Health Conversations on Instagram: A Comparative Study of Textual and Visual Content

ABSTRACT

Amidst technological strides, photo-sharing platforms surge in popularity. Dominated by Facebook and YouTube, Instagram emerges as the third most frequented social platform. User-generated imagery on Instagram yields a wealth of diverse, visually immersive data. This versatile platform empowers interdisciplinary exploration, unveiling novel patterns across science, marketing, psychology, and health. Despite its large user base, Instagram remains an under-researched platform, boasting distinctive demographics compared to peers. We address this gap by investigating health perceptions via Instagram, extracting and comparing textual attributes of posts: user captions, hashtags, and image-generated captions. We found the attributes to have low similarity with user captions and hashtags exhibiting the highest similarity. We generated topics on these attributes and found that shared health ideologies traverse locations except for location-specific nuances. Topics generated from posts often encompass fitness, healthy eating, and exercise, transcending geographical boundaries.

CCS CONCEPTS

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

KEYWORDS

Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

ACM Reference Format:

. 2018. Health Conversations on Instagram: A Comparative Study of Textual and Visual Content. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Social media platforms such as Facebook, Twitter, Instagram, Snapchat, and YouTube have revolutionized the way users share opinions, messages, images, and videos about events around them. Among these platforms, Instagram—a subsidiary of Facebook—has notably redefined image sharing and communication in the social media landscape. Since its inception in 2010, Instagram has experienced exponential growth, reaching over 1 billion monthly active users by 2020 [17], with approximately 100 million images shared daily [1]. It stands as the most popular image-sharing platform, boasting more than twice the number of users compared to Twitter [32].

Users engage with Instagram to share personal experiences and stay updated with people or accounts they follow. In recent years, the platform's utility has expanded beyond personal sharing to encompass diverse applications such as marketing and the dissemination of health information [31]. Brands and businesses leverage Instagram to enhance their visibility and set up shops to sell products to a broader audience [42]. Both Facebook and Instagram offer viable platforms for influencers and businesses to market products or services to their users, with marketers increasingly shifting focus from Facebook to Instagram due to its potential to reach wider and more targeted audiences [26].

The widespread popularity of Instagram has prompted 71% of U.S. businesses to use it for promoting their companies or communicating their messages [1], with around 83% of users discovering new products or services on the platform [9]. In the United Kingdom, Instagram has been reported as the most popular social media platform, with 67% of users increasing their usage and making decisions to purchase products or services based on their interactions [13]. The platform's significant popularity among young adults makes it an excellent sales channel for businesses aiming to target younger consumers.

An Instagram post is an image or video uploaded by the user, which generally consists of seven parts: username, image, caption, hashtags, likes, comments, and location, as shown in Figure 1. Users upload images or videos accompanied by captions—textual descriptions associated with the visual content. Hashtags, denoted by the symbol '#', are keywords or topics that relate to the content, emotions, or moments associated with the post. They enable users to search for and discover content on specific topics of interest, fostering community engagement around shared themes. User interaction is further facilitated through likes and comments, allowing individuals to express appreciation or engage in discussions about a post. These interactions serve as metrics indicating the popularity or engagement level of content. Additionally, users can include location information to provide geographical context to their posts.

The COVID-19 pandemic has significantly altered communication patterns, shifting interactions from in-person to online platforms across various sectors. Instagram emerged as a prominent medium for disseminating virus-related information, utilizing hashtags such as #COVID19 and #coronavirus [35]. Social media usage surged globally during this period; for instance, the UK government leveraged influencers on platforms like Instagram to disseminate information about the coronavirus [7], and India witnessed a 59% increase in daily usage of platforms like Instagram [6]. With global lockdowns limiting physical activities and social gatherings, social media became essential for staying connected with the world, family, and friends, as well as for entertainment during unprecedented times.

While several studies have explored the impact of communication through Instagram on its users, limited attention has been

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Figure 1: Features of an Instagram post

given to analyzing the relationship between the textual features of Instagram posts and their associated images. Understanding this relationship can provide valuable insights into how users convey messages and interact with content on the platform. It can also aid in developing tools such as hashtag recommendation systems for images, enhancing user experience and engagement.

Moreover, examining the shared information on Instagram to understand global reactions to significant topics like COVID-19 can reveal new insights and identify topics of interest helpful in managing future health crises. For example, Susewind et al. [41] analyzed trends in park visitation through COVID-19-related posts on Instagram but found a lack of adequate data, highlighting the need to explore the abundant posts on Instagram more thoroughly.

Despite the vast amount of data available on Instagram, there is a paucity of research focusing on the interplay between image content and textual features such as captions and hashtags, especially in the context of global events like the COVID-19 pandemic. This study aims to fill this gap by investigating the relationships among these components and analyzing how they vary across different locations.

In this work, we address the following research questions:

(1) **RQ1:** Are the content of images, captions, and hashtags in Instagram posts related to each other?
(2) **RQ2:** What topics are generated by Instagram users from different countries related to health, healthy lifestyle, and COVID-19, and do users' perspectives on these health topics differ based on location?
(3) **RQ3:** Which textual feature of an image (caption or hashtags) generates topics most relevant to the main hashtags?

To answer these questions, we extracted Instagram posts containing the hashtags #health, #healthylifestyle, and #COVID19. We filtered posts by location to investigate the role of geographical context in the topics of interest communicated through Instagram. By comparing these topics, we aim to identify important themes for health marketers or businesses on Instagram, enabling them to tailor their marketing strategies to the most popular topics or target audiences based on location.

In summary, our contributions are as follows:

(1) We present one of the first studies to examine the relationship between the content of Instagram images and their associated textual features (captions and hashtags).
(2) We investigate the health-related topics generated by Instagram users from different locations, analyzing how perspectives on health topics vary geographically.
(3) We analyze which textual feature of an image (caption or hashtags) generates topics most relevant to the main hashtags, informing strategies for effective communication on Instagram.

2 RELATED WORK

The advent of social media platforms has transformed the landscape of communication, with Instagram emerging as a pivotal medium for visual storytelling and information dissemination. Research has extensively explored various facets of Instagram, including its role in health communication, marketing, user behavior, and cultural expression. In this section, we review prior studies related to these areas, highlighting the gaps that our research aims to address. Instagram's visual-centric nature has been leveraged for health education and promotion. Kamel Boulos *et al.* [22] underscored the platform's potential in conveying rich health information, particularly to younger audiences. Studies have examined how health organizations utilize Instagram to disseminate information. For instance, Alkazemi *et al.* [2] analyzed posts by Gulf Cooperation Council Ministries of Health, suggesting improvements for more effective communication. Similarly, Kim *et al.* [23] assessed the Centers for Disease Control and Prevention's use of Instagram, finding that embedding text into images was less engaging and recommending sharper, clearer images featuring expressive faces to enhance user interaction.

Research has also delved into specific health topics on Instagram. Lee *et al.* [25] conducted a content analysis of posts with the hashtag #mentalhealth, revealing that the majority focused on general wellness practices and therapy advertisements. They posited that Instagram could serve as a bridge between health professionals and individuals facing health challenges. Muralidhara and Sung [30] applied topic modeling to posts tagged with #health, identifying 47 distinct health-related topics, and highlighted Instagram's potential as a source of public health information. The influence of Instagram on user lifestyles and behaviors has been a focal point in recent studies. Chung *et al.* [10] explored the platform's effect on healthy eating habits, finding that supportive content on Instagram promotes healthier lifestyles. Santarossa *et al.* [37] examined the #fitspo trend, noting that posts often convey positive emotions and that personal accounts garner more popularity than non-personal ones. Cohen *et al.* [11] analyzed body-positive accounts, observing that diverse body representations can offer alternative perspectives on beauty standards, particularly for young women.

Moreover, researchers have utilized Instagram data to predict user behavior and personality traits. Habibi and Salim [20] developed models linking Instagram posts to user characteristics. Reece and Danforth [33] identified markers of depression in user images before clinical diagnosis, demonstrating that machine learning models can detect such markers more effectively than human assessments. Several studies have focused on the relationship between textual and visual content on Instagram. Ferwerda *et al.* [14] investigated personality prediction using visual and content features from user posts, finding limited added value when combining features.

Argyrou *et al.* [3] explored the connection between images and user-generated hashtags, applying topic modeling to filter irrelevant hashtags and confirming that relevant hashtags align with image topics. Giannoulakis and Tsapatsoulis [18] examined the explanatory power of user-provided hashtags, discovering a significant correlation between user and participant-selected hashtags. Fiallos and Wong [15] analyzed a subset of images tagged with #allyouneedisecuador, employing Microsoft Cognitive Services to retrieve visual descriptions and identifying relevant topics through clustering.

The geographical context of Instagram posts has been relatively underexplored. Chang *et al.* [8] analyzed Instagram data across U.S. cities to study cultural differences and trends, noting distinct themes based on location. Singh *et al.* [39] used geotagged images from New York City to estimate demographic diversity, comparing their findings with census data to propose cost-effective methods for studying diversity through social media. Leaver and Highfield [24] investigated how users express emotions related to birth and death on Instagram, finding that the platform facilitates more open emotional expression compared to others. However, comprehensive analyses of how user perspectives on health topics vary geographically remain scarce. While prior research has significantly advanced our understanding of Instagram's role in health communication, lifestyle influence, and user behavior prediction, several gaps persist. Notably, limited attention has been given to the intricate relationships between the visual content of images and their accompanying textual features—captions and hashtags. Moreover, the geographical variation in these relationships, particularly in the context of global events like the COVID-19 pandemic, has not been thoroughly examined.

Our study aims to bridge these gaps by analyzing the interconnectedness of images, captions, and hashtags in Instagram posts related to health, healthy lifestyle, and COVID-19 across different countries. By investigating whether these elements are related and how they vary geographically, we seek to provide deeper insights into user perceptions and communication strategies on Instagram.

3 DATA

The dataset employed in this study was assembled using publicly available Instagram images sourced through Instaloader [21], a tool that systematically collects images, captions, and associated metadata. Three primary hashtags, namely #health, #healthylifestyle, and #covid19, were targeted to capture a broad spectrum of health-related user-generated content shared on the platform. The first two hashtags, #health and #healthylifestyle, represent a general and lifestyle-focused perspective on health topics [5], while #covid19 uniquely captures pandemic-era health-related discourse during the global pandemic of 2020–2021. This selection aligns with our research goals to compare general health content against pandemic-specific narratives. All data were publicly available and retrieved following Instagram's terms of service. User identities were anonymized where possible, and any personally identifiable information was excluded. While retained for spatial analysis, geographic data were not linked to individual users in any personally identifying manner.

In total, the initial corpus encompassed 1,869,025 images. Table 1 presents the distribution of these images across the three hashtags, along with the number of corresponding posts and the geographic tagging information. Each retrieved image included metadata such as upload date, user caption, number of likes and comments, user-supplied hashtags, and geolocation information when available. Table 2 delineates these features. Prior to analysis, captions were tokenized and lowercased. Noise removal steps included removing URLs, emojis, and extraneous whitespace. Hashtags within captions were extracted separately for thematic clustering. The preliminary filtering removed 392,358 images that lacked meaningful captions or contained only the main hashtag as the caption. Subsequently, non-English entries were excluded based on automatic language detection using the spaCy library [40], removing 749,764 images. We focused on English-language posts to ensure consistency in linguistic analysis and avoid biases introduced by translation. Language detection was performed using spaCy's language model, validated on a subset of randomly sampled captions, achieving over 95% accuracy. Geotagging proved critical for our analysis's spatial dimension; however, 594,140 images had no associated location data and were thus discarded. As indicated in Table 1, the final dataset retains a substantial volume of posts in English, facilitating the linguistic and thematic analyses.

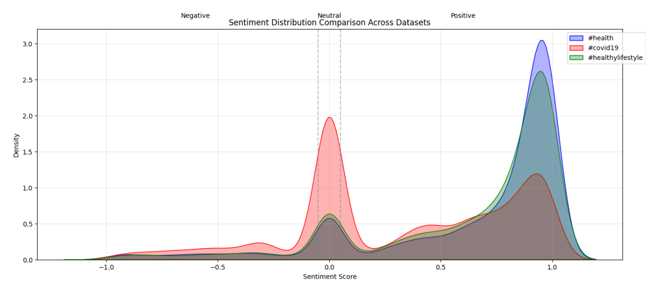

Figure 2: Sentiment analysis of the user captions

A sentiment analysis of the English captions was conducted to probe the underlying affective tone associated with each hashtag. Figure 2 provides the aggregated sentiment distributions. For #health, the overall sentiment skewed prominently toward positivity, with approximately 84.6% of the captions classified as positive and only 6.8% negative. A similar trend emerged for #healthylifestyle, showing 83.5% positive captions and 6.5% negative. In contrast, the #covid19 category presented a more balanced spread of sentiment, with 55.7% positive, 31.4% neutral, and 13.0% negative. These variations indicate that while general health and lifestyle-related discussions on Instagram are mostly optimistic and encouraging, pandemic-related discourse reflects more complex emotional undertones, likely arising from global uncertainty and anxiety.

The geographic distribution of these posts revealed the United States (US) as the most geotagged location, followed by United Kingdom (UK) and India (IN). Figure 3 highlights the top 25 countries with the highest volume of geotagged content. The #covid19 posts showcased a slightly different geographic pattern, with Indonesia (ID) and Brazil (BR) also contributing significantly to the pandemic-related discourse. This divergence suggests that while some health-related discussions are globally dispersed, the pandemic narrative engages an even broader geographical community.

Table 1: Summary of Processed Images and Posts by Hashtag in the Study

Hashtags	No. of Images	No. of Posts	GeoTaggedImages	GeoTaggedPosts	EnglishGeoPosts
#health	610,882	338,299	79,160	19,567	15,578
#healthylifestyle	608,396	373,676	55,476	13,923	10,568
#covid19	651,281	375,205	117,738	26,768	12,860

Table 2: Description of Features in the Instagram Dataset

Feature	Description
Date	The upload date of the image as specified by the user.
Caption	The text by the user as a description of the post, excluding hashtags.
Likes	The number of likes the image received.
Comments	The number of comments the image received.
Hashtags	The hashtags included by the user in the post.
Country Code	The country with the image's geotag, as specified by the uploader.
Caption (Generated)	A caption automatically generated by a model for the image.

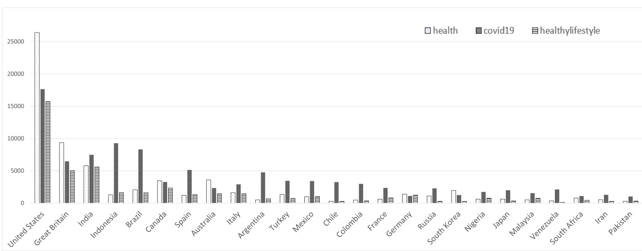

Figure 3: Total posts of the top 25 locations

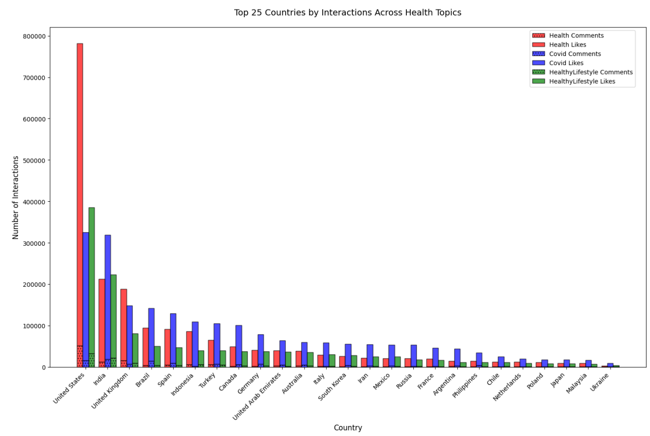

Figure 4: Interactions for each group of dataset of the top 25 locations

The analysis of engagement levels across hashtags and geographical regions (Figure 4) reveals distinctive patterns, with the United States demonstrating overwhelming dominance, particularly in the #health category with 730,901 likes. While developed nations like the US and the UK lead in general health content, a notable shift occurs in COVID-19-related engagement where emerging economies like India and Brazil show proportionally higher activity. India ranks second in pandemic-related discussions with around 310,048 COVID-related likes, while Brazil follows with roughly 127,617 likes, both showing significantly higher COVID-19 engagement compared to their general health content. Western European countries, including Spain, Germany, and France, display moderate but consistent engagement across all three hashtags, though at considerably lower levels (50,000-100,000 interactions) compared to Anglo-American countries. This global distribution of interactions underscores how health-related content resonance varies significantly based on geographic and thematic factors, particularly highlighting the contrast between developed nations' focus on general health content and emerging economies' heightened engagement with pandemic-related discussions.

4 METHODS

The overall approach for this study is structured into two primary components (Figure 5): (1) analysis of textual features and (2) topic modeling based on geographical locations. Given the inherent challenges of directly utilizing image data for modeling or analysis alongside textual features, images are first transformed into textual representations. This transformation facilitates the integration of visual and textual data within a unified analytical framework.

To extract meaningful features from images, we employed the Bootstrapping Language-Image Pre-training (BLIP) model [27]. BLIP is a state-of-the-art architecture designed for vision-language tasks, effectively bridging the gap between visual and textual data by generating high-quality image captions and embeddings that capture both the visual content and contextual information of each image. Prior to feature extraction, all images underwent a preprocessing phase to ensure consistency and quality. This involved resizing the images to a standard resolution of 224 x 224 pixels and applying Gaussian smoothing for denoising. These preprocessing steps were essential to reduce computational complexity and enhance the quality of the input data fed into the BLIP model. The BLIP model processes each image by first generating contextual embeddings that encapsulate the visual features. These embeddings are then utilized to produce descriptive captions that accurately reflect the content of the image. Unlike traditional captioning models that rely on predefined datasets, BLIP leverages large-scale pretraining on diverse image-text pairs, enabling it to generate more nuanced and contextually relevant captions. Each image is thus associated with a *"BLIP caption"*, providing a textual representation of its visual content. The study utilizes three distinct textual features derived from Instagram posts: *(1) BLIP caption, (2) user hashtag*, and *(3) user caption.* These features offer diverse perspectives on the content of each image, encompassing both automatically generated descriptions and user-generated metadata. The BLIP captions provide standardized descriptions based on the image content, while

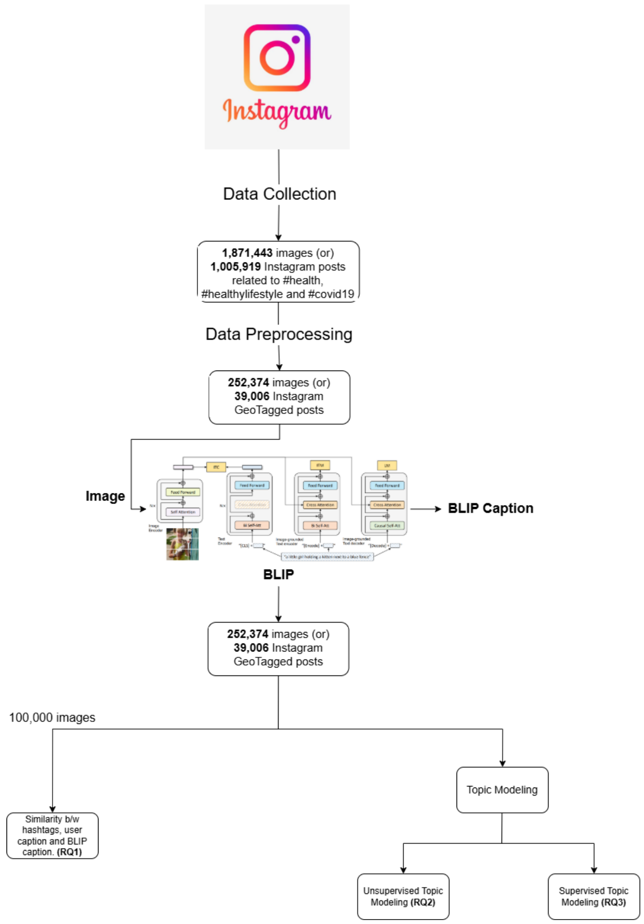

Figure 5: System architecture for the study

user hashtags and captions offer contextual and subjective information as provided by the users themselves.Examples of these features are given in Table 3.

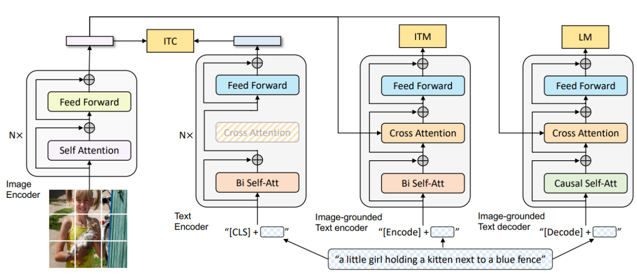

Figure 6: The BLIP model's architecture processes an input image through vision encoders and language decoders to generate a caption describing the scene.

4.1 Analysis of Textual Features

In the initial phase of the study, we conducted a comparative analysis of the three textual features to assess their semantic similarities and differences. This analysis was performed on a random subset of 100,000 images from the collected Instagram dataset. These images were selected to be irrelevant to the main hashtags to ensure an unbiased comparative analysis. We employed Semantic Textual Similarity (STS) using the SentenceTransformers framework [34] to quantify the semantic similarity between the textual features. SentenceTransformers, built on top of PyTorch and utilizing Bidirectional Encoder Representations from Transformers (BERT) networks, is a powerful tool for generating sentence embeddings that capture the semantic meaning of text. Specifically, we utilized the "all-mpnet-base-v2" model, which has demonstrated high performance on the STS benchmark, achieving the highest Spearman's rank correlation compared to other models. The preprocessing of textual data involved tokenization, conversion to lowercase, removal of punctuation and stopwords, lemmatization to reduce words to their base forms, standardization of verbs to the present tense, and stemming to retain only the root words. These steps ensured that the textual data was clean and standardized, facilitating accurate similarity computations. Sentence embeddings generated by the model positioned semantically similar words closer together in the vector space. By computing the cosine similarity between pairs of embeddings, we obtained similarity scores ranging from 0 (no similarity) to 1 (identical semantics). This quantitative assessment provided insights into each textual feature's relative effectiveness and coherence in representing the underlying image content.

4.2 Topic Modeling

The second component of the study involved generating topics for the main hashtags filtered by country. This process was carried out using two distinct techniques: unsupervised and semi-supervised modeling. Both techniques leverage advanced transformer-based models to enhance the accuracy and coherence of topic generation, moving beyond traditional methods such as Latent Dirichlet Allocation (LDA).

4.2.1 Unsupervised Topic Modeling. In the unsupervised approach, our objective was to create clusters of images for each geographical location based solely on image features. To achieve this, we utilized the Vision Transformer (ViT) [12], a cutting-edge architecture that has demonstrated superior performance in image recognition tasks compared to traditional convolutional neural networks like VGG16. ViT processes images by dividing them into fixed-size patches, which are then linearly embedded and fed into a transformer encoder. This methodology captures both local and global contextual information, resulting in high-dimensional embeddings that effectively represent the visual content of each image. Following feature extraction, the high-dimensional image embeddings were processed using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [28], a density-based clustering algorithm renowned for its ability to identify clusters of varying shapes and densities without the need to predefine the number of clusters. This capability addresses the limitations encountered with k-means clustering, where selecting an appropriate number of clusters was both challenging and computationally intensive.

For each location, up to one thousand random images associated with each main hashtag were selected to ensure scalability and

(a) Images clustered correctly

(b) Images clustered poorly

Figure 7: An example of clustering of images based on image features

manageability across the top 25 locations studied. HDBSCAN was applied to these embeddings with parameters tuned to allow the formation of a maximum of ten clusters per location when necessary. This automated determination of cluster numbers based on the intrinsic data structure streamlined the clustering process and enhanced computational efficiency. Once the clusters were established, we employed BERTopic [19], an advanced topic modeling technique that integrates transformer-based embeddings with dynamic clustering algorithms to generate coherent and meaningful topics. BERTopic utilizes transformer models such as BERT or RoBERTa to create contextual embeddings for the textual data, which are then reduced to a lower-dimensional space using Uniform Manifold Approximation and Projection (UMAP) [29]. Subsequently, HDBSCAN is employed to identify clusters of semantically similar hashtags within each image cluster. BERTopic effectively captures the nuanced semantic relationships within the hashtags, producing topics characterized by influential words and their associated weights. These weights indicate the relevance of each word to the respective topic, ensuring that the generated topics are both meaningful and representative of the underlying data.

4.2.2 Semi-Supervised Topic Modeling. The semi-supervised approach incorporates prior knowledge to guide the topic generation process, enhancing the relevance and specificity of the resulting topics. For this purpose, we utilized BERTopic's capability to incorporate seed words, allowing the model to generate topics that converge toward predefined themes. In this study, we seeded the model with predefined topics related to #health, #healthylifestyle, and #covid19. These seed words were selected based on their relevance to the main hashtags and the research objectives. BERTopic leverages these seeds to influence the topic distribution, ensuring that the generated topics align with the specified themes. This targeted approach enables a more focused analysis of how specific themes manifest across different locations and textual features. The semi-supervised technique was applied directly to the three textual features—BLIP captions, user hashtags, and user captions—allowing for a comparative analysis of topic relevance and coherence across different sources of textual information. By examining the generated topics, we assessed which textual feature provided the most pertinent and insightful representation of the main hashtags, as well as how these topics varied across different geographical locations.

4.3 Preprocessing of Textual Data

Prior to conducting topic modeling, we implemented a thorough preprocessing pipeline to standardize and optimize the input data quality. Our preprocessing workflow consisted of several key steps. First, we performed tokenization and basic text cleaning by converting all text to lowercase, removing punctuation marks, and eliminating common stopwords that don't contribute meaningful content. Next, we applied word normalization techniques, including lemmatization, to convert words to their dictionary base forms (e.g., "running" to "run," "better" to "good") and employed stemming from reducing words to their root forms (e.g., "connection" to "connect"). The combination of these preprocessing techniques was essential for creating consistent input data, reducing noise in the text, improving the accuracy of subsequent analysis, and enhancing the quality of both similarity measurements and topic modeling results. This comprehensive preprocessing approach ensured that our text analysis could identify meaningful patterns and topics more effectively.

4.4 Implementation Details

All models and algorithms were implemented using Python, leveraging established libraries and frameworks to facilitate efficient and effective processing. The BLIP model was accessed via the Hugging Face Transformers library, ensuring seamless integration and utilization of pre-trained weights. Fine-tuning of the BLIP model was performed using PyTorch, with the model trained on a subset of our image-caption pairs to enhance its performance on domain-specific data. Vision Transformer (ViT) and BERTopic were similarly implemented using appropriate libraries, with HDBSCAN and UMAP accessed through their respective Python packages.

The computational infrastructure comprised high-performance GPUs to handle the intensive processing requirements of transformer-based models and large-scale clustering operations. This setup was essential to manage the extensive datasets involved in the study, enabling timely and efficient analysis.

5 RESULTS

In the first part of the study, we computed three similarity scores among the three textual features: user caption vs. hashtags (Sim1), user caption vs. BLIP caption (Sim2), and hashtags vs. BLIP caption (Sim3). Figure 8 shows the distribution of these similarity scores. BLIP captions provide richer and more context-aware descriptions of the images, potentially allowing for improved alignment between the image content and user-generated text.

Our results indicate that the relationship between user captions and hashtags (Sim1) is the strongest. The highest Sim1 score achieved is 0.80 for a post where the user caption and hashtags are highly descriptive of the same content, as shown in Table 3. The average Sim1 score is 0.34, confirming that user-generated text features—crafted by the same author—tend to share a substantial semantic overlap.

In contrast, the semantic similarities involving BLIP captions (Sim2 and Sim3) are notably lower. Despite BLIP's advanced vision-language modeling, these automatically generated captions focus on describing the physical attributes of the scene rather than capturing the subjective, context-rich elements that users often include in their captions and hashtags. As a result, both Sim2 (caption vs. BLIP caption) and Sim3 (hashtags vs. BLIP caption) show average similarity scores below 0.1, with their highest values being 0.77 and 0.65, respectively. Interestingly, Sim2 and Sim3 values closely resemble each other, highlighting that BLIP captions differ substantially from both user captions and hashtags in a similar manner.

Notably, the lowest value across all similarities is zero, reflecting instances where the BLIP caption describes objects or scenes absent in user-generated text or where user captions and hashtags reference abstract concepts not directly visible in the image. While BLIP's improved contextual generation provides more nuanced image descriptions than previous captioning models, these findings suggest that user-generated text still encapsulates personal, thematic, or time-sensitive narratives that remain difficult for automated models to fully replicate.

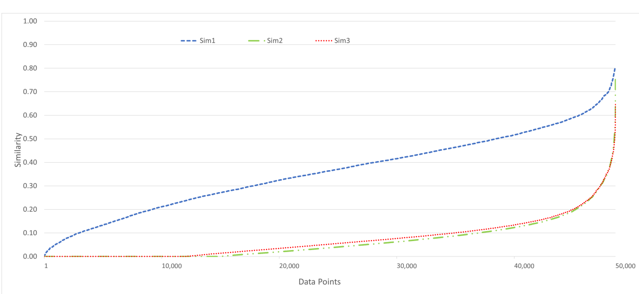

Figure 8: The similarity scores computed between the three textual features.

5.1 Unsupervised Topic Modeling

For the main hashtags—#health, #healthylifestyle, and #covid19—our revised approach employed Vision Transformer (ViT) embeddings and HDBSCAN clustering to group images based on visual similarity. BERTopic was then applied to hashtags within each cluster to extract coherent topics. This pipeline improved upon earlier methods by automatically determining the number of clusters and capturing more subtle, context-rich embeddings of images.

5.1.1 Health-related Topics. When examining #health posts, distinct themes emerged across different countries. The United States produced topics related to exercise routines, dietary patterns (including references to superfoods or keto diets), and localized health initiatives. Great Britain showcased similar interests in balanced nutrition and wellness communities, but also included local references (e.g., particular fitness studios or health campaigns). India's health-related topics frequently emphasized yoga, holistic practices, and community wellness groups. Indonesia leaned towards mindfulness, meditation, and indigenous health concepts, while Brazil balanced discussions around exercise trends (e.g., Zumba, calisthenics) and natural foods with local linguistic expressions of wellness.

5.1.2 Healthy Lifestyle Topics. For #healthylifestyle posts, topics varied widely but consistently referenced personal well-being, family-oriented health choices, the importance of mental health, and sustainable living habits. The United States and Great Britain often highlighted meal planning, vegan options, and active lifestyles integrated into daily routines. India and Brazil generated a blend of modern and traditional approaches, including references to cycling communities, outdoor sports, and local health markets. Indonesia showcased yoga retreats, spa cultures, and naturally sourced health remedies.

5.1.3 COVID-19-Related Topics. With #covid19 posts, a universal set of pandemic-driven themes surfaced: masks, social distancing, stay-at-home activities, and public health updates. Despite this global commonality, each region integrated its cultural lens. For instance, India's topics included references to local health advisories and community-driven initiatives (e.g., volunteer groups), while Brazil mentioned local campaigns and resilience stories. The United States and Great Britain introduced discussions around public policies, vaccination drives, and community support networks, while Indonesia reflected on local preventive measures, alternative therapies, and religious or cultural interpretations of well-being during the pandemic. Table 1 illustrates examples of top topics derived from the #health hashtag for each country, including representative keyword sets that emerged from BERTopic. These samples highlight the interplay between universal health interests (fitness, nutrition, mental health) and region-specific cultural elements.

These topics demonstrate that although global health themes are consistent across geographies—such as the emphasis on exercise, balanced diets, and mental health—regional cultural references, local diets, and linguistic diversity also shape health discourse on social media.

Table 3: Examples of similarities between the three textual features

No.	Post	Caption	Hashtags	BLIP caption	Sim1	Sim2	Sim3
1	[16]	Out for breakfast @marmaladerhos this morning,.... a almond milk gingerbread latte.	#breakfast #morning.... #foodie #yummy #tasty	A plate with a sandwich and a cup of coffee.	0.800	0.380	0.400
2	[38]	Thanks @getoutandliveout for putting this on and to Pat for the excellent instruction.	#getoutand-live.... #healthy-lifestyle #yogi	A man is playing frisbee in a field.	0.536	0.042	0.021
3	[4]	Another sad thing about Covid....this guest bathroom and how it looked before!	#bathrooms-ofinstagram #bath-room.... #covid_19	A bathroom with a sink, mirror, and toilet.	0.154	0.103	0.653
4	[36]	COOKED VEGETABLE and BROTH (A HEALTHY DINNER)	#dohafood #mydohalife.... #do-haqatar #thepearlqatar	A bowl of broccoli and carrots with broth.	0.285	0.774	0.287

5.2 Semi-Supervised Topic Modeling

In the second technique, we employed a semi-supervised approach to generate topics from the three textual features—BLIP captions, user captions, and user hashtags—after filtering the data by geographical location. This approach was guided by predefined seed words associated with our primary themes (#health, #healthylifestyle, and #covid19), enabling the model to produce more focused, thematically consistent topics. As anticipated, terms related to the COVID-19 pandemic were prevalent across all locations, reflecting the global nature of the crisis. Words such as "covid," "pandemic," "time," "week," "cases," and "virus" emerged frequently, along with user-generated trends and social cues like "fff" (follow for follow) and platform-specific chatter ("WhatsApp"). Although these terms help capture the pandemic's ubiquity, many of them are too general to represent coherent subtopics on their own.

By integrating BLIP captions into the semi-supervised process, we aimed to discover location-specific nuances while maintaining thematic alignment. Similar to the unsupervised technique, BLIP captions often introduced descriptive yet straightforward concepts based on the visual elements in each image (e.g., "holding," "sitting," "standing," "man," "woman," "white," "clock," "tie"). However, unlike the user-generated textual features, BLIP captions did not consistently add deeply contextualized or culturally nuanced terms. Instead, they tended to anchor topics in the visual realm, identifying tangible objects or scenarios that could loosely connect to the seeded health or pandemic-related themes. Nevertheless, the semi-supervised approach occasionally surfaced unique topics that diverged from the purely health-oriented seeds. These included references to emerging social movements, lifestyle changes, and societal shifts during the pandemic. For instance, words like "fashion," "friends," "auto detailing," "blm," and "lockdown" appeared in certain clusters, reflecting how users worldwide documented diverse aspects of their daily lives through both images and text. While some of these words are not strictly related to health or the pandemic, their presence highlights the semi-supervised model's ability to capture broader contextual elements that emerge organically within user-generated data.

Table 4: Health Topics Popularity Across Countries

Subtopic	US	GB	India	Indonesia	Brazil
Bodybuilding	+++	++	+	++	+
Self Care	++	++	+	•	•
Fitness	++	+	++	-	+++
Nutrition	•	++	•	-	-
Community	-	-	+	+	++
Lifestyle	•	++	-	++	++
Vegan	+++	•	•	•	+++
Weight Loss	•	•	++	+	+
Mental Health	++	+++	+	+++	+
Medical Awareness	+	-	++	-	++
Training	++	•	•	+	-
Nature	+	+	+++	+	-
Running	++	•	-	-	+

Note: - = no mention (0%); • = low (1-30%); + = moderate (31-60%); ++ = high (61-80%); +++ = very high (81-100%)

Table 5: Healthy Lifestyle Topics Popularity Across Countries

Subtopic	US	GB	India	Indonesia	Brazil
Face Mask	+++	+++	+	++	++
Family	+++	++	+	•	-
Photography	++	++	-	•	++
Travel	++	++	-	-	-
Nature	++	+	-	-	•
Love	+	+	+	++	++
Services	++	++	++	+	+
Politics	-	-	-	+	•
Sports	+	+	-	•	-
Fashion	+	•	•	++	-
Protests	++	-	-	+	-
Safety	++	++	++	+	++
Awareness	+	+	+	+	-
Fitness	•	-	+	+	-

Note: - = no mention (0%); • = low (1-30%); + = moderate (31-60%); ++ = high (61-80%); +++ = very high (81-100%)

6 DISCUSSION

This study provides insights into how health-related discourse manifests across Instagram images and their associated textual features—user captions, user hashtags, and BLIP-generated captions—across multiple geographical locations. Our findings indicate that user captions and hashtags generally exhibit stronger semantic alignment with one another than either does with the automatically generated BLIP captions. This makes intuitive sense, as user-generated texts originate from the same source and often

Table 6: COVID-19 Topics Popularity Across Countries

Subtopic	US	GB	India	Indonesia	Brazil
Face Masks	++	++	++	++	++
COVID Safety	++	++	+	++	+
Awareness	+	•	-	•	-
Mental Health	+	+	++	+	•
Sanitation	++	+	++	+	•
Photography	•	•	-	-	•
Health	+	•	+	•	•
Pandemic	-	+	•	-	•
Travel	++	+	-	-	•
Family	++	•	•	+	-
Safety	++	+	+	•	+
Services	+	+	•	-	-
Quarantine Life	•	-	•	•	•
Politics	-	-	-	•	•

Note: - = no mention (0%); • = low (1-30%); + = moderate (31-60%); ++ = high (61-80%); +++ = very high (81-100%)

reflect personal experiences, localized meanings, and thematic connections that transcend the literal content of the image. In contrast, BLIP captions, while more contextually aware than many earlier vision-language models, still focus primarily on describing visible elements in the image (e.g., "woman," "yoga mat," "fountain") rather than capturing culturally or contextually nuanced information.

The thematic analysis of health-related hashtags—#health, #healthylifestyle, and #covid19—revealed both global consistencies and local divergences. Worldwide, discussions around #covid19 were relatively homogeneous, centering on protective measures like mask-wearing, health guidelines, and home-based activities. However, for #health and #healthylifestyle, there was notable variation among countries. In some regions, health was strongly tied to physical exercise, bodybuilding, and nutrition, while in others it centered on mindfulness, yoga, or traditional wellness practices. This highlights the culturally dependent nature of health discourse, suggesting that what constitutes "health" and "healthy living" is shaped by local customs, values, and environmental contexts.

The tables summarizing topic popularity also underscore the fluidity of these concepts. Terms related to mental health, community, and nature appeared across regions, but their relative importance varied. For instance, mental health themes were more emphasized in certain locations, while others highlighted body-centric fitness or the importance of a plant-based diet. Similarly, #covid19 discussions included universal terms like "face masks" and "covid safety," yet each country's discourse incorporated localized elements reflecting distinct phases of the pandemic or regional public health strategies.

Our findings from both unsupervised and semi-supervised topic modeling approaches were largely consistent. Although semi-supervised modeling, guided by thematic seeds, did not dramatically alter the nature of the topics identified, it helped highlight relevant health and lifestyle themes within a complex and diverse corpus. BLIP captions integrated into the semi-supervised approach provided a foundational understanding of image content, yet they rarely introduced deeper cultural or contextual cues. This suggests that, while advanced vision-language models can offer valuable baseline descriptors, they still rely on user-generated textual features for richer thematic interpretation.

From an applied perspective, these insights can inform strategies for content curation, public health messaging, and brand engagement on social media. Hashtags, in particular, proved effective for identifying thematically coherent topics, offering researchers and practitioners a useful tool for labeling and understanding large-scale visual data. However, careful filtering and validation are necessary, as hashtags can sometimes be off-topic or strategically selected for popularity rather than accuracy.

7 CONCLUSION AND FUTURE WORK

This study aimed to enhance our understanding of how health-related topics, as represented by the hashtags #health, #healthylifestyle, and #covid19, manifest on Instagram. We examined a globally diverse dataset and found that, while the discourse around #covid19 was broadly similar across locations—emphasizing protective measures, awareness, and well-being—topics under #health and #healthylifestyle displayed marked variations. Local cultures, languages, and health ideologies influenced how users represented health through images and textual elements. Our analysis revealed that user-generated hashtags offered the most thematically coherent insights into underlying health narratives, surpassing both user captions and machine-generated BLIP captions. Hashtags often served as concise, context-rich indicators of image content, health practices, and localized wellness trends. Despite the enhanced contextual capabilities of the BLIP model, automatically generated captions remained more literal and less culturally nuanced than human-created text. Nonetheless, BLIP captions proved valuable for establishing a baseline understanding of image content and can potentially support automated content analysis pipelines. The complementary strengths of user hashtags, user captions, and BLIP captions highlight the importance of integrating multiple modalities to gain a holistic view of health discourse on social media.

Several avenues exist for future research. First, a user study could provide critical insights into the value and accuracy of extracted topics, hashtags, and captions from a human perspective, thereby informing the refinement of automated methods. Additionally, further exploration of advanced vision-language models or alternative captioning approaches could enhance the granularity and cultural sensitivity of generated captions. Incorporating other deep learning architectures or fine-tuned language models may yield captions that better capture subtle contextual cues and localized health practices. Moreover, extending the analysis to non-English textual data and geographically distinct cultural contexts would yield a more comprehensive global perspective. Finally, improving data preprocessing and filtering techniques for hashtags—such as removing spam or irrelevant tags—could bolster the reliability of these textual features as proxies for image content. Together, these improvements and extensions can deepen our understanding of how health-related communication evolves online and guide the development of more robust, context-aware systems for labeling and analyzing large-scale social media imagery.

ACKNOWLEDGMENTS

REFERENCES

[1] Omnicore Agency. 2022. Instagram by the Numbers (2020): Stats, Demographics & Fun Facts. https://www.omnicoreagency.com/instagram-statistics/.

[2] Mariam F. Alkazemi, Jeanine P. D. Guidry, Ezaddeen Almutairi, and Marcus Messner. 2020. #Arabhealth on Instagram: Examining Public Health Messages to Arabian Gulf State Audiences. *Health Communication* 37, 1 (2020), 39–47.

[3] A. Argyrou, S. Giannoulakis, and N. Tsapatsoulis. 2018. Topic modelling on Instagram hashtags: An alternative way to Automatic Image Annotation?. In *13th International Workshop on Semantic and Social Media Adaptation and Personalization*. 61–67.

[4] Hirosh Asuhar. 2020. Instagram post by fooddiners. https://www.instagram.com/p/CF2RsHQlE0u/.

[5] Best-Hashtags. 2022. Hashtags for #health. http://best-hashtags.com/hashtag/health/.

[6] A. Bhattacharya. 2020. For Indians under lockdown, social media is the go-to source for news and entertainment. https://scroll.in/article/958805/for-indians-under-lockdown-social-media-is-the-go-to-source-for-news-and-entertainment.

[7] Elvira Bolat. 2020. Why the UK government is paying social media influencers to post about coronavirus. *The Conversation* (Sept. 2020).

[8] Steve Chang. 2016. Instagram post data analysis. *arXiv* (2016).

[9] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health Surveillance* 6, 2 (May 2020), e19273.

[10] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A. Munson. 2017. *When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating.* Association for Computing Machinery, New York, NY, USA, 1674–1687. https://doi.org/10.1145/3025453.3025747

[11] Rachel Cohen, Lauren Irwin, Toby Newton-John, and Amy Slater. 2019. #bodypositivity: A content analysis of body positive accounts on Instagram. *Body image* 29 (2019), 47–57.

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929. https://arxiv.org/abs/2010.11929 Accessed: [date of access].

[13] FashionNetwork.com US. 2021. Instagram was UK's top social media platform in lockdown, key for shopping. https://us.fashionnetwork.com/news/Instagram-was-uk-s-top-social-media-platform-in-lockdown-key-for-shopping,1237799.html.

[14] Bruce Ferwerda and Marko Tkalcic. 2018. Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features?. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization.* Association for Computing Machinery, 157–161.

[15] A. Fiallos, K. Jimenes, C. Fiallos, and S. Figueroa. 2018. Detecting Topics and Locations on Instagram Photos. In *2018 International Conference on eDemocracy eGovernment.* 246–250.

[16] Foodie. 2020. Instagram post by goodfoodddsss. https://www.instagram.com/p/CFzEhF0l81u/.

[17] Instagram from Meta. 2022. About the Instagram Company. https://about.instagram.com/about-us.

[18] S. Giannoulakis, N. Tsapatsoulis, and K. Ntalianis. 2017. Identifying Image Tags from Instagram Hashtags Using the HITS Algorithm. In *IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology.* 89–94.

[19] Maarten Grootendorst. 2020. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2107.00630. https://arxiv.org/abs/2107.00630 Accessed: [date of access].

[20] Muhammad Habibi and Puji Winar Cahyo. 2019. Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform. *Indonesian Journal of Computing and Cybernetics Systems* 13, 4 (2019), 399–408.

[21] Instaloader. 2020. Download Instagram Photos and Metadata. https://instaloader.github.io/.

[22] Maged N. Kamel Boulos, Dean M. Giustini, and Steve Wheeler. 2016. Instagram and WhatsApp in Health and Healthcare: An Overview. *Future Internet* 8, 3 (July 2016), 37.

[23] Yunhwan Kim and Jang Hyun Kim. 2020. Using photos for public health communication: A computational analysis of the Centers for Disease Control and Prevention Instagram photos and public responses. *Health Informatics Journal* 26, 3 (Sept. 2020), 2159–2180.

[24] Tama Leaver and Tim Highfield. 2018. Visualising the ends of identity: Pre-birth and post-death on Instagram. *Information, Communication & Society* 21, 1 (2018), 30–45.

[25] Namyeon Lee, Kelli Buchanan, and Mansoo Yu. 2020. Each post matters: A content analysis of #mentalhealth images on Instagram. *Journal of Visual Communication in Medicine* 43, 3 (2020), 128–138.

[26] Chushi Li. 2020. Facebook vs. Instagram in 2020: All You Need to Know. https://www.socialbakers.com/blog/instagram-vs-facebook-advertising-differences-and-best-practices.

[27] Junnan Li, Dongxu Li, Caiming Yu, and Bin Li. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. arXiv preprint arXiv:2201.12086. https://arxiv.org/abs/2201.12086 Accessed: [date of access].

[28] Leland McInnes, John Healy, and Steve Astels. 2017. HDBSCAN: Hierarchical Density Based Clustering. *The Journal of Open Source Software* 2, 11 (2017), 205. https://doi.org/10.21105/joss.00205

[29] Leland McInnes, John Healy, and James Melville. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint arXiv:1802.03426. https://arxiv.org/abs/1802.03426 Accessed: [date of access].

[30] Sachin Muralidhara and Michael J. Paul. 2018. #Healthy Selfies: Exploration of Health Topics on Instagram. *JMIR Public Health and Surveillance* 4, 2 (2018), e10150.

[31] WHO on Instagram. 2022. World Health Organization (@who) Instagram photos and videos. https://www.instagram.com/who/.

[32] Esteban Ortiz-Ospina. 2019. The rise of social media. https://ourworldindata.org/rise-of-social-media.

[33] Andrew G. Reece and Christopher M. Danforth. 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science* 6, 1 (Aug. 2017), 15.

[34] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv preprint arXiv:1908.10084. https://arxiv.org/abs/1908.10084 Accessed: [date of access].

[35] Alessandro Rovetta and Akshaya Srikanth Bhagavathula. 2020. Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags. *Journal of Medicine Internet Research* 22, 8 (Aug. 2020), e20673.

[36] Azucena Saavedra. 2020. Instagram post by azucena.saavedra. https://www.instagram.com/p/CFX8u79DOgn/.

[37] Sara Santarossa, Paige Coyne, Carly Lisinski, and Sarah J. Woodruff. 2019. #fitspo on Instagram: A mixed-methods approach using Netlytic and photo analysis, uncovering the online discussion and author/image characteristics. *Journal of Health Psychology* 24, 3 (2019), 376–385.

[38] Elicia Simon. 2020. Instagram post by coacheliciasimon. https://www.instagram.com/p/CFyXz5LnQNT/.

[39] Amandeep Singh, Malka N. Halgamuge, and Beulah Moses. 2019. An Analysis of Demographic and Behavior Trends Using Social Media: Facebook, Twitter, and Instagram. *Social Network Analytics* (2019), 87.

[40] spaCy. 2016. Spacy-langdetect · spaCy Universe. https://spacy.io/universe/project/spacy-langdetect.

[41] Sana Suse, Afra Mashhadi, and Spencer A Wood. 2021. Effects of the COVID-19 Pandemic on Park Visitation Measured by Social Media. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21)*. Association for Computing Machinery, New York, NY, USA, 179–182. https://doi.org/10.1145/3462204.3481754

[42] Instagram Business Team. 2019. How to sell your products on Instagram. https://business.instagram.com/blog/how-to-sell-your-products-on-instagram?locale=en_GB.