What Do You Meme? – Identifying Characteristics and User Perceptions of Suicide Memes in Social Media

Liuliu Chen[1], Jo Robinson[2, 3], Mike Conway[1]

[1]School of Computing and Information Systems, University of Melbourne, Australia
[2]Centre for Youth Mental Health, University of Melbourne, Australia
[3]Orygen, Australia
liuliuc@student.unimelb.edu.au, jo.robinson@orygen.org.au, mike.conway@unimelb.edu.au

Abstract

Suicide memes have recently attracted increasing attention from the public. The visual stimulus and the humorous nature of memes, which seem contradictory to serious suicide-related topics, lead to a lack of understanding of how this new expression differs from general serious suicide notes. Focusing on the `r/SuicideMeme` Reddit community, we present a computational quantitative study to better understand suicide memes. Our findings revealed users' visual preferences for suicide memes: fewer and smaller faces, less colorful, sharper and higher contrast images, along with sadder, more fearful, angrier, and higher-arousal overlaying text. We found that these features impact user interactions to different degrees, with — for example — the existence of faces associated with a 10.6% decrease in the expected number of comments. Using topic modeling, we also discovered that user perceptions of suicide memes in `r/SuicideMeme` were mostly favourable. Further, we developed prediction models to identify suicide memes that achieved an accuracy of 80.04% and an F1 score of 80.75%. To the best of our knowledge, this is the first computational quantitative empirical study on suicide memes on social media. We believe our study provides a deeper understanding of the characteristics and motivations behind suicide memes, and insights for designing more effective moderation strategies for such content.

Content Warning: This paper discusses the sensitive topic of suicide and contains suicide-related content.

Introduction

Social media has received significant attention from researchers over the last two decades and is increasingly utilized to help understand mental health status and outcomes, such as depression, anxiety, and suicidality (Le Glaz et al. 2021). As of July 2024, the number of social media users worldwide had reached 5.17 billion (DataReportal 2024), making up 63.7% of the total global population. The vast and diverse content on social media provides significant potential for researchers to uncover and understand behavioral and linguistic markers of mental health status, which in turn can help cast light on mental health processes and risk factors (Chancellor and De Choudhury 2020).

Social media also benefits from its multimodality, allowing users to express themselves in various forms, such as text, images, videos, and emojis. This rich multimodality not only can capture more nuanced and dynamic emotions but also facilitates communicative interactions in online communities (Jovanovic and Van Leeuwen 2018). Specifically, the Internet meme (meme) is a distinctive format that has emerged in social media, that consists of a visual representation (images, GIFs), typically in the form of a humorous or ironic joke that is recontextualized from popular trends (Gupta et al. 2021). Since memes are highly reproducible, easily adapted, and rapidly disseminated (Davison 2012), they have become pervasive on social media and have evolved into a new and accepted form of communication. More than 60% of social media users share memes (Tama-Rutigliano 2018), and they are even more popular among young people, as they serve as "a vessel of communication, a signifier of the comedic zeitgeist, and a device for channeling the inherent anxieties of youth." (Habib 2020).

Suicide, the third leading cause of death among 15-29-year-olds (WHO 2024), has long been a concerning topic on social media, given the existence of online communities where individuals with suicidal thoughts both share their suicide ideation and seek help (Robinson et al. 2016). While these expressions are often serious as opposed to humorous, we do observe that darkly humorous suicide memes have become a common phenomenon across social media platforms. Here, we define "suicide memes" as internet memes that include elements or content related to suicide in images (whether through visual objects or overlay text), as shown in Figure 1. For instance, the subreddit `r/SuicideMeme`, which was created in 2017 for sharing memes containing suicide-related content, has gained considerable attention. It now reaches 18,000 active members and ranks in the top 5% of Reddit communities by size (Reddit 2024b).

However, despite the growing popularity and impact of suicide memes, current research on this novel content remains limited. Most suicide-related studies on social media are solely focused on text or network features, and the role of visual images has rarely been explored (Badian et al. 2023). Although some studies have included images in developing models for suicide risk detection (Badian et al. 2023; Chatterjee et al. 2022), these studies were usually focused on images more generally rather than specific suicide memes. It is unclear how the use of images for suicide-related communication differs from using text, or how these suicide

Figure 1: Synthetic examples of suicide memes.

images differ from other images in general. Only recently have studies begun to focus on suicide memes, exploring the perceptions and attitudes towards suicidality in the context of meme culture (Smith and Linker 2021; Nicomedes et al. 2024; Weiser and Alam 2022). These studies identified a potential mixed impact of suicide memes. While the public may view suicide memes as harmful content, potentially triggering suicidal ideation, some researchers argue that the humorous nature of memes might diminish their suicidogenic potential (Schott 2023). For example, Nicomedes et al. (2024) have found individuals not experiencing suicidal ideation expressed negative reactions to suicide memes, while those with suicidal ideation found them amusing, soothing or cathartic. However, given that these studies were conducted on relatively small sample sizes through surveys and questionnaires, there remains a gap in understanding and analyzing suicide memes comprehensively.

In addition, gaining a better understanding of the characteristics of suicide memes is important for social media platforms to design effective moderation strategies for suicide-related content. However, current moderation strategies for suicide-related content are often focused on detecting and limiting users' access to what are assumed to be straightforwardly harmful posts (Zhang et al. 2024); active moderation of images is limited, only focusing on images that explicitly indicate suicide risks (e.g., bleeding wounds, cutting wrists), and does not take into account the increasing popularity of suicide memes. In some circumstances, suicide memes, given that they are often dependent on humor and the use of visual metaphors, are more difficult to identify. Thus, it is necessary to investigate the unique characteristics of suicide memes and develop specific detection models for them. A further complicating factor is that suicide memes may have different impacts — helpful or harmful — for different cohorts, and therefore adopting a policy of removing all suicide-related memes has the potential unintended consequence of limiting access to what, for some individuals at least, may be actively helpful content. We need to further understand how suicide memes impact users in order to design more nuanced and dynamic moderation strategies.

Therefore, this study aims to address the above gaps by answering the following research questions:

- **RQ1**: Do the features of suicide and non-suicide memes differ and if so, how?
- **RQ2**: To what extent do features in suicide memes associate with user interactions within the online community?
- **RQ3**: What are the perceptions of and attitudes toward suicide memes within the online community?
- **RQ4**: How accurately can suicide memes be identified among memes?

To address these questions, we created a suicide meme dataset from Reddit subreddit `r/SuicideMeme` (N=1967), and a control group of general memes from `r/memes` (N=2380). We extracted multiple visual and textual features and applied statistical hypothesis tests to discover distinguishable features between the two groups. Further focusing on suicide memes and their online community, we explored how these features impact user interactions, and how users perceive these memes using negative binomial regression models and topic modeling. Finally, we experimented with different machine learning models to predict suicide memes using the extracted features.

Main Findings Overall, we found the following:

- Compared to general memes, suicide memes tend to have lower entropy, fewer faces with smaller regions, but higher sharpness and contrast balance in the images. The text overlaid expressed higher sadness, fear, anger, arousal and word counts (RQ1).
- The presence of faces in suicide memes is significantly associated with a 10.6% decrease in the expected number of comments, and emotions and arousal in the text also contribute to user interactions at different levels (RQ2).
- Individuals within the suicide meme online community held generally favourable views about suicide memes since they allowed them to release their burdens with humor and in a relatable manner, but there were also some negative feelings about these memes (RQ3).
- Our prediction models, which use extracted features for predicting suicide memes among memes, have achieved an accuracy of 80.04% and F1 of 80.75%, outperforming the baseline by 32.82% and 34.07%, respectively (RQ4).

Contributions To the best of our knowledge, this is the first study to use computational methods to understand suicide memes. This work could serve as a basis for further exploration of this relatively new form of suicide-related expression. By uncovering the distinct characteristics and users' perceptions of suicide memes, our findings can provide valuable insights for fostering safer online discussions about suicide, designing more targeted intervention and communication strategies for individuals, and informing the development of effective, nuanced moderation strategies for suicide-related content on social media.

Background and Related Work

Internet Memes

The term "meme" was first proposed by the biologist Richard Dawkins in 1976 as "the unit of cultural transmission that can replicate itself through imitation" in human evolution (Dawkins 2016). More recently, and in the context of social media use Davison (2012) defined memes as "a piece of culture, typically a joke, which gains influence through online transmission". On social media platforms, memes can take many forms and serve various purposes, such as light entertainment, political promotion, or as a vehicle for self-expression (Literat and Kligler-Vilenchik 2019). In the last decade, memes have become increasingly popular, with 75% of 13 to 35-year-olds sharing memes (YPulse 2019), a development that has attracted interest from many research areas such as marketing (Malodia et al. 2022), politics (Dean 2019), and healthcare (Wasike 2022; Langford et al. 2022). Researchers have highlighted its role in collective identity (Dynel and Chovanec 2021). Users in online communities often create memes characterized by humor and inside jokes, which typically contain exclusive community knowledge that is not familiar to general users in an easily identifiable fashion (DeCook 2018). Through memes, users establish a relational identity that bonds with other users in the community communicating "I'm part of this", or "I feel you", thus building a greater attachment and connection to their community (Newton et al. 2022).

Suicide Memes

Suicide memes can be found in multiple formats and social media platforms, typically with humor and irony. Figure 1 shows examples of such suicide memes. They typically include content expressing a desire for suicide, feelings of depression and anxiety, or even imagined scenarios involving specific suicide methods. Suicide memes are common on social media. The subreddit `r/SuicideMeme` is in the top 5% of Reddit communities by size. Even on platforms like Facebook and Instagram, which aim to disable searches for content using suicide-related keywords, there are still numerous pages and accounts sharing suicide memes, many with several thousand followers (Chateau 2020).

Despite the growing presence of suicide memes, there is a significant gap in understanding the sharing and functions of these memes within both suicide memes and general communities. Past studies have explored memes related to mental health issues, such as depression (Akram et al. 2020; Yadav et al. 2023), anxiety (Wasike 2022; Akram et al. 2021), but with little attention given to the possibility that these memes may also be suicide-related. Therefore, these studies often do not explore the specific dynamics of suicide memes, leaving a gap in understanding their unique characteristics and potential impact. Recently, researchers have begun to focus specifically on suicide memes, exploring their impact and how they are perceived. For instance, Weiser and Alam (2022) found that suicide memes might decrease sensitivity to suicide, potentially normalizing the concept among users. Perez (2019) suggested that suicide memes reflect the "antifuture" attitude of those posting them, while Nicomedes et al. (2024) argued that suicide memes can be therapeutic and soothing. These findings suggest a currently unexplored dual role for suicide memes with both negative and positive effects on users.

Suicide Prevention on Social Media

There is much research on suicide prevention on social media, and most relevant to our study is (1) user disclosure and interactions, and (2) detection and moderation.

User Disclosure and Interactions The disclosures and interactions of individuals with mental health issues on social media has received much attention from researchers in recent years (Naslund et al. 2020). Benefiting from the anonymity and accessibility of social media, studies have shown that these platforms offer a rich space for individuals to express negative emotions related to suicide, share their personal experiences and thoughts, and seek support from the community (Davies et al. 2024). However, this online suicide disclosure also carries certain risks. Studies have examined the phenomenon of negative emotion contagion through online interactions (Goldenberg and Gross 2020; Mueller and Abrutyn 2015). For example, a group of users showed a higher level of emotional distress after being exposed to other's suicide attempts and were more likely to report suicidality (Mueller and Abrutyn 2015).

Moreover, the ease with which memes can be rapidly transmitted between users may further exacerbate the contagion of negative emotions (Guadagno et al. 2013). For example, Akil, Ujhelyi, and Logemann (2022) found that exposure to depression memes can increase depressive mood. Given this, it is important to identify and understand distinctions in user disclosures and interactions across different suicide-related online communities. Yet, few studies have focused on online communities for suicide memes, which often present a unique intersection of dark humor and mental health issues. These communities may exhibit different patterns of user interaction and emotional expression.

Detection and Moderation Considering the potentially harmful impact of suicide-related content on social media, researchers are also working on automatically detecting relevant posts and designing corresponding moderation strategies to improve online safety. Researchers have applied natural language processing and machine learning techniques to detect quantifiable characteristics on social media posts and developed models to predict suicide-related content (Ji et al. 2021; Matero et al. 2019; Sawhney et al. 2020). Recently, multimodal approaches including both images and audio were proposed to enhance detection ability (Chatterjee et al. 2022; Badian et al. 2023). Once suicide-related content is detected, the current strategy of most platforms is to limit the user access to such content, either directly deleting them on the platform, or presenting warnings with self-help resources to replace risky content (Zhang et al. 2024).

However, to the best of our knowledge, there are no current detection models or moderation strategies specifically focusing on suicide memes. Compared to general suicide-related images (e.g., self-harm scars), suicidal ideation in suicide memes is often masked by humor, making them harder to detect by existing models. Furthermore, current moderation strategies may not be appropriate for handling suicide memes. While suicide content in suicide memes may negatively impact users (Weiser and Alam 2022), the role of internet memes may also help to build positive connections among individuals experiencing suicidal ideation and reduce suicide risks (Nicomedes et al. 2024; Newton et al. 2022; Schott 2023). As a result, current detection and moderation approaches might be problematic, as they block positive disclosure while failing to minimize harmful content. Thus, it is crucial to better understand suicide memes and design more targeted and dynamic detection and moderation approaches.

Ethics and Responsibility

There are several ethical issues and challenges associated with this work. While the Reddit data is publicly available and our data collection process was consistent with Reddit's terms and conditions (Reddit 2024a), individual consent was not sought from Reddit users. To mitigate this we adopted relevant standards outlined by Benton, Coppersmith, and Dredze (2017) to strengthen the protections for individuals who post to the `r/SuicideMeme` community. These additional protections included storage of data in secure environments, and the use of synthetic quotations in this publication.

While the goal of this work is to advance knowledge regarding suicide prevention, we acknowledge that — as with all research — there is potential for bad actors to utilize our results in ways in which we cannot readily anticipate.

This study was approved by the Human Ethics Committee of the University of Melbourne (No.2024-29894-61458-5).

Dataset

Figure 2 illustrates the data collection process in this study. We selected the `r/SuicideMeme` subreddit on Reddit, an increasingly popular online community that allows users to publicly post suicide memes to express their feelings. We used pushshift.io Reddit API (Baumgartner et al. 2020) to download all submissions from `r/SuicideMeme` and extracted image URLs from each submission. We then downloaded images using the Requests library from image URLs.

Since this study focuses on understanding memes, we applied the following filtering steps to exclude any irrelevant images or memes: (i) the meme must contain both an image and overlaying text; (ii) the text on memes must be in English and readable by OCR; (iii) all the defined features in this study can be extracted from the memes; (iv) identical or near-duplicate memes (same image with fewer than 3-word differences) were removed. After filtering, we obtained 1967 suicide memes as our suicide meme dataset, ranging from April 2018 to December 2022.

To compare the differences between suicide and non-suicide memes, we used `r/memes` subreddit as a control group for collecting non-suicide memes, as the mention of suicide in memes is not allowed based on this subreddit's rules. We applied the same collection steps to download all memes from this subreddit. Since the number of memes collected in `r/memes` is much larger than in `r/SuicideMeme`, we randomly sampled 5000 memes out of all memes, and applied additional filtering rules to match the time range of suicide-meme dataset and confirmed that these memes are non-suicidal (Figure 2). Following these steps, we obtained 2380 memes for the non-suicide memes dataset, ranging from June 2018 to December 2022.

Methodologies

Features Extraction

For each meme, we extracted features from its metadata, visual content and textual content. From metadata, we computed 4 types of features, including *width, height, the number of pixels*, and *the aspect ratio*.

Studies have shown that visual perceptions are highly relevant to mental state perception in various types of psychopathology (Phillips et al. 2003; Green et al. 2008). Recently, a growing body of research has examined the imagery shared on social media. For instance, Bakhshi et al. (2015, 2016) found visual features like exposure, entropy, and color palette contribute to online social engagement and connections. Manikonda and De Choudhury (2017) suggested that minimalist visual cues (e.g., brightness, contrast) selected by individuals may reflect a desire to draw attention to their psychological states in mental health disclosures, which are also known to attract involuntary attention due to visual salience (Prinzmetal, Long, and Leonhardt 2008). Guntuku et al. (2019) further revealed that depressed users preferred images not showing faces, while anxious users usually chose sharper images. In this paper, we build upon the image analysis methods used in these studies to uncover potential involuntary visual attention of users to suicide memes. We computed 10 commonly used visual features based on visual content (pixels), as shown in Table 1.

Features	Description
Asymmetry	Visual imbalance or unevenness of an image, measured by the Euclidean distance between the left and right half of image histograms.
Entropy	Shannon entropy on pixel frequency distribution to quantify image complexity
Brightness	The average value in HSV space
Contrast	The standard deviation of pixel intensities
Contrast balance	The difference between the original and contrast-equalized images to access the evenness of contrast and luminance distribution.
Exposure balance	The absolute skewness of the luminance histogram
Sharpness	How clear the edges and details in an image are, measured by CPBD metric (Sadaka et al. 2008)
Face count	Number of faces contained in the image.
Face region	The ratio of face area to the whole image
Text region	The ratio of text area to the whole image

Table 1: Descriptions of visual features.

To understand the psycho-linguistic patterns associated with the overlaying text on memes, we also extracted textual features using EmoLex (Mohammad and Turney 2013) and NRC-VAD (Mohammad 2018) lexicons, as shown in Table 2. Compared to sentiment analysis methods using deep learning and transformers, We employed lexicon-based methods due to their interpretability, allowing for a more nuanced understanding of word-level emotional expressions. The NRC-VAD and EmoLex lexicons are widely used in studies of mental health discourse (Mohammad 2022), although limitations exist, as they cannot capture the semantic

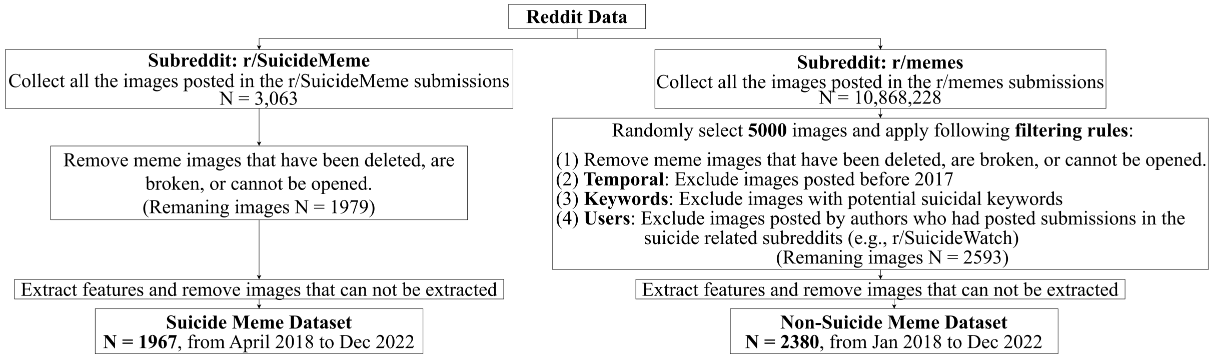


Figure 2: Data collection flowchart.

Features	Methods	Description
Words count	\	The number of words in the overlying text on the meme
Valence	NRC-VAD	Sentiment: negative–positive
Arousal		Excited–calm or active-passive
Dominance		Have full control–no control
Anger	EmoLex: Emotion intensity calculated by lexicons frequencies	Annoyance-anger-rage
Fear		Apprehension–fear–terror
Anticipation		Interest-anticipation–vigilance
Trust		Acceptance–trust–admiration
Surprise		Distraction–suprise–amazement
Sadness		Pensiveness–sadness–grief
Joy		Serenity–joy–ecstacy
Disgust		Boredom–disgust–loathing

Table 2: Descriptions of textual features.

relations or contextual nuances in emotional expressions and often miss slang commonly used on social media.

We employed NRC-VAD lexicons to calculate the scores of word meanings for three dimensions: *valence, dominance, and arousal*, which contains over 20k human-rated English words and is currently the biggest dictionary. EmoLex lexicon was applied to extract the presence of eight basic emotions in the text: *anger, fear, anticipation, trust, surprise, sadness, joy*, and *disgust*, from over 14k English terms collected via crowdsourced human annotation. While NRC-VAD provides intensity-based scores for emotional dimensions, EmoLex captures the presence and frequency of categorical emotional labels.

Statistical Analysis

Based on the extracted 26 features (4 meta-data, 10 visual, and 12 textual), we conducted statistical analysis.

Statistical Hypothesis Tests We applied multiple statistical tests to examine significant differences in features between suicide memes and non-suicide memes, depending on the distribution and variance of feature values. These tests included ANOVA, Kruskal-Wallis, and Mann-Whitney U tests. For each feature, we first determined whether it satisfied the test assumptions (e.g., normal distribution or equal variance), and then selected the statistical test that satisfied all assumptions. A conventional significance threshold of α = .05 was used to determine statistical significance.

Negative Binomial Regression To investigate which features are more likely to influence user interactions with suicide memes, we built two statistical models using defined features as independent variables. The first model used the post scores (the number of upvotes - the number of downvotes) as the dependent variable, while the second used the number of comments. We employed negative binomial regression, a method widely used in social media studies and well-suited for over-dispersed count variables (Bakhshi et al. 2016). We used Pearson's chi-squared test to assess the goodness of both models and applied Benjamini-Hochberg correction to account for multiple comparisons. We then examined the regression coefficients (β) to understand the effect of each independent variable on the scores or comments.

Topic Modeling

To explore how users think about suicide memes within the `r/SuicideMeme` community, we trained a BERTopic model on 6823 comments posted in suicide-meme posts, following Grootendorst (2022)'s methods. To preserve the natural richness of user engagement and ensure finding meaningful topic clusters, we retained all comments without filtering. Unlike traditional topic models like LDA, BERTopic leverages transformer-based models to capture semantics relationships and context, providing potentially more interpretable and dense topic clusters. Although BERTopic assumes each document only contains a single topic, we believe this assumption is appropriate in our case, as comments are typically short and convey a single idea. More detailed sub-model configurations are shown in Appendix A. Notably, although we used ChatGPT to generate initial topic labels for easier interpretability, we manually evaluated each label to ensure alignment with the cluster content and made necessary adjustments to make it more informative.

Since we aimed to explore user attitudes specifically on suicide memes, we manually reviewed each topic and only kept the clusters directly associated with memes or humor. While other topics may also be of interest, they were excluded to maintain focus on understanding suicide memes and the perceptions surrounding them. We then conducted a content analysis on these clusters to gain a deeper understanding of the discussions within them.

Suicide Meme Prediction

To evaluate the effectiveness of the identified features in distinguishing between suicide and non-suicide memes, we conducted experiments using currently popular classification models with different feature combinations.

For baseline models, we trained three widely used pretrained vision models, ViT (Dosovitskiy et al. 2021), RestNet50 (He et al. 2016), and VGG19 (Simonyan and Zisserman 2014), as they are considered to be the most robust state-of-art models for image classification in both Transformer-based and CNN-based architectures (Liu et al. 2024). We trained these models directly using the raw meme images, without any feature extraction, to provide a reference baseline for prediction performance.

We experimented with different combinations of features: visual features only, text features only, and a combination of both. For each combination, we applied various data processing methods, including normalization, standardization, and logarithmic transformation. During training, the datasets were randomly shuffled and split into training, validation, and test sets with a 0.7:0.1:0.2 ratio. For each feature combination, we tested multiple machine learning models, such as Support Vector Classification (SVC), multilayer perceptrons (MLP), and performed a grid search with cross-validation to fine-tune the model parameters using the scikit-learn Python package. To ensure robustness, we performed 10 experiments and evaluated model performance based on the average accuracy and the F1 score.

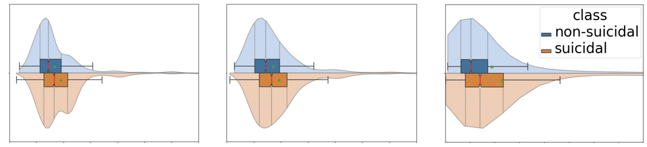
Results

Feature Comparison Between Suicide and Non-Suicide Memes (RQ1)

Out of the 26 features we examined, 23 showed statistically significant differences ($p<.05$). The three features that were not significantly different are *aspect ratio*, *brightness*, *contrast*. For each dimension (meta, visual, textual), we selected the top five features with the most significant differences (based on p-values) for further analysis. The statistical results for all features are presented in Appendix B.

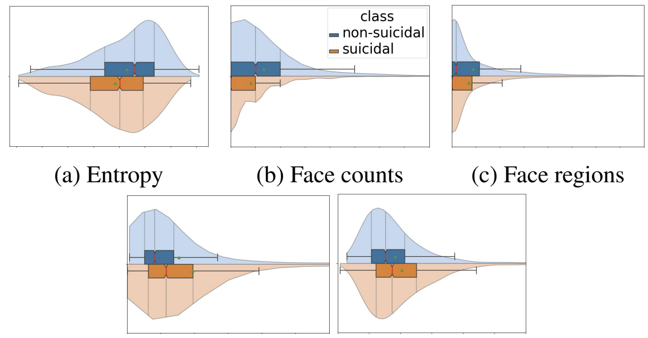
Meta Feature Analysis As shown in Figure 3, suicide memes generally have higher width, height, and pixel numbers, which often relate to image quality. This suggests suicide memes are generally larger and potentially more informative, such as including more textual elements or complex imagery, which might make them more attractive to users.

Visual Features Analysis Among the visual features, *entropy*, *the number of faces*, *face region*, *sharpness*, and *contrast balance* are the five most significantly different features between suicide and non-suicide memes ($p<.001$*** for all 5 features). Suicide memes tend to have lower entropy, which often suggests fewer variations and less intensity in image colors. The sharpness and contrast balance values of suicide memes are significantly higher than the non-



(a) Width (b) Height (c) Number of pixels

Figure 3: Violin plots: meta features.



(a) Entropy (b) Face counts (c) Face regions

(d) Sharpness (e) Contrast balance

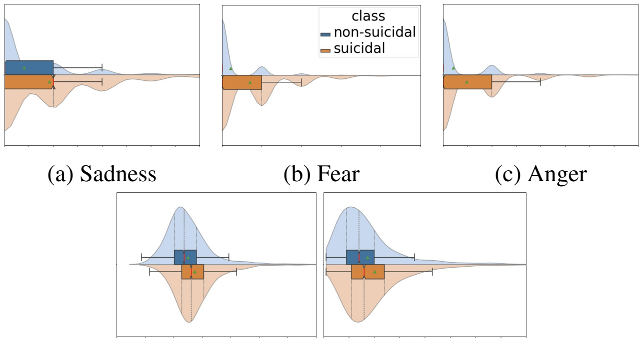
Figure 4: Violin plots: top 5 most significant visual features.

suicide memes. These two features measure the clarity of details and the differences between light and dark areas, which can emphasize textures and create visually strong and dramatic effects. This use of contrast could be a visual metaphor for the extreme emotional states and striking themes behind suicide memes. Interestingly, suicide memes also tend to feature fewer faces and smaller face regions, which might reflect a sense of social isolation or diminished focus on the human element for users who posted suicide memes.

Textual Features Analysis We identified five most significant different textual features are *sadness*, *fear*, *anger*, *arousal*, and *word counts* ($p<.001$*** for all), shown in Figure 5. Compared to general memes, the text on suicide memes often expresses more sadness, fear, and anger, which might indicate the dark and negative feelings of the users who post them. Additionally, the arousal score in suicide memes is generally higher. Previous research has shown that heightened arousal levels can increase the likelihood of suicide (Law and Anestis 2021; Ribeiro et al. 2015). Studies have also identified physiological differences in negative affective states between low arousal (e.g., sadness) and high arousal (e.g., anger) (Marci et al. 2007), which contribute to different levels of pain tolerance (Carter et al. 2002). Suicide memes also tend to have higher word counts, suggesting that these memes may contain more detailed expressions of emotions or more informative messages.

Impact of Suicide Meme Features on User Interactions (RQ2)

We constructed two negative binomial regression models to explore the effect of extracted features on the scores and the number of comments that suicide memes received. Excluding features without significant effects ($p<.05$), Table



(a) Sadness (b) Fear (c) Anger

(d) Arousal (e) Words count

Figure 5: Violin plots: top 5 most significant textual features.

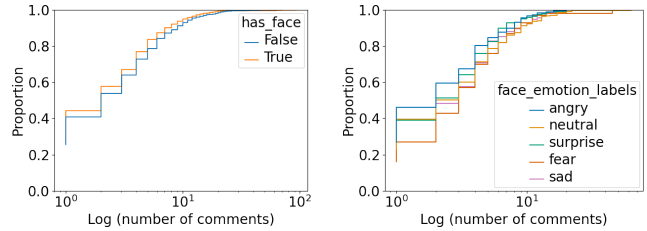
3 shows the two models with features that were significant in at least one of the models. We further compared suicide memes that attracted the highest and lowest number of comments and scores, as shown in Appendix C.

We can see that suicide memes, that include faces, are more likely to decrease the number of received comments by 10.6% (β=-0.112, IRR=0.894). This finding aligns with the results from the previous section, where suicide memes tend to contain fewer faces. We further analyzed the scores and the number of comments across the existence of face and each face emotion category. Figure 6 shows the ECDF plots. We can see that suicide memes without faces receive more comments compared to those with faces, and faces with fear and sadness on suicide memes tend to attract more comments than those with other emotion categories. Other visual features, although they showed significance in the models, their coefficients were too small to have meaningful effects on user interactions and therefore could be neglected.

Regarding the textual features, the emotions of sadness, anger, and fear are positively associated with both score and comment models. Among them, anger has the largest effect on comments (β=0.136, IRR=1.146), while its effect on score model is relevantly small (β=0.067, IRR=1.069). The anticipation emotion is positive correlated with a small increase in the number of comments (β=0.079, IRR=1.082), whereas the surprise emotion links to the decrease the score of suicide memes (β=-0.078, IRR=0.925). Arousal is significantly associated with a 30.1% increase in the likelihood of receiving higher scores (β=0.263, IRR=1.301). These results are consistent with the previous findings, which show suicide memes typically have higher values of sadness, fear, anger, and arousal. However, the impact of the number of words on these two models is very limited and contradictory between the score model and the comment model.

Understanding Perceptions and Attitudes on Suicide Memes (RQ3)

We trained a BERTopic model on 6823 comments associated with suicide meme posts, which resulted in 73 clusters. Appendix D provides an overview of all topic clusters. We manually checked these clusters and only focused on the topics about memes and humor, in which 14 topics were identified.



(a) Face existence (b) Face emotions

Figure 6: ECDF plots: the differences in distributions of the number of comments.

Table 4 presents these topics and their representative text. We further conducted content analysis on each topic cluster and summarized what users frequently discussed about suicide memes: gratitude and positive sentiment; humor and irony; relatability and shared experiences.

Gratitude and Positive Sentiment Many topic clusters demonstrate a positive sentiment toward the posting of suicide memes, particularly in topics 4, 12, and 36. Although general public might think suicide memes are dark and negative, users in these communities find them lighter than their own circumstances, leading to a sense of relief. They often express appreciation for these memes, with some even taking active steps to repost or share them with others.

Humor and Irony The discussion about humor and irony is also a prominent theme in `r/SuicideMeme`, as reflected in topic clusters such as topics 2, 29, 32. While memes are generally humorous, the humor in suicide memes seems darker and more ironic, given that the content and topics are suicide-related and more serious. For example, some users express that these suicide memes make them laugh, while others may find them disturbing, though sometimes in a cathartic way. The irony in these memes might serve as a coping strategy, allowing users to confront their feelings in a way that is "safer" and less overwhelming.

Relatability and Shared Experiences In topic clusters like 8, 16, 19, 25, and 31, the most frequently mentioned phrases are "same", "relatable", "me in real life". These phrases highlight the potential function of suicide memes in enhancing social and emotional connection within the suicidal community, helping individuals with similar mental health issues feel understood and less alone in their struggles. However, negative emotion contagion might also exist. These suicide memes could also be a stimulus that triggers stress or discomfort in some individuals, such as expressions like "I dont like this" or "Don't attack me" in topics 62, 65.

Predicting Suicide Memes (RQ4)

To explore the prediction of suicide memes, we experimented with various feature combinations and preprocessing methods with multiple machine learning algorithms, such as MLP, Logistic Regression, Decision Tree and SVC. We found that SVC showed marginally superior performance over the others. Table 5 presents the prediction per-

		Score model			Comment model		
		β	**std.e**	***p***	β	**std.e**	***p***
(Intercept)		-4.46	0.036	***	-6.01	0.062	***
Meta features	Width	-1.32e-04	7.59e-05	***	-4.02e-06	1.31e-04	–
	Height	-1.58e-04	2.73e-05	***	4.26e-04	1.14e-04	***
	Number of pixels	3.14e-07	1.23e-07	*	-3.65e-07	2.19e-07	.
Visual features	Exposure balance	0.002	4.51e-03	–	0.014	0.006	*
	Asymmetry	1.02e-07	3.58e-07	**	-5.52e-07	6.28e-07	–
	Sharpness	-1.35e-09	5.96e-10	*	1.60e-09	1.03e-09	–
	Contrast balance	-1.88e-07	2.63e-07	***	-3.26e-07	4.52e-07	.
	Has face	-0.058	0.035	–	**-0.112**	0.049	*
	Text region	0.247	0.154	–	**0.619**	0.253	*
Textual features	**Sadness**	0.055	0.032	**	**0.089**	0.025	***
	Anger	**0.067**	0.031	**	**0.136**	0.051	***
	Fear	0.034	0.016	*	**0.099**	0.025	***
	Anticipation	-0.010	0.022	–	0.079	0.035	*
	Surprise	-0.078	0.038	*	-0.147	0.076	–
	Arousal	**0.263**	0.119	*	0.030	0.200	–
	Word counts	-0.005	8.67e-04	***	0.003	0.001	*

Table 3: Negative Binomial Regression for the scores and the number of comments. (*:$p<0.05$, **:$p<0.005$, ***:$p<0.001$)

formances on the test set for different feature combinations. We also reported additional evaluation metrics and hyperparameters of our best-performing model in Appendix E.1.

We found that all feature combinations outperformed the baseline models, suggesting all extracted features are effective in predicting suicide memes. Textual features were more important than visual features for this classification task. Overall, the fusion of visual and textual features provides the best performance with an accuracy of 80.04% and an F1 score of 80.75%, which indicates the complementary information contained in these two modalities.

Qualitative Error Analysis We selected the best-performing model trained on the Reddit dataset (i.e., SVC with visual and textual features) for further analysis. Overall, 65% of misclassified cases are suicide memes that the model failed to identify (See Appendix E.2 for confusion matrix). These suicide memes are typically metaphorical and implicit. For example, in Figure 7 (a), suicide ideation has to be inferred from both text and image context, which the extracted features failed to capture. As shown in Figure 7 (b), the misclassified non-suicide meme typically contains emotional expression related to mental health disorders (e.g., depression) or descriptions of death (e.g., drowning).

Generalization To evaluate the generalization of our findings, we used the Google Search API to collect suicide and non-suicide memes from additional social media platforms: Instagram, Facebook, and X (formerly Twitter) (details in Appendix E.3.1). Following the same methods, we extracted proposed features and conducted statistical analysis (Appendix E.3.2), providing preliminary evidence that most of the identified distinguishing features between suicide and non-suicide memes are consistent across platforms.

Using the same model configurations as the best-performing model trained on the Reddit dataset, we trained three classifiers independently on each dataset. Table 6

(a) Examples of misclassified suicide memes

(b) Examples of misclassified non-suicide memes

Figure 7: Examples of misclassified cases

presents an overview of collected datasets and corresponding model performance (see Appendix E.3.3 for additional evaluation metrics). Overall, the performances on the Instagram and X datasets are similar to those on the Reddit dataset, while performances on the Facebook dataset show some decline. These results provide preliminary evidence suggesting that the identified distinguishing features in our study may be effective for suicide meme identification across platforms. However, further examination is needed to understand the influence of Google's ranking algorithm on the datasets collected via the Google Search API.

Discussion

Principal Findings

This study examined the phenomenon of suicide memes on Reddit. We extracted and identified several distinctive

No.Topic	Count	Topic	Representative text
2	235	Appraisal to humorous suicide memes	"Literally laughed and cried on this meme, though some just think it's fxxked up."; "A joke for the win! That's what I live for."; "I like this suicide meme"
4	191	Gratitude and appreciation	"That is a best! One of the greatest memes I've ever seen so far.", "Thank you so mcuh! Todays been the first day for a long time that ive felt even the lightest bit okay despite the crappy circumstances. It means everything to me."
8	147	Relatable real-life reflections	"Just me rn."; "Me in real life."; "How dare you portray me so accurately!"
12	106	Reposting engagement	"Hey! Just repost..."; "Lolll it deserves a repost"; "I like this sub"
16	102	Shared experiences	"Damn same, I'm not the only one"; "Cool bro, we're the same"; "Same, buddy"
19	124	Resonation to everyday struggle	"It's me daily"; "Me everyday from 12 y.o."; "That is pretty much my everyday"
21	81	Relatable agreement	"This is so realllll and I hate it sm."; "How can this be so true loll"
25	76	Group solidarity	"We are in the same gang"; "I feel the same and like ;-;"
29	71	Humorous laugh reactions	"Wow this is so funny"; "Thanks. That really got me and made me laughing"
31	47	Highly relatable tweets and memes	"Well they're hilarious since they're relatable to me"; "Hmmm, so relatable"
32	75	Humorous Internet memes	"Lmao how can this be so real, thanks."; "I screamed omg hahahahaha"
36	58	Laughter and lighthearted reactions	"Lolll so true. I'm gonna send it to the GC"; "That's exactly why I'm here haha"
62	32	Dislike photo captions	"I don't like this image since I'm in it"; "Im feeling low after a peak seeing this"
65	59	Feeling attacked unexpectedly	"Why'd you have to call me out like that?"; "Don't attack me using this"; "I'm feeling attacked"

Table 4: Topics related to suicide memes (the representative text was synthesized due to ethical considerations but retains the main idea of the original sentences).

Models	Feature Combination	Accuracy(%)	F1(%)
ViT	Raw meme images	57.92	56.97
Resnet50	Raw meme images	56.52	53.09
VGG19	Raw meme images	60.26	60.23
SVC	Visual features	$67.44_{0.0156}$	$66.26_{0.0173}$
SVC	Textual features	$75.40_{0.0144}$	$77.94_{0.0186}$
SVC	Visual+textual features	$80.04_{0.0170}$	$80.75_{0.0100}$

Table 5: Models performance at predicting suicide memes. (Noted accuracy and F1 reported are averaged over 10 experimental runs; subscripts show standard deviation)

	Data Overview		Performances	
	Suicide	Non-suicide	Accuracy(%)	F1(%)
Instagram	594	667	$82.33_{0.0186}$	$82.04_{0.0196}$
Facebook	499	545	$73.68_{0.0140}$	$73.20_{0.0161}$
X/Twitter	104	143	$83.89_{0.0138}$	$82.82_{0.0149}$

Table 6: Data overview and model performances on Instagram, Facebook, and X datasets (accuracy and F1 scores averaged over 10 runs; subscripts indicate standard deviation).

characteristics that differentiate suicide memes from general memes (RQ1). The differences in visual elements might show unintentional preferences of individuals who post suicide memes, potentially relating to their suicide ideation. Overall, these users prefer sharper, higher-contrast, and less colorful memes with fewer faces. This is consistent with existing findings on visual elements in mental health disclosure, which suggest that such preferences may reflect social isolation and a lack of emotional balance (Manikonda and De Choudhury 2017; Guntuku et al. 2019). Moreover, suicide memes typically have higher image quality. The differences in textual features align with previous research showing that suicide-related text generally contains higher levels of sadness, fear, and anger, along with higher arousal scores (Law and Anestis 2021; Coppersmith et al. 2016).

We further identified that the existence of faces and several textual features played a crucial role in influencing user interactions with the community (RQ2). Specifically, having faces on suicide memes is significantly associated with a 10.6% decrease in the number of comments received. Different emotions and arousal in the text also contributed to the scores and comments at various levels. This suggests that while individuals may have certain visual preferences for suicide memes, it might be the contents and themes embedded in the text that drive deeper interactions. Our results indicate that these contents resonate more strongly with other users and make them more likely to give comments to share their thoughts and feelings. However, these interactions always involve suicide-related information, raising the risk of contagiously spreading negative emotions and suicidal thoughts on social media platforms – which may be particularly concerning given that memes can often go viral (Fu et al. 2013). Thus, the dangers of suicide memes can not be ignored, despite their seemingly entertaining form.

In addition, we uncovered user attitudes toward suicide memes (RQ3). The results highlight the potential function of humorous and ironic meme formats in suicide disclosures. Most users expressed a positive attitude toward suicide memes, recognizing their dark humor and irony. Research has suggested that humor could be a coping mechanism for dealing with mental health issues (Henman 2001), and our results support this in the context of suicide. People found suicide memes that are humorous and ironic are less serious and allow them to momentarily disengage from their burdened minds (Goldstein 2013). The sense of feeling relatable and understood also helps them feel less alone and more connected with others. Besides, while not explicitly indicated in the text clusters, the form of memes appears to be more acceptable in public disclosures – less serious but still informative, where we observed two clusters are related to the active reposting and sharing actions of users. This may indicate that engaging with suicide memes may present potential benefits for individuals in suicide communities. However, humor might also have a harmful effect since it could serve to minimize suffering and pain (Goldstein 2013). We also found some topic clusters mentioned feeling slightly triggered as the content hits too close to home.

Finally, combining visual and textual features, our prediction model achieved the best performance in distinguishing suicide memes from general memes (RQ4), with an 80.04% accuracy and an 80.75% F1 score. This outperformed the baseline by 32.82% and 34.07% respectively, which demonstrates informative complementarities of each modality in prediction and their effectiveness. While this approach improves the accuracy of identifying suicide memes and could inform effective moderation, additional factors should be considered, as our findings also suggest that suicide memes may be beneficial to certain groups and contribute to online communities positively.

Implications

Our work and findings have important implications for fostering safer suicide-related discussions and informing content detection and moderation on social media.

Informing Safer Online Suicide-content Discussion This study provides valuable insights for understanding the characteristics of suicide memes and the motivations for posting them, which could help regulators, policy-makers, and social media platforms develop contextually sensitive moderation and communication strategies for engaging with individuals in suicide meme communities. Our findings indicate humor in suicide memes could benefit individuals who post them, but also might provoke negative emotions in those being exposed. This highlights the need for guidelines that balance the benefits and risks of humor in suicide-related online interactions and inform safer practices.

Informing Design of Suicide-content Moderation Suicide content moderation on social media is a complex process that requires coordinated efforts from multiple actors, including model developers, moderators and operators of social media platforms, and users. Our findings contribute to more effective suicide meme detection, achieving an accuracy of 80.04%, with preliminary evidence suggesting the potential of cross-platform generalization. This research is a critical first step in enhancing suicide meme moderation, which will help platform operators identify suicide memes more effectively and accurately.

Further, some current approaches to suicide content moderation may be insufficiently nuanced (i.e. simply restricting all user access to suicide content). Our topic modeling analysis highlights the dual role of suicide memes – they could help foster supportive connections, but may also act as harmful triggers for individuals. These results suggest that more targeted and dynamic moderation strategies are needed. Instead of binary "allow or restrict" decisions for all suicide memes, moderators could further access the context and intent behind them to better define to what extent such suicide memes are harmful and how should they be moderated.

While the model's performance is promising, the extent to which it can be integrated into suicide content moderation requires careful consideration. For example, multiple thresholds should be considered to support different moderation strategies, from automatic intervention for high-confidence cases to further human review for more ambiguous cases. This approach could help reduce false positives while ensuring the prompt moderation on potentially harmful suicide memes. Our error analysis also highlights some challenges in this process. Some suicide memes are metaphorical and require interpretation from both visual and textual cues, which existing models failed to capture. Additionally, it remains difficult to distinguish between suicide memes and general memes featuring death-related descriptions or negative emotional expressions. These issues suggest the need for further categorizing suicide memes based on feature variations. Future studies could focus on refining the model's ability to handle such complexities, while human oversight for borderline cases may help mitigate errors.

Limitations

This work has several limitations. First, the suicide memes were collected only from `r/SuicideMeme`, and compared to a randomly sampled subset of general memes. Future studies should aim for more general and comprehensive suicide meme disclosures from other online communities and platforms. Second, we assumed that all memes from `r/SuicideMeme` were suicide related without any further clinical evaluations. Although our approach was developed in consultation with suicide prevention professionals, our underlying assumptions may not be correct in all cases. Additionally, we acknowledge that suicide is a culturally sensitive issue. Given that Reddit's user demographic is predominantly young males from the U.S., our findings may not fully capture cultural differences in suicide-related expressions. Despite these limitations, we believe that our study provides meaningful insights into suicide disclosure through memes.

Conclusion

In summary, this study presents the first computational examination of suicide memes in the context of an online community. We identified multiple significant different features between suicide and general memes and examined how these features influence user interactions. The topic modeling results revealed an overall favourable attitude toward suicide memes. We also developed prediction models to assess the feasibility of identifying suicide memes. Our findings suggest that suicide memes may have both positive and harmful effects, highlighting the need for more targeted communication strategies within suicide meme communities, as well as more nuanced algorithms to better moderate these suicide memes.

References

Akil, A. M.; Ujhelyi, A.; and Logemann, H. A. 2022. Exposure to Depression Memes on Social Media Increases Depressive

Mood and It Is Moderated by Self-Regulation: Evidence from Self-Report and Resting EEG Assessments. *Frontiers in Psychology*, 13: 880065.

Akram, U.; Drabble, J.; Cau, G.; Hershaw, F.; Rajenthran, A.; Lowe, M.; Trommelen, C.; and Ellis, J. G. 2020. Exploratory Study on the Role of Emotion Regulation in Perceived Valence, Humour, and Beneficial Use of Depressive Internet Memes in Depression. *Scientific Reports*, 10(1): 899. Publisher: Nature Publishing Group.

Akram, U.; Irvine, K.; Allen, S. F.; Stevenson, J. C.; Ellis, J. G.; and Drabble, J. 2021. Internet Memes Related to the COVID-19 Pandemic as a Potential Coping Mechanism for Anxiety. *Scientific Reports*, 11(1): 22305. Publisher: Nature Publishing Group.

Badian, Y.; Ophir, Y.; Tikochinski, R.; Calderon, N.; Klomek, A. B.; Fruchter, E.; and Reichart, R. 2023. Social Media Images Can Predict Suicide Risk Using Interpretable Large Language-Vision Models. *The Journal of Clinical Psychiatry*, 85(1): 50516. Publisher: Physicians Postgraduate Press, Inc.

Bakhshi, S.; Shamma, D. A.; Kennedy, L.; Song, Y.; de Juan, P.; and Kaye, J. J. 2016. Fast, Cheap, and Good: Why Animated GIFs Engage Us. In *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '16)*, CHI '16, 575–586. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-3362-7.

Bakhshi, S.; Shamma, D. A.; Kennedy, L. S.; and Gilbert, E. 2015. Why We Filter Our Photos and How It Impacts Engagement. In *Proc. Int. Conf. Web Soc. Media (ICWSM '15)*.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The Pushshift Reddit Dataset. In *Proc. Int. AAAI Conf. Web Soc. Media (ICWSM '20)*, volume 14, 830–839.

Benton, A.; Coppersmith, G.; and Dredze, M. 2017. Ethical Research Protocols for Social Media Health Research. In Hovy, D.; Spruit, S.; Mitchell, M.; Bender, E. M.; Strube, M.; and Wallach, H., eds., *Proc. 1st ACL Workshop on Ethics in Nat. Lang. Process. (EthicsNLP '17)*, 94–102. Valencia, Spain: ACL.

Carter, L. E.; McNeil, D. W.; Vowles, K. E.; Sorrell, J. T.; Turk, C. L.; Ries, B. J.; and Hopko, D. R. 2002. Effects of Emotion on Pain Reports, Tolerance, and Physiology. *Pain Research and Management*, 7(1): 21–30.

Chancellor, S.; and De Choudhury, M. 2020. Methods in Predictive Techniques for Mental Health Status on Social Media: A Critical Review. *npj Digital Medicine*, 3(1): 1–11. Publisher: Nature Publishing Group.

Chateau, L. 2020. "Damn I Didn't Know Y'all Was Sad? I Thought It Was Just Memes": Irony, Memes and Risk in Internet Depression Culture. *M/C Journal*, 23(3). Number: 3.

Chatterjee, M.; Kumar, P.; Samanta, P.; and Sarkar, D. 2022. Suicide Ideation Detection from Online Social Media: A Multi-Modal Feature-Based Technique. *International Journal of Information Management Data Insights*, 2(2): 100103.

Coppersmith, G.; Ngo, K.; Leary, R.; and Wood, A. 2016. Exploratory Analysis of Social Media Prior to a Suicide Attempt. In Hollingshead, K.; and Ungar, L., eds., *Proc. 3rd Workshop on Comput. Ling. Clin. Psychol. (CLPsych '16)*, 106–117. San Diego, CA, USA: Association for Computational Linguistics.

DataReportal. 2024. Global Social Media Statistics. https://datareportal.com/social-media-users. [Accessed 07-09-2024].

Davies, P.; Veresova, M.; Bailey, E.; Rice, S.; and Robinson, J. 2024. Young People's Disclosure of Suicidal Thoughts and Behavior: A Scoping Review. *Journal of Affective Disorders Reports*, 16: 100764.

Davison, P. 2012. 9. The Language of Internet Memes. In Mandiberg, M., ed., *The Social Media Reader*, 120–134. New York University Press. ISBN 978-0-8147-6302-5.

Dawkins, R. 2016. *The Selfish Gene*. Oxford University Press. ISBN 978-0-19-878860-7.

Dean, J. 2019. Sorted for Memes and Gifs: Visual Media and Everyday Digital Politics. *Political Studies Review*, 17(3): 255–266. Publisher: SAGE Publications.

DeCook, J. R. 2018. Memes and Symbolic Violence: #ProudBoys and the Use of Memes for Propaganda and the Construction of Collective Identity. *Learning, Media and Technology*, 43(4): 485–504.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *Proc. Int. Conf. Learn. Represent. (ICLR '21)*.

Dynel, M.; and Chovanec, J. 2021. Creating and Sharing Public Humour across Traditional and New Media. *Journal of Pragmatics*, 177: 151–156.

Fu, K.-w.; Cheng, Q.; Wong, P. W. C.; and Yip, P. S. F. 2013. Responses to a Self-Presented Suicide Attempt in Social Media. *Crisis*, 34(6): 406–412. Publisher: Hogrefe Publishing.

Goldenberg, A.; and Gross, J. J. 2020. Digital Emotion Contagion. *Trends in Cognitive Sciences*, 24(4): 316–328. Publisher: Elsevier.

Goldstein, J. H. 2013. *The Psychology of Humor: Theoretical Perspectives and Empirical Issues*. Academic Press. ISBN 978-1-4832-8854-3.

Green, M. J.; Waldron, J. H.; Simpson, I.; and Coltheart, M. 2008. Visual Processing of Social Context during Mental State Perception in Schizophrenia. *Journal of Psychiatry & Neuroscience : JPN*, 33(1): 34–42.

Grootendorst, M. 2022. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv preprint arXiv:2203.05794*.

Guadagno, R. E.; Rempala, D. M.; Murphy, S.; and Okdie, B. M. 2013. What Makes a Video Go Viral? An Analysis of Emotional Contagion and Internet Memes. *Computers in Human Behavior*, 29(6): 2312–2319.

Guntuku, S. C.; Preotiuc-Pietro, D.; Eichstaedt, J. C.; and Ungar, L. H. 2019. What Twitter Profile and Posted Images Reveal about Depression and Anxiety. *Proc. Int. AAAI Conf. Web Soc. Media (ICWSM '19)*, 13: 236–246.

Gupta, A.; Gupta, D.; Chaudhary, P.; Raje, P.; and Chandiramani, K. 2021. Exploring Trivialisation of Mental Health Issues using Internet Memes in Young Adults. *The International Journal of Indian Psychology*, 9: 868–881.

Habib, A. 2020. The Philosophy of Meme Culture. https://nuvomagazine.com/magazine/spring-2020/the-philosophy-of-meme-culture. [Accessed 07-09-2024].

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR '16)*, 770–778.

Henman, L. 2001. Humor as a Coping Mechanism: Lessons from POWs. *Humor-international Journal of Humor Research - HUMOR*, 14: 83–94.

Ji, S.; Pan, S.; Li, X.; Cambria, E.; Long, G.; and Huang, Z. 2021. Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications. *IEEE Transactions on Computational Social Systems*, 8(1): 214–226.

Jovanovic, D.; and Van Leeuwen, T. 2018. Multimodal Dialogue on Social Media. *Social Semiotics*, 28(5): 683–699. Publisher: Routledge _eprint: https://doi.org/10.1080/10350330.2018.1504732.

Langford, B. J.; Laguio-Vila, M.; Gauthier, T. P.; and Shah, A. 2022. Go V.I.R.A.L.: Social Media Engagement Strategies in Infectious Diseases. *Clinical Infectious Diseases*, 74(Supplement_3): e10–e13.

Law, K. C.; and Anestis, M. D. 2021. Testing Whether Suicide Capability Has a Dynamic Propensity: The Role of Affect and Arousal on Momentary Fluctuations in Suicide Capability. *Frontiers in Psychology*, 12. Publisher: Frontiers.

Le Glaz, A.; Haralambous, Y.; Kim-Dufor, D.-H.; Lenca, P.; Billot, R.; Ryan, T. C.; Marsh, J.; Devylder, J.; Walter, M.; Berrouiguet, S.; et al. 2021. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *Journal of Medical Internet Research*, 23(5): e15708.

Literat, I.; and Kligler-Vilenchik, N. 2019. Youth Collective Political Expression on Social Media: The Role of Affordances and Memetic Dimensions for Voicing Political Views. *New Media & Society*, 21(9): 1988–2009.

Liu, C.; Dong, Y.; Xiang, W.; Yang, X.; Su, H.; Zhu, J.; Chen, Y.; He, Y.; Xue, H.; and Zheng, S. 2024. A Comprehensive Study on Robustness of Image Classification Models: Benchmarking and Rethinking. *International Journal of Computer Vision*, 1–23.

Malodia, S.; Dhir, A.; Bilgihan, A.; Sinha, P.; and Tikoo, T. 2022. Meme Marketing: How Can Marketers Drive Better Engagement Using Viral Memes? *Psychology & Marketing*, 39(9): 1775–1801.

Manikonda, L.; and De Choudhury, M. 2017. Modeling and Understanding Visual Attributes of Mental Health Disclosures in Social Media. In *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '17)*, 170–181.

Marci, C. D.; Glick, D. M.; Loh, R.; and Dougherty, D. D. 2007. Autonomic and Prefrontal Cortex Responses to Autobiographical Recall of Emotions. *Cognitive, Affective, & Behavioral Neuroscience*, 7(3): 243–250.

Matero, M.; Idnani, A.; Son, Y.; Giorgi, S.; Vu, H.; Zamani, M.; Limbachiya, P.; Guntuku, S. C.; and Schwartz, H. A. 2019. Suicide Risk Assessment with Multi-level Dual-Context Language and BERT. In Niederhoffer, K.; Hollingshead, K.; Resnik, P.; Resnik, R.; and Loveys, K., eds., *Proc. 6th Workshop Comput. Ling. Clin. Psychol. (CLPsych '19)*, 39–44. Minneapolis, Minnesota: Association for Computational Linguistics.

Mohammad, S. M. 2018. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proc. Ann. Conf. Assoc. Comput. Linguist. (ACL '18)*. Melbourne, Australia.

Mohammad, S. M. 2022. Best Practices in the Creation and Use of Emotion Lexicons. *arXiv preprint arXiv:2210.07206*.

Mohammad, S. M.; and Turney, P. D. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3): 436–465.

Mueller, A. S.; and Abrutyn, S. 2015. Suicidal Disclosures among Friends: Using Social Network Data to Understand Suicide Contagion. *Journal of Health and Social Behavior*, 56(1): 131–148. Publisher: SAGE Publications Inc.

Naslund, J. A.; Bondre, A.; Torous, J.; and Aschbrenner, K. A. 2020. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. *Journal of Technology in Behavioral Science*, 5: 245–257.

Newton, G.; Zappavigna, M.; Drysdale, K.; and Newman, C. E. 2022. More Than Humor: Memes as Bonding Icons for Belonging in Donor-Conceived People. *Social Media+ Society*, 8(1): 20563051211069055.

Nicomedes, C. J. C.; Sasot, C. F.; Santos, G. F.; Distor, J. M. S.; Marzan, P. B.; and Manda, A. R. 2024. A Convergent-mixed Method Study on the Attitudes and Perception Towards Suicide Memes and Suicidality. *The Open Psychology Journal*, 17(1).

Perez, J. 2019. *Suicide memes: Internet users' anti-future expressions*. Ph.D. thesis, Georgetown University in Qatar.

Phillips, M. L.; Drevets, W. C.; Rauch, S. L.; and Lane, R. 2003. Neurobiology of Emotion Perception I: The Neural Basis of Normal Emotion Perception. *Biological Psychiatry*, 54(5): 504–514.

Prinzmetal, W.; Long, V.; and Leonhardt, J. 2008. Involuntary Attention and Brightness Contrast. *Perception & Psychophysics*, 70: 1139–1150.

Reddit. 2024a. Reddit Privacy Policy. https://www.reddit.com/policies/privacy-policy. [Accessed 07-09-2024].

Reddit. 2024b. SuicideMeme. best meme. https://www.reddit.com/r/SuicideMeme/. [Accessed 07-09-2024].

Ribeiro, J. D.; Bender, T. W.; Buchman, J. M.; Nock, M. K.; Rudd, M. D.; Bryan, C. J.; Lim, I. C.; Baker, M. T.; Knight, C.; Gutierrez, P. M.; et al. 2015. An Investigation of the Interactive Effects of the Capability for Suicide and Acute Agitation on Suicidality in a Military Sample. *Depression and anxiety*, 32(1): 25–31.

Robinson, J.; Cox, G.; Bailey, E.; Hetrick, S.; Rodrigues, M.; Fisher, S.; and Herrman, H. 2016. Social Media and Suicide Prevention: A Systematic Review. *Early intervention in psychiatry*, 10(2): 103–121.

Sadaka, N. G.; Karam, L. J.; Ferzli, R.; and Abousleman, G. P. 2008. A No-Reference Perceptual Image Sharpness Metric Based on Saliency-Weighted Foveal Pooling. In *Proc. 15th IEEE Int. Conf. Image Process. (ICIP '08)*, 369–372. IEEE.

Sawhney, R.; Joshi, H.; Gandhi, S.; and Shah, R. R. 2020. A Time-Aware Transformer Based Model for Suicide Ideation Detection on Social Media. In Webber, B.; Cohn, T.; He, Y.; and Liu, Y., eds., *Proc. Conf. Empir. Methods Nat. Lang. Process. (EMNLP '20)*, 7685–7697. Online: Association for Computational Linguistics.

Schott, G. R. 2023. *Death as Entertainment: Young People and Death Awareness*. Taylor & Francis.

Simonyan, K.; and Zisserman, A. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.

Smith, N.; and Linker, S. 2021. Suicide-Memes as Exemplars of the Everyday Inauthentic Relationship with Death. *Mortality*, 26: 1–16.

Tama-Rutigliano, K. 2018. Council Post: Memes: A Digital Marketing Tool For Every Industry — forbes.com. https://www.forbes.com/sites/forbescommunicationscouncil/2018/08/10/memes-a-digital-marketing-tool-for-every-industry/. [Accessed 07-09-2024].

Wasike, B. 2022. Memes, Memes, Everywhere, nor Any Meme to Trust: Examining the Credibility and Persuasiveness of COVID-19-Related Memes. *Journal of Computer-Mediated Communication*, 27(2): zmab024.

Weiser, R.; and Alam, N. 2022. Meme Culture and Suicide Sensitivity: a Quantitative Study. *Humanities and Social Sciences Communications*, 9(1): 1–10. Publisher: Palgrave.

WHO. 2024. Suicide — who.int. https://www.who.int/news-room/fact-sheets/detail/suicide. [Accessed 13-01-2025].

Yadav, S.; Caragea, C.; Zhao, C.; Kumari, N.; Solberg, M. A.; and Sharma, T. 2023. Towards Identifying Fine-Grained Depression Symptoms from Memes. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

YPulse. 2019. Topline: Social Media Behavior. https://www.ypulse.com/report/2019/02/20/topline-social-media-behavior2/. [Accessed 07-09-2024].

Zhang, C. C.; Zaleski, G.; Kailley, J. N.; Teng, K. A.; English, M.; Riminchan, A.; and Robillard, J. M. 2024. Debate: Social Media Content Moderation May Do More Harm than Good for Youth Mental Health. *Child and Adolescent Mental Health*, 29(1): 104–106.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes.
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes.
   (g) Did you discuss any potential misuse of your work? Yes. See Ethics and Responsibility.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. See Ethics and Responsibility.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? We did not include the code and data, but we have provided methodology details to reproduce the experimental results. Data will be shared upon request.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, we provided training details in the Methodologies and Appendices sections
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, we reported the standard deviations for the trained models across multiple experiments.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No. However, we recognized the potential for false positives in suicide meme detection.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes. We cited existing models appropriately in the paper.
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? No.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA. Since we use Reddit data, it is impractical to obtain user consent.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. See the Ethics and Responsibility section.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA.
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes. See Ethics and Responsibility.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA

Appendices

A: BERTopic Configurations

BERTopic steps	Models	Parameters
Embeddings	Sentence Transformer: all-MiniLM-L6-v2	\
Dimensionality reduction	UMAP	N_neighbors = 15; n_component = 5
Clustering	HDBSCAN	Min_cluster_size = 20; Metric = 'euclidean' Cluster_selection_method = 'eom'
Tokenizer	CountVectorizer	Min_df = 2; Ngram_range = (1, 3)
Weighting scheme	c-TF-IDF	Seed_words = ["meme", "humor", "joke"] Seed_multiplier = 2
Fine-tune representations	ChatGPT	Prompt = """ I have a topic that contains the following documents: [DOCUMENTS] The topic is described by the following keywords: [KEYWORDS] Based on the information above, extract a short but highly descriptive topic label of at most 5 words. Make sure it is in the following format: topic: topic label """

Table 7: Submodel configurations of BERTopic.

B: Statistical Hypothesis Test Results

	Features	P-value
Meta Features	Width	6.52e-23***
	Height	5.98e-20***
	Number of pixels	2.63e-24***
	The aspect ratio	0.27

Table 8: Statistical hypothesis test results: meta features.

	Features	P-value
Visual Features	Asymmetry	1.25e-8***
	Brightness	0.11
	Contrast	0.07
	Contrast balance	6.74e-9***
	Entropy	5.31e-33***
	Exposure balance	5.32e-9***
	Sharpness	1.27e-15***
	Face count	3.21e-24***
	Face region	1.33e-15***
	Text region	2.58e-11***

Table 9: Statistical hypothesis test results: visual features.

	Features	P-value
Textual features	Words count	1.15e-30***
	Valence	1.08e-11***
	Arousal	4.13e-29***
	Dominance	0.004**
	Anger	3.18e-40***
	Anticipation	2.85e-3***
	Disgust	9.55e-7***
	Fear	2.78e-87***
	Sadness	1.75e-101***
	Trust	6.97e-11***
	Surprise	5.57e-07***
	Joy	2.45e-10***

Table 10: Statistical hypothesis test results: textual features.

C: Comparisons of Suicide Meme with Different Levels of User Engagement

To further understand the association between suicide memes and user engagement, we present the top suicide memes corresponding to each user engagement level (i.e., scores, comments), as shown in Figure 8. For memes with the lowest number of comments or lowest scores, the same suicide memes were identified, as shown in Figure 8(c). Additionally, there is an overlap for one suicide meme (SpongeBob) between the highest number of comments (Figure 8(a)) and the highest scores (Figure 8(b)).

Based on the examples, we observe several common characteristics of suicide memes that attract higher use engagement: they tend to have longer captions and seem darker – often expressing explicit suicide ideation, with some even referencing specific suicide methods, such as pills or images of nooses. We also observe the use of animated figures in these suicide memes. On the other hand, suicide memes with lower scores and comments tend to have fewer words and less emotionally intense imagery (e.g., depictions of suicide methods). Some memes contain less direct expression of a wish to be dead. Although one meme mentioned overdose as a potential suicide method, its primary message focuses on something preventing the user from doing so.

(a) Suicide memes with highest comments

(b) Suicide memes with highest scores

(c) Suicide memes with lowest scores, also with lowest comments

Figure 8: Examples of suicide meme with different levels of user engagement.

D: Topic Modeling

Although this paper focuses on topics related to suicide memes, we also examined other identified topics. Table 11 presents the top 10 topic clusters unrelated to suicide memes. Overall, these topics are centered around suicide discussions. The most prevalent topic involves mental health support and coping (topic 1) mental health support and coping (topic 1). Several topics reflected users' expression of a desire to death and suicide ideation (topic 0, 3). Notably, many topic clusters contained discussions about specific suicide methods, such as overdose (topic 1), guns (topic 6), and ropes (topic 14), with detailed comparisons and considerations (topic 5). There are also some discussions regarding their failure attempts (topic 11). This raise concerns that the varying severity of such discussions might be obscured by the humor in suicide memes. It is necessary to further investigate suicide memes to better identify contents and discussions associated with higher risks.

E: Suicide Memes Prediction

E.1: Parameters and Evaluation Metrics

In Table 12, we present the hyperparameters of our best-performing model trained on the Reddit dataset, SVC with both visual and textual features. While the main paper reports Accuracy and F1 scores, this appendix includes additional metrics such as recall and precision, as shown in Table 13.

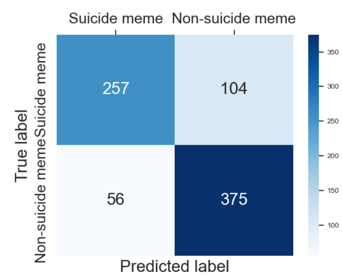


Figure 9: Confusion matrix on test set (model: SVC with visual and textual features)

E.2: Qualitative Error Analysis

E.3: Model Generalization Test

E.3.1: Data Collection We used the Google Search API to collect datasets from three additional social media platforms: Instagram, Facebook, and X, with the parameters and search queries detailed in Table 14. For each collected suicide and non-suicide meme, we applied the same filtering criteria used for the Reddit dataset, as described in the Dataset section. Table 15 presents an overview of the data after filtering.

Notably, limitations may exist, as the current ranking algorithm used by Google remains unclear, and the applied search domain returned results predominantly from the United States. It is unclear whether this may introduce bias and affect the evaluation of the generalization performance of our findings.

No.Topic	Count	Topic	Representative text
-1	2791	Mental health support and coping	"If this is not a joke, seek support from family, friends or even professionals. Try to figure it out why you see suicide as a viable answer."
0	210	Suicidal people with suicide thoughts	"I guess I've gotten used to these thoughts by now, as it's been years. I lie in bed every night thinking about suicide. It's just that sometimes, it becomes really hard to suppress those dark thoughts.", "You could probably find around five suicide notes hidden around my house just this past month."
1	208	Lethal effects of overdose	"Lol overdosing on sleeping pills won't work – just don't try it.", "The fatal amount of caffeine for humans is around 6 grams, though it varies from person to person. Overdosing on stimulants might not be a pleasant way to go though, as you might survive but now with horrible heart disease."
3	174	Desire for peaceful death	"I wish I could be dead but I'm scared", "I'd always choose a quick, painless death over living the life I have now, if I had the choice."
5	134	Suicide contemplation and consideration	"I'll probably choose hanging or overdose", "I guess number two or four. It would be really painful if an overdose doesn't work immediately, though you'll probably just die if jumping off a building as long as it's tall enough."
6	124	Gun violence	"Lollll shooter-master... can you aim for my head pwease pwease?", "Anyone know the price of a shotgun?"
10	85	Dealing with depression	"You're feeling lonely with depression, but karma makes you feel needed.", "Quit halfway through. I was too depressed to finish and figured it'd probably end up just as bad anyway."
11	81	Suicide attempts	"It is continuous suffering after failed attempts", "How could you fail? Some methods are so guaranteed to die without even leaving the house."
14	71	Discussion about hanging	"I'd do it, but I have no idea where to hang the rope.", "FYI a hangman's knot isnt actually the best knot for hanging yourself."
15	68	Family relationships	"My crazy parents said they'll kill me if I ever thought about ending myself. Ah yes, that's exactly what I wanted.", "Lucky you! My family is extremely unwilling to cooperate."

Table 11: Top 10 topic clusters unrelated to suicide memes (The representative text was synthesized due to ethical considerations but retains the main idea of the original sentences).

Steps	Methods	Parameters	Values
Data processing	StandardScaler()	with_mean	True
		with_std	True
Prediction model	SVC()	Kernel	"linear"
		Gamma	0.01
		C	10

Table 12: Parameters of best-performing model: SVC with both visual and textual features.

Models	Feature Combination	Recall(%)	Precision(%)
ViT	Raw meme images	56.96	56.99
Resnet50	Raw meme images	55.68	55.73
VGG19	Raw meme images	61.88	62.12
SVC	Visual features	$65.95_{0.0271}$	$67.48_{0.0124}$
SVC	Textual features	$75.74_{0.0132}$	$79.33_{0.0130}$
SVC	Visual+textual features	$79.80_{0.0234}$	$81.30_{0.0154}$

Table 13: Models performance at predicting suicide memes. (Noted accuracy and F1 reported are averaged over 10 experimental runs; subscripts show standard deviation)

E.3.2: Features Comparison Following the same methods, we extracted the proposed features from these three datasets and conducted statistical tests to examine significant differences in the meta, visual, and textual features that were identified as most significant in the Reddit dataset. Figures 10, 11, and 12 present the distribution of these features as violin plots for Instagram, Facebook, and X, respectively.

For the Instagram dataset, most features showed significant differences between suicide memes and non-suicide memes, except for *width*, *height*, and *contrast balance*, which p values were greater than 0.5. The features with significant differences followed a similar distribution to the Reddit dataset.

For the Facebook dataset, we did not find a significant difference in feature *width*, *height*, *contrast balance*, *face counts*, and *face region*, while other features showed significant difference. The features with significant differences followed a similar distribution to the Reddit dataset.

For suicide memes on X, we observed the same patterns: significant differences in most features, except for *width* and *contrast balance*. These significant features followed a similar distribution to those in the Reddit dataset.

Overall, most of the visual and textual features of suicide memes from `r/SuicideMeme` are also consistent on Instagram, Facebook, and X. Specifically, features such as lower *entropy*, higher level of *sharpness*, *sadness*, *feature*, *anger*, *arousal*, *words counts* were consistently observed across all platforms. The fewer *face counts* and reduced *face region* were also observed on Instagram and X. These consistent patterns suggest that visual and textual features of suicide memes could be extended to broader social media platforms.

While minor variations exist — such as *width* and *con-*

	Domain	Language	Search Queries Suicide Meme	 Non-suicide Meme
Instagram			”suicide meme site:instagram.com”	”meme site:instagram.com”
Facebook	”google.com”	lang_us (English only)	”suicide meme site:facebook.com”	”meme site:facebook.com”
X			”suicide meme site:X.com”	”meme site:X.com”

Table 14: Parameters of data collection using Google Search API.

	Suicide Meme	Non-suicide Meme
Instagram	594	667
Facebook	499	545
X	104	143

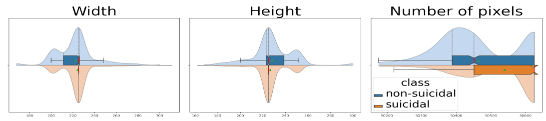
Table 15: Overview of Instagram, Facebook, and X datasets.

trast balance showing no significant differences on all three platforms, potentially due to differences in platform demographic or image processing techniques — they did not appear to affect the model's performance on the Instagram and X datasets. Interestingly, the classification performance on the Facebook dataset showed a decline, which may be due to the lack of significant difference in *face counts* and *face region* features between suicide memes and non-suicide memes on Facebook.

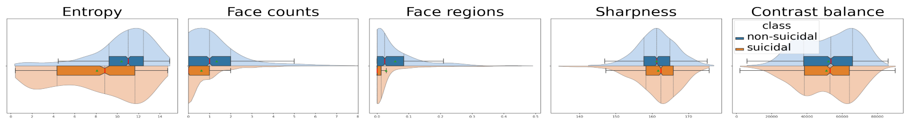
E.3.3: Evaluation Metric We also reported the recall and precision scores of the classification model trained independently on the Instagram, Facebook, and X datasets. As shown in Table 16.

	Recall(%)	Precision(%)
Instagram	$81.90_{0.0194}$	$83.02_{0.0172}$
Facebook	$73.24_{0.0145}$	$74.44_{0.0157}$
X/Twitter	$82.01_{0.0115}$	$85.18_{0.0159}$

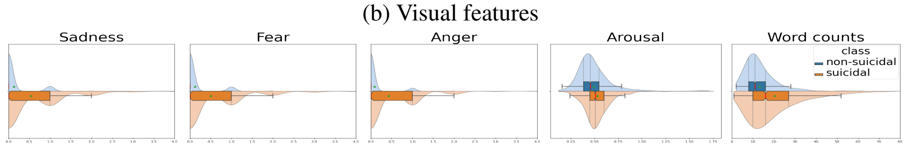
Table 16: Recall and precision scores of the classification model on Instagram, Facebook, and X datasets. (Noted accuracy and F1 reported are averaged over 10 experimental runs; subscripts show standard deviation)



(a) Meta features

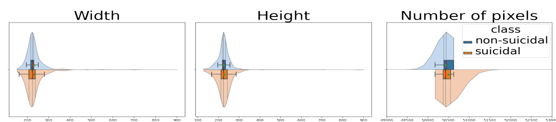


(b) Visual features

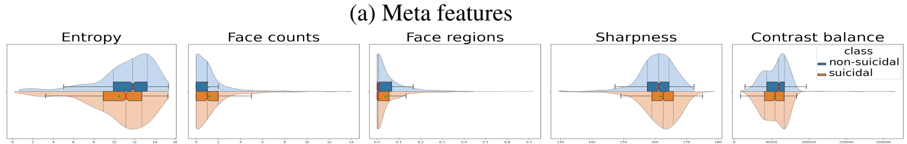


(c) Text features

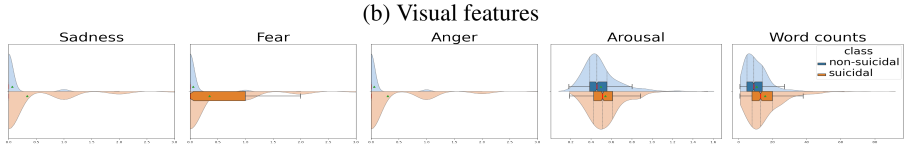
Figure 10: Violin plots: Instagram.



(a) Meta features

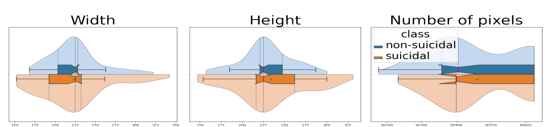


(b) Visual features

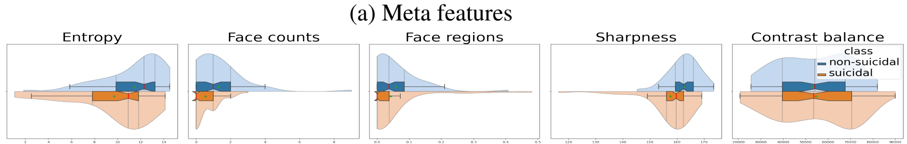


(c) Text features

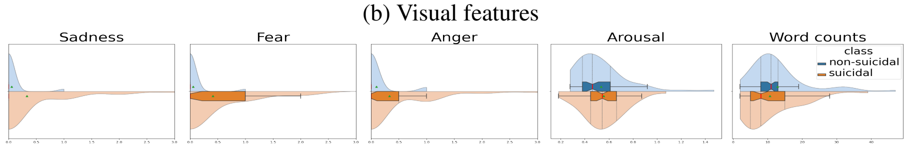
Figure 11: Violin plots: Facebook.



(a) Meta features



(b) Visual features



(c) Text features

Figure 12: Violin plots: X.