An Automated Analysis of YouTube Videos on Childbirth: Quality, Sentiment, and Engagement

Chen Mordehai *Department of Software and Information Systems Engineering*
Ben-Gurion University of the Negev
mordeche@post.bgu.ac.il

1 Introduction

Childbirth is a critical life event that often prompts individuals to seek information and support online. YouTube, one of the world's largest video-sharing platforms, has emerged as a major resource for childbirth-related content. While it offers diverse perspectives—ranging from personal birth stories to professional medical advice—concerns persist about the reliability and quality of this information [4, 6, 7]. Given the platform's widespread reach and the potential impact of its content on public health, there is a growing need for scalable methods to analyze, evaluate, and categorize the information presented in childbirth videos. In this work, we propose an automated pipeline that not only collects and processes a large dataset of YouTube videos but also applies natural language processing (NLP) techniques to: Identify and extract medical recommendations, Assess the sentiment of transcribed video content, Classify videos according to content type (e.g., medical recommendations, personal experiences), and Examine thumbnail images for additional insights, such as caption-based sentiment. By integrating multiple NLP classifiers, topic modeling approaches, and image-captioning pipelines, our methodology provides a holistic view of childbirth-related content on YouTube. Specifically, we address the following research questions: **(1)** How is childbirth-related information distributed across different queries and topics? **(2)** What is the overall quality and sentiment of medical recommendations found in these videos?

To foster transparency and reproducibility, we make our entire **source code publicly available on GitHub** at `https://github.com/ChenMordehai/YouTube-Childbirth-Analysis`.

2 Related Work

The proliferation of health-related content on social media platforms, particularly YouTube, has prompted numerous studies evaluating the quality and reliability of such information. Yet the quality and reliability of this content remain subjects of significant concern. A number of studies have explored these issues using systematic methodologies and standardized evaluation tools. For instance, O'Kelly et al. (2024) [6] conducted a detailed analysis of opioid use disorder content by retrieving the top five most viewed videos per search term during a systematic search conducted between June and July 2022. Their evaluation, based on the Global Quality Scale (GQS), JAMA Benchmark Criteria, and DISCERN tool, revealed that the overall quality of these videos was low (with a median GQS of only 2), emphasizing the urgent need for more evidence-based material from public health experts.

In a similar effort to assess online health information, Li et al. (2024) [4] investigated videos related to tachycardia. By initially gathering 150 videos with the keyword "tachycardia" and narrowing the selection to 113 videos that met their inclusion criteria, they applied a suite of metrics—including JAMA, modified DISCERN, GQS, and a tachycardia-specific scale—to determine content quality. Although their results highlighted that longer videos and those produced by academic or healthcare professionals tended to score higher, the overall reliability and educational value of the tachycardia-related content remained suboptimal.

Expanding the focus to pediatric health, Rudisill et al. (2023) [7] examined YouTube as a source of information on pediatric scoliosis. Their analysis of 153 unique videos showed that despite a substantial cumulative view count (28.5 million views), the reliability and educational quality—as measured by JAMA, GQS, and a Pediatric Scoliosis-Specific Score (PSS)—were generally low.

These investigations collectively highlight a common trend: while YouTube serves as a significant platform for disseminating health information across various medical domains, the overall quality and reliability of this information remain inconsistent. In their study, Lum-Wang et al. (2024) [5] evaluated how patients with lung cancer utilize online resources, including social media, as complementary to clinical discussions. The study found that while patients frequently turn to the internet for additional information, the quality and reliability of the content vary widely. This inconsistency can lead to misinformation, affecting patient perceptions and potentially influencing treatment decisions. Similar concerns are echoed in the work by Aurlene et al. (2024) [1] and van Dijk et al. (2024) [8]. In their study, Aurlene et al. evaluated 120 YouTube videos on denture care, assessing their usefulness and reliability. The findings revealed that 65.8% of the videos were classified as having low usefulness, and 74.2% scored 2 or below on the modified DISCERN tool, indicating low reliability. Notably, videos uploaded by dentists or dental hygienists had significantly higher usefulness scores compared to those from other sources. However, such professional contributions were limited, underscoring a gap in authoritative content on the platform. Similarly, van Dijk et al. investigated 70 YouTube videos concerning external cephalic version (ECV), a procedure to turn breech-presenting fetuses to a head-first position. Their analysis showed that only 14% of the videos were deemed useful, with all useful videos originating from educational channels or healthcare professionals. In contrast, over half of the not useful videos were uploaded by birth attendants and vloggers, who, despite achieving higher audience engagement, often lacked comprehensive medical information. This disparity highlights the scarcity of professional input in online content related to ECV. Moreover, research on digital media usage among specific patient populations underscores both the widespread reliance on digital platforms for health information and the potential for collaboratively created content to enhance patient education. For instance, a survey conducted by Hussain et al. (2023) [3] explored the use of digital media during pregnancy and birth among women in Western Victoria, Australia. The study found that pregnancy applications were the most utilized medium, followed by websites, social media, YouTube, podcasts, online discussion forums, and labor applications. The primary motivations for digital media use included information seeking, connecting with others for social support, and reassurance. These findings highlight the integral role digital media plays in the pregnancy experience and suggest that healthcare institutions should integrate digital media into antenatal education and care to meet the evolving needs of women. In a similar vein, the SIMBA CoMICs initiative led by Elhariry et al. (2024) [2] demonstrated the effectiveness of training medical students to create accurate medical information for social media dissemination. This project focused on producing evidence-based, peer-reviewed short videos on polycystic ovary syndrome (PCOS) and thyroid disorders, which were then shared across platforms like TikTok, Instagram, YouTube, and Twitter. The initiative not only aimed to combat misinformation and improve health literacy but also provided medical students with valuable experience in content creation and public engagement. The study reported positive audience engagement metrics and highlighted the potential of such collaborative efforts in enhancing the quality of health information available online.

3 Methods

3.1 Data Collection and Preprocessing

Query Generation and Video Scraping We generated 446 search queries pertinent to childbirth topics[1]. The generation of these queries was facilitated using `OpenAI's ChatGPT o3-mini model`[2], which assisted in formulating diverse and relevant search terms. Subsequently, we employed Python's `scrapetube`[3] package to scrape YouTube for videos corresponding to each query. For every retrieved video, we collected metadata including `video_id`, `title`, `channel_name`, `channel_id`, `published_time`, `duration`, `views`, `video_url`, `channel_url`, and `thumbnail_url`.

Audio Extraction and Transcription Using the `YoutubeDL`[4] package, we downloaded the audio track for each video. The audio was then transcribed automatically with the `insanely-fast-whisper`[5] pack-

[1] https://github.com/ChenMordehai/YouTube-Childbirth-Analysis/blob/main/Appendix/all_generated_queries.txt

[2] https://openai.com/index/openai-o3-mini/

[3] https://pypi.org/project/scrapetube/

[4] https://github.com/ytdl-org/youtube-dl

[5] https://github.com/Vaibhavs10/insanely-fast-whisper

age, ensuring rapid processing of a large volume of data.

3.2 Data Analysis

Preprocessing and Filtering Removed videos with a duration of less than 15 seconds. Used the `langdetect`[6] library to filter transcriptions, ensuring only English-language videos were retained. Converted relative timestamps (e.g., "2 years ago") to absolute publication dates.

Exploratory Data Analysis Computed descriptive statistics, including video duration distribution, view count distribution, and publication year analysis. Performed sentiment analysis on the transcribed text using the `TextBlob`[7] library, categorizing content into positive, neutral, or negative sentiment.

Medical Recommendation vs. Personal Experience Classification Employed a `T5-based`[8] classifier to categorize transcriptions into *Medical Recommendation*, *Personal Experience*, or *Other*. Analyzed the distribution of classifications and their relationship to engagement metrics such as view counts and video duration.

Topic Modeling and Query Clustering Applied Latent Dirichlet Allocation (LDA) on the transcriptions to identify prevalent discussion topics. Used sentence embeddings with K-means clustering to group related search queries.

Recommendation Quality and Classification Extracted individual recommendation sentences from transcriptions using NLP-based sentence tokenization. Used a `T5-based` classifier to categorize clinical recommendations as *correct*, *ambiguous*, or *false*. Analyzed the distribution of recommendation types and their association with video metadata (e.g., source credibility, engagement levels).

Pattern and Correlation Analysis Examined relationships between video attributes (e.g., duration, publication year, views) and the accuracy of medical recommendations. Explored correlations between sentiment polarity, recommendation correctness, and user engagement.

[6]`https://pypi.org/project/langdetect/`

[7]`https://textblob.readthedocs.io/en/dev/`

[8]`https://huggingface.co/google-t5/t5-base`

Additional Analyses *Thumbnail Image Analysis:* Used an image captioning pipeline(**nlpconnect/vit-gpt2-image-captioning**[9]) to generate descriptive text for video thumbnails, followed by sentiment analysis on these captions. *Medication Extraction:* Applied the `Med7`[10] NLP model to extract medication mentions from transcriptions and analyzed their frequency of occurrence.

4 Result Analysis

4.1 Overall Statistical Analysis

Our dataset comprises a large collection of childbirth-related YouTube videos, reflecting diverse queries, publication dates, and channel sources. A total of 16,516 unique channels are represented in our dataset, indicating that childbirth content is produced by a wide range of creators. These channels include professional healthcare providers, parent-focused vloggers, and general interest channels, among others. Figure 1 shows that short videos (fewer than 10 minutes) dominate the dataset, numbering nearly 18,000. Mid-length videos (10–30 minutes) also constitute a substantial portion ($\sim$ 8,000 videos), while longer videos (30 minutes or more) become progressively less common. This skew suggests that content creators frequently opt for concise formats to capture viewer attention.

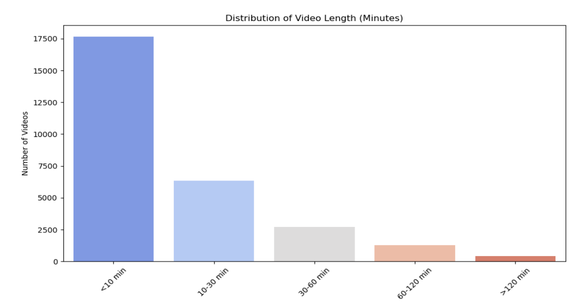

Figure 1: Distribution of Video Length (Minutes).

The number of videos per query varies, as il-

[9]`https://huggingface.co/nlpconnect/vit-gpt2-image-captioning`

[10]`https://github.com/kormilitzin/med7`

lustrated in Figure 2. There are 238 videos per query(median). Additionally, a separate analysis of the top 10 hashtags by total views (Figure 3) reveals that certain topics gain high view counts, while simultaneously having a large amount of related videos in the dataset.

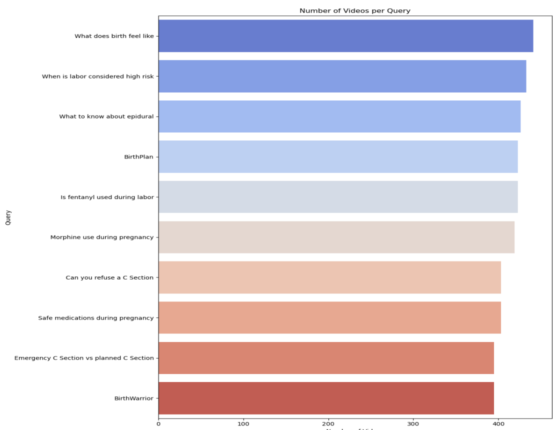

Figure 2: Number of Videos per Query(Top 10))

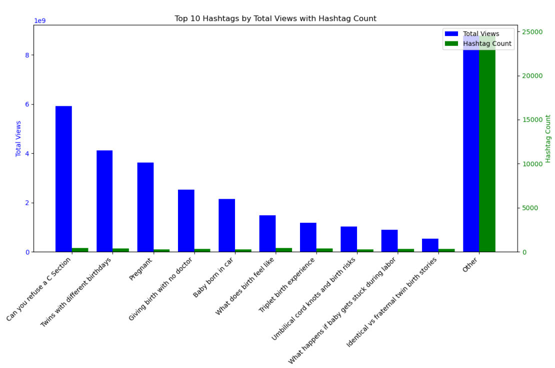

Figure 3: Top 10 Hashtags by Total Views with Hashtag Count.

4.2 Semantic Analysis

As shown in Figure 4, the sentiment distribution clusters around a slightly positive mean, indicating that the majority of videos tend to have neutral-to-positive language. Figure 5 suggests that many childbirth-related videos maintain a reassuring or uplifting tone.

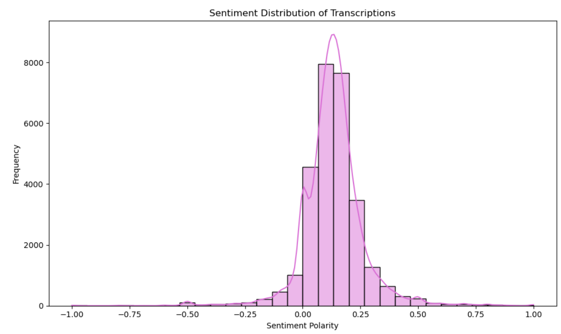

Figure 4: Sentiment Distribution of Transcriptions.

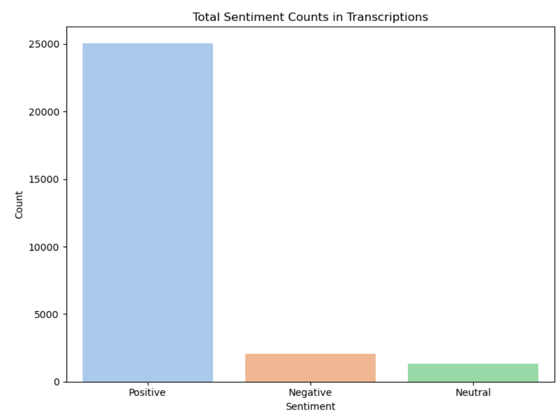

Figure 5: Total Sentiment Counts in Transcriptions.

Figure 6 shows that *Other* content is the most common category(13,527), followed by *Personal Experience*(9,507), and then *Medical Recommendation*(5,402). This distribution implies that, while personal experiences are frequently shared, a substantial portion of the videos also include more generalized or unrelated content, and a smaller (but still significant) fraction focuses on direct medical advice.

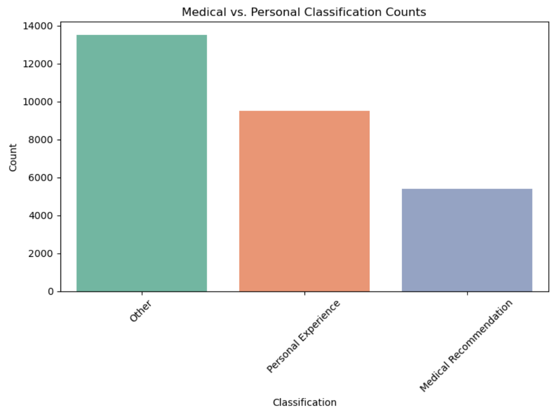

Figure 6: Medical vs. Personal Classification Counts.

Within the subset of *Medical Recommendation* content, the vast majority of medical recommendations are classified as balanced. Very few videos focus solely on pros or solely on cons. We assessed the sentiment of video thumbnails. Figure 7 shows that the sentiment of thumbnails generally centers around neutral or slightly negative values, this contrast between transcription sentiment and thumbnail sentiment can be attributed to stylistic choices or design elements used in thumbnails (e.g., medical settings or serious expressions).

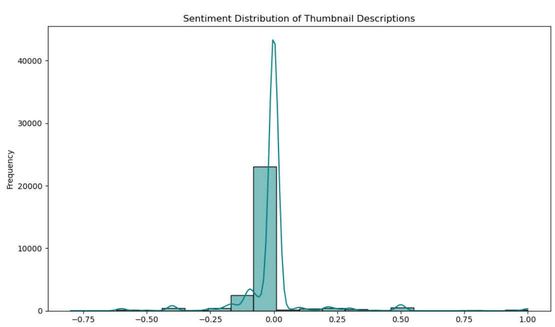

Figure 7: Sentiment Distribution of Thumbnail Descriptions.

Figure 8 illustrates the top ten medications (or medication-related terms) identified in the video transcriptions. Notably, the highest-frequency mentions relate to opioids (e.g., "opioid," "opioids," "fentanyl," and "morphine"), which collectively suggest that pain management is a prominent topic in childbirth-related videos. Oxytocin and antibiotics also feature prominently, underscoring discussions around labor induction, postpartum care, and infection control. The appearance of terms like "gweithio" or "butter" may stem from linguistic variations, colloquial references, or potential misclassifications within the model. Overall, the high frequency of opioid-related terms highlights a particular interest in analgesic strategies during labor and postpartum recovery, reflecting broader concerns about pain relief and medication safety in childbirth contexts.

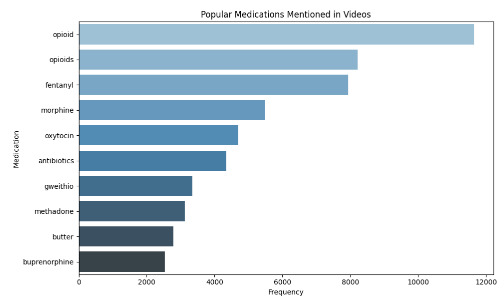

Figure 8: Popular Medications Mentioned in Videos.

5 Conclusions and Future Directions

This study outlines an automated pipeline for collecting and analyzing YouTube videos on childbirth. The analysis indicates that while transcribed content generally shows a neutral tone, the visual elements (i.e., thumbnail descriptions) tend to be more reserved. The current methodology allows for distinguishing between personal experiences and medical recommendations; however, limitations in classification accuracy and interpretation of visual cues remain. Future work should focus on refining classification models and improving the robustness of sentiment analysis. Additionally, extending the approach to other digital platforms and incorporating updated data may help provide a more comprehensive view of online childbirth information.

6 Acknowledgment

The authors acknowledge the assistance of ChatGPT[11] and contributions from the Stack Overflow community[12], which were instrumental in addressing technical issues and refining the phrasing of this paper.

References

[1] Nesa Aurlene, Sabiha Shaheen Shaik, Virginia Dickson-Swift, and Santosh Kumar Tadakamadla. Assessment of usefulness and reliability of youtube™ videos on denture care. *International Journal of Dental Hygiene*, 22(1):106–115, 2024.

[2] Maiar Elhariry, Kashish Malhotra, Kashish Goyal, Marco Bardus, SIMBA Team, CoMICs, and Punith Kempegowda. A simba comics initiative to cocreating and disseminating evidence-based, peer-reviewed short videos on social media: Mixed methods prospective study. *JMIR medical education*, 10:e52924, 2024.

[3] Ilyana Mohamed Hussain, Nicki Hartney, and Linda Sweet. A survey of australian women's digital media usage in pregnancy and labour and birth. *BMC Pregnancy and Childbirth*, 23(1):688, 2023.

[4] Zexi Li, Chunyi Yan, Xiaojun Lyu, Fanghui Li, and Rui Zeng. Assessing quality and reliability of online videos on tachycardia: a youtube video-based study. *BMC Public Health*, 24(1):2620, 2024.

[5] Sandy Lum-Wang, Paris-Ann Ingledew, Hannah Carolan, Emma M Dunne, Elisa Chan, Ying Wang, Mitchell Liu, and Brandon S Chai. An evaluation of the usage of internet resources by patients with lung cancer. *Journal of Cancer Education*, pages 1–7, 2024.

[6] Bridget O'Kelly, Perry Holmes, Anna Cheng, Joshua D Lee, and Babak Tofighi. Dissemination of health content through social networks: Youtube and opioid use disorders. *Journal of Substance Use and Addiction Treatment*, 165:209475, 2024.

[7] Samuel S Rudisill, Nour Z Saleh, Alexander L Hornung, Shadi Zbeidi, Roohi M Ali, Zakariah K Siyaji, Junyoung Ahn, Michael T Nolte, Gregory D Lopez, and Arash J Sayari. Youtube as a source of information on pediatric scoliosis: a reliability and educational quality analysis. *Spine Deformity*, 11(1):3–9, 2023.

[8] Merle R van Dijk, Anne-Fleur van der Marel, Leonie E van Rheenen-Flach, Wessel Ganzevoort, Etelka Moll, Fedde Scheele, and Joost Velzel. Youtube as a source of patient information on external cephalic version: Cross-sectional study. *JMIR Formative Research*, 8(1):e50087, 2024.

[11] `https://chatgpt.com/`
[12] `https://stackoverflow.com/`