection models can achieve reasonable performance (Zhou and Zafarani 2020). However, these solutions often rely on complex model architectures and a large amount of well-annotated training samples to learn useful features and patterns for identifying misinformation (Weinzierl and Harabagiu 2021). Thus, it is impractical for such solutions to detect misinformation in an emergent healthcare domain that often lacks ground-truth labels. Moreover, a few recent knowledge-driven solutions also incorporate healthcare-related knowledge facts (i.e., entities and their relations) from medical documents (e.g., medical research publications and fact-checking articles) in a specific domain to improve the healthcare misinformation detection performance (Kou et al. 2022a). While such knowledge-driven approaches can complement the lack of medical knowledge, they are inadequate for detecting misinformation in an emergent healthcare domain that not only lacks medical documents but also presents a certain domain discrepancy with other domains. Therefore, the early detection of emergent healthcare misinformation remains a challenging problem.

In this paper, we propose a knowledge-based domain adaptive solution to detect misinformation in an emergent healthcare domain. While some emergent healthcare domains, such as the recent outbreaks of Mpox and Polio, are related to known pathogens (e.g., poxvirus, poliovirus), we observe that the misinformation from such emergent healthcare domains is often more relevant to the recent news/infor-

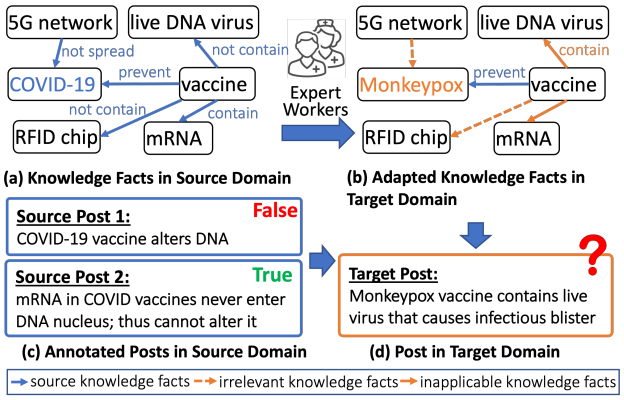


Figure 1: Domain Adaptive Emergent Healthcare Misinformation Detection

mation than the historical resources about the disease. For example, a popular misleading post about Mpox claims that "the Monkeypox disease is from the chimpanzee adenovirus used in the COVID vaccine." Such kind of misinformation cannot be identified with the resources in the Mpox literature that were obtained before the COVID-19 pandemic. Motivated by the above observation, our work aims to leverage the rich and timely resources (e.g., annotated data, medical reports) from a relevant healthcare domain (e.g., COVID-19) to identify misinformation in an emergent healthcare domain. Figure 1 shows an example of the domain adaptive emergent healthcare misinformation detection problem. In particular, we develop a crowdsourcing-based strategy to explore the medical knowledge of expert workers (i.e., crowd workers with medical expertise) to adapt the abundant resources from a *source domain* (i.e., the healthcare domain with sufficient annotated data and medical knowledge) to detect misleading posts in an emergent *target domain* (i.e., the healthcare domain that is short of annotated data and medical knowledge). However, two important challenges are identified in developing our solution.

Complex and Noisy Knowledge in Source Domain. A source domain (e.g., COVID-19) often contains a large amount of literature resource (e.g., research publications, fact-checking articles) from which the knowledge facts can be extracted for misinformation detection (Cui et al. 2020). A straightforward solution is to directly use the knowledge facts from the source domain to detect misinformation in the target domain. However, such a solution often ignores the complex nature of knowledge facts in the source domain, which often contains many knowledge facts that are irrelevant to the target domain. For example, the knowledge facts extracted from fact-checking articles often contain many non-medical entities, such as "5G network" and "RFID chip" in Figure 1(a), that are often irrelevant to medical science and are rarely seen in misinformation from the domains other than COVID-19. Such irrelevant knowledge facts can be of little help in detecting misinformation in the target domain. Thus, the first challenge is how to leverage complex and noisy knowledge facts from the source domain to facilitate misinformation detection in the target domain.

Inapplicable Knowledge from Source to Target Domain. While some knowledge facts from the source domain can be generalized to identify misinformation in the target domain, there often exists a non-trivial amount of knowledge facts in the source domain that are relevant but inapplicable to the target domain. For example, "vaccine" $\xrightarrow{\text{not contain}}$ "live DNA virus" (Figure 1(a)) is a widely accepted knowledge fact in the source domain of COVID-19. However, the knowledge fact is inapplicable to the target domain of Mpox, where the vaccine does contain live vaccinia virus (a DNA virus) that can cause serious vaccine adverse events among people with immunocompromising conditions. Such inapplicable knowledge facts have to be identified and corrected to detect misinformation in the target domain. However, it often requires expertise from medical experts to fully examine the inconsistency of knowledge facts between different domains, which is both labor-intensive and time-consuming (Kou et al. 2022a). Therefore, the second challenge to be addressed is how to efficiently utilize the limited amount of domain experts to correct inapplicable knowledge facts in the source domain and adapt the corrected facts to examine the truthfulness of posts in the target domain.

To address the above challenges, we develop CrowdAdapt, a novel crowdsourcing-based domain adaptive approach that effectively identifies and adapts relevant knowledge facts from the source healthcare domain to accurately detect misinformation in the emergent target healthcare domain. To address the first challenge, we design a context-driven knowledge fact extraction module to explicitly identify the relevant knowledge facts from the rich yet noisy knowledge facts in the source domain. To address the second challenge, we develop a consistency-aware knowledge updating strategy that incorporates the expertise of medical experts to adapt the knowledge facts in the source domain for detecting misinformation in the emergent target domain. To the best of our knowledge, CrowdAdapt is the first *crowdsourcing-based domain adaptive solution* for early healthcare misinformation detection in the emergent healthcare domain. The proposed CrowdAdapt can be further applied to a broader range of emerging domains (e.g., environment, drugs, criminal events) towards the early detection of misinformation on social media. We evaluate the performance of the CrowdAdapt framework using two real-world domain adaption case studies by leveraging the resources in the domain of COVID-19 (i.e., source domain) to detect misinformation in Mpox and Polio (i.e., target domains). Evaluation results demonstrate the superiority of CrowdAdapt over the state-of-the-art baselines in accurately detecting emergent healthcare misinformation.

Related Work

Healthcare Misinformation. A significant amount of efforts have been made to tackle the problem of health misinformation detection (Zhou and Zafarani 2020; Shang et al. 2021). For example, Cui *et al.* leveraged the co-attention mechanism to retrieve relevant information from a medical knowledge base to identify misinformation related to

cancer and diabetes (Cui et al. 2020). Weinzierl *et al.* utilized a domain-specific language model to detect COVID-19 vaccine misinformation on social media (Weinzierl and Harabagiu 2021). A fundamental limitation of existing healthcare misinformation detection solutions is that they mainly rely on a sufficient amount of ground-truth labels to supervise the training of an effective misinformation classification model. Therefore, such approaches often fall short of detecting emergent healthcare misinformation due to the lack of timely ground truth labels. In this paper, we present CrowdAdapt, a domain adaptive emergent healthcare misinformation detection solution that explores the domain discrepancy between different domains to adapt annotated data and medical knowledge in the source domain for detecting healthcare misinformation in the emergent target domain.

Domain Adaptation. Domain adaptation is a new learning technique that has recently been applied to mitigate the domain discrepancy issue in misinformation detection (Yue et al. 2023; Zeng et al. 2024). For example, Zhang *et al,* proposed a BERT-based domain adaptative model that designs a domain classifier to detect multimodal fake news in different domains (Zhang et al. 2020). Yue *et al.* developed a contrastive learning based misinformation detection framework by generating the pseudo-labels of the data in the target domain to improve the domain adaptation performance (Yue et al. 2022). These solutions mainly focus on capturing the domain shift between the source and target domains to reduce the model's reliance on the annotated target data. However, such solutions largely ignore the domain discrepancy of the knowledge facts associated with the posts, which is particularly important for detecting healthcare misinformation. To address such a limitation, CrowdAdapt develops a knowledge-driven domain adaptation mechanism to effectively incorporate medical knowledge facts for the early detection of emergent healthcare misinformation.

Crowdsourcing Knowledge Graph. Recent advances in crowdsourcing have been applied to facilitate the construction and modeling of knowledge graphs (Shang et al. 2022a). For example, Li *et al.* proposed a domain knowledge graph construction methodology that leverages the crowdsourcing efforts and text mining techniques to jointly construct an E-commerce knowledge graph for explainable product recommendation in online shopping (Li et al. 2020). Al-Khatib *et al.* designed an argumentation knowledge graph approach that acquires argumentation-based annotations from crowd workers to assist argumentative question answering (Al-Khatib et al. 2020). However, existing crowdsourcing knowledge graph solutions often assume a sufficient amount of resources (e.g., research publications, medical articles) can be leveraged to extract knowledge facts and construct the knowledge graph for the downstream tasks. Therefore, these solutions cannot be directly applied to obtain the medical knowledge for detecting misinformation in an emergent domain where a very limited amount of knowledge resources (e.g., research publications, fact-checking articles) is available. In contrast, CrowdAdapt designs a crowdsourcing-based knowledge-updating strategy that leverages the medical knowledge from expert workers to update and adapt knowledge facts in the source domain to effectively detect misinformation in the target domain.

Problem Statement

The problem of domain adaptive healthcare misinformation detection aims at adapting a misinformation classification model learned from the training data in the *source domain* to detect emergent healthcare misinformation in a *target domain*. We first introduce a few key concepts in the problem statement and formally formulate the domain adaptive healthcare misinformation detection problem.

Definition 1 Domain (d): A domain d is defined as a healthcare topic of interest (e.g., COVID-19, Mpox). In particular, we consider two types of domains in our study:

- **Source Domain** ($d = s$): a *high-resource* domain with adequate annotated data and medical knowledge.
- **Target Domain** ($d = t$): a *low-resource* domain often related to an emergent healthcare topic where very limited annotated data and medical knowledge are available.

Definition 2 Post (p): A post p is defined as a piece of text (e.g., "COVID-19 vaccine alters DNA" in Figure 1) where the depicted content is relevant to a healthcare domain (e.g., COVID-19). Specifically, we denote $P_s = \{p_1^s, p_2^s, \cdots, p_M^s\}$ and $P_t = \{p_1^t, p_2^t, \cdots, p_N^t\}$ to be the sets of source posts and target posts from the source domain s and target domain t, respectively.

Definition 3 Source Article (l): A source article $l \in L$ refers to an online article that is related to the source domain (i.e., COVID-19). In our study, we mainly focus on two types of source articles, including the *news articles* that are collected from credible online news publishers (e.g., CDC, Mayo Clinic), and *fact-checking articles* from fact-checking organizations (e.g., FactCheck.org, Politifact).

Definition 4 Ground-truth Label (y): We define the ground-truth label y_p of each post $p \in P_s \cup P_t$ as a binary label (i.e., $y_p \in \{0, 1\}$). Specifically, a post p is *misleading* (i.e., $y_p = 0$) if it contains entirely or partially false or unverified information which may contribute to both imminent and long-term harm to public health and safety (Wang et al. 2019). Otherwise, the post is considered as *non-misleading* ($y_p = 1$). We also denote Y_s and Y_t as ground-truth labels for posts in P_s and P_t, respectively.

Definition 5 Crowdsourcing Platform (C): A crowdsourcing platform refers to an online platform where *requesters* can request various services from *crowd workers* with diversified expertise via compensated *crowdsourcing tasks* (Turk 2016).

Definition 6 Expert Workers: The expert workers are the group of crowd workers who are verified by the crowdsourcing platform to have professional healthcare knowledge and are capable of crowdsourcing tasks that require healthcare expertise. We will introduce the details of the crowdsourcing task in the next section.

The goal of our crowdsourcing-based domain adaptive healthcare misinformation detection problem is to optimize

the misinformation detection performance on the target domain by leveraging the collaborative efforts from the crowdsourcing platform C, annotated source posts (P_s, Y_s) and source articles L. Formally, our problem is formulated as an adaptive binary classification problem that classifies each post in the target domain ($p_n^t \in P_t$) into two categories (i.e., misleading or non-misleading) with the objective:

$$\arg\max_{\hat{y}_n^t} \Pr(\hat{y}_n^t = y_n^t | P_s, P_t, Y_s, L, C), \ \forall 1 \leq n \leq N \quad (1)$$

where y_n^t and $\hat{y}_n^t$ are the ground-truth and estimated label of the target post p_n^t, respectively.

Solution

An overview of the CrowdAdapt framework is shown in Figure 2. In particular, CrowdAdapt consists of three main modules: 1) a *Graph-based Knowledge Encoder (GKE)* module that constructs a graph-based medical knowledge information network to explicitly model the medical knowledge facts and extract the useful knowledge facts related to the posts from different domains; 2) a *Domain-invariant Representation Learning (DRL)* module that aims to jointly learn the domain-invariant representation of the posts and their relevant knowledge facts extracted by the GKE module; and 3) a *Crowdsourcing-based Knowledge Updater (CKU)* module that incorporates the medical expertise from expert workers to verify and correct the uncertain medical knowledge facts extracted from GKE and accurately detect misleading posts in the target domain.

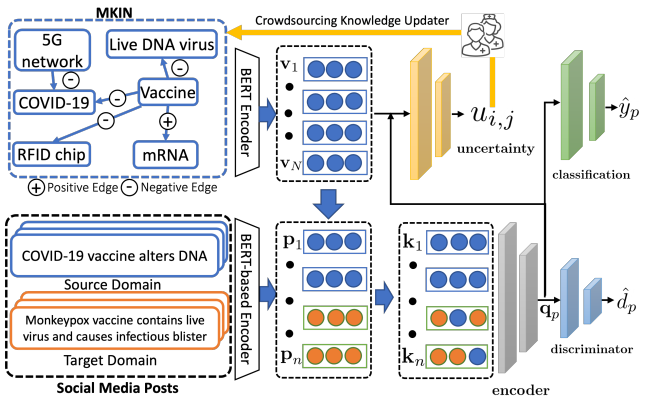


Figure 2: Overview of the CrowdAdapt Framework

Graph-based Knowledge Encoder

The graph-based knowledge encoder module designs a graph-based knowledge information network to explicitly explore the relationship between different healthcare-related entities and extract useful healthcare knowledge facts that are relevant to the posts from a given domain. We observe that existing domain adaptive misinformation detection solutions mainly focus on leveraging the data annotations (i.e., labeled posts) in the source domain to reduce the model's reliance on the data annotations in the target domain (Li et al. 2021; Zhang et al. 2020). However, such solutions largely ignore the healthcare knowledge facts associated with the posts, which is particularly important for identifying misleading posts in emergent healthcare domains. Therefore, to mitigate such a limitation, we develop a graph-based medical knowledge information network to explicitly extract the medical knowledge information from the widely available articles in the source domain (i.e., source articles) to facilitate the detection of misinformation in the target domain. We first define the medical knowledge information network (MKIN) as follows.

Definition 7 Medical Knowledge Information Network: We define the medical knowledge information network as a direct graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ and $\mathcal{E}$ refer to the *nodes* and *edges* that are defined below, respectively.

Definition 8 Node: We define a node v as a semantic entity (e.g., "vaccine" in Figure 2) that is extracted from a source article. In particular, we denote a set of N nodes as $\mathcal{V} = \{v_1, v_2, \cdots, v_N\}$.

Definition 9 Edge: We define an edge e as the semantic relation between a pair of relevant nodes in MKIN. Specifically, we consider two types of edges in our study, i.e., $e \in \{e^+, e^-\}$, where e^+ represent the "positive" relation between a pair of entities (e.g., the "contain" relation between "vaccine" and "mRNA" in Figure 2) and e^- represent the "negative" relation between a pair of entities (e.g., the "not spread" relation between "5G network" and "COVID-19" in Figure 2). We denote a set of M edges in MKIN as $\mathcal{E} = \{e_1, e_2, \cdots, e_M\}$. In addition, we also define two *binary* adjacency matrices A^+ and A^- to explicitly indicate the pairwise positive and negative relations of all nodes in $\mathcal{G}$, respectively. In particular, $A^+_{i,j} = 1$ and $A^-_{i,j} = 1$ indicates the positive and negative relation between node v_i and v_j, respectively. Otherwise, $A^+_{i,j}$ and $A^-_{i,j}$ are 0, indicating no relation between node v_i and v_j.

Definition 10 Knowledge Triple: We also define a knowledge triple $t = (v, e, v')$ as a pair of relevant nodes v and v' that are connected via an edge e in $\mathcal{G}$.

With the medical knowledge information network $\mathcal{G}$ constructed as above, our next objective is to learn the context-aware semantic representation of each node in MKIN by exploring its semantic dependency on other relevant nodes in MKIN. In particular, we first develop a BERT-based semantic encoder to extract the semantic representation of each node in MKIN. Formally, let $v_i = [w_1, w_2, \cdots, w_{n_i}]$ be the semantic entity of node $v_i \in \mathcal{V}$, where w_k for $1 \leq k \leq n_i$ is the k^{th} word in node v_i. We first adopt a pre-trained BERT model (Devlin et al. 2018) to retrieve the word embedding $\mathbf{u}_k$ of each word w_k, where $\mathbf{u}_k \in \mathbb{R}^d$ and d is the dimension of the word embedding. We then apply the mean-pooling and max-pooling to the word embeddings of each node and concatenate the pooled embeddings to obtain the final node embedding $\mathbf{v}_i \in \mathbb{R}^{2d}$ that aggregates the semantic representation of each node $v_i \in \mathcal{V}$. We also define a node embedding matrix $V \in \mathbb{R}^{N \times 2d}$ as the matrix that contains the node embeddings of all nodes in MKIN.

While the node embeddings can capture the semantic meaning of each node in MKIN, it remains a challenge to effectively extract the key knowledge triples from MKIN to identify the misinformation in the target domain. This is because MKIN is constructed from a number of articles in the source domain and often contains many knowledge triples that are irrelevant to the topics discussed in the posts from a target domain. For example, the knowledge triple ("vaccine", ⊖, "RFID chip") in Figure 2 is of little help for identifying the misleading post "Mpox vaccine contains live virus and causes infectious blister" in the target domain of Mpox. To address such a challenge, we design a post-based knowledge triple refinement strategy to explicitly capture the critical knowledge triples that are relevant to the given post. For example, the knowledge triples related to the "live DNA virus" can be captured in MKIN to facilitate the detection of the misleading post that the live DNA virus in Mpox vaccine causes infectious blister. Thus, we explicitly measure the semantic relevance between a post and each node in $\mathcal{V}$ to obtain the knowledge triples that are more relevant to a given post. In particular, we adopt the same BERT-based encoding strategy to encode each post $p \in \{P_s, P_t\}$ and denote the encoded vector representation of p as $\mathbf{p} \in \mathbb{R}^{1\times 2d}$. Finally, the post-based knowledge triple refinement strategy to obtain the refined adjacency matrices $\hat{A}^+$ and $\hat{A}^-$ as follows.

$$\hat{A}_p^+ = f\left((V\mathbf{p}^\top W_a^+) \odot A^+\right); \hat{A}_p^- = f\left((V\mathbf{p}^\top W_a^-) \odot A^-\right) \tag{2}$$

where $V \in \mathbb{R}^{N\times 2d}$ is node embedding matrix. $f(\cdot)$ is the softmax function, and W_a^+ and W_a^- are learnable weights.

Domain-invariant Representation Learning

Given the fact that there are often no ground-truth labels for the post in an emergent target domain to supervise the learning, our next objective is to learn the domain-invariant representations of the posts and their relevant medical knowledge triples in the refined MKIN from the GKE module by only using the available ground-truth labels of the posts from the source domain. Existing domain adaptive learning frameworks mainly focus on the domain discrepancy of the post content between the source and target domains and target to map the post content from different domains to a domain-invariant feature space. However, such solutions ignore the domain discrepancy of medical knowledge information in MKIN, which is critical to identify misleading posts in different healthcare domains. To overcome such a limitation, we present a joint representation learning framework to jointly learn the domain-invariant representations of the posts from different domains as well as their relevant knowledge triples. In particular, we first aggregate the medical knowledge information from the refined MKIN by propagating the node information based on their relations captured in the refined adjacency matrices $\hat{A}^+$ and $\hat{A}^-$ (Eq. 2). Formally, the knowledge aggregation process is defined as

$$\hat{\mathbf{v}}_i = \sigma\left(\sum_{\mathbf{v}_j\in\mathcal{N}_i^+} \frac{1}{\omega_i^+} W^+\mathbf{v}_j + \sum_{\mathbf{v}_j\in\mathcal{N}_i^+} \frac{1}{\omega_i^-} W^-\mathbf{v}_j + \mathbf{v}_i\right) \tag{3}$$

where $\hat{\mathbf{v}}_i$ is the learned node representation of $v_i \in \mathcal{V}$ with the knowledge information propagated from neighborhood nodes of v_i. $\sigma(\cdot)$ is the non-linear ReLU activation function. $\mathbf{v}_i$ and $\mathbf{v}_j$ are the node embeddings of v_i and v_j in $\mathcal{V}$, respectively. $\mathcal{N}_i^+$ and $\mathcal{N}_i^-$ refer to the set of neighborhood nodes of $v_i \in \mathcal{V}$ under the edge e^+ and e^-, respectively. W^+, W^-, and W are the learnable weight parameters. ω_i^+ and ω_i^- are the normalization factors of W^+ and W^-, respectively. We further aggregate the node representation for $v_i \in \mathcal{V}$ to obtain the representation of knowledge triples that are relevant to a post $p \in \{P_s, P_t\}$ based on the score measured in $\hat{A}_p^+$ and $\hat{A}_p^-$ (Eq. 2), followed by an average operation. The aggregated knowledge triple representation is defined as $\mathbf{t}_p = [\mathbf{t}_p^+||\mathbf{t}_p^-]$, where $||$ denotes the concatenation operation. $\mathbf{t}_p^+$ and $\mathbf{t}_p^-$ are the knowledge triple representations computed based on the learned node representation $\hat{\mathbf{v}}_i$ of all $v_i \in \mathcal{V}$ and the refined adjacency matrices $\hat{A}_p^+$ and $\hat{A}_p^-$, respectively.

Using the aggregated representations of the knowledge triples, we design a discriminative encoder network with an adversarial loss to jointly learn the knowledge-enriched representation of the posts from the source and target domains while minimizing the domain divergence of the learned features. In particular, the discriminative encoder network consists of two main components: 1) a two-layer *encoder network* that aims to learn the key information from the input posts and relevant knowledge triples in MKIN; 2) a two-layer *discriminator network* that targets at accurately distinguishing the domain of the encoded posts and their relevant knowledge triples. Formally, the encoder network and discriminator network are defined as follows.

$$\mathbf{q}_p = \textbf{encoder}([\mathbf{p}||\mathbf{t}_p]) \quad \text{and} \quad \hat{d}_p = \textbf{discriminator}(\mathbf{q}_p) \tag{4}$$

where $\mathbf{p}$ and $\mathbf{t}_p$ are the encoded vector representation and the aggregated knowledge triple representation of posts p, respectively. $\mathbf{q}_p$ is the knowledge-enriched representation of a post p and $\hat{d}_p$ is the estimated domain of $\mathbf{q}_p$.

With the discriminative encoder network defined above, we adopt the adversarial loss to effectively regulate the encoder network (Eq. 4) to learn the domain-invariant representation from the posts and their relevant knowledge triples that cannot be distinguished by the discriminator network (Eq. 4). Formally, the adversarial loss is defined as follows:

$$\mathcal{L}_{adv} = \sum_{p\in\{P_s,P_t\}} -d_p \log(\hat{d}_p)_1 - (1-d_p)\log(1-(\hat{d}_p)_0) \tag{5}$$

where d_p and $\hat{d}_p$ are the true and estimated domain of post $p \in \{P_s, P_t\}$, respectively.

The latent representation learned from the discriminative encoder network effectively captures the domain-invariant knowledge-enriched features of the posts from the source and target domains. Such domain-invariant features with minimized domain discrepancy can be leveraged to detect misleading posts regardless of the domain of the posts. In particular, we employ a two-layer classification network to accurately predict the truthfulness of each post. Formally,

the classification network is defined as $\hat{y}_p = MLP(\mathbf{q}_p)$. We optimize the classification network with cross-entropy loss:

$$\mathcal{L}_{cla} = -\sum_{p \in P_s} (1 - y_p) \log(1 - (\hat{y}_p)) + y_i \log(\hat{y}_p) \quad (6)$$

where y_p is the ground-truth label of the source post $p \in P_s$.

The overall learning objective $\mathcal{L}$ is to jointly optimize the discriminative encoder network and the classification network by maximizing the adversarial loss $\mathcal{L}_{adv}$ and minimizing the classification loss $\mathcal{L}_{cla}$ as $\mathcal{L} = \mathcal{L}_{cla} - \lambda \mathcal{L}_{adv}$, where λ is a hyperparameter to be tuned for optimizing the trade-off between $\mathcal{L}_{adv}$ and $\mathcal{L}_{cla}$.

Crowdsourcing-based Knowledge Updater

The crowdsourcing-based knowledge updater (CKU) module is designed to leverage the medical expertise of the domain experts to verify and correct any uncertain knowledge triples in MKIN from the GKE module that may only be applicable in the source domain but cannot be directly adapted to detect misinformation in the target domain. We observe that the MKIN constructed from the articles in the source domain also contains the knowledge triples that are not applicable to examining the truthfulness of posts in the target domain. For example, the knowledge triple ("vaccine", ⊖, "live DNA virus") in MKIN from the source domain of COVID-19 (Figure 2) could lead to the incorrect prediction result on the true claim that "Mpox vaccine contains live virus that causes infectious blister," due to the conflicting fact that the Mpox vaccine is made with attenuated live DNA virus while the COVID-19 vaccine is not. Therefore, it is critical to identify and correct such inapplicable knowledge triples in MKIN to ensure that the medical knowledge obtained from the source domain can be applied to accurately detect misinformation in the target domain.

To address the above problem, we design a crowdsourcing-based knowledge updating strategy that incorporates the efforts of expert workers (i.e., domain experts from the crowdsourcing platform) to effectively identify and correct the knowledge triples in MKIN to accurately detect misinformation in the target domain. However, verifying the correctness of knowledge triples in a specific domain often requires background knowledge from domain experts who are often expensive and may not always be available (Kou et al. 2022a). Therefore, it is impractical to ask expert workers to annotate all knowledge triples in MKIN. To this end, we design a post-driven knowledge triple retrieval process to identify a set of uncertain knowledge triples in MKIN that can be sent to the expert workers to verify their applicability in the target domain. Intuitively, the knowledge-enriched domain-invariant representation of a post learned in the DRL module contains the semantic features of relevant knowledge triples in MKIN for examining the truthfulness of the post, which can also be leveraged to estimate the relationship between a pair of nodes in MKIN. For example, the knowledge-enriched representation of the post "COVID-19 vaccine alters DNA" can capture the critical knowledge features extracted from the knowledge triple ("vaccine", ⊖, "live DNA virus") for examining the truthfulness of the post. Such knowledge-enriched representation is expected to confidently infer the relationship (i.e., edge label) between the corresponding nodes in the knowledge triple. Therefore, we train an MLP-based edge classifier to classify the edge label $e^p_{i,j}$ between a pair of nodes (v_i, v_j) in MKIN based on the context of a post p. In particular, we consider three categories of $e^p_{i,j}$, including "positive", "negative", and "no relation" as the edges identified in MKIN, and define the edge classifier as $\Pr(\hat{e}^p_{i,j}) = MLP([\mathbf{q}_p || \mathbf{v}_i || \mathbf{v}_j])$ where $\hat{e}^p_{i,j}$ is the estimated edge label of $e^p_{i,j}$. $\mathbf{q}_p$ is the domain-invariant representation of a post $p \in P_s$ and $\mathbf{v}_i$ and $\mathbf{v}_j$ are the BERT-encoded representation of nodes $v_i, v_j \in \mathcal{V}$, respectively. We optimize the edge classifier with cross-entropy loss between $e^p_{i,j}$ and $\hat{e}^p_{i,j}$.

We then measure the overall uncertainty of each knowledge triple in MKIN in the target domain based on the entropy of the estimated edge labels obtained from the edge classifier. Formally, let $t_{i,j} = (v_i, e_{i,j}, v_j)$ be the knowledge triple containing nodes v_i and v_j, and the uncertainty of each knowledge triple $t_{i,j}$ is computed as

$$u_{i,j} = -\sum_{p \in P_t} \Pr(\hat{e}^p_{i,j}) \times \log \Pr(\hat{e}^p_{i,j}) \quad (7)$$

We further retrieve the top K knowledge triples with the highest uncertainty scores and K is a tunable hyperparameter to be determined based on the model performance and budget. The retrieved knowledge triples are then sent to the expert workers for applicability verification. We show the details of the crowdsourcing task in the Evaluation section. Finally, we update MKIN with the expert-verified knowledge triples (i.e., the knowledge triples verified by the crowd experts) and further optimize the discriminative encoder network and classification network in DRL to accurately detect misinformation in the target domain.

Evaluation

In this section, we evaluate the healthcare misinformation detection performance of CrowdAdapt in various domain adaptation scenarios. In particular, we adopt *COVID-19* as the source domain, and choose *Mpox* and *Polio* as the target domains to evaluate the domain adaptation effectiveness of CrowdAdapt. COVID-19 has been a popular healthcare domain of misinformation since the beginning of the global pandemic, and many efforts have been made to combat the spread of COVID-19 misinformation. The recent outbreaks of Mpox (in May 2022) and Polio (in July 2022) are trending healthcare domains that have attracted a non-trivial amount of misinformation but lack sufficient timely resources for misinformation detection. Therefore, we consider Mpox and Polio as our target healthcare domains in our study. Evaluation results from extensive experiments show that CrowdAdapt achieves significant performance gains compared to state-of-the-art baselines in terms of early healthcare misinformation detection accuracy.

Datasets

Source Articles We focus on two types of source articles, including the *medical news articles* and *fact-checking ar-*

ticles. The medical news articles are online articles from reliable medical news publishers (e.g., CDC, Mayo Clinic) discussing the up-to-date medical information (e.g., official guidance, treatments, precautions) related to the source domain. The fact-checking articles are the reports published by professional journalists and scholars on mainstream fact-checking websites (e.g., FactCheck.org, Politifact) concerning the misinformation related to the source domain. We finally collected 259 source articles in our study.

Posts We collect social media posts from both the source and target domains to study the domain adaptation performance of CrowdAdapt.

Source Posts. The goal of CrowdAdapt is to leverage existing annotated datasets in the source domain (i.e., source posts) to detect misinformation in an emergent healthcare domain that has limited or no annotated data (i.e., target posts). Therefore, we use five widely adopted public COVID-19 misinformation datasets with ground-truth labels as the source post datasets, including Constraint (Patwa et al. 2021), COVIDRumor (Cheng et al. 2021), MM-CoVar (Chen, Chu, and Subbalakshmi 2021), ANTi-Vax (Hayawi et al. 2022), and CMU-MisCov19 (Memon and Carley 2020). We use the ground-truth labels provided in each dataset and remove invalid posts that are duplicate or cannot be retrieved. We note that existing COVID-19 misinformation datasets (i.e., Constraint, COVIDRumor, MM-CoVar, ANTiVax) primarily annotate the source posts into binary classes (i.e., misleading or non-misleading). Following such a conventional practice, we adopt the original binary labels in each dataset for our experiments. For the dataset with non-binary ground-truth labels, such as CMU-MisCov19 that also categorizes COVID-19 posts into topic-based classes (e.g., "True Prevision", "False Fact or Prevention"), we further group these non-binary labels into binary classes in terms of their veracity meaning. A summary of the source post datasets is presented in Table 1. While we focus on binary classification in our experiments, we also acknowledge that healthcare misinformation detection is a complex problem where certain posts may not be sufficiently classified into binary classes. We believe the designed framework of CrowdAdapt can be further extended to address the multi-class healthcare misinformation problem. The details about the generalization of CrowdAdapt will be discussed in the Discussion section.

Dataset	# Posts	# Misleading	# Non-misleading
Constraint	10,700	5,600	5,100
COVIDRumor	5,505	3,661	1,844
MMCoVaR	2,791	1,315	1,476
ANTiVax	12,326	4,156	8,170
CMU-MisCov19	3,114	1,269	1,845

Table 1: Summary of Source Post Datasets

Target Posts. We collect the target posts from Twitter[1] based on the relevant keywords in the Mpox and Polio domains. For each dataset, we randomly select 500 posts as the test set to evaluate the early misinformation detection performance and use the remaining data for the unsupervised training in CrowdAdapt. We invite three independent healthcare experts at our institution to annotate the target posts in the test sets and obtain the ground-truth labels based on their majority votes to ensure the label quality. We summarize the target post datasets in Table 2.

	Mpox	**Polio**
# Posts	9,156	12,893
# Annotated Posts	500	500
# Misleading	168	141
# Non-misleading	332	359
Date Range	5/1-5/31, 2022	7/1-7/31, 2022

Table 2: Summary of Target Post Datasets

Baselines and Experiment Setup

Baselines and Implementation Details We compare CrowdAdapt with state-of-the-art baselines in domain adaptive and knowledge graph based misinformation detection.

- **BDANN** (Zhang et al. 2020): BDANN is a BERT-based domain adaptation solution for multimodal fake news detection. We exclude the visual features in BDANN and leverage the BERT-based feature extraction model trained on the source posts to classify target posts.
- **MDA-WS** (Li et al. 2021): MDA-WS is a weakly supervised domain adaptive fake news detection framework that leverages labeled source domain news articles and the word frequency based weak labels of target domain news articles to detect fake news in the target domain.
- **EANN** (Wang et al. 2018): EANN is an event adversarial network framework that learns transferable features from source news events for fake news detection on emerging news events.
- **DETERRENT** (Cui et al. 2020): DETERRENT is a graph attention network solution that utilizes relational medical knowledge to detect misleading healthcare news.
- **CompGCN** (Vashishth et al. 2020): CompGCN is an advanced multi-relational knowledge graph solution that exploits the entity and their relations to extract key information from graph data.

To ensure a fair comparison, we keep the source and target posts to all compared methods the same in our evaluation. In addition, for the knowledge graph based methods (i.e., DETERRENT, CompGCN, CrowdAdapt), we use the same MKIN constructed in CrowdAdapt as the medical knowledge graph for classifying misleading posts. We

[1]https://developer.twitter.com/en/docs/twitter-api

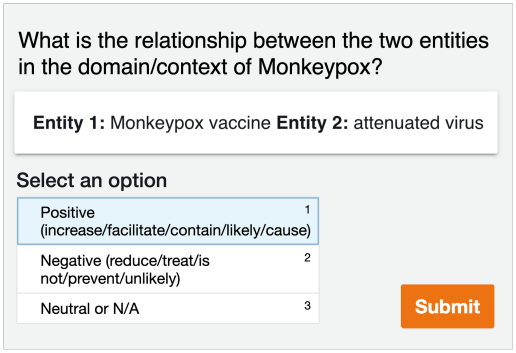


Figure 3: Example of Knowledge Triple Verification Task

strictly follow the model configurations of all baselines as documented in the original papers and carefully tune the hyperparameters to obtain the best results. In our experiments, we utilize all the source posts and unlabeled target posts for the unsupervised training of the encoder network and the domain discriminator network. Additionally, we use the labeled source posts for the supervised training of the classification network. We adopt the commonly used metrics for classification evaluation, including *Accuracy (Acc.)*, *Precision (Prec.)*, *Recall*, and *F1 Score (F1)*.

In our model implementation, we set the dimensions of the node embeddings and post embeddings as 768. The total number of epochs is set to 80 with a batch size of 32. We adopt an initial learning rate of 0.0001 with a decay of 0.95. We set the total number of retrieved uncertain knowledge triples K as 100. We run the experiments on Ubuntu 20.04 with four NVIDIA A40.

Crowdsourcing Platform We choose Amazon Mechanical Turk (MTurk) as the crowdsourcing platform to acquire expert knowledge in the target domain from healthcare professionals. MTurk is one of the largest crowdsourcing platforms that provides 24/7 crowdsourcing services from a large number of crowd workers with diversified expertise. In particular, we recruit the expert workers who have been verified by MTurk as "healthcare experts" to participate in our study (Turk 2016). In addition, we also developed a domain screening test for each studied target domain to ensure the qualification of the expert workers. The qualified expert workers will be assigned to the knowledge triple verification tasks (Figure 3). To ensure the quality of the response, we only select the qualified expert workers with 95% or higher Human Intelligence Task (HIT) rate. To reduce the potential bias in the crowdsourcing responses, we recruit 5 expert workers for each knowledge triple verification task and apply the majority voting to resolve any conflicts between the responses. The inter-rater agreement of the responses for the Mpox and Polio datasets are 0.74 and 0.71 in terms of the kappa score, and 0.87 and 0.85 in terms of intraclass correlation coefficient (ICC), respectively. A kappa score above 0.60 and an ICC above 0.75 indicate substantial agreement among the annotators (Chaturvedi and Shweta 2015). We pay $0.47 per knowledge triple in our experiment, including the payment to both the expert worker and MTurk.

Constraint				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.554	0.443	0.587	0.505
MDA-WA	0.624	0.667	0.614	0.639
EANN	0.576	0.596	0.606	0.601
DETERRENT	0.636	0.682	0.643	0.662
CompGCN	0.622	0.651	0.609	0.630
CrowdAdapt	**0.640**	**0.688**	**0.652**	**0.670**
COVIDRumor				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.658	0.638	0.733	0.683
MDA-WA	0.628	0.626	0.748	0.682
EANN	0.534	0.558	0.563	0.560
DETERRENT	0.672	0.668	0.727	0.696
CompGCN	0.648	0.659	0.677	0.668
CrowdAdapt	**0.682**	**0.702**	**0.793**	**0.745**
MMCoVaR				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.532	0.483	0.467	0.475
MDA-WA	0.604	0.627	0.619	0.623
EANN	0.548	0.471	0.474	0.472
DETERRENT	0.628	0.617	0.649	0.633
CompGCN	0.588	0.625	0.603	0.614
CrowdAdapt	**0.642**	**0.699**	**0.682**	**0.691**
ANTiVax				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.606	0.593	0.612	0.602
MDA-WA	0.592	0.590	0.601	0.596
EANN	0.584	0.558	0.564	0.557
DETERRENT	0.638	0.676	0.618	0.645
CompGCN	0.618	0.637	0.607	0.621
CrowdAdapt	**0.648**	**0.693**	**0.681**	**0.687**
CMU-MisCov19				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.634	0.658	0.641	0.649
MDA-WA	0.681	0.661	0.692	0.676
EANN	0.592	0.613	0.596	0.604
DETERRENT	0.697	0.683	0.689	0.686
CompGCN	0.676	0.669	0.702	0.685
CrowdAdapt	**0.727**	**0.708**	**0.736**	**0.722**

Table 3: Detection Performance in Target Domain - Mpox

Detection Performance

We first compare the misinformation detection performance of CrowdAdapt with all baseline schemes for detecting misleading posts in the Mpox and Polio target domains. The evaluation results on the Mpox and Polio datasets are shown in Table 3 and Table 4, respectively. We observe that CrowdAdapt consistently outperforms all compared baselines on all source datasets for detecting misinformation in both Mpox and Polio datasets. For example, on the Mpox dataset, CrowdAdapt achieves a 1.2%, 7.1%, 9.1%, 6.5%, and 5.2% performance improvements against the best-performing baseline (i.e., DETERRENT) in terms of the F1 score on the Constraint, COVIDRumor, MMCoVar, ANTiVax, and CMU-MisCov19, respectively. We also observe similar performance gains on the Polio dataset. The performance gains can be attributed to the crowdsourcing-

Constraint				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.642	0.681	0.613	0.645
MDA-WA	0.636	0.673	0.625	0.649
EANN	0.644	0.618	0.652	0.634
DETERRENT	0.674	0.687	0.667	0.677
CompGCN	0.652	0.635	0.661	0.648
CrowdAdapt	**0.692**	**0.706**	**0.687**	**0.697**
COVIDRumor				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.702	0.688	0.701	0.694
MDA-WA	0.658	0.678	0.646	0.661
EANN	0.664	0.681	0.629	0.654
DETERRENT	0.688	0.671	0.693	0.682
CompGCN	0.660	0.657	0.673	0.665
CrowdAdapt	**0.722**	**0.709**	**0.744**	**0.726**
MMCoVaR				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.616	0.602	0.617	0.609
MDA-WA	0.602	0.635	0.591	0.612
EANN	0.626	0.617	0.631	0.624
DETERRENT	0.664	0.641	0.677	0.659
CompGCN	0.642	0.639	0.655	0.647
CrowdAdapt	**0.712**	**0.703**	**0.726**	**0.714**
ANTiVax				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.634	0.619	0.657	0.638
MDA-WA	0.652	0.643	0.659	0.651
EANN	0.676	0.651	0.685	0.667
DETERRENT	0.672	0.674	0.698	0.686
CompGCN	0.668	0.683	0.659	0.671
CrowdAdapt	**0.706**	**0.693**	**0.716**	**0.704**
CMU-MisCov19				
Baseline	Accuracy	Precision	Recall	F1
BDANN	0.669	0.675	0.684	0.679
MDA-WA	0.681	0.696	0.672	0.684
EANN	0.676	0.663	0.680	0.671
DETERRENT	0.708	0.689	0.703	0.696
CompGCN	0.692	0.661	0.686	0.673
CrowdAdapt	**0.731**	**0.728**	**0.719**	**0.723**

Table 4: Detection Performance in Target Domain - Polio

based domain adaptive knowledge verification strategy in CrowdAdapt that leverages the medical knowledge of expert workers to examine and correct the knowledge triples in MKIN for the accurate detection of misleading posts in the target domain. In addition, the significant performance improvements over knowledge-agnostic domain adaption solutions also highlight the importance of medical knowledge in detecting misinformation in emergent healthcare domains.

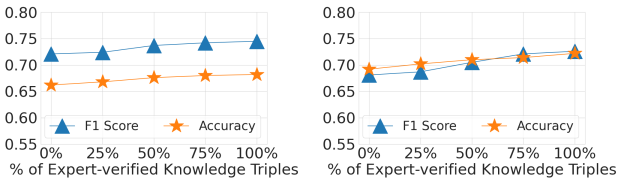
Ablation Study

We study the importance of the key components in the CrowdAdapt framework. In particular, we consider three variants of CrowdAdapt, including 1) **CrowdAdapt\G** that excludes the MKIN and only extracts the domain-invariant representation from the post content to detect misinformation, 2) **CrowdAdapt\P** that removes the post-based knowledge refinement and only applies the mean-pooling layer to obtain the knowledge representation from MKIN, 3) **CrowdAdapt\U** that excludes knowledge triples verified and corrected by expert workers, and only uses the original knowledge triples in MKIN to guide the misinformation detection.

The results of the ablation study on the Mpox and Polio datasets are summarized in Table 5. We observe that CrowdAdapt achieves its best misinformation detection performance when it incorporates all key components in the framework. In particular, we observe that the incorporation of the expert-verified knowledge triples in MKIN greatly enhances the domain adaptive misinformation detection on the target domain which further validates the effectiveness of crowdsourced expert knowledge in CrowdAdapt.

Effect of Expert-verified Knowledge Triples

We further investigate the effect of expert-verified knowledge triples on the detection performance of CrowdAdapt in the target domain. In particular, we vary the number of expert-verified knowledge facts to be annotated by the expert workers from 0% to 100% of the K retrieved knowledge triples in the CKU module. The results are reported in Figure 4. We use the COVIDRumor dataset as the dataset in the source domain of COVID-19 and evaluate the domain adaptive misinformation detection performance in the target domains of both Mpox and Polio. We observed similar performance gains on other COVID-19 datasets and omitted the evaluation results due to the page limit. In particular, we observe the overall performance of CrowdAdapt improves as the number of expert-verified knowledge triples increases and gradually plateaus after the number of expert-verified knowledge triples reaches 75% of the retrieved knowledge triples. A possible reason is that, as we retrieve additional knowledge triples from MKIN to be verified by the domain experts, the newly retrieved knowledge triples have lower uncertainty scores (i.e., the entropy of the prediction results of edge classifier), which are less likely to be corrected by domain experts and contribute less to CrowdAdapt for identifying misleading posts in the target domain.



(a) Mpox (b) Polio

Figure 4: Effect of Expert-verified Knowledge Triples

Discussion

In this study, we focus on the domain-adaptive emergent healthcare misinformation detection problem. Following the commonly adopted problem settings in misinformation detection (Zhou and Zafarani 2020), we formulate our problem as a binary classification task where the goal is to determine whether a social media post is misleading (i.e., containing entirely or partially false or unverified information)

Target Domain	Method	Constraint		COVIDRumor		MMCoVaR		ANTiVax	
		Accuracy.	F1	Accuracy	F1	Accuracy.	F1	Accuracy.	F1
MPox	CrowdAdapt	**0.640**	**0.670**	**0.682**	**0.745**	**0.642**	**0.691**	**0.648**	**0.687**
	CrowdAdapt\G	0.602	0.647	0.644	0.713	0.612	0.653	0.616	0.659
	CrowdAdapt\P	0.616	0.655	0.658	0.726	0.620	0.664	0.628	0.663
	CrowdAdapt\U	0.624	0.661	0.662	0.721	0.632	0.676	0.634	0.671
Polio	CrowdAdapt	**0.692**	**0.706**	**0.722**	**0.726**	**0.712**	**0.714**	**0.706**	**0.704**
	CrowdAdapt\G	0.668	0.677	0.684	0.691	0.688	0.697	0.678	0.687
	CrowdAdapt\P	0.672	0.683	0.688	0.679	0.696	0.701	0.684	0.688
	CrowdAdapt\U	0.676	0.681	0.692	0.681	0.702	0.709	0.692	0.698

Table 5: Results of Ablation Study

or not. However, we acknowledge that healthcare misinformation detection is a challenging problem where the misinformation can be categorized into different classes based on specific criteria, such as the degree of misleadingness (e.g., misleading, partially misleading, unsure), subject matter (e.g., conspiracy theories, fake cure), stance (e.g., agree, disagree, no stance), and perceived health risk (e.g., highly severe, possibly severe) (Memon and Carley 2020). To address such a challenge, we can extend CrowdAdapt by incorporating a multi-class classifier, which would enable the classification of each post into non-binary categories of interest. In particular, the CrowdAdapt framework can be optimized by replacing the current classification loss function (Equation 7) with the multi-class cross-entropy loss, i.e., $\mathcal{L}_{cla} = -\sum_{p \in P_s} \sum_{k=1}^{K} y_p^{(k)} \log \hat{y}_p^{(k)}$ where $y_p^{(k)} \in \{0, 1\}$ indicates whether the true label of a post p is class k ($y_p^{(k)} = 1$) or not ($y_p^{(k)} = 0$), and $\hat{y}_p^{(k)} \in (0, 1]$ is the predicted probability of a post p being in class k.

Moreover, we recognize another limitation inherent in binary classification for misinformation detection is that the use of binary label classes (i.e., misleading and non-misleading) may overlook the uncertainty in labeling due to the ambiguity of a post. This ambiguity often arises from the varying contexts and ideological perspectives among the audience (Jia et al. 2022). For instance, a post with suspicion of the slow and poor handling of HIV drugs can be classified as misleading based on linguistic characteristics (e.g., language style, sentiment). However, such a claim may be valid and legitimate, especially from the viewpoint of HIV-vulnerable communities (e.g., LGBTQ+). As a consequence, a binary misinformation detection model trained with such inherently uncertain labels may impose a limit on the misinformation detection performance. A potential solution to address the label uncertainty challenge is to integrate the misinformation classifier with the uncertainty estimation mechanism (Das, Basak, and Dutta 2022) to explicitly quantify the uncertainty of the model output (i.e., misleading or non-misleading). In particular, instead of using fixed binary labels for model optimization, we can consider the multiple annotations of each data sample as a probability distribution and optimize the classification framework with the probability-based Kullback-Leibler (KL) divergence loss to estimate the probability distribution (i.e., uncertainty) of the predictions. In addition, explainable machine learning approaches (Cui et al. 2020) may also aid in explaining and justifying the classification results, providing crucial support for healthcare-related decision-making across diverse communities.

In addition, we observe that the misinformation in different healthcare domains (e.g., COVID-19, Mpox, Polio) often presents their unique characteristics. For instance, while both the source domain (i.e., COVID-19) and target domain (i.e., Mpox, Polio) in our study are all health-related, the relative significance of the same topic (e.g., vaccine) can vary with respect to misinformation. In particular, within the COVID-19 domain, vaccine-related misinformation has emerged as a critical concern, which often highlights the negative effects of vaccine (e.g., DNA alternation, impact on women's fertility), and thus significantly increases COVID-19 vaccine hesitancy and amplifies the risk for vulnerable populations. In contrast, misinformation related to Mpox or Polio is often less relevant to the Mpox or Polio vaccine. Instead, the misinformation is more relevant to the alleged origins (e.g., LGBTQ+ communities) or transmission surrounding these diseases (e.g., sexual transmission), which is more likely to cause homophobic assertions. Such an observation of domain discrepancy further suggests the necessity of developing domain-adaptive misinformation detection solutions that explicitly address such discrepancy. In addition, the domain discrepancy also suggests the impact and public's tolerance of the same misinformation topic can vary across different domains. For example, vaccine hesitancy may be more pronounced in the COVID-19 domain compared to the Mpox or Polio domains. To address the domain discrepancy on the misinformation impact and tolerance, one possible approach is to leverage few-shot learning (Yue et al. 2023) that incorporates a small number of misinformation samples in each target domain, along with their quantified impact and tolerance, to enhance the model's capability of identifying the most consequential misinformation in different domains.

Another challenge of crowdsourcing-based domain-specific misinformation detection lies in the unknown expertise of the crowd workers. In the experiments, we recruit expert workers with premium qualifications in the healthcare domain to finish the knowledge triple verification task. While majority voting and interrater agreement are considered to reduce the uncertainty in the crowdsourcing responses, it is still possible that some expert workers have less relevant experience or knowledge to provide accurate annotations for certain knowledge triples. To lift the assumption that the crowdsourcing responses are equally valid and accurate, a potential solution is to explicitly quantify the confidence and certainty of each crowdsourcing response, such as asking expert workers to provide their confidence level in each response. Such confidence-aware knowledge triples can be further integrated into the Crowdsourcing-based Knowledge Updater module via the uncertainty-aware information aggregation strategy (Feng, Wang, and Ding 2021) to reduce the overall uncertainty of representation learned from the Medical Knowledge Information Network.

Scalability is an important factor for misinformation detection solutions, especially given the explosive amount of social media data input and emerging domains. First, the efficiency of analyzing healthcare-related social media posts is critical for providing timely prediction results in the early detection of misinformation in emergent healthcare domains. In particular, the time complexity of CrowdAdapt in the inference phase only grows linearly with respect to the number of social media posts to be classified in the target domain. To address the scalability challenge of classifying a number of posts in an emergent domain, a possible solution is to implement CrowdAdapt on distributed GPU clusters or cloud computing platforms to improve computing efficiency. We plan to address the scalability challenge in our future work. Second, the scalability of adapting CrowdAdapt to detect emergent healthcare misinformation across a wide array of health-related domains, especially for the ones that are novel and unseen before. Although the evaluation results have demonstrated the effectiveness and superiority of CrowdAdapt in detecting emergent healthcare misinformation in the target domain (i.e., Mpox, Polio), the overall misinformation detection performance still experiences a modest decline compared to the misinformation performance in the source domain (i.e., COVID-19). A possible reason is that the knowledge facts in the source domain might not encompass all relevant entities in the target domain. For example, entities related to sexually transmitted diseases (STDs), frequently seen in the misinformation related to Mpox, may not necessarily be covered by the knowledge facts in the COVID-19 domain due to the different transmission methods of the two diseases. To mitigate this limitation, we plan to incorporate the resources (e.g., annotated data and medical documents) from a more diverse set of source domains (e.g., HIV, HPV) in our future work to expand the coverage of knowledge facts and improve the domain-adaptive misinformation detection performance of CrowdAdapt. In addition, to further enhance the ability of CrowdAdapt in identifying novel and creative misinformation in emergent domains, we plan to incorporate contrastive learning (Yue et al. 2022), a self-supervised learning method that can learn discriminative features from data in different domains without requiring any data annotations from the target domains. Specifically, we will integrate the overall loss function of CrowdAdapt with the self-supervised contrastive loss to guide the framework to capture the latent features that are generalizable to novel and unseen data in the emergent healthcare domains.

Conclusion

In this paper, we study the problem of early misinformation detection in an emergent healthcare domain. We present CrowdAdapt, a crowdsourcing-based domain adaptive framework that effectively explores a high-resource source domain to accurately detect misinformation in an emergent target domain. We further leverage the expertise of expert workers to explicitly correct the inapplicable knowledge facts from the source to the target domain to improve the domain adaptation performance on misinformation detection. We conduct two real-world case studies of the domain adaptive misinformation detection from COVID-19 to Mpox and Polio using five COVID-19 misinformation datasets. Evaluation results show that CrowdAdapt achieves substantial performance gains compared to state-of-the-art baselines in accurately detecting misleading social media posts in both target domains.

Ethical Statement

Emergent healthcare misinformation is a critical issue on social media and online communities. While major social media platforms have made efforts to combat the spread of online misinformation, we note that the early detection of emergent healthcare misinformation remains a challenge to be fully addressed. We envision the success of this work will greatly mitigate the propagation of emergent healthcare misinformation and improve the information credibility of online social media. The research protocol was approved by the Institutional Review Board (IRB) at our institution. In our experiments, the datasets were obtained from publicly accessible websites and data repositories. To ensure the ethics of the study and preserve user privacy, we only collected the post information (i.e., post content and tweet ID). We did not obtain the user identity information associated with each post (e.g., username, user profile). To comply with the ethical standards and terms of service, we will only release the tweet IDs for the datasets collected in this study.

Acknowledgements

This research is supported in part by the National Science Foundation under Grant No. IIS-2202481, CHE-2105032, IIS-2130263, CNS-2131622, CNS-2140999. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

Al-Khatib, K.; Hou, Y.; Wachsmuth, H.; Jochim, C.; Bonin, F.; and Stein, B. 2020. End-to-end argumentation knowledge graph construction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 7367–7374.

Chaturvedi, S.; and Shweta, R. 2015. Evaluation of inter-rater agreement and inter-rater reliability for observational data: an overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3): 20–27.

Chen, M.; Chu, X.; and Subbalakshmi, K. 2021. MMCoVaR: Multimodal COVID-19 Vaccine Focused Data Repository for Fake News Detection and a Baseline Architecture for Classification. *arXiv preprint arXiv:2109.06416*.

Cheng, M.; Wang, S.; Yan, X.; Yang, T.; Wang, W.; Huang, Z.; Xiao, X.; Nazarian, S.; and Bogdan, P. 2021. A COVID-19 rumor dataset. *Frontiers in Psychology*, 12.

Cui, L.; Seo, H.; Tabar, M.; Ma, F.; Wang, S.; and Lee, D. 2020. DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Das, S. D.; Basak, A.; and Dutta, S. 2022. A heuristic-driven uncertainty based ensemble framework for fake news detection in tweets and news articles. *Neurocomputing*, 491.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Feng, B.; Wang, Y.; and Ding, Y. 2021. Uag: Uncertainty-aware attention graph neural network for defending adversarial attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 7404–7412.

Hayawi, K.; Shahriar, S.; Serhani, M. A.; Taleb, I.; and Mathew, S. S. 2022. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public health*.

Jia, C.; Boltz, A.; Zhang, A.; Chen, A.; and Lee, M. K. 2022. Understanding Effects of Algorithmic vs. Community Label on Perceived Accuracy of Hyper-partisan Misinformation. *Proceedings of the ACM on Human-Computer Interaction*.

Kou, Z.; Shang, L.; Zhang, Y.; and Wang, D. 2022a. Hc-covid: A hierarchical crowdsource knowledge graph approach to explainable covid-19 misinformation detection. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP): 1–25.

Kou, Z.; Shang, L.; Zhang, Y.; Yue, Z.; Zeng, H.; and Wang, D. 2022b. Crowd, Expert & AI: A Human-AI Interactive Approach Towards Natural Language Explanation Based COVID-19 Misinformation Detection. In *IJCAI*, 5087–5093.

Li, F.-L.; Chen, H.; Xu, G.; Qiu, T.; Ji, F.; Zhang, J.; and Chen, H. 2020. AliMeKG: Domain knowledge graph construction and application in e-commerce. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2581–2588.

Li, Y.; Lee, K.; Kordzadeh, N.; Faber, B.; Fiddes, C.; Chen, E.; and Shu, K. 2021. Multi-Source Domain Adaptation with Weak Supervision for Early Fake News Detection. In *2021 IEEE International Conference on Big Data*, 668–676.

Memon, S. A.; and Carley, K. M. 2020. Characterizing covid-19 misinformation communities using a novel twitter dataset. *arXiv preprint arXiv:2008.00791*.

Patwa, P.; Sharma, S.; Pykl, S.; Guptha, V.; Kumari, G.; Akhtar, M. S.; Ekbal, A.; Das, A.; and Chakraborty, T. 2021. Fighting an infodemic: Covid-19 fake news dataset. In *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation*, 21–29.

Shang, L.; Kou, Z.; Zhang, Y.; Chen, J.; and Wang, D. 2022a. A privacy-aware distributed knowledge graph approach to qois-driven covid-19 misinformation detection. In *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*, 1–10. IEEE.

Shang, L.; Kou, Z.; Zhang, Y.; and Wang, D. 2021. A multimodal misinformation detector for covid-19 short videos on tiktok. In *2021 IEEE international conference on big data (big data)*, 899–908. IEEE.

Shang, L.; Zhang, Y.; Yue, Z.; Choi, Y.; Zeng, H.; and Wang, D. 2022b. 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)*, 34–41. IEEE.

Turk, A. M. 2016. Introducing Premium Qualifications.

Vashishth, S.; Sanyal, S.; Nitin, V.; and Talukdar, P. 2020. Composition-based Multi-Relational Graph Convolutional Networks. In *ICLR*.

Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm international conference on knowledge discovery & data mining*, 849–857.

Wang, Y.; McKee, M.; Torbica, A.; and Stuckler, D. 2019. Systematic literature review on the spread of health-related misinformation on social media. *Social science & medicine*.

Weinzierl, M. A.; and Harabagiu, S. M. 2021. Automatic detection of COVID-19 vaccine misinformation with graph link prediction. *Journal of biomedical informatics*, 124.

Yue, Z.; Zeng, H.; Kou, Z.; Shang, L.; and Wang, D. 2022. Contrastive Domain Adaptation for Early Misinformation Detection: A Case Study on COVID-19. *arXiv preprint arXiv:2208.09578*.

Yue, Z.; Zeng, H.; Zhang, Y.; Shang, L.; and Wang, D. 2023. Metaadapt: Domain adaptive few-shot misinformation detection via meta learning. *arXiv preprint arXiv:2305.12692*.

Zeng, H.; Yue, Z.; Shang, L.; Zhang, Y.; and Wang, D. 2024. Unsupervised Domain Adaptation Via Contrastive Adversarial Domain Mixup: A Case Study on COVID-19. *IEEE Transactions on Emerging Topics in Computing*.

Zhang, T.; Wang, D.; Chen, H.; Zeng, Z.; Guo, W.; Miao, C.; and Cui, L. 2020. BDANN: BERT-based domain adaptation neural network for multi-modal fake news detection. In *2020 international joint conference on neural networks*, 1–8.

Zhou, X.; and Zafarani, R. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5): 1–40.

Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, please see the Discussion section.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, please see the Abstract and Introduction sections.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, please see the Solution section.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, please see the Discussion section.
   (e) Did you describe the limitations of your work? Yes, please see the Discussion section.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, please see the Ethical Statement section.
   (g) Did you discuss any potential misuse of your work? Yes, please see the Ethical Statement section.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, please see the Ethical Statement section.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we have carefully reviewed the guidelines and make sure our work conforms to them.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, please see the Implementation Details subsection.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, please see the Implementation Details subsection.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, please see the Implementation Details subsection.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, please see the Evaluation section.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes, please see the Discussion section.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, please see the Evaluation section.
   (b) Did you mention the license of the assets? Yes, please see the Evaluation section.
   (c) Did you include any new assets in the supplemental material or as a URL? NA
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, please see the Discussion section.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, please see the Evaluation section.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? Yes, please see the Evaluation section.
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes, please see the Ethical Statement section.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Yes, please see the Evaluation section.

(d) Did you discuss how data is stored, shared, and deidentified? Yes, please see the Ethical Statement section.