


www.jpis.az

Disinformation detection in the medical domain: current approaches, limitations, and future directions

Vagif Mammadaliyev¹, Vusal Shahbazov²

^{1,2}Institute of Information Technology, B. Vahabzadeh str., 9A, AZ1141, Baku, Azerbaijan

¹vagifmammadaliyev@gmail.com, ²vusa.013@gmail.com

ARTICLE INFO

<https://doi.org/10.25045/jpis.v17.i1.09>

Article history:

Received 03 September 2025

Received in revised form

07 November 2025

Accepted 16 January 2026

Keywords:

Medical disinformation

Health misinformation

Automatic misinformation detection

Natural language processing

Machine learning

Knowledge graphs

Misinformation datasets

ABSTRACT

Medical disinformation poses a serious threat to medical demographic security by distorting health behavior at scale, often amplified by confusion among individuals seeking reliable medical information across diverse topics. These distortions can increase vaccine hesitancy, encourage unproven or harmful practices such as ingesting bleach as a purported COVID-19 treatment, and delay evidence-based care. Medical disinformation also erodes trust in health institutions and contributes to cumulative harms, including increased morbidity and mortality, widening health disparities, and, in some cases, real-world violence linked to conspiracy narratives. Despite rapid advances in automated detection methods, the evidence base remains fragmented, obscuring dominant approaches, required resources, and critical research gaps. This paper presents a systematic review of medical disinformation detection research. Major modeling paradigms and reported evaluation evidence are synthesized, encompassing traditional machine learning, deep learning and transformer-based models, knowledge graph approaches, and fact-checking pipelines, together with the datasets and medical knowledge resources that support them. Commonly used feature types are categorized, their strengths and limitations are assessed, persistent weaknesses in resources and detection pipelines are identified, and targeted recommendations are offered to improve future systems and support more reliable medical informatics that strengthens medical demographic security.

1. Introduction

The proliferation of misinformation and disinformation in the medical and health domains represents a pressing global challenge, exacerbated by the widespread accessibility of online platforms and the scale at which health-related content can be monitored, amplified, and evaluated (Schlicht et al. 2023; Abdullayeva 2025). While often used interchangeably, misinformation refers to false or inaccurate information spread without intent to deceive, whereas disinformation is false information deliberately intended to mislead or cause harm (Hameleers 2022; Siani et al. 2024). In parallel with the expansion of digital health services and app-based care, policy and implementation efforts have also highlighted risks related to false or misleading claims about digital

health tools, underscoring the broader relevance of health misinformation in digital platforms (Graefen & Fazal 2025). For the purposes of this research, both terms are considered together and used interchangeably due to their shared capacity to undermine medical-demographic security.

Population health security (PHS) refers to a state in which citizens, society, and the environment are protected from threats that can harm physical and mental health (WHO). PHS emphasizes safeguarding people from diseases, injuries, and adverse environmental exposures through a set of measures aimed at disease prevention, the creation of safe living conditions, and improved access to medical care.

Medical-demographic security (MDS) is a broader concept. According (Mammadova et al. 2025), MDS is defined as a state in which the population's public health is protected from

threats arising from internal and external factors that affect the quality, structure, and size of the population. In the context of Healthcare 4.0, MDS is achieved through the development of a monitoring and control system for morbidity, mortality, population aging, reproductive health, and other medical-demographic indicators (MDIs). Such a system enables systematic tracking of MDI trends and supports public health decision-making aimed at mitigating threats that disrupt natural reproduction and the living conditions of the population.

The COVID-19 pandemic exemplified the crisis through the "infodemic"—a term coined by the World Health Organization to describe the overwhelming flood of both accurate and misleading content—which precipitated real-world harms, including hospitalizations from ingesting bleach-based substances purported to combat the virus.

The spread of such misinformation complicates the identification of reliable information (Imamverdiyev and Sukhostat 2023), fosters vaccine hesitancy (Xiang and Lehmann 2021), promotes adoption of unproven therapies, and delays evidence-based care, culminating in elevated morbidity and mortality (Sell et al. 2021; Smith et al. 2023). It disproportionately burdens various demographic groups, exacerbating health disparities (Kısa and Kısa 2024; Senteio et al. 2025), suboptimal decision-making, and psychological strain across populations (Kauk et al. 2024; Senteio et al. 2025), while eroding trust in health institutions and authorities (Falyuna 2022). These dynamics destabilize public health systems and demographic structures, with implications for medical-demographic security through distorted health behaviors, widened inequities, and cumulative population-level harms (Falyuna 2022).

Although much research targets disinformation on core medical topics like COVID-19, pernicious campaigns often extend beyond strict health domains, inflicting profound indirect effects on public health and demographics. Conspiracy theories alleging that 5G networks cause or exacerbate COVID-19 via radiation proliferated rapidly on social media (Ahmed et al. 2020; Barve and Saini 2023), inciting destructive acts such as arson attacks and vandalism on essential mobile phone masts (Ahmed et al. 2020) and further eroding confidence in experts, thereby amplifying vaccine hesitancy (Ahmed et al. 2020; Langguth et al.

2022). Similarly, the tobacco industry's decades-long disinformation—misrepresenting risks, touting illusory "safer" products, and fabricating scientific uncertainty—prolonged high smoking rates, exacting enduring demographic costs including millions of premature deaths and entrenched inequities (Tan and Bigman 2020).

These cases illustrate how medical and adjacent disinformation warps risk perceptions, behavioral norms, and institutional trust, compounding threats to medical-demographic security. They highlight the imperative for advanced automated detection frameworks, addressing fragmented evidence bases, resource limitations, and generalization challenges to bolster reliable medical informatics (Schlicht et al. 2023). Fig. 1. depicts examples of widely circulated health misinformation claims.

<p>Vaping and 'light' tobacco products are harmless or much safer than smoking, so they don't really harm health.</p>	<p>COVID-19 vaccines cause infertility / contain microchips for tracking.</p>
<p>5G causes or spreads COVID-19.</p>	<p>Unproven 'natural' cures can treat or cure cancer (e.g., baking soda, miracle mineral solution, high-dose vitamins) and are better than standard therapy.</p>

Fig. 1. Illustrative examples of disinformation

In response, recent years have witnessed a marked increase in scholarly efforts to develop automated detection methods, particularly leveraging natural language processing and machine learning techniques (Barve and Saini 2023; Schlicht et al. 2023). Prior to 2019, resources for health misinformation detection were scarce, with only isolated datasets available; however, the pandemic spurred the creation of over 20 COVID-19-focused datasets and numerous studies applying advanced models such as transformers (Sanaullah et al. 2022; Schlicht et al. 2023). While much of this work centers on COVID-19, emerging research addresses broader medical topics, highlighting the need for generalized approaches beyond pandemic-specific contexts (Sharifpoor et al. 2025).

While a growing body of review literature has examined automated health and medical misinformation detection, existing surveys predominantly focus on cataloguing learning algorithms or benchmarking model performance, often within narrow topical scopes such as COVID-19. Comparatively limited attention has

been devoted to cross-paradigm synthesis at the level of feature representations, or to system-level integration of knowledge-driven verification with general misinformation detection methods. These gaps motivate a consolidated analysis that emphasizes feature-centric comparison, resource limitations, and integrative detection architectures capable of operating under evidential uncertainty, providing context for the detailed review and synthesis presented in the following sections.

2. Related work

Numerous approaches have been developed to address the challenge of automated disinformation and misinformation detection in the medical and health domains. Predominant among these are traditional machine learning methods, including supervised classifiers like support vector machines and random forests, alongside modern deep learning techniques such as convolutional neural networks, bidirectional long short-term memory networks, and transformer-based models (Sotto and Viviani 2022; Hussna et al. 2024). These methods have demonstrated promising qualitative performance in identifying misleading health-related content, particularly on social media platforms and news sources, by leveraging features ranging from linguistic and stylistic cues to domain-specific medical embeddings (Zhao et al. 2020; Sotto and Viviani 2022).

However, prior studies consistently identify key limitations inherent to these approaches. Datasets often suffer from insufficient size, narrow topical scopes—frequently limited to COVID-19—and constraints in linguistic diversity, cultural adaptability, and sample balance, which hinder model generalization across broader health contexts (Ayoub et al. 2021; Hussna et al. 2024). Feature engineering pipelines, while effective for certain textual and multimodal inputs, struggle to comprehensively capture all relevant semantic, contextual, and affective elements due to token limits, embedding constraints, and the inherent design of machine learning architectures (Sotto and Viviani 2022; Schlicht et al. 2023). Furthermore, many systems incorporate human-in-the-loop mechanisms, such as expert annotation for ground truth labeling or hybrid validation in fact-checking pipelines, which enhance reliability but impose scalability challenges and elevate costs, particularly for real-

time deployment (Nabožny et al. 2022; Martinez-Rico et al. 2024).

Emerging efforts address these opacity issues through explainable AI techniques, such as LIME integrated with BiLSTM and SHAP with DistilBERT, which provide transparent rationales for predictions in COVID-19 and health misinformation detection (Ayoub et al. 2021; Hussna et al. 2024). These methods improve trust, debugging, and societal safeguards against harmful mispredictions, while deep learning's "black-box" nature persists as a key challenge requiring transparency (Nabožny et al. 2022; Schlicht et al. 2023). Notably, XAI represents the most user-friendly approach from a legal perspective, facilitating regulatory compliance (e.g., EU AI Act mandates for high-risk systems) and accountability through auditable explanations (Mollas et al. 2023; Ramachandram et al. 2025).

Despite substantial research activity and many promising experimental results, real world deployment of medical misinformation detection systems remains challenging due to limited generalizability, shifting health narratives, and practical constraints in scalability and reliability. This persistent gap between benchmark performance and operational effectiveness underscores the need for a consolidated synthesis of current approaches and evidence to identify what works, where it fails, and how future systems can be strengthened. The aim of this review is to consolidate recent advances in medical disinformation detection by synthesizing prevailing approaches, the datasets and knowledge graphs on which they rely, and the principal feature families used in practice. To address these gaps, a systematic review of 70 recent publications is conducted, complemented by an in depth analysis of the 45 most relevant studies. A structured comparison of feature categories is presented, limitations in current datasets and knowledge resources are critically examined, and targeted recommendations are offered, including directions for extending feature design, to support the development of more robust medical informatics solutions.

3. Materials and methods

3.1. Review methodology

To conduct this systematic review, relevant publications were initially identified by searching scientific databases using the terms "medical

disinformation", "medical misinformation", "disinformation detection" and "fake news detection in medical domain". The search was restricted to papers published from 2019 onwards, given that this year marked a significant increase in health misinformation, largely attributed to the onset and progression of the COVID-19 pandemic (Sell et al. 2021; Xiang and Lehmann 2021; Sanaullah et al. 2022; Barve and Saini 2023; Schlicht et al. 2023), as further illustrated in Fig. 2. Following this initial collection, papers were preliminarily scored based on their relevance to the research objectives, utilizing large language models. The highest-scoring papers then underwent a rigorous manual analysis to select those most pertinent for inclusion in this review article.

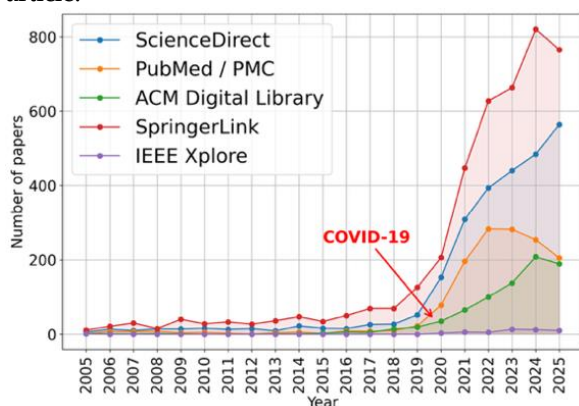


Fig. 2. 'Medical Disinformation' Search Results Across Academic Platforms (2005-2025)

3.2. Problem solution

3.2.1 Methods and frameworks for medical disinformation detection

Beyond early machine-learning approaches, a substantial line of work exploits structured knowledge and graph-based models for health misinformation detection. The DETERRENT model introduces a knowledge-guided graph attention network that links health articles to entities in a medical knowledge graph and propagates evidence across article–entity and entity–entity edges to classify cancer and diabetes misinformation (Cui et al. 2020). While DETERRENT reports substantial gains over text-only baselines on two curated datasets, its performance is tightly coupled to the coverage and correctness of the underlying knowledge graph and ignores temporal dynamics and user interaction signals such as comments or expert responses.

In a similar spirit, the TMD-GLP method for automatic detection of COVID-19 vaccine

misinformation casts the problem as graph link prediction by organizing the COVAXLIES dataset of tweets into a Misinformation Knowledge Graph and leveraging knowledge embedding models to score links between posts and known misinformation targets (Weinzierl and Harabagiu 2021). This neural architecture outperforms classification baselines on vaccine-related Twitter data but remains limited to a narrow domain, short social-media posts, and ignores conversation threads.

More recent work on building a framework for fake news detection in the health domain introduces the KEANE dataset and a multi-stage pipeline that first identifies check-worthy sentences, then performs sentence-level fact-checking before aggregating to article-level veracity labels (Martinez-Rico et al. 2024). Although this framework combines Transformers, classical classifiers, MetaMap-based concept extraction and knowledge-graph resources, the authors highlight important limitations: incomplete knowledge bases, reduced interpretability of Transformer models, heuristic article-level aggregation that can underperform end-to-end models, and considerable computational cost for large-scale deployment.

A second cluster of studies focuses on automated fact-checking pipelines and explainable verdicts for health claims. The Content Similarity Measure approach treats online health URLs as candidates to be verified against a set of trusted sources; it computes lexical, semantic, sentiment and domain-specific similarity features between target and reference content, and then classifies articles using threshold rules and machine-learning models (Barve and Saini 2023). This fact-checking-based framework achieves strong performance across multiple health datasets but is constrained by the coverage of its reference sources and operates primarily at document level, with limited ability to reason about fine-grained claims or multimodal cues. Explainable Automated Fact-Checking for Public Health Claims goes further by introducing the PUBHEALTH dataset, where each claim is labeled with a veracity decision and a free-text justification. Transformer-based models can reach competitive claim-level accuracy but struggle to generate faithful explanations and the dataset remains confined to English claims and to existing fact-checking outlets (Kotonya and Toni 2020). HealthFC frames medical fact-checking as a natural-language-inference task, linking consumer

health claims to evidence sentences from systematic reviews and classifying them as supported, refuted or not-enough-information (Vladika et al. 2023). Evidence-aware models improve over simple text classifiers, but the authors emphasize that high-quality expert annotations are costly and that coverage is limited to topics with existing systematic reviews, leaving many emerging or low-resource conditions unaddressed. Check-COVID similarly constructs a benchmark where news claims are verified against COVID-19 abstracts, using a RoBERTa-based pipeline for abstract retrieval, rationale selection and veracity prediction (Wang et al. 2023). While macro-F1 scores above 80 on oracle settings demonstrate that current models can fact-check many composed and extracted claims, performance drops sharply when realistic retrieval is used and when temporal or numerical reasoning is required, highlighting the difficulty of grounding health claims in scientific literature at scale.

Many works still cast health misinformation detection as supervised text classification over social-media or news content, often introducing new datasets. Dissecting the Infodemic provides an in-depth analysis of COVID-19 misinformation on Twitter/X, benchmarking traditional machine-learning models and deep neural networks on manually labeled tweets (Hussna et al. 2024). The study shows that transformer-based models substantially outperform linear baselines and that domain-adapted language models are particularly effective, but also notes overfitting risks, limited cross-topic generalization and a strong focus on English COVID-19 content. Luo et al. propose deep learning models for ternary classification in COVID-19 infodemic detection, explicitly modeling “uncertain” content in addition to true and false claims (Luo et al. 2024). Their experiments on an English COVID-19 dataset indicate that several deep architectures outperform classical classifiers, and that including an uncertainty class improves robustness to borderline or incomplete statements; nonetheless, performance varies widely across models, training data remain modest in size, and results are again restricted to a single language and crisis. Detecting Misleading Information on COVID-19 applies an ensemble of classical methods and neural architectures to over three million annotated tweets, using a three-level voting strategy to label posts as accurate, inaccurate or unverifiable. The approach achieves strong

performance on the authors’ large-scale corpus, but requires continuous retraining as new narratives emerge and inherits labeling noise from automatically propagated annotations (Elhadad et al. 2020).

Automatic detection of COVID-19 vaccine misinformation with graph link prediction also targets vaccine narratives but frames the task as inferring links between user posts and curated scientific claims; while this graph formulation improves performance over purely textual baselines (Weinzierl and Harabagiu 2021), it still depends on a manually curated set of “ground truth” statements and does not directly address out-of-graph or evolving rumors.

Complementing these task-specific models, several papers offer broader syntheses of the field. A recent systematic review on automatic detection of health misinformation surveys approaches across diseases, platforms and modalities, concluding that most existing systems rely on supervised learning over user-generated text, with relatively little attention to cross-domain generalization, multilingual settings or integration of knowledge graphs, user networks and temporal features (Schlicht et al. 2023).

Another overview of health misinformation detection in the social web adopts a data-science perspective, cataloguing features, learning algorithms and evaluation practices, and highlighting that many pipelines still use generic text classifiers with limited domain adaptation, explainability, or user-centric evaluation (Sotto and Viviani 2022). A systematic review of machine-learning applications for COVID-19 misinformation further shows that deep learning and transformer models dominate recent work, but that most studies remain confined to a handful of English datasets, focus on binary classification, and rarely study longitudinal impact or real-world deployment (Sanauallah et al. 2022).

Finally, research on reliable misinformation mitigation with GPT-4 and smaller language models demonstrates that large language models can achieve or surpass state-of-the-art accuracy on several benchmark veracity-classification datasets in zero-shot or lightly prompted settings, while also being able to express epistemic uncertainty. At the same time, the authors show that performance degrades on harder, context-dependent cases and that careful calibration and evaluation are needed to avoid overconfident but wrong judgments, especially in high-stakes domains like health (Pelrine et al. 2023).

A distinct strand of work introduces human-in-the-loop and expert-support frameworks for medical misinformation. Nabożny et al. propose filtering classifiers that prioritize non-credible medical statements for expert review, substantially increasing efficiency in identifying misinformation across health topics while depending on training data and potentially biasing surfaced content (Nabożny et al. 2022).

Mendes et al. present an end-to-end pipeline for COVID-19 treatment claims on Twitter that extracts and ranks trending claims by check-worthiness before classifying stance toward false treatments; human evaluation confirms it detects about half of new misleading claims before news debunking, with two-thirds of flagged tweets as policy violations and ~124 violations confirmed per annotator-hour, though it requires oversight, is platform-specific, and may overlook rare rumors (Mendes et al. 2023).

Other studies examine domain-specific medical narratives with lightweight features. Fridman et

al. analyze social media posts on unproven cancer therapies, training classifiers on linguistic, stylistic, and readability cues to achieve up to 73% accuracy in flagging misleading content; while interpretable, results are modest, cancer-focused, and rely on manual "unproven" labeling (Fridman et al. 2025). Table 1 summarizes one representative high-performing approach for each major method family considered in this review: traditional machine learning, deep learning, knowledge-graph-based models, large language models (LLMs), and combined or hybrid systems. For each family, we report a widely cited exemplar method together with the task setting and its headline accuracy or F1-score as reported by the original authors. These figures are not directly comparable across datasets or label schemes, but they illustrate the relative maturity of different modeling paradigms and highlight that high performance can be achieved by both classical and more recent architectures when evaluated in suitably constrained scenarios.

Table 1. Representative high-performing approaches to medical misinformation detection, grouped by method family

Method group	Representative paper / model	Task (roughly)	Headline metric	Limitations
Traditional ML	SVM on COVID-19 Fake News Dataset (Patwa et al. 2021)	Binary COVID fake-news detection	Acc / F1 \approx 93.32%	Surface-level textual features, statistical correlations, absence of factual verification, specifically prepared datasets, lack of explainability
Deep Learning	BERT in ternary COVID-19 infodemic detection (Luo et al. 2024)	3-way COVID infodemic classification	Acc \approx 81.01%	Limited or imbalanced datasets, overfitting, absence of factual verification, lack of explainability, limited contextual understanding
Knowledge Graph	DETERRENT (KG-guided GAT) (Cui et al. 2020)	Binary health misinformation (diab/can)	Acc 92.06% / 96.52%	Dependency on expertise, knowledge update difficulty, lack of complementary information (i.e. apart from knowledge base)
LLM	GPT-4 Binary on LIAR-New (Pelrine et al. 2023)	General claim veracity (LIAR-New)	Acc 81.2% (all) / 91.0% (Possible)	Computationally intensive and resource-demanding, limited token encoding, hallucinations, overconfidence without knowledge, evaluation difficulty
Combined / Hybrid	CSM algorithm (multi-dataset) (Martinez-Rico et al. 2024)	Health URL fact-checking, multi-corpus	Acc up to 91.06% (internal), 87–89% on CoAID/ReCOVery/FakeHealth	Error propagation in multi-stage systems, evaluation difficulty, dependence on reliable external sources, computationally intensive

3.3 Datasets

Several authors contribute specialized datasets related to disinformation in medical domain. A consolidated overview of the most widely used datasets in the domain of medical and health-related misinformation is provided. These resources represent a diverse collection of general

medical, COVID-19-specific, vaccination-focused, and other health-related disinformation datasets, each designed to support various computational tasks such as classification, stance detection, credibility estimation, and narrative analysis. The datasets differ substantially in size, linguistic

coverage, annotation methodology, and data modalities—ranging from social media posts and news articles to fact-checked claims and knowledge-enhanced representations. By compiling these datasets in a unified format, Table 2 highlights the breadth of empirical resources available to the research community and underscores the evolution of misinformation

datasets toward larger, more heterogeneous, and more rigorously validated benchmarks. This overview also serves as a foundational reference for comparing methodological approaches reviewed in this paper and for identifying gaps related to multilingual coverage, veracity annotation quality, and domain specificity in medical misinformation research.

Table 2. Overview of publicly available datasets for medical and health-related misinformation research

Dataset	Primary Topic	Size / Entries	Format / Data Type	Source / Link
CoAID (Cui and Lee 2020)	COVID-19	~300,000	Articles, posts, fact-checks	https://github.com/cuilimeng/CoAID
FakeHealth (Dai et al. 2020)	General health	~4,000	Articles with metadata	https://github.com/yaqingwang/FakeHealth
ANTI-Vax (Hayawi et al. 2021)	Vaccination	700,000+	Tweets	https://github.com/sarahhayawi/ANTI-Vax
ReCOVery (Zhou et al. 2020)	COVID-19	~2,000 news articles	Articles with credibility labels	https://github.com/baohq1999/ReCOVery
ArCOV19-Rumors (Haouari et al. 2020)	COVID-19	10,000+	Arabic tweets	https://huggingface.co/datasets/asas-ai/ArCOV19-Rumors
CHECKED (Yang et al. 2021)	General health	6,000+	Claims + fact-checks	https://github.com/cuilimeng/CHECKED
Medical Credibility Corpus (Schlicht et al. 2023)	General medical	1,000+	Statements, credibility labels	https://github.com/alenaobozny/medical_credibility_corpus

3.4 Knowledge Resources

In the domain of medical misinformation detection, knowledge resources—particularly structured medical knowledge graphs and bases—play a pivotal role by encoding entities such as diseases, symptoms, treatments, drugs, and their semantic relations as derived from authoritative biomedical sources. These resources differ fundamentally from the plain-text datasets

outlined earlier, which primarily serve as labeled corpora for supervised classification tasks such as stance detection or veracity labeling. Instead, knowledge graphs provide a stable relational backbone that models medically plausible concepts and associations. The common structure of knowledge graphs is illustrated schematically in fig. 3, which shows simple medical examples of entities connected by relations.

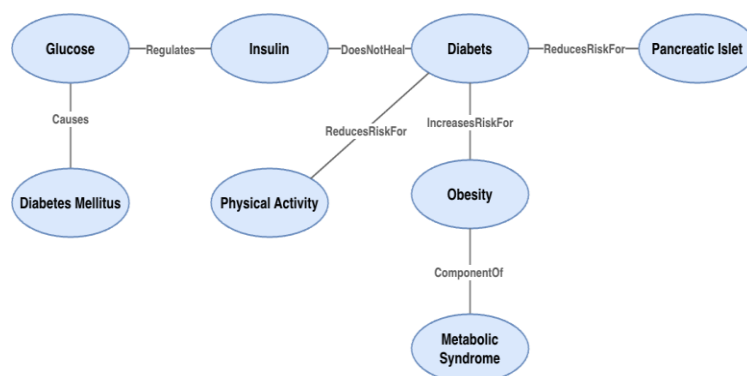


Fig. 3. Example of medical knowledge graph triples representing entity–relation–entity facts in the medical domain.

When integrated into misinformation detection systems, they support entity normalization across heterogeneous terminologies, enrich unstructured text through links to canonical concepts, and enable reasoning over interconnected facts that can expose implausible or inconsistent medical claims. Table 3 summarizes the main medical knowledge graphs. Complementing these structured knowledge resources are fact-checking and verification datasets in the health domain. These datasets typically consist of individual, well-defined claims paired with evidence passages from credible sources such as

scientific articles, clinical studies, or expert fact-checking reports. Each claim is annotated with a veracity label (for example, supported, refuted, or not-enough-information), and in some cases accompanied by short textual justifications. The claim–verdict–evidence structure is illustrated in Fig. 4. In the first case, the claim “Melatonin helps against jet lag” is accompanied by supporting evidence derived from scientific studies, leading to a Supported verdict, which indicates that the available evidence substantiates the claim

Table 3. Medical knowledge graphs / knowledge bases for misinformation detection

Knowledge Graphs/Bases	Description	Size/Coverage	Key Features
UMLS Metathesaurus (Nguyen et al. 2021)	Multi-lingual biomedical vocabulary integrating 100+ controlled vocabularies	1M+ concepts, 5M+ names	Hierarchical & associative relationships, semantic types
KnowLife (Cui et al. 2020)	Biomedical KB from scientific literature and health portals	500K+ facts, 25K entities, 591K triples	13 relations with positive/negative types, 93% precision
Wikidata (Waagmeester et al. 2020)	Collaborative open-source multilingual knowledge graph	1.65B triples, 116M items	Community-edited, CC0 license, largest open KG
YAGO (Suchanek et al. 2023)	High-quality KB from Wikipedia, WordNet, Wikidata, GeoNames	132M facts (v4.5)	95% accuracy, temporal/spatial dimensions, human-readable

In contrast, the claim “Brain training boosts intelligence” is paired with evidence that questions its validity, resulting in a Refuted verdict, where the evidence contradicts the asserted statement.

Finally, the claim “Caffeine reduces attention” is associated with no retrieved evidence, producing a Not Enough Information verdict, which reflects insufficient or missing data to draw a reliable conclusion.

These examples highlight the fundamental differences between verification outcomes: claims may be confirmed, contradicted, or remain unresolved depending on the presence, relevance, and strength of supporting evidence. Such distinctions are essential for modeling realistic fact-checking scenarios, particularly in domains like health, where incomplete or conflicting information is common. Such datasets enable models to perform evidence retrieval, assess the relationship between a claim and its supporting or contradicting texts, and generate explanations that mirror expert verification processes. Table 4 summarizes the fact-checking and verification datasets.

Claim: Melatonin helps against jet lag
Evidence: ... Overall, the studies show that melatonin may help ...
Verdict: Supported
Claim: Brain training boosts intelligence
Evidence: ... A review from 2013 casts doubt ...
Verdict: Refuted
Claim: Caffeine reduces attention
Evidence: NULL
Verdict: Not Enough Information

Fig. 4. Schematic representation of fact-checking and verification datasets as claim–verdict–evidence triples.

3.5 Results

Feature engineering represents a fundamental stage in medical disinformation detection, as feature representations directly influence model robustness and generalizability. Content-based features, such as lexical, syntactic, and semantic cues, are commonly used with traditional

machine-learning and deep-learning models, where effectiveness is often limited by shallow semantics and domain dependence. Source-based, propagation-based, and behavior-based features are more frequently associated with statistical and graph-based approaches, enabling the exploitation of credibility signals and interaction patterns but remaining constrained by data availability and platform access. Knowledge-based features underpin knowledge-graph and fact-checking pipelines, where performance is strongly conditioned by knowledge coverage, entity normalization, and the integration of external medical resources.

Despite the prominence of SVM-based, deep-learning, and transformer-based paradigms in medical disinformation detection, system performance is strongly conditioned by the feature representations provided to these models. Feature engineering was therefore analyzed as a distinct dimension of the literature and grouped into five categories: content-based, source-based, propagation-based, behavior-based, and knowledge-based features. Content-based features align most directly with SVMs and neural text models (including transformers), but are often limited by shallow semantics, multilingual variation, and rapidly evolving medical terminology (Sotto and Viviani 2022; Schlicht et al. 2023). Source-based features are typically integrated into statistical or hybrid pipelines to proxy credibility, yet their utility is constrained by incomplete metadata and limited automation

(Enyan et al. 2020; Martinez-Rico et al. 2024). Propagation-based and behavior-based features are most naturally exploited by graph-based and temporal models, but are frequently underused due to platform restrictions and data scarcity that impede cascade- and coordination-level analysis (Zhao et al. 2020; Hussna et al. 2024). Knowledge-based features underpin knowledge-graph and fact-checking pipelines through entity linking and relational verification, but their effectiveness depends on knowledge coverage and is particularly challenged during emerging crises (Cui et al. 2020; Chen et al. 2022). Overall, the reviewed evidence indicates that “model dominance” cannot be interpreted independently of representation choices, which determine robustness and generalizability across settings.

To present the results of the analysis on feature representations in medical and health-related misinformation detection, existing studies were reviewed and the features were categorized into five principal types: content based, source based, propagation based, behavior based, and knowledge based. **Table 5** delineates the key strengths and limitations of each type, such as content based features' proficiency in detecting linguistic anomalies and persuasive rhetoric alongside their vulnerability to shallow semantics, or knowledge based features' robust entity linking tempered by sparse coverage in crises—while assessing the severity of these limitations for system robustness and generalizability.

Table 4. Fact-checking and verification datasets

Fact-Checking Databases / Verification Datasets	Description	Size/Coverage	Key Features
HealthFC (Vladika et al. 2023)	Bilingual (German/English) health claims dataset with medical expert verdicts backed by evidence from clinical trials and systematic reviews.	750 claims (bilingual)	Expert veracity labels, rationale sentences, level of evidence
SciFact (Wadden et al. 2022)	Scientific claim verification dataset with expert-written claims paired with evidence-containing abstracts from research literature.	1.4K claims with evidence abstracts	SUPPORTS, REFUTES, NOT_ENOUGH_INFO + rationales
FEVER (Nie et al. 2020)	Large-scale fact extraction and verification dataset from Wikipedia. Widely used benchmark for transfer learning in fact-checking research.	Large-scale (Wikipedia-based)	SUPPORTS, REFUTES, NOT_ENOUGH_INFO
healthfeedback.org (Dammu et al. 2024)	Expert-curated health fact-checking database from Health Feedback organization.	784 entries	Expert-reviewed
FakeCOVID (Shahi and Nandini 2020)	Large multilingual COVID-19 dataset with granular labels from multiple fact-checking organizations.	7,621 data points	Multi-label (Correct, Fake, False, Misleading, Mixed)

4. Discussion

Looking back at the existing methods summarized in table 1, traditional machine learning algorithms achieve strong performance on specific, narrowly defined datasets and domains. However, these models rarely sustain comparable results in real-world settings. Their feature representations typically encode little explicit medical knowledge, or at best capture highly specialized and narrow aspects of the domain. As soon as the topic, platform, or population shifts, these models struggle to generalize beyond the conditions under which they were originally trained.

Deep learning and large language model (LLM)-based approaches tend to exhibit better generalization, particularly when applied to previously unseen contexts. By learning richer semantic and contextual representations, they are more capable of capturing patterns associated with medical misinformation that were not explicitly hand-crafted as features. Nevertheless, even these advanced models benefit substantially from access to external knowledge resources. In principle, integrating medical knowledge bases can temper purely statistical judgments with

medically grounded constraints. In practice, building and maintaining such knowledge bases requires sustained effort from domain experts, and several of the reviewed studies highlight the technical difficulty of effectively incorporating these resources into end-to-end detection pipelines. Across the reviewed literature, many authors emphasize that achieving higher reliability and accuracy typically requires humans to remain in the loop, leaving most systems effectively semi-automatic. Expert annotators are needed to construct ground-truth labels, to curate and update knowledge bases, and in some cases to validate or override model predictions in operational settings. This reliance on human oversight is particularly evident in fact-checking frameworks, which generally produce one of three logical outcomes for a given claim: supported, refuted, or not enough information. While the first two categories yield actionable decisions, the third exposes a persistent zone of ambiguity. When available evidence is insufficient, the system cannot safely classify the claim as either true or false, yet leaving it entirely unresolved is often unsatisfactory from a practical perspective.

Table 5. Feature types used in medical misinformation detection: strengths, limitations, and severity of their impact

Feature Type	Strengths	Limitations	Severity
Content-based	Analyzes grammatical, syntactic, and semantic patterns; detects persuasive rhetoric, sentiment/emotional cues, readability issues, and linguistic anomalies typical for misleading medical content. Enables fine-grained linguistic profiling.	Often shallow without domain semantics; struggles with implicit claims, multilingual variation, and emerging terminology. Over-reliance on text-only signals limits multimodal and temporal reasoning (Sotto and Viviani 2022; Schlicht et al. 2023)	Medium
Source-based	Proxies expertise via metadata, enhancing user reliability detection (Enyan et al. 2020; Sotto and Viviani 2022)	Overlook full automation challenges (Martinez-Rico et al. 2024)	Medium-High
Propagation-based	Reveals viral anomalies in tweet cascades (Hussna et al. 2024).	Lack due to data scarcity or platform limits, ignoring temporal spikes or bot coordination (Zhao et al. 2020; Hussna et al. 2024)	Medium
Behavior-based	Behavioral cues like interaction patterns can outperform linguistics in some online communities (Zhao et al. 2020)	Lack due to data scarcity or platform limits, ignoring temporal spikes or bot coordination (Zhao et al. 2020; Hussna et al. 2024)	Medium
Knowledge-based	Knowledge graphs provide verifiable relations, linking entities for healthcare claims (Cui et al. 2020)	Integration is rare, constrained by graph coverage for emerging crises (Cui et al. 2020; Chen et al. 2022)	High

In such “not enough information” cases, additional processing becomes necessary. One promising direction, reflected in several of the reviewed approaches, is to complement fact-checking outputs with models that learn general patterns of online disinformation. These include

manipulative or polarizing tone, heightened emotional language, and semantic inconsistencies or oversimplifications that recur in misleading health content. By combining fact-checking outcomes with these content-oriented signals, systems can at least provide a more nuanced

assessment of risk, even when definitive evidence is unavailable, rather than collapsing all uncertainty into a single undifferentiated category.

Motivated by these observations, a hybrid approach to medical misinformation detection is proposed that jointly exploits knowledge bases and multiple feature families, including content-, source-, and propagation-based signals. The goal is to combine the strengths of knowledge-driven verification with the broader coverage of general misinformation detection methods.

Fig. 5 and 6 illustrate two high-level architectures that embody this idea. Fig. 5 presents a sequential pipeline in which an initial fact-checking module operates as the primary decision mechanism. When the fact-checking component can confidently return a supported or refuted verdict, its output is delivered directly to the user. When it returns not enough information, the pipeline triggers an additional processing stage that applies general misinformation detection models based on content, source, and propagation features. In this way, the system prioritizes explicit knowledge when available, while still leveraging learned patterns of disinformation to provide a more informative assessment under evidential uncertainty.

Fig. 6, by contrast, depicts a parallel architecture. Here, fact-checking and general misinformation detection are executed concurrently, without immediately privileging the output of either component. The final decision is produced by an ensemble step that combines the fact-checking verdict with signals derived from content-, source-, and propagation-based models. This design avoids relying solely on the fact-checking module at any stage, which is advantageous when knowledge bases are incomplete or when claims are only partially covered by existing resources.

Both architectures remain fully consistent with the insights from the review: knowledge verification should be central whenever reliable evidence is available, but robust medical misinformation detection also depends on learning broader patterns of misleading communication and exploiting a diverse set of feature types. These architectures could also prove advantageous in non-medical domains and for general disinformation detection.

5. Conclusion

Medical disinformation poses sustained risks to public health and, at scale, undermines medical-demographic security by distorting health behaviors, eroding institutional trust, and compounding population-level harms.

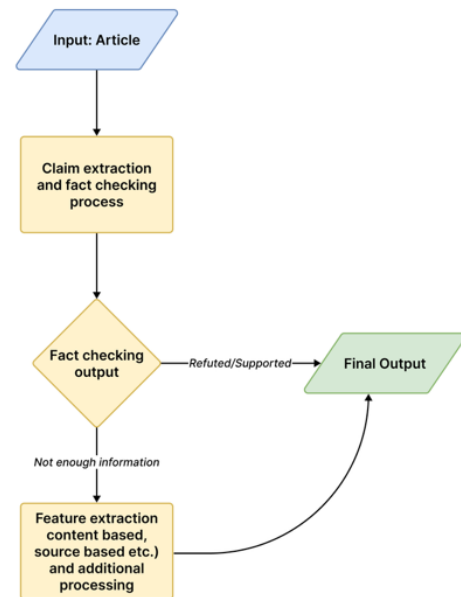


Fig. 5. Sequential hybrid architecture for medical misinformation detection

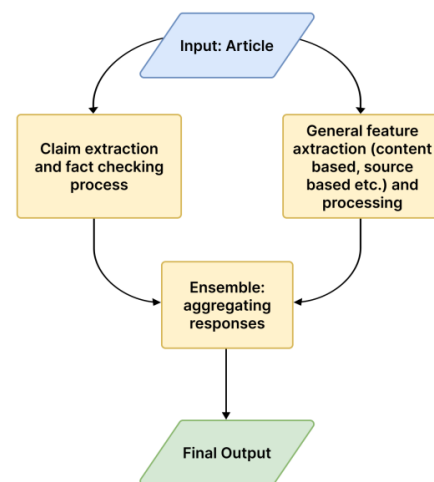


Fig. 6. Parallel hybrid ensemble architecture for medical misinformation detection

This review synthesized recent work on automated medical disinformation detection, consolidating evidence across modeling approaches, datasets, feature design, and structured knowledge resources.

Across surveyed studies, traditional machine learning, deep learning, knowledge-graph-enhanced models, and LLM-based approaches

often achieve strong results on constrained benchmarks; however, performance and reliability remain closely tied to narrow topics, platforms, and languages. The empirical evidence base is further constrained by datasets that are predominantly COVID-19-centric and platform-specific, limiting cross-domain generalization and leaving adjacent health-relevant disinformation only partially covered. Knowledge graphs and fact-checking resources enable more principled veracity assessment, but their coverage remains incomplete, they are costly to maintain, and they can be difficult to integrate into end-to-end pipelines. At the feature level, existing systems continue to rely heavily on content-based and source-based signals, while propagation-, behavior-, and knowledge-based signals are comparatively underutilized despite their potential to improve robustness and realism.

Taken together, these findings indicate that progress in medical disinformation detection, from a population-level perspective, will depend less on any single modeling paradigm and more on integrated system design: knowledge-aware verification when reliable evidence is available, complementary feature families when evidence is sparse, and careful handling of “not enough information” cases. In practical deployments, this integration should be paired with explainable AI and human oversight to support accountability and safe decision-making in high-stakes settings. By clarifying the resource landscape and its limitations and by outlining concrete directions for richer feature integration, this review aims to support the development of more reliable medical informatics systems that better protect medical-demographic security.

Acknowledgements

Gratitude is expressed to the authors of this idea, corresponding member Corresponding Member of Azerbaijan National Academy of Sciences Masuma Mammadova and Associate professor Lyudmila Sukhostat.

References

- Abdullayeva S, Online media monitoring and evaluation: comparative approaches (2025) *Problems of Information Society*, 16(2), 107-115. doi: 10.25045/jpis.v16.i2.12.
- Ahmed W, Vidal-Alaball J, Downing J, Seguí FL (2020) COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *Journal of Medical Internet Research* 22(5), 1201–1209. <https://doi.org/10.2196/19458>
- Akhavein D, Sheel M, Abimbólá Şeýe (2025) Health security— Why is ‘public health’ not enough? *Global Health Research and Policy* 10(1). <https://doi.org/10.1186/s41256-024-00394-7>
- Al-Mugti HS, Aldeghalbey AA, Swaif KA, Alrashdi HH, Mahdi EM, Alharbi MB, Alsaidi AS, Algathradi NY, Alanazi SM, Alsameh NS, Kariri A, Alasmari EA, Alqarni KA, Asiri EJ, Alhasan JH (2023) Saudi Health System and Health Security Structure: A Scope Review Study Addressing the National Need for Governing the Health Security. *Cureus* 15(10). <https://doi.org/10.7759/cureus.47376>
- Ayoub J, Yang XJ, Zhou F (2021) Combat COVID-19 infodemic using explainable natural language processing models. *Information Processing & Management* 58(4). <https://doi.org/10.1016/j.ipm.2021.102569>
- Barve Y, Saini JR (2023) Detecting and classifying online health misinformation with ‘Content Similarity Measure (CSM)’ algorithm: an automated fact-checking-based approach. *The Journal of Supercomputing* 79(8), 9127–9156. <https://doi.org/10.1007/s11227-022-05032-y>
- Chen C, Wang H, Shapiro MA, Xiao Y, Wang F, Shu K (2022) Combating Health Misinformation in Social Media: Characterization, Detection, Intervention, and Open Issues. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2211.05289>
- Cui L, Lee D (2020) CoAID: COVID-19 Healthcare Misinformation Dataset. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2006.00885>
- Cui L, Seo H, Tabar M, Ma F, Wang S, Lee D (2020) DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation, 492–502. <https://doi.org/10.1145/3394486.3403092>
- Dai E, Sun Y, Wang S (2020) Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. *Proceedings of the International AAAI Conference on Web and Social Media* 14(1), 853–862. <https://doi.org/10.1609/icwsm.v14i1.7350>
- Dammu PPS, Naidu H, Dewan M, Kim Y, Roosta T, Chadha A, Shah C (2024) ClaimVer: Explainable Claim-Level Verification and Evidence Attribution of Text Through Knowledge Graphs, 13613–13627. <https://doi.org/10.18653/v1/2024.findings-emnlp.795>
- Elhadad MK, Li KF, Gebali F (2020) Detecting Misleading Information on COVID-19. *IEEE Access* 8, 165201–165215. <https://doi.org/10.1109/access.2020.3022867>
- Enyan D, Yiwei S, Wang S (2020) FakeHealth [Data set]. Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.48550/arXiv.2002.00837>
- Falyuna N (2022) Science disinformation as a security threat and the role of science communication in the disinformation society. *Scientia et Securitas* 3(1), 69–78. <https://doi.org/10.1556/112.2022.00086>
- Fridman I, Boyles D, Chheda R, Baldwin-SoRelle C, Smith AB, Lafata JE (2025) Identifying Misinformation About Unproven Cancer Treatments on Social Media Using User-Friendly Linguistic Characteristics: Content Analysis. *JMIR Infodemiology* 5. <https://doi.org/10.2196/62703>
- Graefen B. and Fazal N., (2025) “From global best practices to national implementation: digital health strategies for Azerbaijan, *Problems of Information Society*, 16(2), 39-46, DOI: 10.25045/jpis.v16.i2.05.
- Hameleers M (2022) Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory* 33(1), 1–10. <https://doi.org/10.1093/ct/qtac021>
- Hauouri F, Hasanain M, Suwaileh R, Elsayed T (2020) ArCOV-19: The First Arabic COVID-19 Twitter Dataset with

- Propagation Networks. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2004.05861>
- Hayawi K, Shahriar S, Serhani MA, Taleb I, Mathew SS (2021) ANTI-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public Health* 203, 23–30. <https://doi.org/10.1016/j.puhe.2021.11.022>
- Hussna AU, Alam MdGR, Islam R, Alkhamees BF, Hassan MM, Uddin MdZ (2024) Dissecting the infodemic: An in-depth analysis of COVID-19 misinformation detection on X (formerly Twitter) utilizing machine learning and deep learning techniques. *Heliyon* 10(18). <https://doi.org/10.1016/j.heliyon.2024.e37760>
- Imamverdiyev Y, Sukhostat L (2023) COVID-19: Cybersecurity Issues in Times of Pandemic. *Electronic Government, an International Journal* 1(1), 569–590. <https://doi.org/10.1504/eg.2024.10060533>
- Kauk J, Humprecht E, Kreysa H, Schweinberger SR (2024) Large-scale analysis of online social data on the long-term sentiment and content dynamics of online (mis) information. *Computers in Human Behavior* 165. <https://doi.org/10.1016/j.chb.2024.108546>
- Kısa S, Kısa A (2024) A Comprehensive Analysis of COVID-19 Misinformation, Public Health Impacts, and Communication Strategies: Scoping Review. *Journal of Medical Internet Research* 26. <https://doi.org/10.2196/56931>
- Kotonya N, Toni F (2020) Explainable Automated Fact-Checking for Public Health Claims. <https://doi.org/10.18653/v1/2020.emnlp-main.623>
- Langguth J, Filkuková P, Brenner S, Schroeder DT, Pogorelov K (2022) COVID-19 and 5G conspiracy theories: long term observation of a digital wildfire. *International Journal of Data Science and Analytics* 15(3), 329–346. <https://doi.org/10.1007/s41060-022-00322-3>
- Luo J, Baz DE, Shi L (2024) Utilizing deep learning models for ternary classification in COVID-19 infodemic detection. *Digital Health* 10. <https://doi.org/10.1177/20552076241284773>
- Mammadova M, Jabrayilova Z, Mammadaliyev V (2025) Medical-Demographic Identity of Territorial Units in the Healthcare 4.0 Environment (in
- Mammadova M., Jabrayilova Z., Mammadaliyev V. (2025) Medical-Demographic Identity of Territorial Units in the Healthcare 4.0 Environment. *Proc. of the II Republican Scientific-Practical Conference “Digital Medicine 4.0: Challenges, Opportunities and Perspectives”*, Baku, Azerbaijan, pp. 226–231 (in Russian). <https://doi.org/10.25045/SPCDH4.0.2025.47>
- Martinez-Rico JR, Araujo L, Martínez-Romo J (2024) Building a framework for fake news detection in the health domain. *PLOS ONE* 19(7). <https://doi.org/10.1371/journal.pone.0305362>
- Mendes E, Chen Y, Xu W, Ritter A (2023) Human-in-the-loop Evaluation for Early Misinformation Detection: A Case Study of COVID-19 Treatments. <https://doi.org/10.18653/v1/2023.acl-long.881>
- Mollas I, Bassiliades N, Tsoumakas G (2023) Truthful meta-explanations for local interpretability of machine learning models. *Applied Intelligence* 53(22), 26927–26948. <https://doi.org/10.1007/s10489-023-04944-3>
- Nabożny A, Balcerzak B, Morzy M, Wierzbicki A, Savov P, Warpechowski K (2022) Improving medical experts' efficiency of misinformation detection: an exploratory study. *World Wide Web* 26(2), 773–798. <https://doi.org/10.1007/s11280-022-01084-5>
- Nguyen V, Yip HY, Bodenreider O (2021) Biomedical Vocabulary Alignment at Scale in the UMLS Metathesaurus, 2672–2683. <https://doi.org/10.1145/3442381.3450128>
- Nie Y, Bauer L, Bansal M, Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R, Khouja J, Zhou Y, Zhao T, Jiang M, Binou J, Ma H, Santus E, Schulte H, Serra G, Utsuro T, Zhao J (2020) Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER). <https://doi.org/10.18653/v1/2020.fever-1>
- Osareme OJ, Muonde M, Maduka CP, Olorunsogo TO, Omotayo O (2024) Demographic shifts and healthcare: A review of aging populations and systemic challenges. *International Journal of Science and Research Archive*, 11(1), 383–395. <https://doi.org/10.30574/ijrsra.2024.11.1.0067>
- Patwa P, Sharma S, Pykl S, Guptha V, Kumari G, Akhtar MS, Ekbal A, Das A, Chakraborty T (2021) Fighting an Infodemic: COVID-19 Fake News Dataset. *Communications in Computer and Information Science*. Springer. https://doi.org/10.1007/978-3-030-73696-5_3
- Pelrine K, Imouza A, Thibault C, Reksoprodjo M, Gupta C, Christoph J, Godbout J, Rabbany R (2023) Towards Reliable Misinformation Mitigation: Generalization, Uncertainty, and GPT-4. <https://doi.org/10.18653/v1/2023.emnlp-main.395>
- Ramachandram D, Joshi H, Zhu JD, Gandhi D, Hartman L, Raval A (2025) Transparent AI: The Case for Interpretability and Explainability. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2507.23535>
- Sanaullah A, Das A, Kabir MA, Shu K (2022) Applications of machine learning for COVID-19 misinformation: a systematic review. *Social Network Analysis and Mining*, 12. <https://doi.org/10.1007/s13278-022-00921-9>
- Schlicht IB, Fernandez E, Chulvi B, Rosso P (2023) Automatic detection of health misinformation: a systematic review. *Journal of Ambient Intelligence and Humanized Computing* 15, 2009–2021. <https://doi.org/10.1007/s12652-023-04619-4>
- Sell TK, Hosangadi D, Smith E, Trotochaud M, Vasudevan P (2021) National Priorities to Combat Misinformation and Disinformation for COVID-19 and Future Public Health Threats: A Call for a National Strategy. <https://centerforhealthsecurity.org/sites/default/files/2023-02/210322-misinformation.pdf>
- Senteio C, Fields SD, Singh RKP, Kamoga RMN, Andrews E, Gandsman D, Halton C, Rysinova V, Snow S (2025) Overcoming health misinformation in marginalized groups: a systematic review. *International Journal for Equity in Health*, 24(1). <https://doi.org/10.1186/s12939-025-02657-2>
- Shahi GK, Nandini D (2020) FakeCovid- A Multilingual Cross domain Fact Check Dataset for COVID-19. Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.3965871>
- Sharifpoor E, Okhovati M, Ghazizadeh-Ahsae M, Beigi MA (2025) Classifying and fact-checking health-related information about COVID-19 on Twitter/X using machine learning and deep learning models. *BMC Medical Informatics and Decision Making*, 25(1). <https://doi.org/10.1186/s12911-025-02895-y>
- Siani A, Joseph M, Dacin C (2024) Susceptibility to scientific misinformation and perception of news source reliability in secondary school students. *Discover Education* 3(1). <https://doi.org/10.1007/s44217-024-00194-8>
- Smith R, Chen KM, Winner D, Friedhoff S, Wardle C (2023) A Systematic Review Of COVID-19 Misinformation Interventions: Lessons Learned. *Health Affairs*, 42(2). <https://doi.org/10.1377/hlthaff.2023.00717>
- Sotto SD, Viviani M (2022) Health Misinformation Detection in the Social Web: An Overview and a Data Science Approach. *International Journal of Environmental*

- Research and Public Health 19(4), 2173–2193. <https://doi.org/10.3390/ijerph19042173>
- Suchanek FM, Alam M, Bonald T, Paris P-H, Soria JB (2023) YAGO 4.5: A Large and Clean Knowledge Base with a Rich Taxonomy. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2308.11884>
- Tan ASL, Bigman CA (2020) Misinformation About Commercial Tobacco Products on Social Media— Implications and Research Opportunities for Reducing Tobacco-Related Health Disparities 110(3), 281–283. <https://doi.org/10.2105/ajph.2020.305910>
- Vladika J, Schneider P, Matthes F (2023) HealthFC: Verifying Health Claims with Evidence-Based Medical Fact-Checking. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2309.08503>
- Waagmeester A, Stupp GS, Burgstaller-Muehlbacher S, Good BM, Griffith M, Griffith OL, Hanspers K, Hermjakob H, Hudson T, Hybiske K, Keating S, Manske M, Mayers M, Mietchen D, Mitraka E, Pico AR, Putman T, Riutta A, Queralt-Rosinach N, Schriml LM, Shafee T, Slenter D, Stephan R, Thornton K, Tsueng G, Tu R, Ul-Hasan S, Willighagen E, Wu C, Su AI (2020) Wikidata as a knowledge graph for the life sciences. eLife 9. <https://doi.org/10.7554/elife.52614>
- Wadden D, Lo K, Kuehl B, Cohan A, Beltagy I, Wang LL, Hajishirzi H (2022) SciFact-Open: Towards open-domain scientific claim verification. <https://doi.org/10.18653/v1/2022.findings-emnlp.347>
- Wang G, Harwood K, Chillrud L, Ananthram A, Subbiah M, McKeown K (2023) Check-COVID: Fact-Checking COVID-19 News Claims with Scientific Evidence. <https://doi.org/10.18653/v1/2023.findings-acl.888>
- Weinzierl M, Harabagiu SM (2021) Automatic detection of COVID-19 vaccine misinformation with graph link prediction. Journal of Biomedical Informatics 124. <https://doi.org/10.1016/j.jbi.2021.103955>
- Xiang D, Lehmann LS (2021) Confronting the misinformation pandemic. Health Policy and Technology 10(3). <https://doi.org/10.1016/j.hlpt.2021.100520>
- Yang C, Zhou X, Zafarani R (2021) CHECKED: Chinese COVID-19 fake news dataset. Social Network Analysis and Mining 11(1). <https://doi.org/10.1007/s13278-021-00766-8>
- Zhao Y, Da J, Yan J (2020) Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. Information Processing & Management 58(1). <https://doi.org/10.1016/j.ipm.2020.102390>
- Zhou X, Mulay A, Ferrara E, Zafarani R (2020) ReCOVvery. 3205–3212. <https://doi.org/10.1145/3340531.3412880>

How to cite: Vagif Mammadaliyev, Vusal Shahbazov (2026). Disinformation detection in the medical domain: current approaches, limitations, and future directions. Problems of Information Society, 1, 81–94. <https://doi.org/10.25045/jpis.v17.i1.09>