Sifting Truth from Spectacle! A Multimodal Hindi Dataset for Misinformation Detection with Emotional Cues and Sentiments

Raghvendra Kumar, Pulkit Bansal, Raunak Kumar Singh and Sriparna Saha

Abstract—**Misinformation is a rising threat across diverse media and languages. While English detection has advanced, Hindi—a globally spoken language—lags in research. In response, we introduce a novel multimodal Hindi dataset containing 6,544 article-image pairs designed to strengthen misinformation detection capabilities. Existing datasets largely focus on English and tend to be unimodal; we, therefore, develop a multimodal dataset in Hindi, combining both text and images. This dataset is rigorously curated to ensure bias-free content, minimizing predispositions that could skew misinformation classification. Additionally, we have incorporated emotion and sentiment annotations for each news sample, making it a pioneering effort. Through extensive evaluations, we demonstrate the dataset's effectiveness across diverse NLP applications, utilizing models such as IndicBART, IndicBERT, multilingual-BERT, and Vision-Transformer for misinformation classification across various configurations with a final integrated setup combining text, images, sentiments, and emotions. This comprehensive approach underscores the dataset's versatility and depth, facilitating robust misinformation detection across different modalities and feature combinations. Furthermore, we assess the readability scores of the news articles, adding a unique dimension to misinformation research. This work represents a significant advancement toward combating misinformation, fostering resilient model development, and promoting information integrity in Hindi-speaking regions, setting a foundation for future efforts in low-resource languages.**

Index Terms—**Misinformation Detection, Multimodal Hindi Dataset, Sentiment and Emotion Analysis.**

I. Introduction

In the digital era, multimodality has become increasingly important due to the rise of multimedia content and the need for more engaging and accessible communication. Numerous studies [1]–[4] have underscored that the majority of consumer internet traffic is driven by multimedia content, highlighting the escalating importance of multimodality in online communication. For instance, in education, multimodality enhances language learning by allowing engagement through words, images, gestures, sounds, and physical space. In marketing, it enables tailored messaging across preferred channels and devices, such as a company using a mix of written content, infographics, and videos to showcase sustainable products while offering tools to calculate carbon footprints. Multimodality, therefore, has become crucial in the digital era, impacting education, marketing, literacy development, etc. Furthermore, language versatility has also gained prominence, with studies highlighting the interconnected nature of these two phenomena and the importance of addressing their challenges and opportunities. For example, research in [5] explored the implications of social, cultural, and linguistic diversity alongside the increased visibility of technology in educational settings. Likewise, the extensive surveys conducted by Dabre et al. [6], and Erdem et al. [7] revealed the intricate relationship between multilingualism and multimodality, offering insights into the dynamic landscape of the digital era within a diverse and technologically-driven world.

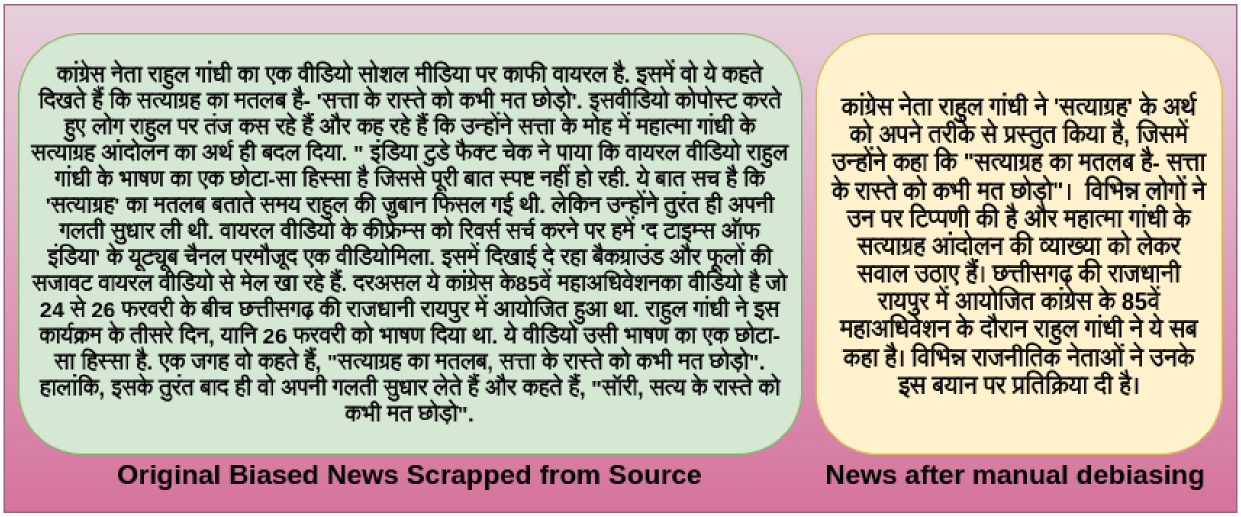
Alongside multimodality and multilingualism, the role of sentiments and emotions has also become increasingly significant. Emotional and sentiment analysis enables more refined understanding and responses across digital content, enhancing user engagement by recognizing the affective dimension of communication [8]–[10]. In educational settings, sentiment-aware tools can adapt content based on learner responses [11], while in marketing, emotion-driven insights help tailor customer interactions, creating more personalized and impactful experiences [12]. As digital platforms evolve, addressing the affective elements of communication—capturing sentiments and emotions—becomes essential for fostering deeper connection and enhancing the effectiveness of multimodal, multilingual interactions.

Recent advancements in Natural Language Processing (NLP), especially Large Language Models (LLMs), have profoundly transformed the field [13]–[15]. These models enable machines to understand and generate human-like text on an unprecedented scale. However, their technical capabilities also raise ethical concerns, particularly regarding the potential for misinformation and malicious content generation. The term "**Misinformation**" encompasses a spectrum of misleading or false information disseminated intentionally or unintentionally and takes various forms: *rumours*, which spread rapidly on social networks without verifiable evidence; *fake news*, fabricated to deceive for political or financial gain; *clickbait*, exaggerated to generate traffic at the expense of accuracy; *satire and parody*, while intended for entertainment, may be misconstrued as factual information; and *deepfakes* [16], using AI to create realistic yet fake audio or video content, posing significant manipulation threats. Understanding diverse forms

Raghvendra Kumar and Sriparna Saha are with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India (e-mail: {raghvendra_2221cs27,sriparna}@iitp.ac.in).

Pulkit Bansal is with the Department of Mathematics and Computing, Indian Institute of Technology Patna, India (e-mail: pulkit_2101mc48@iitp.ac.in).

Raunak Kumar Singh, with the Department of Computer Science, Darbhanga College of Engineering, Darbhanga, India (e-mail: 2raunaksingh@gmail.com)



कांग्रेस नेता राहुल गांधी का एक वीडियो सोशल मीडिया पर काफी वायरल है. इसमें वो ये कहते दिखते हैं कि सत्याग्रह का मतलब है- 'सत्ता के रास्ते को कभी मत छोड़ो'. इसवीडियो कोपोस्ट करते हुए लोग राहुल पर तंज कस रहे हैं और कह रहे हैं कि उन्होंने सत्ता के मोह में महात्मा गांधी के सत्याग्रह आंदोलन का अर्थ ही बदल दिया. " इंडिया टुडे फैक्ट चेक ने पाया कि वायरल वीडियो राहुल गांधी के भाषण का एक छोटा-सा हिस्सा है जिससे पूरी बात स्पष्ट नहीं हो रही. ये बात सच है कि 'सत्याग्रह' का मतलब बताते समय राहुल की जुबान फिसल गई थी. लेकिन उन्होंने तुरंत ही अपनी गलती सुधार ली थी. वायरल वीडियो के कीफ्रेम्स को रिवर्स सर्च करने पर हमें 'द टाइम्स ऑफ इंडिया' के यूट्यूब चैनल परमौजूद एक वीडियोमिला. इसमें दिखाई दे रहा बैकग्राउंड और फूलों की सजावट वायरल वीडियो से मेल खा रहे हैं. दरअसल ये कांग्रेस के85वें महाअधिवेशनका वीडियो है जो 24 से 26 फरवरी के बीच छत्तीसगढ़ की राजधानी रायपुर में आयोजित हुआ था. राहुल गांधी ने इस कार्यक्रम के तीसरे दिन, यानि 26 फरवरी को भाषण दिया था. ये वीडियो उसी भाषण का एक छोटा-सा हिस्सा है. एक जगह वो कहते हैं, "सत्याग्रह का मतलब, सत्ता के रास्ते को कभी मत छोड़ो". हालांकि, इसके तुरंत बाद ही वो अपनी गलती सुधार लेते हैं और कहते हैं, "सॉरी, सत्य के रास्ते को कभी मत छोड़ो".

कांग्रेस नेता राहुल गांधी ने 'सत्याग्रह' के अर्थ को अपने तरीके से प्रस्तुत किया है, जिसमें उन्होंने कहा कि "सत्याग्रह का मतलब है- सत्ता के रास्ते को कभी मत छोड़ो"। विभिन्न लोगों ने उन पर टिप्पणी की है और महात्मा गांधी के सत्याग्रह आंदोलन की व्याख्या को लेकर सवाल उठाए हैं। छत्तीसगढ़ की राजधानी रायपुर में आयोजित कांग्रेस के 85वें महाअधिवेशन के दौरान राहुल गांधी ने ये सब कहा है। विभिन्न राजनीतिक नेताओं ने उनके इस बयान पर प्रतिक्रिया दी है।

Original Biased News Scrapped from Source
News after manual debiasing

Fig. 1. The left side shows the original news with veracity indicators, while the right side displays the manually cleaned version, free from any bias.

of misinformation is crucial for developing effective mitigation strategies. *Throughout our paper, we collectively refer to these as misinformation.*

Research Gap & Motivation: Research in misinformation detection has mostly focused on text, with few multimodal approaches and limited studies in non-English languages. Existing datasets are primarily in English, overlooking non-English-speaking regions' linguistic and cultural diversity. This gap underscores the need for specialized datasets in widely spoken languages like Hindi [17] to promote inclusive research. While text-only datasets are beneficial, adding visual elements like images mirrors real-world scenarios where misinformation combines text and visuals. Moreover, understanding the emotional and sentiment-driven aspects of misinformation is crucial, as it can significantly influence public perception and engagement. Sentiments and emotions embedded in content shape how misinformation spreads and is perceived [18], [19]. Our Hindi multimodal dataset not only includes text and images but also incorporates sentiment and emotion annotations, enriching the analysis by addressing the affective components of misinformation.

Furthermore, a unique aspect of our dataset lies in how we handle the source of the data. While most misinformation datasets rely on fact-checking websites to curate content, which often directly indicates the veracity of articles and images, this can introduce bias. In contrast, we have taken extra care to manually remove any such indicators of veracity from both the text and images, ensuring that the news content is completely free of bias. Our experimental results demonstrate a clear distinction between biased and unbiased datasets, with noticeable removal of skewness, highlighting the effectiveness of our approach. This comprehensive resource helps in understanding the multifaceted nature of misinformation in Hindi, providing new opportunities to combat it effectively. *The forthcoming list highlights the significant* ***contributions of our work:***

- *We introduce a multimodal Hindi dataset specifically curated for misinformation detection. The dataset consists of 6,544 pairs of news articles and images, both carefully neutralized to eliminate any potential distortions, as shown in Figure 1 and 2. Each entry is annotated with one of three basic sentiments and one of seven emotions based on Paul Ekman's framework [20], including neutrality. These entries are then classified into misinformation or genuine content categories.*
- *To benchmark the effectiveness of our dataset, we apply state-of-the-art NLP models tailored for Indian languages, including IndicBERT [21], IndicBART [22], and multilingual BERT [23], alongside the Vision Transformer (ViT) [24], ResNet-50 [25] and VGG-19 [26] for image analysis. These models were chosen for their advanced attention-based embeddings, which are well-suited to the nuances of multilingual and multimodal data.*
- *In addition, we explore a variety of multimodal fusion techniques [27]–[29], such as early fusion and late fusion, while comparing the performance of deep learning models with traditional classifier-based approaches.*
- *As an extra layer of value, we manually annotate the news articles into multiple categories, including Politics, International Relations, Health, Economics, Technology, and Culture & Entertainment, among others. Furthermore, we provide Hindi readability scores [30] for both genuine and misinformation articles to offer a deeper understanding of the linguistic complexity in each category.*

II. Related Works

Misinformation has existed for decades, with historical accounts dating back to the 1920s, as noted by Lippmann [31]. This exploration examines opinion formation and the media's role in shaping public perception, providing foundational insights into misinformation spread. The 1950s, during the Cold War, saw significant scrutiny of propaganda's impact on public opinion [32]. However, in contemporary times, technological

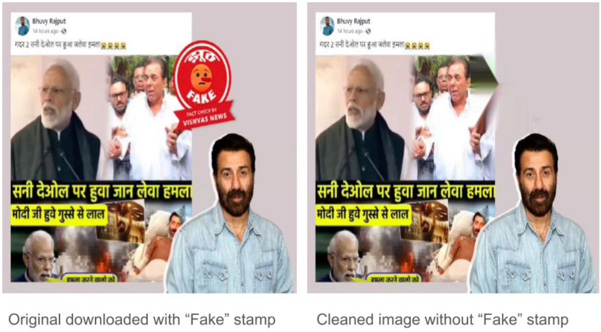

FAKE
FACT CHECK BY VISHVAS NEWS
सनी देओल पर हुवा जान लेवा हमला
मोदी जी हुवे गुस्से से लाल
Original downloaded with “Fake” stamp
Cleaned image without “Fake” stamp

Fig. 2. Image transformation showing image after removing fake news verification stamp.

advancements and social media have transformed misinformation’s propagation and consumption. Traditionally spread through newspapers and television, misinformation was often geographically confined. Now, social media enables rapid, global dissemination. While these historical works may not explicitly employ the term “misinformation,” they address related concepts, offering valuable insights into the study of information dissemination and manipulation. In the following sections, we explore related datasets, particularly highlighting multilingual or non-English datasets. Despite being notably fewer in number compared to their English counterparts, as evidenced by various comprehensive surveys [33]–[36], these datasets play a crucial role in the fight against misinformation. We also include a brief subsection on datasets focused on combating COVID-related misinformation, acknowledging their prevalence during the pandemic, even though our dataset primarily addresses post-COVID generic news content. Lastly, we highlight significant contributions from affective computing in the fight against misinformation.

A. Unimodal Misinformation Datasets

Posadas-Durán et al. [37] introduced a Spanish fake news dataset with 491 true and 480 fake articles across 9 topics, establishing a baseline for Spanish-language detection. Ma et al. [38] provided a dataset of 2,313 rumours and 2,351 non-rumours in Chinese from Weibo. Arabic datasets included stance and fact-checking labels, as seen in [39]–[41]. Stance-annotated datasets for Danish and Croatian appeared in [42]–[44]. For Hindi, Bhardwaj et al. [45], Kumar et al. [46], and Sharma et al. [47] introduced datasets for hostility and fake news detection. Similar work has been conducted by researchers in [48]–[51]; however, these studies overlook bias. Their results, boasting over 98% accuracy, clearly illustrate the presence of the very bias we aim to address. ***Limitations:*** *While unimodal datasets remain valuable, the rise of vision-language models underscores the need for multimodal datasets. Our dataset combines both text and visuals to address the evolving demands of modern research and technology. Furthermore, none of these works consider the role of sentiments and emotions in combating misinformation.*

B. Multimodal Misinformation Datasets

Reis et al. [52] introduced a dataset of fact-checked images circulated on WhatsApp during the 2018 Brazilian and 2019 Indian elections, highlighting misinformation spread on messaging platforms. Jin et al. [53] presented a multimedia rumour dataset from Weibo and Twitter, classifying 9,528 rumour and 13,924 non-rumour text-image pairs, demonstrating the effectiveness of their attention-based RNN for multimodal rumour detection. Papadopoulou et al. [54] introduced a dataset of user-generated videos that have been debunked or verified. Similarly, Kazemi et al. [55] presented a dataset of verified claims from WhatsApp tipline reports and public group messages, including Hindi language content. Two notable multimodal datasets addressing misinformation in the Indian context are IFND [56] and fakenewsindia [57]. ***Limitations:*** *However, it’s worth noting that the text within these datasets is in English. This underscores the scarcity of multimodal datasets in Hindi for addressing misinformation. Our dataset aims to fill this gap.*

C. Covid-19 Misinformation Datasets

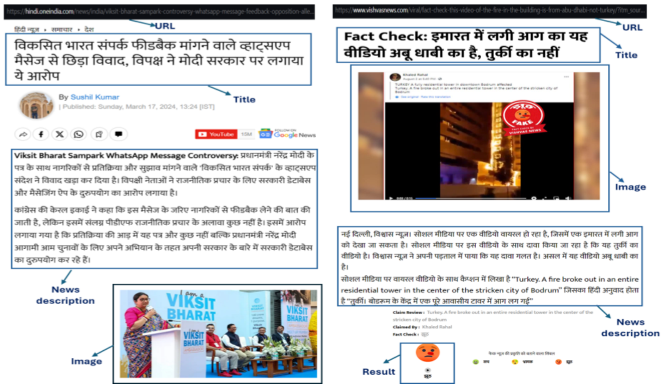
Shahi et al. [58] created a dataset of 5,182 fact-checked COVID-19 articles in 40 languages. Haouari et al. [59] introduced 9.4K annotated Arabic COVID-19 tweets. Li et al. [60] developed MM-COVID, with 3,981 fake and 7,192 truthful instances across 6 languages. In India, [61] curated a Hindi and Bengali fake news dataset for COVID-19 tweets, while [62] introduced COVID-19-FAKES, a bilingual Arabic/English dataset for Twitter misinformation. Yang et al. [63] provided a dataset of 2,120 microblogs with multimedia and social network data. Alam et al. [64] focused on English and Arabic tweets with over 500 retweets each. ***Limitations:*** *While most COVID-based datasets originate from Twitter and are characterized by shorter and less coherent text, ours are sourced from news websites. Consequently, our dataset features articles with additional information and clarity.*

D. Affective Computing for Combating Misinformation

Lutz et al. [65] analyzed linguistic cues in fake news, finding that cognitive engagement rises with article length and analytic language boosts affective processing, aiding platform design for critical thinking. In [66], researchers used textual novelty and emotion prediction for misinformation detection with entailment and emotion models. Another study [67] showed that low heart rate variability and brief fixations increase fake news misjudgment due to affective responses. Bakir et al. [68] linked emotions with fake news spread, advocating media literacy to counteract “information disorder.” Ghanem et al. [69] introduced FakeFlow, a neural model capturing emotional manipulation, outperforming existing methods across datasets. [70] used virality theory, finding that positive emotion, sentiment alignment, and sensationalism impact fake news spread. Similarly, [71] combined user attributes and emotion analysis in misinformation detection with improved results. Kumar et al. [72] applied LLMs with sentiment and emotion analysis, also using ChatGPT to simulate distortions, distinguishing real and AI-generated misinformation. For further insights, readers may refer to additional notable works, reviews, and surveys, including [73]–[78]. ***Limitations:*** *All previously mentioned works focus predominantly on the English language, whereas our work centres on Hindi as the primary language.*

III. Our Dataset

The following sections outline our dataset's construction process, bias elimination, sentiment and emotion annotations, image refinement, categorization strategy, and effectiveness demonstrated through extensive experimental evaluations.



https://hindi.oneindia.com/news/india/viksit-bharat-sampark-controversy-whatsapp-message-feedback-opposition-alle...

हिंदी न्यूज • समाचार • देश

विकसित भारत संपर्क फीडबैक मांगने वाले व्हाट्सएप मैसेज से छिड़ा विवाद, विपक्ष ने मोदी सरकार पर लगाया ये आरोप

By Sushil Kumar
| Published: Sunday, March 17, 2024, 13:24 [IST]

Viksit Bharat Sampark WhatsApp Message Controversy: प्रधानमंत्री नरेंद्र मोदी के पत्र के साथ नागरिकों से प्रतिक्रिया और सुझाव मांगने वाले 'विकसित भारत संपर्क' के व्हाट्सएप संदेश ने विवाद खड़ा कर दिया है। विपक्षी नेताओं ने राजनीतिक प्रचार के लिए सरकारी डेटाबेस और मैसेजिंग ऐप के दुरुपयोग का आरोप लगाया है।

कांग्रेस की केरल इकाई ने कहा कि इस मैसेज के जरिए नागरिकों से फीडबैक लेने की बात की जाती है, लेकिन इसमें संलग्न पीडीएफ राजनीतिक प्रचार के अलावा कुछ नहीं है। इसमें आरोप लगाया गया है कि प्रतिक्रिया की आड़ में यह पत्र और कुछ नहीं बल्कि प्रधानमंत्री नरेंद्र मोदी आगामी आम चुनावों के लिए अपने अभियान के तहत अपनी सरकार के बारे में सरकारी डेटाबेस का दुरुपयोग कर रहे हैं।

https://www.vishvasnews.com/viral/fact-check-this-video-of-the-fire-in-the-building-is-from-abu-dhabi-not-turkey/?itm_sour...

Fact Check: इमारत में लगी आग का यह वीडियो अबू धाबी का है, तुर्की का नहीं

Khaled Rahal
August 2 at 5:40 PM ·
TURKEY A fully residential tower in downtown Bodrum affected
Turkey. A fire broke out in an entire residential tower in the center of the stricken city of Bodrum
See original · Rate this translation

नई दिल्ली, विश्वास न्यूज़। सोशल मीडिया पर एक वीडियो वायरल हो रहा है, जिसमें एक इमारत में लगी आग को देखा जा सकता है। सोशल मीडिया पर इस वीडियो के साथ दावा किया जा रहा है कि यह तुर्की का वीडियो है। विश्वास न्यूज़ ने अपनी पड़ताल में पाया कि यह दावा गलत है। असल में यह वीडियो अबू धाबी का है।

सोशल मीडिया पर वायरल वीडियो के साथ कैप्शन में लिखा है "Turkey. A fire broke out in an entire residential tower in the center of the stricken city of Bodrum" जिसका हिंदी अनुवाद होता है "तुर्की। बोडरूम के केंद्र में एक पूरे आवासीय टावर में आग लग गई"

Claim Review : Turkey. A fire broke out in an entire residential tower in the center of the stricken city of Bodrum
Claimed By : Khaled Rahal
Fact Check : झूठ

फेक न्यूज की प्रकृति को बताने वाला सिंबल
● सच ● भ्रामक ● झूठ

Fig. 3. The sample format of webpages from OneIndia (left and genuine) and Vishvas News (right and misinformation), which contributed the most articles to our dataset, along with the extracted features during the scraping process.

A. Websites Used for Dataset Formation

We thoroughly explored various websites providing fact-checked news articles containing textual content and images. After careful consideration, we selected *VishvasNews, Aajtak, AmarUjala, DailyHunt* and *OneIndia* as our primary sources for fact-checking.

RQ-1: Why did we choose these specific websites and not others? These websites were chosen for their clarity in classifying news as misinformation or genuine. Their optimal structure and absence of unwanted ads simplified scraping and clear verdicts aided ground truth classification. As reputable fact-checking platforms in India, their commitment to accuracy reinforces their credibility.

RQ-2: How do we ensure minimal overlap between news articles from different websites? To address this concern, we employed both manual and automatic methods. In the manual approach, we randomly selected 100 samples from each website and observed significant stylistic differences in writing between the sites. This variance is expected, as journalists' styles naturally differ. *Additionally, this is reinforced by the variation in the Hindi readability scores obtained for each source, as shown in Table II.* For the automatic method, we extracted 500 random headlines from articles from all websites (except AmarUjala, 126 data points) and computed ROUGE-1 [79] and ROUGE-L [79] scores between them. Notably, the calculated scores were consistently low. This analysis solidified our finding that there was minimal to no overlap between the articles sourced from these websites.

B. Textual Data Scraping and Image Retrieval

Misinformation News Data Scraping: Each misinformation sample from all websites (except *DailyHunt*) includes an image, a concise description, and a determination of whether the news is true or false. Utilizing *BeautifulSoup*, we scraped the title of each article, its corresponding image, news description, and the true (genuine) or false (misinformation) label, as illustrated in Figure 3.

Genuine News Data Collection: Due to the prevalence of misinformation, we balanced our dataset by extracting genuine news articles from the Hindi news section of the *OneIndia.com* and *DailyHunt* website. Like the misinformation-containing articles collection process, these articles, too, contained an image, a brief description, and news content. Using *BeautifulSoup*, we retrieved the title, URL of the corresponding image, and brief description, labeling the news as Genuine/True.

Cleaning of Verified Images: As observed in Figure 3, **misinformation/fake news images (and not the genuine images)** often contain a veracity claim stamped on the image itself. To address this, we used the OpenCV library for such template removal, followed by smoothing the affected section to maintain a natural appearance. This process ensures that the images do not reveal their authenticity, simulating real-life scenarios where fake images are not explicitly labelled. Figure 2 illustrates the transformation of an original image, initially stamped as fake, to its appearance after the stamp's removal. The left side shows the image with the fake news verification stamp, while the right side displays the image post-stamp removal.

RQ-3: Does removing the 'fake' veracity stamp from images introduce bias, potentially affecting the accuracy of misinformation classification? To ensure no bias is introduced, we conducted misinformation classification solely on the images themselves using both traditional machine learning and deep learning methods. As detailed in the results section (Table VII), an interesting finding is that both machine learning and deep learning methods yielded moderate accuracy. If bias had been introduced, we would expect much higher accuracy rates. This consistency across methods supports our assurance that no bias has been introduced.

RQ-4: Why are images necessary? Wouldn't text alone be sufficient? We address this with an example, illustrated in Figure 4. The left image shows a counter that could belong to various settings, such as a bank or a movie theatre. However, the accompanying text clarifies it as a railway ticket counter. In the right example, the text describes a 500 INR note with an asterisk, indicating it's counterfeit, while the corresponding image helps visualize what this fake note looks like. This illustrates our motivation for creating a multimodal Hindi dataset, where text and images work in tandem to provide context and enhance understanding.

C. Manual Bias Elimination of News Topics

To ensure a neutral presentation of news topics, a team of three proficient annotators was hired to perform manual bias elimination on circulated news articles. These annotators, proficient in the Hindi language, were compensated at a rate

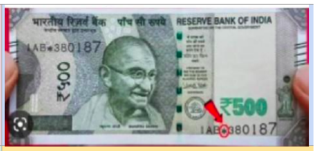
	
Scenario: When Text Enhances the Clarity of an Image	Scenario: When an Image Adds Further Insight to Text
Hindi Text: हाल ही में रेलवे ने स्पष्टीकरण जारी कर बताया कि ट्रेन से यात्रा करने वाले बच्चों के लिए टिकट की बुकिंग से संबंधित नियम में कोई बदलाव नहीं किया गया है। वहीं अब ऑफलाइन टिकट काउंटरों को लेकर रेलवे ने नई जानकारी साझा की है। भारतीय रेलवे ने टिकट काउंटरों से बारे में किए गए दावों को फर्जी करार दिया है। **In English:** Recently, the railways issued a clarification stating that there has been no change in the rules regarding ticket booking for children traveling by train. Additionally, the railways have now shared new information regarding offline ticket counters. Indian Railways has declared the claims made about the ticket counters as fake.	**Hindi Text**: नोटबंदी और फिर 2000 रुपए के नोटों का चलन बंद होने के बाद अब 500 के नए नोट चर्चा का विषय बने हुए हैं। वो भी ऐसे नोट जिनके नंबर पैनल में स्टार (*) चिह्न अंकित है। इन नोटों के वैध होने पर सवाल सोशल मीडिया पर उठने लगे हैं। **In English**: After demonetization and the discontinuation of the 2000-rupee notes, the new 500-rupee notes have become a topic of discussion. Specifically, notes with a star (*) symbol in their number panel are raising questions about their validity on social media.

Fig. 4. Example of text enhancing image understanding and image adding context to text.

of USD 1.20 per ten articles, and the task was equally divided among them. Each annotator initially worked independently on their assigned articles, focusing on presenting the information without any indication of bias. After completing their work, each annotation underwent a peer-review process in which the other two team members reviewed it for adherence to the neutrality guidelines. An annotation was approved only when both reviewers confirmed it met the standard of objective representation, ensuring the reader could not discern whether the news was genuine or fake upon reading. This method minimized personal bias, resulting in consistent and impartial descriptions of the news events.

The objective of the task for each annotator was to extract the main news content being circulated in a neutral and comprehensive manner, avoiding any language that may imply the news's authenticity or lack thereof. Annotators should refrain from using phrases or terminology that would indicate the information's questionable nature, such as terms suggesting that the news is a "claim," "misleading," or "viral." Instead, annotators should present the information in a descriptive and complete way, ensuring that readers cannot determine whether the news is true or false. The goal was to capture only the information as it is being shared without addressing its factual accuracy. This ensures a neutral extraction focused solely on the circulated content, free from additional commentary or fact-checking cues.

D. Annotation of Bias-Free News Topics

We enlisted five independent PhD students from the Humanities department, proficient in Hindi and knowledgeable in specific content areas, to manually annotate news articles across the following topics: Politics, Culture & Entertainment, International Relations, Health, Economics, Technology, Social Issues, and Others. To further enrich the annotations, the annotators labelled each article with basic Paul Ekman emotions (anger, fear, joy, sadness, surprise, disgust) and a neutrality label if the article did not exhibit any particular emotional tone. In addition, each article was tagged with one of three sentiment categories—positive, negative, or neutral—based on its overall sentiment. Each article's topic, sentiment and emotion were determined through a majority voting system, with ties resolved by a faculty member from the same department.

To ensure consistency and reliability, the annotators underwent a comprehensive training session where they were briefed on annotation guidelines, the definitions and scope of each category, and provided with examples for both topic and emotion/sentiment classifications. Following this, a pilot annotation phase allowed them to annotate a subset of articles and discuss any discrepancies or challenges, thereby aligning their understanding and approach before proceeding to the main task. Figure 5 illustrates the count of news articles (misinformation and genuine) across different topics and the distribution of articles sourced from the five websites used for scraping.

RQ-5: Why did we choose the mentioned Emotions and Sentiments? To gain deeper insights into the psychological impact of misinformation on readers, we annotated key emotions—anger, fear, joy, sadness, surprise, and disgust—alongside neutrality and sentiment labels (positive, negative, neutral). These emotions are grounded in Paul Ekman's model, which identifies fundamental emotional responses universally recognized across cultures. By focusing on these core emotions, we aim to capture how emotional cues within misinformation might evoke reactions that shape readers' perceptions or amplify inherent biases. This emotion framework has been extensively validated and applied in affective computing and NLP research [72], [80]–[82], adding robustness to our approach. Additionally, we included sentiment annotations to assess the overall tone of the articles, allowing us to examine how positive, negative, or neutral framing might impact credibility and reader engagement. Sentiment analysis has proven useful across various domains [83]–[85]. Together, these emotion and sentiment labels provide a structured approach to understanding how misinformation leverages emotional engagement, potentially influencing reader trust and comprehension.

E. Final Dataset Organization

After collecting data from various articles on the aforementioned websites, we structured the dataset in CSV format for easy accessibility. This dataset includes ten key pieces of information: Title, Article URL, Image URL, Original Article, Bias-Free Description, Topic, Readability Score, Emotion, Sentiment and Label. Additionally, we provide a Python script to download images directly from the provided URLs as well as the scrapping code for the entire article.

F. Dataset Insights

Our dataset comprises 6,544 news article-image pairs, equally divided between misinformation and genuine news (3272 each). Figure 5 displays the count of news articles by topic and the distribution of articles from the five websites used. Similarly, Figure 6 illustrates the distribution of emotions and sentiments across annotated articles. Additionally, we present further statistical details of our dataset in tabular format in Table I. We used the Indic-NLP-Library tokenizer [86].

Hindi Readability Score (HRS): Sinha et al. [30] proposed various models to assess readability scores in Bangla and

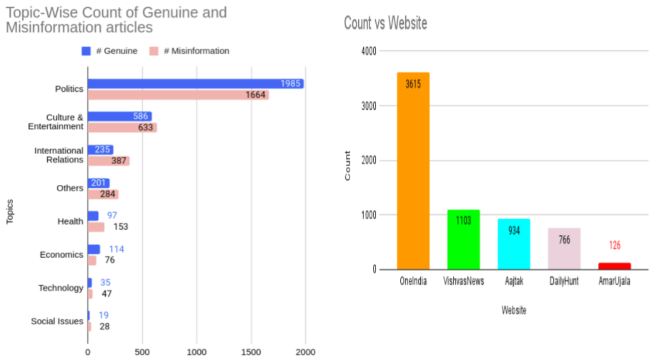


Fig. 5. Source-wise and Topic-wise Distribution of News Articles

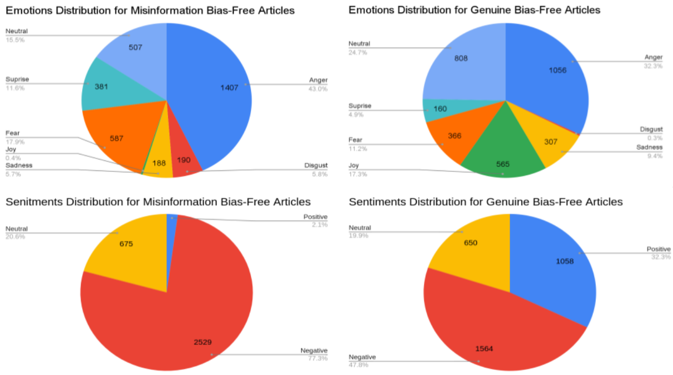


Fig. 6. Illustration showcasing the breakdown of emotions and sentiments within the annotated articles.

TABLE I
THE TABLE SHOWS THE MAXIMUM, MINIMUM, AND AVERAGE VALUES FOR TOKENS, WORDS, UNIQUE WORDS, AND SENTENCES PER ORIGINAL ARTICLE AND BIAS-FREE ARTICLE.

Original Scrapped Articles				
Metrics	**# Tokens**	**# Words**	**# Unique Words**	**# Sentences**
Max	8653	7589	1305	206
Min	69	28	28	6
Average	550	498	223	24
Manually Bias Removed Articles				
Metrics	**# Tokens**	**# Words**	**# Unique Words**	**# Sentences**
Max	252	227	145	14
Min	20	17	17	1
Average	94	86	64	5

TABLE II
COMPARISON OF HINDI READIBILITY SCORES (HRS) FOR FALSE AND TRUE BIAS-FREE ARTICLES ACROSS VARIOUS TOPICS.

Source Wise	**HRS ↓**	**Topic Wise ↓ & HRS →**	**False**	**True**
OneIndia	7.94	*Politics*	7.55	8.15
VishvasNews	7.49	*Culture & Entertainment*	7.45	8.06
Aajtak	7.47	*International Relations*	7.50	8.11
DailyHunt	8.04	*Others*	7.33	8.10
AmarUjala	7.81	*Health*	7.53	8.10
Class Wise		*Economics*	7.89	8.04
False	7.50	*Technology*	7.30	8.26
True	8.11	*Social Issues*	7.39	7.74

Hindi, considering factors such as average sentence length, average word length, average syllables per word, and polysyllabic word frequency. We used one of their recommended Hindi readability scores (HRS), which is calculated as follows:

$$\text{HRS} = -2.34 + 2.14 \times \text{AWL} + 0.01 \times \text{PSW}$$

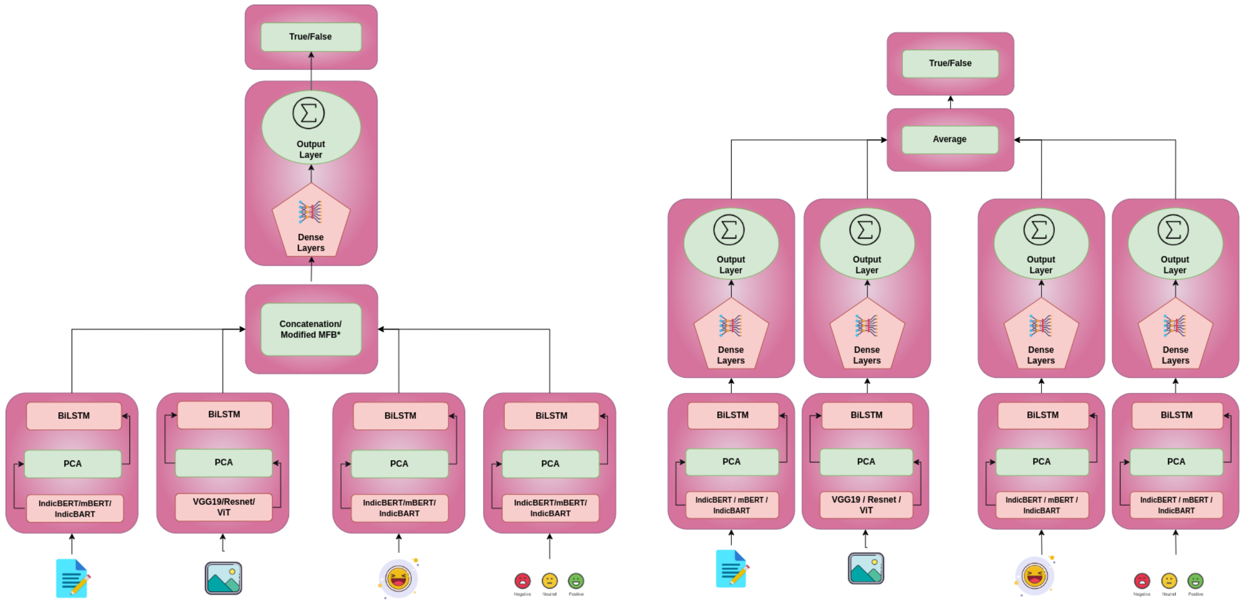
where AWL represents average word length, and PSW denotes the number of polysyllabic words containing more than two syllables; also, higher HRS values indicate greater readability. Table II presents the Hindi Readability Score (HRS) for various news sources and topics, comparing the scores for false and true news within each category. The analysis reveals considerable variations among the sources, with Aajtak exhibiting the lowest HRS and DailyHunt showing the highest, suggesting differing levels of readability in their reporting. Topics such as Technology, Social Issues, and Others score lower HRS than Politics, International Relations, and Economics, especially when considering fake bias-free articles. For true bias-free articles, Technology and Politics receive the highest HRS, while Social Issues receive the lowest, highlighting variability in reporting quality across different subjects. *Moreover, misinformation articles consistently score lower HRS than genuine articles across all topics, suggesting that the deliberate falsification of news—often by introducing unnecessary complexities, drama, or provocative elements—may inadvertently compromise the readability and clarity of the content.* To conclude, the distribution of news samples is as follows: OneIndia contributes 2,505 true and 1,110 false articles; VishvasNews and Aajtak provide 1,103 and 934 false articles, respectively; DailyHunt offers 766 true articles; AmarUjala adds 1 true and 125 false articles. This diverse sourcing enables robust analysis by reflecting various reporting practices.

RQ-6: How can we leverage dataset metrics to inform decision-making in NLP research? The statistics in Table I highlight our dataset's diversity. The wide range of metrics like tokens, words, unique words, and sentences reflect varied news articles catering to different informational needs. Consistent average values indicate balanced content distribution, suitable for named entity recognition, text summarization, sentiment analysis, and more, enabling insightful NLP research.

RQ-7: What are the advantages of using a news article dataset for misinformation detection compared to shorter, less coherent texts like tweets? Our dataset, comprising news articles classified as misinformation (false) or genuine (true), offers significant advantages over shorter texts like tweets. The coherent nature of news articles allows for a deeper analysis of language patterns, context, and narrative structures. Unlike tweets, news articles provide detailed explanations, background information, and supporting evidence, enabling comprehensive analysis and the identification of subtle misinformation cues. Authored by professional journalists and undergoing editorial scrutiny, these articles enhance credibility and reliability. Therefore, our dataset is a valuable resource for developing robust misinformation detection models.

IV. EXPERIMENTS: SETUP & BASELINES

For text embedding extraction, we utilized state-of-the-art transformer models suited for Indian languages such as IndicBART (`ai4bharat/IndicBART`) [22], IndicBERT



Early Fusion / Multimodal factorized bilinear pooling*

Late Fusion

Fig. 7. Utilized Deep Learning Techniques: MFB Variant, Late Fusion, and Early Fusion.

(`ai4bharat/indic-bert`) [21], and multilingual-BERT (`bert-base-multilingual-cased`) [23]. For image embeddings, we employed ResNet-50 [25], VGG-19 [26] and Vision Transformer (ViT) (`google/vit-base-patch16-224`) [24]. *To incorporate attention-based embedding generation, we employed these aforementioned models.* To ensure compatibility between text and image embeddings, we applied PCA to match the 768-dimensional output of BERT-related models. Extending this approach to IndicBART embeddings ensured consistency across all custom deep-learning embedding models. Given the uniformity of our dataset, we experimented using 5-fold cross-validation to provide a balanced and reliable performance assessment by utilizing all data for both training and testing across multiple splits.

TABLE III
THE DIFFERENT HYPERPARAMETERS UTILIZED IN THE EXPERIMENTS.

Model	Hyperparamters Value
RBF SVM	gamma=0.45, C=3.7
Random Forest	criterion='gini, min_samples_split=2
Naive Bayes	var_smoothing=1e-9
AdaBoost	n_estimator=50, learning_rate=1
QDA	reg_param=0, tol=1e-4
Deep Learning Models	batch_size=32, epochs=10, dropout=0.05

Traditional ML Approach: The experiment employed a dictionary of traditional machine learning classifiers as indicated in Table IV, V & VI. Hyperparameter tuning was performed on a subset of the dataset to identify optimal settings for robust model performance. For instance, the RBF SVM classifier used 'gamma=0.45' and 'C=3.7' to balance model complexity and regularization. We also explored various hyperparameter values, such as 'n_estimators=1000' for the Random Forest model, to further enhance results. This process ensured that the models were finely tuned to capture the underlying patterns in the data while avoiding overfitting. Additional hyperparameter details are provided in Table III. We emphasize that the same hyperparameter values were applied in both our primary experiments (unbiased dataset) and supplementary experiments (biased dataset).

Deep Learning Approach: In our experiments, we employed both early and late fusion techniques for dataset classification, as shown in Figure 7. **Early fusion** (feature-level fusion) combines raw data from multiple modalities at the initial processing stage, treating the fused data as a unified modality for single-model processing. This is advantageous when modalities are similar and well-aligned. Conversely, **Late fusion** processes each modality separately using specialized models and combines predictions later, mitigating the impact of differing feature counts. Additionally, inspired by **multimodal factorized bilinear (MFB) pooling** from Yu et al. [29], we experimented with a variant to establish a baseline for our deep learning experiments.

We used TensorFlow and Keras to build deep learning models with a bidirectional LSTM architecture for multimodal data integration. Fine-tuning hyperparameters enhanced model efficacy, with the Adam optimizer and binary cross-entropy loss function ensuring robust learning. Evaluation metrics included accuracy, precision, recall, and F1-score. For late fusion, separate models for image and text data, each using a bidirectional LSTM layer followed by dense layers with ReLU activation, were trained and then fused. Implementing the MFB technique involved element-wise multiplication of bidirectional LSTM outputs followed by power normalization, square root, and L2 normalization, with resulting features

TABLE IV

INDICBART RESULTS FOR CONFIGURATIONS COMBINING TEXT WITH IMAGE EMBEDDINGS: TEXT AND IMAGE, TEXT WITH IMAGE AND EMOTIONS, TEXT WITH IMAGE AND SENTIMENTS, AND TEXT WITH IMAGE, SENTIMENTS, AND EMOTIONS.

Text Embedding IndicBART												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	88.39%	90.07%	86.04%	88.01%	90.18%	92.21%	87.59%	89.84%	85.41%	89.14%	80.34%	84.51%
AdaBoost	87.32%	89.01%	84.89%	86.90%	88.24%	89.66%	86.20%	87.89%	86.67%	87.62%	85.12%	86.35%
QDA	87.20%	84.31%	91.13%	87.59%	88.20%	84.31%	93.60%	88.71%	86.29%	80.77%	94.91%	87.27%
Naive Bayes	62.99%	60.25%	74.33%	66.55%	71.47%	70.25%	73.55%	71.86%	74.64%	81.43%	63.22%	71.18%
Rbf SVM	53.69%	52.63%	54.79%	53.69%	52.03%	50.00%	50.70%	50.35%	53.15%	54.79%	54.05%	54.42%
Early Fusion	89.87%	91.60%	85.71%	88.56%	90.00%	92.19%	85.51%	88.72%	90.75%	92.25%	84.40%	88.15%
Late Fusion	89.49%	91.20%	85.07%	88.03%	89.77%	91.60%	85.71%	88.56%	89.81%	92.31%	84.51%	88.24%
MFB	88.68%	90.11%	86.89%	88.44%	90.66%	92.86%	88.10%	90.40%	92.77%	94.39%	90.96%	92.64%
Text Embedding: IndicBART with Emotions												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	85.29%	85.51%	85.29%	85.28%	85.29%	85.51%	85.29%	85.28%	85.75%	85.76%	85.75%	85.75%
AdaBoost	90.22%	90.29%	90.22%	90.22%	90.22%	90.29%	90.22%	90.22%	89.53%	89.54%	89.53%	89.53%
QDA	76.93%	83.07%	76.93%	75.87%	76.93%	83.07%	76.93%	75.87%	64.40%	78.91%	64.40%	59.48%
Naive Bayes	64.94%	66.11%	64.94%	64.37%	64.94%	66.11%	64.94%	64.37%	65.05%	66.26%	65.05%	64.48%
Rbf SVM	51.45%	51.52%	49.28%	50.37%	52.78%	52.46%	45.07%	48.48%	51.97%	48.44%	52.54%	50.41%
Early Fusion	92.68%	93.74%	91.48%	92.57%	**93.87%**	95.42%	92.16%	93.75%	93.78%	94.19%	93.31%	93.73%
Late Fusion	91.44%	89.44%	93.98%	91.64%	93.28%	90.89%	**96.19%**	93.46%	92.50%	90.75%	94.66%	92.65%
MFB	89.46%	91.36%	87.18%	89.22%	91.11%	92.53%	89.43%	90.94%	93.52%	95.38%	91.46%	93.36%
Text Embedding: IndicBART with Sentiments												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	84.84%	85.06%	84.84%	84.82%	85.37%	85.65%	85.37%	85.35%	86.02%	86.03%	86.02%	86.02%
AdaBoost	88.62%	88.66%	88.62%	88.61%	90.22%	90.29%	90.22%	90.22%	89.53%	89.54%	89.53%	89.53%
QDA	74.22%	81.61%	74.22%	72.69%	77.04%	83.18%	77.04%	75.99%	64.71%	79.02%	64.71%	59.92%
Naive Bayes	64.82%	65.97%	64.82%	64.26%	64.94%	66.11%	64.94%	64.37%	65.05%	66.26%	65.05%	64.48%
Rbf SVM	53.97%	47.89%	61.82%	53.97%	53.29%	53.01%	53.01%	53.01%	53.13%	54.76%	55.42%	55.09%
Early Fusion	92.51%	94.11%	90.71%	92.36%	93.72%	94.76%	92.67%	93.65%	93.64%	94.11%	93.20%	93.61%
Late Fusion	91.37%	91.29%	91.55%	91.38%	93.14%	92.58%	93.82%	93.18%	92.60%	92.72%	92.51%	92.60%
MFB	88.42%	89.56%	87.01%	88.25%	90.94%	92.10%	89.56%	90.80%	93.23%	**95.64%**	90.70%	93.05%
Text Embedding: IndicBART with Emotions and Sentiments												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	85.56%	86.17%	85.56%	85.51%	85.87%	86.41%	85.87%	85.82%	86.25%	86.37%	86.25%	86.24%
AdaBoost	88.96%	89.01%	88.96%	88.96%	89.84%	89.89%	89.84%	89.83%	89.76%	89.77%	89.76%	89.76%
QDA	78.04%	78.78%	78.04%	77.87%	75.63%	82.45%	75.63%	74.35%	73.72%	82.27%	73.72%	71.94%
Naive Bayes	68.26%	72.06%	68.26%	66.93%	68.30%	72.12%	68.30%	66.97%	68.45%	72.17%	68.45%	67.17%
Rbf SVM	52.63%	55.56%	50.00%	52.63%	53.57%	68.18%	44.12%	53.57%	52.52%	52.11%	53.62%	52.86%
Early Fusion	92.99%	94.41%	91.40%	92.87%	93.72%	93.68%	93.76%	93.71%	**93.87%**	94.16%	93.55%	**93.84%**
Late Fusion	91.92%	90.57%	93.63%	92.05%	93.34%	92.51%	94.32%	93.40%	92.51%	91.26%	94.02%	92.61%
MFB	89.50%	90.25%	88.64%	89.42%	91.00%	91.58%	90.33%	90.93%	93.72%	94.73%	92.62%	93.63%

TABLE V

INDICBERT RESULTS FOR CONFIGURATIONS COMBINING TEXT WITH IMAGE EMBEDDINGS: TEXT AND IMAGE, TEXT WITH IMAGE AND EMOTIONS, TEXT WITH IMAGE AND SENTIMENTS, AND TEXT WITH IMAGE, SENTIMENTS, AND EMOTIONS.

Text Embedding IndicBERT												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	79.64%	79.03%	80.19%	79.60%	80.06%	79.33%	80.80%	80.06%	83.04%	83.66%	81.73%	82.68%
AdaBoost	82.09%	81.32%	82.88%	82.09%	83.19%	83.77%	81.96%	82.85%	82.70%	82.26%	82.96%	82.61%
QDA	80.40%	76.74%	86.74%	81.43%	80.44%	74.55%	91.90%	82.32%	77.73%	71.10%	**92.75%**	80.50%
Naive Bayes	69.25%	71.89%	62.30%	66.75%	69.21%	71.74%	62.45%	66.78%	72.38%	75.13%	66.15%	70.36%
Rbf SVM	51.45%	51.52%	49.28%	50.37%	52.78%	52.46%	45.07%	48.48%	51.97%	48.44%	52.54%	50.41%
Early Fusion	85.68%	85.30%	86.33%	85.77%	87.65%	89.54%	85.56%	87.37%	87.67%	88.48%	86.87%	87.58%
Late Fusion	84.98%	87.46%	81.77%	84.47%	88.37%	90.51%	85.79%	88.06%	87.18%	87.61%	86.63%	87.09%
MFB	82.47%	83.94%	80.45%	82.09%	87.13%	89.83%	83.84%	86.69%	88.34%	90.79%	85.42%	87.98%
Text Embedding: IndicBERT with Emotions												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	81.74%	81.83%	81.74%	81.73%	82.31%	82.39%	82.31%	82.31%	84.38%	84.38%	84.38%	84.38%
AdaBoost	83.61%	83.61%	83.61%	83.61%	85.87%	85.87%	85.87%	85.87%	85.64%	85.64%	85.64%	85.64%
QDA	76.32%	78.85%	76.32%	75.83%	70.55%	71.22%	70.55%	70.27%	65.74%	77.85%	65.74%	61.72%
Naive Bayes	69.21%	73.26%	69.21%	67.91%	69.14%	73.06%	69.14%	67.86%	69.40%	73.40%	69.40%	68.14%
Rbf SVM	53.97%	47.89%	61.82%	53.97%	53.29%	53.01%	53.01%	53.01%	53.13%	54.76%	55.42%	55.09%
Early Fusion	87.50%	88.76%	86.00%	87.30%	88.83%	90.06%	87.39%	88.66%	89.70%	89.17%	90.41%	89.77%
Late Fusion	88.03%	88.70%	87.31%	87.96%	**90.62%**	91.23%	90.01%	**90.56%**	90.10%	90.92%	89.13%	90.00%
MFB	84.93%	85.64%	83.95%	84.75%	87.85%	89.61%	85.64%	87.54%	89.18%	90.88%	87.07%	88.93%
Text Embedding: IndicBERT with Sentiments												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	81.82%	81.87%	81.82%	81.82%	82.35%	82.39%	82.35%	82.35%	85.26%	85.26%	85.26%	85.26%
AdaBoost	84.76%	84.76%	84.76%	84.76%	85.49%	85.50%	85.49%	85.48%	85.41%	85.41%	85.41%	85.41%
QDA	75.17%	79.87%	75.17%	74.22%	60.12%	76.39%	60.12%	53.11%	67.69%	78.84%	67.69%	64.37%
Naive Bayes	75.06%	75.08%	75.06%	75.05%	75.74%	75.78%	75.74%	75.73%	79.87%	79.93%	79.87%	79.85%
Rbf SVM	52.63%	55.56%	50.00%	52.63%	53.57%	68.18%	44.12%	53.57%	52.52%	52.11%	53.62%	52.86%
Early Fusion	87.30%	88.84%	85.35%	87.03%	89.43%	89.98%	88.69%	89.32%	89.73%	89.61%	89.96%	89.77%
Late Fusion	88.43%	88.64%	88.23%	88.41%	90.40%	**91.78%**	88.81%	90.24%	89.15%	90.63%	87.34%	88.94%
MFB	84.00%	84.87%	82.94%	83.82%	87.45%	88.51%	86.06%	87.26%	89.79%	90.35%	89.16%	89.73%
Text Embedding: IndicBERT with Emotions and Sentiments												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	82.66%	82.76%	82.66%	82.65%	83.08%	83.14%	83.08%	83.07%	85.26%	85.26%	85.26%	85.26%
AdaBoost	84.49%	84.49%	84.49%	84.49%	86.25%	86.25%	86.25%	86.25%	86.55%	86.56%	86.55%	86.55%
QDA	75.13%	75.17%	75.13%	75.12%	70.66%	80.00%	70.66%	68.30%	67.65%	78.82%	67.65%	64.32%
Naive Bayes	70.24%	73.50%	70.24%	69.26%	70.02%	73.20%	70.02%	69.03%	71.31%	74.69%	71.31%	70.37%
Rbf SVM	54.66%	53.66%	55.70%	54.66%	55.06%	54.79%	51.28%	52.98%	55.64%	64.15%	45.95%	53.54%
Early Fusion	87.84%	88.88%	86.67%	87.69%	89.50%	89.76%	89.32%	89.47%	90.01%	91.07%	88.74%	89.87%
Late Fusion	88.68%	88.73%	88.60%	88.66%	90.49%	91.10%	89.85%	90.45%	90.22%	90.93%	89.48%	90.13%
MFB	84.76%	85.50%	83.80%	84.61%	88.13%	90.18%	85.57%	87.80%	89.36%	90.61%	87.91%	89.18%

TABLE VI

mBERT RESULTS FOR CONFIGURATIONS COMBINING TEXT WITH IMAGE EMBEDDINGS: TEXT AND IMAGE, TEXT WITH IMAGE AND EMOTIONS, TEXT WITH IMAGE AND SENTIMENTS, AND TEXT WITH IMAGE, SENTIMENTS, AND EMOTIONS.

Text Embedding mBERT												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	86.06%	87.70%	83.58%	85.59%	86.36%	87.72%	84.27%	85.96%	88.46%	90.61%	85.58%	88.03%
AdaBoost	85.26%	86.18%	83.65%	84.90%	85.79%	86.62%	84.35%	85.47%	86.13%	86.77%	84.97%	85.86%
QDA	86.36%	85.61%	87.12%	86.36%	85.94%	81.79%	92.14%	86.66%	82.89%	76.82%	**93.75%**	84.44%
Naive Bayes	81.36%	84.31%	76.64%	80.29%	81.47%	84.88%	76.18%	80.29%	84.11%	89.44%	77.02%	82.77%
Rbf SVM	49.66%	51.28%	52.63%	51.95%	50.39%	46.27%	53.45%	49.60%	50.66%	50.70%	47.37%	48.98%
Early Fusion	89.98%	91.42%	88.27%	89.80%	90.98%	93.57%	88.11%	90.70%	91.56%	92.74%	90.25%	91.44%
Late Fusion	87.24%	88.19%	86.10%	87.09%	90.19%	91.14%	89.02%	90.06%	89.76%	89.64%	89.95%	89.79%
MFB	85.01%	86.44%	83.15%	84.72%	88.43%	90.25%	86.28%	88.17%	90.92%	91.54%	90.24%	90.86%
Text Embedding: mBERT with Emotions												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	87.51%	87.52%	87.51%	87.51%	87.70%	87.70%	87.70%	87.70%	90.03%	90.04%	90.03%	90.03%
AdaBoost	88.01%	88.01%	88.01%	88.01%	88.08%	88.11%	88.08%	88.08%	87.85%	87.88%	87.85%	87.85%
QDA	78.80%	82.85%	78.80%	78.17%	72.12%	80.37%	72.12%	70.18%	69.82%	80.39%	69.82%	67.07%
Naive Bayes	69.40%	70.52%	69.40%	69.04%	69.14%	71.44%	69.14%	68.36%	79.72%	80.21%	79.72%	79.65%
Rbf SVM	52.50%	45.71%	62.75%	52.89%	52.59%	55.56%	49.18%	52.17%	52.87%	53.49%	57.50%	55.42%
Early Fusion	90.97%	91.50%	90.54%	90.91%	91.89%	**95.00%**	88.72%	91.59%	92.02%	93.73%	90.10%	91.77%
Late Fusion	90.88%	89.33%	92.90%	91.07%	92.15%	92.24%	92.11%	92.14%	91.95%	92.42%	91.56%	91.90%
MFB	87.32%	88.72%	85.54%	87.08%	89.06%	89.83%	88.12%	88.95%	91.52%	93.35%	89.42%	91.33%
Text Embedding: mBERT with Sentiments												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	81.82%	81.87%	81.82%	81.82%	88.16%	88.16%	88.16%	88.16%	90.30%	90.30%	90.30%	90.30%
AdaBoost	84.76%	84.76%	84.76%	84.76%	88.66%	88.70%	88.66%	88.65%	87.85%	87.88%	87.85%	87.85%
QDA	75.17%	79.87%	75.17%	74.22%	74.29%	81.36%	74.29%	72.84%	69.90%	80.42%	69.90%	67.17%
Naive Bayes	75.06%	75.08%	75.06%	75.05%	69.33%	70.53%	69.33%	68.93%	79.72%	80.21%	79.72%	79.65%
Rbf SVM	52.38%	48.61%	60.34%	53.85%	52.52%	52.11%	53.62%	52.86%	52.29%	51.81%	56.58%	54.09%
Early Fusion	91.05%	91.37%	90.88%	91.02%	**92.88%**	94.64%	90.93%	92.71%	92.31%	92.68%	91.98%	92.30%
Late Fusion	91.00%	91.33%	90.59%	90.94%	92.13%	93.37%	90.73%	92.02%	91.73%	92.19%	91.26%	91.70%
MFB	85.90%	86.18%	85.55%	85.83%	89.03%	90.08%	87.69%	88.86%	91.50%	93.27%	89.52%	91.34%
Text Embedding: mBERT with Emotions and Sentiments												
Image Embeds:	**VGG-19**				**RESNET-50**				**ViT**			
Classifiers: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	82.66%	82.76%	82.66%	82.65%	87.97%	87.97%	87.97%	87.97%	90.15%	90.15%	90.15%	90.14%
AdaBoost	84.49%	84.49%	84.49%	84.49%	88.66%	88.67%	88.66%	88.65%	87.82%	87.82%	87.82%	87.81%
QDA	75.13%	75.17%	75.13%	75.12%	74.90%	81.85%	74.90%	73.53%	70.02%	80.47%	70.02%	67.32%
Naive Bayes	70.24%	73.50%	70.24%	69.26%	69.63%	72.85%	69.63%	68.61%	70.93%	74.42%	70.93%	69.94%
Rbf SVM	54.61%	51.90%	56.94%	54.30%	54.96%	48.39%	62.50%	54.55%	55.80%	50.68%	59.68%	54.81%
Early Fusion	91.03%	91.10%	91.03%	91.02%	92.60%	94.53%	90.44%	92.42%	91.78%	92.07%	91.79%	91.80%
Late Fusion	90.53%	88.55%	93.07%	90.72%	92.83%	92.64%	93.20%	**92.88%**	92.27%	92.01%	92.74%	92.31%
MFB	86.26%	88.69%	83.12%	85.78%	88.72%	90.52%	86.53%	88.46%	91.53%	93.15%	89.72%	91.38%

TABLE VII

RESULTS FOR TEXT ONLY AND IMAGE ONLY WITH VARIOUS EMBEDDINGS COMBINATION

Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	65.11%	62.75%	77.29%	69.26%	67.96%	77.24%	71.43%	74.22%	67.51%	58.77%	69.07%	63.51%
AdaBoost	61.15%	67.30%	64.25%	65.74%	63.04%	77.88%	50.29%	61.11%	65.83%	65.99%	62.18%	64.03%
QDA	64.78%	73.68%	67.38%	70.39%	67.45%	69.06%	60.10%	64.27%	64.98%	84.00%	63.32%	72.21%
Naive Bayes	54.48%	55.91%	60.29%	58.02%	58.54%	54.73%	58.20%	56.41%	57.36%	42.24%	59.65%	49.45%
Rbf SVM	55.10%	52.00%	56.52%	54.17%	64.51%	70.70%	65.68%	68.10%	59.17%	55.68%	62.02%	58.68%
DL	69.23%	85.22%	70.50%	**77.17%**	**70.38%**	66.12%	**88.97%**	75.86%	70.11%	79.14%	66.67%	72.37%
Image Embeddings: →	**Resnet-50**				**VGG-19**				**Vision-Transformer**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	51.07%	51.77%	47.58%	49.59%	51.15%	51.72%	46.82%	49.15%	64.44%	78.39%	69.29%	73.56%
AdaBoost	50.61%	51.21%	49.55%	50.36%	50.34%	50.72%	53.56%	52.10%	64.11%	49.60%	83.78%	62.31%
QDA	49.77%	50.91%	18.96%	27.63%	51.26%	51.54%	55.76%	53.57%	62.38%	65.27%	52.91%	58.45%
Naive Bayes	51.95%	52.84%	46.45%	49.44%	50.57%	51.67%	30.53%	38.38%	58.63%	68.56%	63.03%	65.68%
Rbf SVM	51.22%	44.78%	56.60%	50.00%	51.97%	53.23%	50.77%	51.97%	51.66%	46.58%	50.00%	48.22%
DL	65.92%	66.67%	74.59%	70.41%	67.71%	75.29%	68.59%	71.78%	68.53%	**91.03%**	56.35%	69.61%

passed through dense layers (512, 256, 128, 64, 32, 16). The model was trained with the Adam optimizer and binary cross-entropy loss over 10 epochs with a batch size of 32. *Baseline scenarios of image and text-only embeddings were also considered.*

To explore the impact of sentiments and emotions, we applied the same fusion techniques to combine sentiment/emotion embeddings with text and image embeddings, using a similar architecture to perform these experiments. Sentiment and emotion embeddings were generated using the same models applied to the text and then processed through a BiLSTM layer. These embeddings were fused and concatenated with the text and image embeddings, after which the model was trained on the combined representation.

V. RESULTS AND ANALYSIS

We discuss the findings and the key insights from our various experiments. Please note that the bold values indicate the best results for the metric and the combinations. For Table IV, fusion methods, particularly Early Fusion and MFB, consistently yield the highest accuracy, precision, recall, and F1 scores across all image embeddings (VGG-19, RESNET-50, and ViT). Early Fusion outperforms other methods, especially with RESNET-50 and ViT, showing its effectiveness in integrating text and image data. The addition of emotion and sentiment embeddings further improves classification performance, especially in bias-free multimodal datasets. Traditional classifiers like RBF SVM and Naive Bayes underperform, highlighting the limitations of these methods with complex, emotion-infused data. Fusion-based classifiers like AdaBoost, however, demonstrate robust performance. Overall, the results emphasize the importance of fusion strategies, with Early Fusion and MFB being the most effective.

In Table V, both early and late fusion methods produce strong performance, with Early Fusion generally outperforming Late Fusion, particularly with AdaBoost. ViT embeddings show superior performance in capturing detailed image features. Emotion-based embeddings enhance performance, with Late Fusion slightly outperforming Early Fusion when paired with ViT. Sentiment-based embeddings further improve accuracy and F1-score, with ViT and AdaBoost showing the most significant improvements. Simpler classifiers like RBF SVM struggle, while fusion approaches like MFB still provide solid results, bridging traditional machine learning and deep learning methods. Overall, incorporating emotion and sentiment embeddings with fusion techniques enhances multimodal classification.

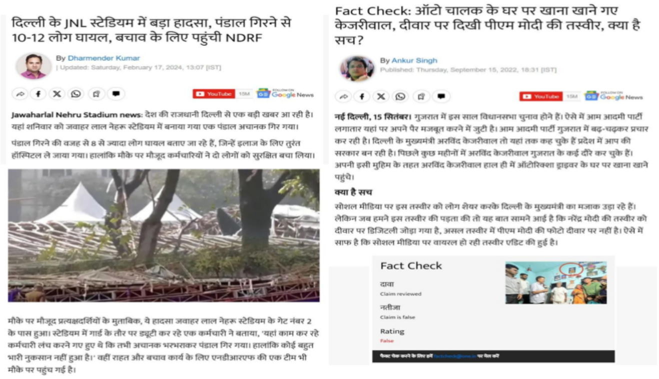
For Table VI, combining mBERT with various image embeddings significantly boosts performance, with Early Fusion again outperforming other configurations across all metrics. The addition of emotional and sentiment embeddings improves accuracy, precision, and recall, particularly with RESNET-50 and ViT. Early Fusion excels in all configurations, but Late Fusion performs well, particularly in recall. Simpler classifiers like Naive Bayes and RBF SVM struggle, reinforcing the need for more complex classifiers in multimodal tasks. Overall, the combination of IndicBERT with emotion and sentiment embeddings, along with Early Fusion, significantly enhances multimodal classification performance. RESNET-50 generally outperforms other image models, although ViT shows competitive results in Early Fusion setups.

Finally, to examine the impact of unimodal approaches, we conducted separate experiments using only text embeddings and only image embeddings, without any emotional or sentiment inputs. As shown in Table VII, the performance of unimodal models is noticeably lower than that of multimodal models. This also supports the assertion that image cleansing does not introduce bias, as the results remain within a moderate accuracy range; if bias were present, the accuracy would likely be much higher. These findings confirm that multimodality significantly enhances misinformation detection for our proposed dataset.

Thus, we can conclude that for classifying multimodal misinformation, integrating text and image data with sentiment and emotion embeddings yields the best performance. This approach effectively captures the unique relationships between textual and visual elements, while the added emotional and sentiment context enhances the model's ability to understand and classify complex, emotionally charged content.

VI. Error Analysis and Conclusion

Error Analysis and Limitations: Given the outstanding performance of the Early Fusion technique with ViT image and IndicBART text embeddings, we conduct an error analysis of false positives and false negatives. Figure 8 examines cases where genuine articles are misclassified as misinformation (false positives) and actual misinformation is overlooked (false negatives). This analysis uncovers misclassification patterns, identifies biases, and informs model improvement strategies. In the false positive sample, the introductory line of the news article mentions a significant event from the capital of India, New Delhi, followed by a discussion of injuries. However, it also states that there were not many losses incurred. This contradictory nature may have made the model classify it as misinformation despite being genuine news. In the false negative sample, the vital misinformation regarding an edited photograph of Prime Minister Narendra Modi is overshadowed by the main context of the news, which is Chief Minister Arvind Kejriwal's visit to an autorickshaw driver's house. In cases where the main context dominates, the model may classify it as a false negative, overlooking the vital misinformation present.



दिल्ली के JNL स्टेडियम में बड़ा हादसा, पंडाल गिरने से 10-12 लोग घायल, बचाव के लिए पहुंची NDRF

By Dharmender Kumar
| Updated: Saturday, February 17, 2024, 13:07 [IST]

Jawaharlal Nehru Stadium news: देश की राजधानी दिल्ली से एक बड़ी खबर आ रही है। यहां शनिवार को जवाहर लाल नेहरू स्टेडियम में बनाया गया एक पंडाल अचानक गिर गया।

पंडाल गिरने की वजह से 8 से ज्यादा लोग घायल बताए जा रहे हैं, जिन्हें इलाज के लिए तुरंत हॉस्पिटल ले जाया गया। हालांकि मौके पर मौजूद कर्मचारियों ने दो लोगों को सुरक्षित बचा लिया।

मौके पर मौजूद प्रत्यक्षदर्शियों के मुताबिक, ये हादसा जवाहर लाल नेहरू स्टेडियम के गेट नंबर 2 के पास हुआ। स्टेडियम में गार्ड के तौर पर ड्यूटी कर रहे एक कर्मचारी ने बताया, 'यहां काम कर रहे कर्मचारी लंच करने गए हुए थे कि तभी अचानक भरभराकर पंडाल गिर गया। हालांकि कोई बहुत भारी नुकसान नहीं हुआ है।' वहीं राहत और बचाव कार्य के लिए एनडीआरएफ की एक टीम भी मौके पर पहुंच गई है।

Fact Check: ऑटो चालक के घर पर खाना खाने गए केजरीवाल, दीवार पर दिखी पीएम मोदी की तस्वीर, क्या है सच?

By Ankur Singh
Published: Thursday, September 15, 2022, 18:31 [IST]

नई दिल्ली, 15 सितंबर। गुजरात में इस साल विधानसभा चुनाव होने हैं। ऐसे में आम आदमी पार्टी लगातार यहां पर अपने पैर मजबूत करने में जुटी है। आम आदमी पार्टी गुजरात में बढ़-चढ़कर प्रचार कर रही है। दिल्ली के मुख्यमंत्री अरविंद केजरीवाल तो यहां तक कह चुके हैं प्रदेश में आप की सरकार बन रही है। पिछले कुछ महीनों में अरविंद केजरीवाल गुजरात के कई दौरे कर चुके हैं। अपनी इसी मुहिम के तहत अरविंद केजरीवाल हाल ही में ऑटोरिक्शा ड्राइवर के घर पर खाना खाने पहुंचे।

क्या है सच

सोशल मीडिया पर इस तस्वीर को लोग शेयर करके दिल्ली के मुख्यमंत्री का मजाक उड़ा रहे हैं। लेकिन जब हमने इस तस्वीर की पड़ता की तो यह बात सामने आई है कि नरेंद्र मोदी की तस्वीर को दीवार पर डिजिटली जोड़ा गया है, असल तस्वीर में पीएम मोदी की फोटो दीवार पर नहीं है। ऐसे में साफ है कि सोशल मीडिया पर वायरल हो रही तस्वीर एडिट की हुई है।

Fact Check

दावा
Claim reviewed

नतीजा
Claim is false

Rating
False

फैक्ट चेक करने के लिए हमें factcheck@one.in पर मेल करें

Fig. 8. Error Analysis: False Positive (left) and False Negative (right) Samples

Conclusion and Future Works: Our study presents a Hindi multimodal dataset for misinformation detection. Multiple ML and DL-based experiments using attention-based embeddings show the effectiveness of combining text and visuals for accurate classification. Our findings highlight the benefits of multimodal fusion techniques and contribute valuable resources to misinformation detection research. Future research should extend the dataset to include modalities like audio or video for more comprehensive misinformation detection. Developing multimodal datasets and models for under-resourced languages is crucial for combating misinformation, offering valuable opportunities to enhance detection systems' effectiveness.

VII. Experiments on Original Biased Dataset

This section presents supplementary experiments conducted on the original biased dataset, where inherent veracity indicators may influence the classification process. The results, showing an accuracy range of 95% to 97% for both traditional machine learning algorithms and deep learning methods in Tables VIII, underscore the significant impact of these biases on model performance. Such high accuracy suggests that the model detects biased patterns rather than genuinely identifying misinformation, thus highlighting the need for debiasing misinformation detection. The methodology and hyperparameter values used are consistent with those applied to the bias-free dataset, reinforcing the critical importance of bias removal for achieving fair and accurate performance across diverse datasets.

TABLE VIII
RESULTS OF MACHINE LEARNING AND DEEP LEARNING APPROACHES ON THE ORIGINAL SCRAPED DATASET.

Machine Learning Approach												
Image Embedding: Resnet-50												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	91.29%	97.84%	84.77%	90.84%	88.96%	90.94%	87.06%	88.96%	94.73%	96.81%	92.03%	94.36%
AdaBoost	95.23%	95.90%	94.67%	95.28%	89.53%	90.73%	88.56%	89.63%	94.69%	94.78%	94.10%	94.44%
QDA	82.85%	**99.00%**	66.99%	79.91%	81.32%	96.80%	65.59%	78.20%	87.20%	80.41%	96.89%	87.88%
Naive Bayes	85.37%	87.40%	83.27%	85.29%	86.75%	89.92%	83.40%	86.53%	93.28%	94.04%	91.79%	92.90%
Rbf SVM	55.23%	53.22%	99.92%	69.45%	55.12%	53.22%	**100.00%**	69.47%	51.18%	49.53%	**100.00%**	66.24%
Image Embedding: VGG-19												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	92.13%	98.26%	85.88%	91.65%	88.92%	88.08%	89.32%	88.70%	94.73%	96.99%	92.45%	94.66%
AdaBoost	94.31%	94.65%	94.00%	94.32%	88.88%	88.37%	88.85%	88.61%	94.61%	94.71%	94.64%	94.67%
QDA	83.08%	98.13%	67.65%	80.09%	86.48%	85.28%	87.28%	86.27%	86.78%	97.85%	75.53%	85.25%
Naive Bayes	69.71%	65.23%	85.19%	73.89%	67.57%	64.71%	73.39%	68.78%	90.57%	89.68%	91.92%	90.79%
Rbf SVM	55.04%	52.81%	**100.00%**	69.12%	51.60%	50.14%	**100.00%**	66.79%	54.28%	52.52%	**100.00%**	68.87%
Image Embedding: Vision-Transformer												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	90.11%	94.19%	85.23%	89.48%	89.88%	92.14%	87.66%	89.84%	95.07%	97.48%	92.45%	94.90%
AdaBoost	93.62%	93.31%	93.81%	93.56%	88.31%	88.91%	88.11%	88.50%	94.23%	94.15%	94.22%	94.19%
QDA	89.19%	95.74%	81.75%	88.19%	79.83%	95.30%	63.65%	76.32%	91.60%	94.47%	88.21%	91.24%
Naive Bayes	78.69%	78.73%	77.88%	78.30%	78.92%	80.64%	77.26%	78.92%	91.98%	93.24%	90.37%	91.78%
Rbf SVM	54.24%	51.91%	**100.00%**	68.34%	54.81%	53.06%	**100.00%**	69.33%	53.97%	51.86%	**100.00%**	68.30%
Baseline: Image Only Embeddings												
Image Embeddings: →	**Resnet-50**				**VGG-19**				**Vision-Transformer**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	92.48%	93.45%	92.08%	92.36%	92.06%	92.19%	91.91%	92.01%	95.13%	95.36%	94.95%	95.09%
AdaBoost	93.87%	93.83%	93.88%	93.85%	90.53%	90.51%	90.48%	90.49%	93.73%	93.71%	93.72%	93.71%
QDA	83.29%	83.23%	83.23%	83.23%	83.98%	83.93%	83.94%	83.93%	87.19%	87.13%	87.19%	87.15%
Naive Bayes	70.20%	72.46%	71.02%	69.89%	64.48%	65.55%	65.07%	64.34%	92.20%	92.18%	92.32%	92.19%
Rbf SVM	53.34%	76.57%	50.44%	35.58%	53.20%	76.54%	50.30%	35.26%	53.34%	76.57%	50.44%	35.58%
Baseline: Text Only Embeddings												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	91.29%	87.67%	74.92%	80.85%	88.69%	89.73%	87.39%	88.54%	93.26%	94.63%	91.84%	93.21%
AdaBoost	92.23%	92.91%	91.70%	92.30%	87.55%	87.43%	87.70%	87.57%	94.23%	93.63%	94.98%	94.30%
QDA	89.53%	98.62%	80.57%	88.69%	87.66%	93.09%	81.36%	86.83%	93.89%	98.82%	88.90%	93.59%
Naive Bayes	69.79%	65.79%	84.70%	74.06%	76.13%	74.50%	79.45%	76.89%	91.67%	90.18%	93.61%	91.87%
Rbf SVM	96.91%	97.64%	96.25%	**96.94%**	92.55%	93.86%	91.06%	92.44%	**95.91%**	98.24%	93.54%	95.83%
Deep Learning Approach												
Image Embedding: Resnet-50												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Methodology Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Early Fusion	95.61%	96.28%	**94.88%**	95.58%	90.72%	95.56%	85.41%	90.20%	**95.95%**	97.47%	94.35%	**95.89%**
Late Fusion	95.87%	**97.62%**	94.04%	95.80%	92.13%	91.37%	93.05%	92.20%	92.55%	95.96%	88.85%	92.26%
MFB	93.70%	95.54%	91.67%	93.57%	85.52%	85.88%	85.03%	85.45%	90.76%	92.99%	88.16%	90.51%
Image Embedding: VGG-19												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Early Fusion	**96.26%**	96.90%	**95.57%**	**96.23%**	88.04%	93.38%	81.89%	87.26%	95.11%	94.84%	95.42%	95.13%
Late Fusion	94.65%	**98.11%**	91.06%	94.45%	87.43%	89.58%	84.72%	87.08%	93.01%	97.39%	88.39%	92.67%
MFB	92.36%	95.27%	89.15%	92.11%	81.36%	85.12%	76.01%	80.31%	85.75%	86.11%	85.26%	85.68%
Image Embedding: Vision-Transformer												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Early Fusion	**97.40%**	**98.21%**	96.56%	**97.38%**	92.21%	89.32%	95.87%	92.48%	96.33%	97.94%	94.65%	96.27%
Late Fusion	96.56%	97.88%	95.19%	96.51%	50.27%	50.19%	71.12%	58.85%	92.90%	95.17%	90.37%	92.71%
MFB	96.72%	97.96%	95.42%	96.67%	91.83%	94.91%	88.39%	91.53%	96.49%	95.79%	**97.25%**	96.51%
Baseline: Image Only Embeddings												
Image Embeddings: →	**Resnet-50**				**VGG-19**				**Vision-Transformer**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Deep Learning Based	**86.55%**	89.06%	**83.35%**	**86.11%**	80.60%	82.12%	78.23%	80.13%	83.23%	**89.98%**	74.79%	81.69%
Baseline: Text Only Embeddings												
Text Embeddings: →	**IndicBART**				**IndicBERT**				**mBERT**			
Classifiers Used: ↓	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Deep Learning Based	90.20%	**92.16%**	87.99%	**90.02%**	**91.32%**	85.50%	**91.79%**	88.54%	89.38%	83.37%	88.40%	85.81%

REFERENCES

[1] S. S. Sundar, "Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads," *Journalism & Mass Communication Quarterly*, vol. 77, no. 3, pp. 480–499, 2000.

[2] L. Liang and Y. Yao, "The influence of multimodality in the digital era for teaching and learning english as a second language," in *2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*. Atlantis Press, 2023, pp. 361–368.

[3] P. Falkowski-Gilski, "On the consumption of multimedia content using mobile devices: A year to year user case study," *Archives of Acoustics*, vol. 45, no. 2, pp. 321–328, 2020.

[4] T. Laor and Y. Galily, "Who's clicking on on-demand? media consumption patterns of generations y & z," *Technology in Society*, vol. 70, p. 102016, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160791X22001579

[5] I. de Saint-Georges and J.-J. Weber, *Multilingualism and multimodality: Current challenges for educational studies*. Springer Science & Business Media, 2013.

[6] R. Dabre, C. Chu, and A. Kunchukuttan, "A survey of multilingual neural machine translation," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–38, 2020.

[7] E. Erdem, M. Kuyu, S. Yagcioglu, A. Frank, L. Parcalabescu, B. Plank, A. Babii, O. Turuta, A. Erdem, I. Calixto *et al.*, "Neural natural language generation: A survey on multilinguality, multimodality, controllability and learning," *Journal of Artificial Intelligence Research*, vol. 73, pp. 1131–1207, 2022.

[8] H. Oh, K.-Y. Goh, and T. Q. Phan, "Are you what you tweet? the impact of sentiment on digital news consumption and social media sharing," *Information Systems Research*, vol. 34, no. 1, pp. 111–136, 2023.

[9] C. U. Huh and H. W. Park, "Setting the public sentiment: Examining the relationship between social media and news sentiments," *Systems*, vol. 12, no. 3, p. 105, 2024.

[10] E. H. Park and V. C. Storey, "Emotion ontology studies: A framework for expressing feelings digitally and its application to sentiment analysis," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–38, 2023.

[11] R. Alatrash, R. Priyadarshini, H. Ezaldeen, and A. Alhinnawi, "Augmented language model with deep learning adaptation on sentiment analysis for e-learning recommendation," *Cognitive Systems Research*, vol. 75, pp. 53–69, 2022.

[12] M. Polignano, F. Narducci, M. de Gemmis, and G. Semeraro, "Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors," *Expert Systems with Applications*, vol. 170, p. 114382, 2021.

[13] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong *et al.*, "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.

[14] S. Yin, C. Fu, S. Zhao, K. Li, X. Sun, T. Xu, and E. Chen, "A survey on multimodal large language models," *arXiv preprint arXiv:2306.13549*, 2023.

[15] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, and D. Roth, "Recent advances in natural language processing via large pre-trained language models: A survey," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–40, 2023.

[16] M. Westerlund, "The emergence of deepfake technology: A review," *Technology innovation management review*, vol. 9, no. 11, 2019.

[17] G. Julian, "What are the most spoken languages in the world," *Retrieved May*, vol. 31, no. 2020, p. 38, 2020.

[18] R. Wang, Y. He, J. Xu, and H. Zhang, "Fake news or bad news? toward an emotion-driven cognitive dissonance model of misinformation diffusion," *Asian Journal of Communication*, vol. 30, no. 5, pp. 317–342, 2020.

[19] C. Martel, G. Pennycook, and D. G. Rand, "Reliance on emotion promotes belief in fake news," *Cognitive research: principles and implications*, vol. 5, pp. 1–20, 2020.

[20] P. Ekman, "Are there basic emotions?" *Psychological Review*, vol. 99, no. 3, 1992.

[21] D. Kakwani, A. Kunchukuttan, S. Golla, G. N.C., A. Bhattacharyya, M. M. Khapra, and P. Kumar, "IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages," in *Findings of EMNLP*, 2020.

[22] R. Dabre, H. Shrotriya, A. Kunchukuttan, R. Puduppully, M. M. Khapra, and P. Kumar, "Indicbart: A pre-trained model for indic natural language generation," *arXiv preprint arXiv:2109.02903*, 2021.

[23] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018. [Online]. Available: http://arxiv.org/abs/1810.04805

[24] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, and P. Vajda, "Visual transformers: Token-based image representation and processing for computer vision," *arXiv preprint arXiv:2006.03677*, 2020.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[27] K. Gadzicki, R. Khamsehashari, and C. Zetzsche, "Early vs late fusion in multimodal convolutional neural networks," in *2020 IEEE 23rd international conference on information fusion (FUSION)*. IEEE, 2020, pp. 1–6.

[28] N. L. Ignazio Gallo, Gianmarco Ria and R. L. Grassa, "Image and text fusion for upmc food-101 using bert and cnns," in *International Conference on Image and Vision Computing New Zealand (IVCNZ 2020)*, Nov 2020, pp. 1–6.

[29] Z. Yu, J. Yu, C. Xiang, J. Fan, and D. Tao, "Beyond bilinear: Generalized multimodal factorized high-order pooling for visual question answering," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 12, pp. 5947–5959, 2018.

[30] M. Sinha, S. Sharma, T. Dasgupta, and A. Basu, "New readability measures for bangla and hindi texts," in *Proceedings of COLING 2012: Posters*, 2012, pp. 1141–1150.

[31] W. Lippmann, *Public Opinion: By Walter Lippmann*. Macmillan Company, 1929.

[32] A. L. George, "Propaganda analysis," *Evanston, Illinois*, vol. 30, 1959.

[33] T. Murayama, "Dataset of fake news detection and fact verification: a survey," *arXiv preprint arXiv:2111.03299*, 2021.

[34] A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: a survey of evaluation datasets," *PeerJ Computer Science*, vol. 7, p. e518, 2021.

[35] A. Ullah, A. Das, A. Das, M. A. Kabir, and K. Shu, "A survey of covid-19 misinformation: Datasets, detection techniques and open issues," *arXiv preprint arXiv:2110.00737*, 2021.

[36] S. Raponi, Z. Khalifa, G. Oligeri, and R. Di Pietro, "Fake news propagation: A review of epidemic models, datasets, and insights," *ACM Transactions on the Web (TWEB)*, vol. 16, no. 3, pp. 1–34, 2022.

[37] J.-P. Posadas-Durán, H. Gómez-Adorno, G. Sidorov, and J. J. M. Escobar, "Detection of fake news in a new corpus for the spanish language," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4869–4876, 2019.

[38] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," 2016.

[39] R. Baly, M. Mohtarami, J. Glass, L. Màrquez, A. Moschitti, and P. Nakov, "Integrating stance detection and fact checking in a unified corpus," *arXiv preprint arXiv:1804.08012*, 2018.

[40] T. Elsayed, P. Nakov, A. Barrón-Cedeno, M. Hasanain, R. Suwaileh, G. Da San Martino, and P. Atanasova, "Overview of the clef-2019 checkthat! lab: automatic identification and verification of claims," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings 10*. Springer, 2019, pp. 301–321.

[41] A. Barrón-Cedeno, T. Elsayed, P. Nakov, G. Da San Martino, M. Hasanain, R. Suwaileh, F. Haouari, N. Babulkov, B. Hamdan, A. Nikolov *et al.*, "Overview of checkthat! 2020: Automatic identification and verification of claims in social media," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11*. Springer, 2020, pp. 215–236.

[42] A. E. Lillie, E. R. Middelboe, and L. Derczynski, "Joint rumour stance and veracity prediction," in *Nordic Conference of Computational Linguistics (2019)*. Linköping University Electronic Press, 2019, pp. 208–221.

[43] J. Nørregaard and L. Derczynski, "Danfever: claim verification dataset for danish," in *Proceedings of the 23rd Nordic conference on computational linguistics (NoDaLiDa)*, 2021, pp. 422–428.

[44] M. Bošnjak and M. Karan, "Data set for stance and sentiment analysis from user comments on croatian news," in *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, 2019, pp. 50–55.

[45] M. Bhardwaj, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Hostility detection dataset in hindi," *arXiv preprint arXiv:2011.03588*, 2020.

[46] S. Kumar and T. D. Singh, "Fake news detection on hindi news dataset," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 289–297, 2022.

[47] D. K. Sharma and S. Garg, "Machine learning methods to identify hindi fake news within social-media," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–6.

[48] J. Badam, A. Bonagiri, K. Raju, and D. Chakraborty, "Aletheia: A fake news detection system for hindi," in *Proceedings of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)*, ser. CODS-COMAD '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 255–259. [Online]. Available: https://doi.org/10.1145/3493700.3493736

[49] R. Sharma and A. Arya, "Lfwe: Linguistic feature based word embedding for hindi fake news detection," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 22, no. 6, Jun. 2023. [Online]. Available: https://doi.org/10.1145/3589764

[50] ——, "Mmhfnd: Fusing modalities for multimodal multiclass hindi fake news detection via contrastive learning," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, Aug. 2024. [Online]. Available: https://doi.org/10.1145/3686797

[51] S. Bansal, N. S. Singh, S. S. Dar, and N. Kumar, "Mmcfnd: Multimodal multilingual caption-aware fake news detection for low-resource indic languages," 2024. [Online]. Available: https://arxiv.org/abs/2410.10407

[52] J. C. Reis, P. Melo, K. Garimella, J. M. Almeida, D. Eckles, and F. Benevenuto, "A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections," in *Proceedings of the international AAAI conference on web and social media*, vol. 14, 2020, pp. 903–908.

[53] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 795–816.

[54] O. Papadopoulou, M. Zampoglou, S. Papadopoulos, and I. Kompatsiaris, "A corpus of debunked and verified user-generated videos," *Online information review*, vol. 43, no. 1, pp. 72–88, 2019.

[55] A. Kazemi, K. Garimella, D. Gaffney, and S. A. Hale, "Claim matching beyond english to scale global fact-checking," *arXiv preprint arXiv:2106.00853*, 2021.

[56] D. K. Sharma and S. Garg, "Ifnd: a benchmark dataset for fake news detection," *Complex & intelligent systems*, vol. 9, no. 3, pp. 2843–2863, 2023.

[57] A. Dhawan, M. Bhalla, D. Arora, R. Kaushal, and P. Kumaraguru, "Fakenewsindia: A benchmark dataset of fake news incidents in india,

collection methodology and impact assessment in social media," *Computer Communications*, vol. 185, pp. 130–141, 2022.

[58] G. K. Shahi and D. Nandini, "Fakecovid–a multilingual cross-domain fact check news dataset for covid-19," *arXiv preprint arXiv:2006.11343*, 2020.

[59] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, "Arcov19-rumors: Arabic covid-19 twitter dataset for misinformation detection," *arXiv preprint arXiv:2010.08768*, 2020.

[60] Y. Li, B. Jiang, K. Shu, and H. Liu, "Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation," *arXiv preprint arXiv:2011.04088*, 2020.

[61] D. Kar, M. Bhardwaj, S. Samanta, and A. P. Azad, "No rumours please! a multi-indic-lingual approach for covid fake-tweet detection," in *2021 grace hopper celebration India (GHCI)*. IEEE, 2021, pp. 1–5.

[62] M. K. Elhadad, K. F. Li, and F. Gebali, "Covid-19-fakes: A twitter (arabic/english) dataset for detecting misleading information on covid-19," in *Advances in Intelligent Networking and Collaborative Systems: The 12th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2020) 12*. Springer, 2021, pp. 256–268.

[63] C. Yang, X. Zhou, and R. Zafarani, "Checked: Chinese covid-19 fake news dataset," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 58, 2021.

[64] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Da San Martino, A. Abdelali, H. Sajjad, K. Darwish *et al.*, "Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 913–922.

[65] B. Lutz, M. Adam, S. Feuerriegel, N. Pröllochs, and D. Neumann, "Which linguistic cues make people fall for fake news? a comparison of cognitive and affective processing," *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. CSCW1, Apr. 2024. [Online]. Available: https://doi.org/10.1145/3641030

[66] R. Kumari, N. Ashok, T. Ghosal, and A. Ekbal, "What the fake? probing misinformation detection standing on the shoulder of novelty and emotion," *Information Processing & Management*, vol. 59, no. 1, p. 102740, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457321002223

[67] B. Lutz, M. T. Adam, S. Feuerriegel, N. Pröllochs, and D. Neumann, "Affective information processing of fake news: Evidence from neurois," *European Journal of Information Systems*, vol. 33, no. 5, pp. 654–673, 2024.

[68] V. Bakir and A. McStay, "Empathic media, emotional ai, and the optimization of disinformation," in *Affective politics of digital media*. Routledge, 2020, pp. 263–279.

[69] B. Ghanem, S. P. Ponzetto, P. Rosso, and F. Rangel, "FakeFlow: Fake news detection by modeling the flow of affective information," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, P. Merlo, J. Tiedemann, and R. Tsarfaty, Eds. Online: Association for Computational Linguistics, Apr. 2021, pp. 679–689. [Online]. Available: https://aclanthology.org/2021.eacl-main.56

[70] K. Nanath, S. Kaitheri, S. Malik, and S. Mustafa, "Examination of fake news from a viral perspective: an interplay of emotions, resonance, and sentiments," *Journal of Systems and Information Technology*, vol. 24, no. 2, pp. 131–155, 2022.

[71] V. Indu and S. M. Thampi, "Misinformation detection in social networks using emotion analysis and user behavior analysis," *Pattern Recognition Letters*, vol. 182, pp. 60–66, 2024.

[72] R. Kumar, B. Goddu, S. Saha, and A. Jatowt, "Silver lining in the fake news cloud: Can large language models help detect misinformation?" *IEEE Transactions on Artificial Intelligence*, 2024.

[73] Z. Liu, T. Zhang, K. Yang, P. Thompson, Z. Yu, and S. Ananiadou, "Emotion detection for misinformation: A review," *Information Fusion*, p. 102300, 2024.

[74] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment analysis for fake news detection," *Electronics*, vol. 10, no. 11, p. 1348, 2021.

[75] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining dual emotion for fake news detection," in *Proceedings of the web conference 2021*, 2021, pp. 3465–3476.

[76] A. Choudhry, I. Khatri, M. Jain, and D. K. Vishwakarma, "An emotion-aware multitask approach to fake news and rumor detection using transfer learning," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 1, pp. 588–599, 2022.

[77] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Information Processing & Management*, vol. 57, no. 2, p. 102025, 2020.

[78] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities," *Expert Systems with Applications*, vol. 153, p. 112986, 2020.

[79] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization branches out*, 2004, pp. 74–81.

[80] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," *IEEE access*, vol. 7, pp. 111 866–111 878, 2019.

[81] W. Graterol, J. Diaz-Amado, Y. Cardinale, I. Dongo, E. Lopes-Silva, and C. Santos-Libarino, "Emotion detection for social robots based on nlp transformers and an emotion ontology," *Sensors*, vol. 21, no. 4, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/4/1322

[82] A. R. Murthy and K. A. Kumar, "A review of different approaches for detecting emotion from text," in *IOP Conference Series: Materials Science and Engineering*, vol. 1110, no. 1. IOP Publishing, 2021, p. 012009.

[83] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and vader sentiment," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 122, no. 16. sn, 2019.

[84] A. Borg and M. Boldt, "Using vader sentiment and svm for predicting customer response sentiment," *Expert Systems with Applications*, vol. 162, p. 113746, 2020.

[85] S. Sohangir, N. Petty, and D. Wang, "Financial sentiment lexicon analysis," in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, 2018, pp. 286–289.

[86] A. Kunchukuttan, "Indic nlp library," https://anoopkunchukuttan.github.io/indic_nlp_library/, accessed: 2024-11-11.

VIII. Biography Section

Raghvendra Kumar is a third-year PhD student in Computer Science and Engineering at Indian Institute of Technology Patna, under the guidance of **Dr. Sriparna Saha**. His research centers on multimodal abstractive summarization, with additional significant involvement in misinformation detection. He has published in notable venues and journals such as CIKM, ICDAR, EACL, IJCNN, IEEE TAI, and IEEE TCSS.

Pulkit Bansal is a fourth-year undergraduate student in the Department of Mathematics and Computing at Indian Institute of Technology Patna. He has conducted his research on misinformation detection under the guidance of **Dr. Sriparna Saha**.

Raunak Kumar Singh is a fourth-year undergraduate student in the Department of Computer Science at Darbhanga College of Engineering, Darbhanga. He has conducted research on misinformation detection during six months internship at Indian Institute of Technology Patna under the guidance of **Dr. Sriparna Saha**.

Dr Sriparna Saha Dr. Sriparna Saha is an Associate Professor at the Indian Institute of Technology, Patna, specializing in Artificial Intelligence, Machine Learning, and Natural Language Processing. She has authored or co-authored over 400 papers. Dr. Saha is a senior IEEE member and a fellow of IET, UK. Among her many honors are the Lt Rashi Roy Memorial Gold Medal from the Indian Statistical Institute, the Google India Women in Engineering Award (2008), the NASI Young Scientist Platinum Jubilee Award (2016), the IEI Young Engineers' Award (2016), the SERB Women in Excellence and Early Career Research Awards (2018), and the Pattern Recognition Letters Editor Award (2023).