Making Sense of Sentiments for Aesthetic Plastic Surgery

Anand Choudhary
Department of Physics
Indian Institute of Technology Roorkee
India
anand_c@ph.iitr.ac.in

Erik Cambria
School of Computer Science and Engineering
Nanyang Technological University
Singapore
cambria@ntu.edu.sg

***Abstract*—With social media pervading all aspects of our life, the opinions expressed by netizens are a gold mine ready to be exploited in a meaningful way to influence all major public domains. Sentiment analysis is a way to interpret this unstructured data using AI tools. It is a well-known fact that there has been a 'Zoom Boom' in the field of aesthetic plastic surgery due to the COVID-19 pandemic and the same has put the focus of attention sharply on our appearance. Polarity detection of tweets published on popular aesthetic plastic surgery procedures before and after the onset of COVID can provide great insights for aesthetic plastic surgeons and the health industry at large. In this work, we develop an end-to-end system for the sentiment analysis of such tweets incorporating a state-of-the-art fine-tuned deep learning model, an ingenious 'keyword search and filter approach' and SenticNet. Our system was tested on a large database of 196,900 tweets and the results were visualized using affectively correct word clouds and also subjected to rigorous statistical hypothesis testing to draw meaningful inferences. The results showed a high level of statistical significance.**

***Index Terms*—Aesthetic Plastic Surgery, Sentiment Analysis**

I. Introduction

In today's era of digitalization, social media has become an indispensable part of our lives, interlinking everyone across the globe. It has transformed the Web to an interactive, user-centric one from a passive, information-centric one [1]. Consequently, today, netizens' interactions on social media constantly generate massive amounts of data related to almost everything happening in the world. The analysis of this data has the potential to generate actionable insights in various domains such as marketing, finance, politics, business, health-care, etc. However, this data is unstructured and thus, not directly machine interpretable [2].

Natural Language Processing (NLP) is a field of AI which aims to address this issue by deciphering the meaning of human language using linguistics as well as statistics. Sentiment analysis, a subfield of NLP, revolves around an affectively intelligent categorization of natural language-based text so as to effectively capture the pulse of the people in the context of some event(s) [3]. Needless to say, such an analysis can be quite helpful in (hitherto less explored) the domain of aesthetic plastic surgery as it can give the surgeons cues for pepping up people's positivity and counteracting their negativity and misinformation so as to ultimately reap rich dividends.

The recent 'Zoom Boom' in aesthetic plastic surgery has been fueled by the booming business of video conferencing solutions such as Zoom during the ongoing COVID-19 pandemic [4]. The practice of 'Work from Home' led to people connecting on video calls for hours on end while constantly viewing their own reflections in action. Consequently, people started noticing new facial lines and forehead wrinkles and the way their noses looked, etc. and started contrasting their looks with those of others while on call: a phenomenon commonly referred to as 'Zoom Dysmorphia' [5].

This further led to a ripple effect with these people also viewing their entire selves in the mirror quite often and self-judging the shape of their bodies: a kind of body dysmorphia. This apparently heightened self-consciousness and consequent dissatisfaction with their looks coupled with an incentive of extra recovery time from surgery as a result of 'Work from Home' led to a lot of people flocking towards getting aesthetic plastic surgeries done as soon as restrictions on elective surgeries were lifted. Additionally, the 'Zoom Boom' also led to a surge in the number of people taking to Twitter to express their opinions on popular aesthetic plastic surgery procedures (APSPs).

All of this gave birth to the foundational idea behind our paper: "Did the sentiments of people regarding all the popular APSPs shift increasingly in the positive direction post the onset of COVID as compared to pre-COVID or not? Also, do these trends, in any way, correlate with the actual hardcore APSP statistics pre and post the onset of COVID?" Our paper aims to investigate these through an in depth sentiment analysis of APSP tweets involving juxtaposition of the public's pre-COVID (before the onset of COVID) and post-COVID (after the onset of COVID) sentiments.

To the best of our knowledge, it is the largest such study on the sentiment analysis of APSP tweets. The main contributions of our paper are listed below.

- Design and implementation of an end-to-end system (Fig. 1) for sentiment analysis of APSP tweets corresponding to 7 of the top APSP procedures worldwide, incorporating all the steps from data collection to visualization and interpretation with Subjectivity Detection being handled by SenticNet and Polarity Detection being handled by a fine-tuned RoBERTa Base model.

- Inclusion of a crucial component for the filtration of those APSP tweets which were actually advertisements and hence, biased by utilizing regular expressions to perform a 'keyword search and filter approach' followed by further refinement using SenticNet.
- A novel technique to form *affectively* correct Word Clouds by carefully dealing with semantically antithetic terms such as 'good' and 'not good', etc. and adjusting for profanity in the APSP tweets.
- Collation of a gold standard dataset[1] of 7000 APSP tweets and their corresponding manually annotated polarities by experts in aesthetic plastic surgery.

II. Related Work

Several studies have been conducted to examine the links between media activity and the plastic surgery practice (in general) and hence, decipher the public's opinions regarding the same. Some of them [6]–[8] bring to the fore a different aspect of public perception regarding plastic surgery; as to how many people know what plastic surgeons really do and present a unanimous conclusion: plastic surgeons are perceived to only be associated with performing aesthetic/cosmetic surgery procedures and are dissociated with their reconstructive prowess. Yet others [9], [10] claim to gauge the people's interests and sentiments regarding plastic surgery (in general) or specific APSPs. However, a common feature of all these studies is that their results are based on analyses conducted on small (in terms of both absolute numbers and geographical extent) sample sizes which renders them statistically insignificant and incapable for generalization to actual trends across the world.

A recent study [11] used hedonometrics in order to analyze plastic surgery tweets from 2012-2016 and hence, provide advice to plastic surgeons to spread positive sentiments regarding their practice on social media. However, as noted in a previous paper [12] (whose algorithm for hedonometrics was used by [11]), the hedonometer only assigns happiness scores (determined via manual evaluations) to unigrams and not to any n-grams (n $\geq$ 2) and simply takes a frequency weighted average of these scores to determine the overall happiness of a given piece of text. This results in a major limitation of this technique in that it is unable to take into account the contextual meaning of words which is bound to result in highly erroneous results. This problem has also been highlighted by a review article [13] which points out that while 'cancer reconstruction' should be ideally affecting an increase in the average happiness score of reconstruction tweets, the study has erroneously suggested exactly the opposite.

Another recent study [14] conducted an infodemiology study of tweets in order to analyze the public's interest in cosmetic surgery procedures during COVID. However, the study suffers from 2 major limitations. Firstly, the authors mention that for the purpose of sentiment analysis, inter alia, they removed the default Natural Language Toolkit (NLTK) stopwords [15] from the tokenized tweets.

[1]https://sentic.net/apsp.zip

It must be noted that the list of stopwords in NLTK includes the word 'not'. Consequently, the removal of the word 'not' can prove to be highly detrimental for the purpose of an accurate sentiment analysis since, for example, it can simply eliminate the difference between semantically antithetic terms such as 'good' and 'not good'. Secondly, the authors employed VADER (Valence Aware Dictionary and sEntiment Reasoner) [16] for determining tweet polarities and their corresponding intensities. For the present task, the use of VADER is highly disadvantageous considering its lexicon is quite limited and does not contain most of terms related to aesthetic plastic surgery such as 'nose', 'job', 'facelift', 'breast', etc. Since VADER assigns a sentiment score equal to zero to any and all unknown words in a tweet, it is but obvious that the sentiment scores of most of the aesthetic plastic surgery tweets would be zero/slightly positive/slightly negative. This is also confirmed by the box plots in figures 4A), 4C) and 4D) of their paper. Additionally, a major limitation of all twitter analyses on plastic surgery tweets till date is that they do not consider the removal of a lot of advertising tweets by doctors, clinics, hospitals, etc. This is indeed a crucial step for accurately gauging the public's opinion considering these advertising tweets would otherwise inadvertently sway the dominant tweet polarity towards the positive side. Our work overcomes all the above-mentioned shortcomings.

III. Methodology

A. Data Collection

We scraped 1,194,823 tweets (in total) from the 1st of January, 2018 to the 31st of December, 2019 (Pre-COVID Dataset) and 1,254,897 tweets (in total) from 1st of January, 2020 to the 31st of December, 2021 (Post-COVID Dataset) on the top 7 APSPs worldwide - Rhinoplasty (Nose Job), Facelift (Invasive and Non-Invasive), Blepharoplasty (Eyelid Surgery), Liposuction, Breast Augmentation (Boob Job), Abdominoplasty (Tummy Tuck) and Botox. These APSPs have been mentioned in the list of the top 5 surgical and non-surgical procedures worldwide as per the latest 2020 International Society of Aesthetic Plastic Surgery (ISAPS) International Survey on Aesthetic/Cosmetic Procedures [17] as well as the list of the top 5 cosmetic surgical procedures and cosmetic minimally invasive procedures as per the latest 2020 American Society of Plastic Surgeons (ASPS) Plastic Surgery Statistics Report [18].

The keywords used for scraping the tweets corresponding to these APSPs using snscrape [19] are enlisted in Table I. We used an upper bar of 30,000 tweets per APSP and at the same time, ensured the scraping of the same maximum number of tweets corresponding to every keyword of the APSP for every half a month. The number of tweets obtained for each APSP for the pre-COVID and post-COVID durations is given in Table II. It can clearly be seen that while there has been an overall increase in the number of APSP tweets in the post-COVID duration as compared to the pre-COVID duration, there has been an increase in the number of tweets pertaining to only Botox (majorly) and Rhinoplasty.

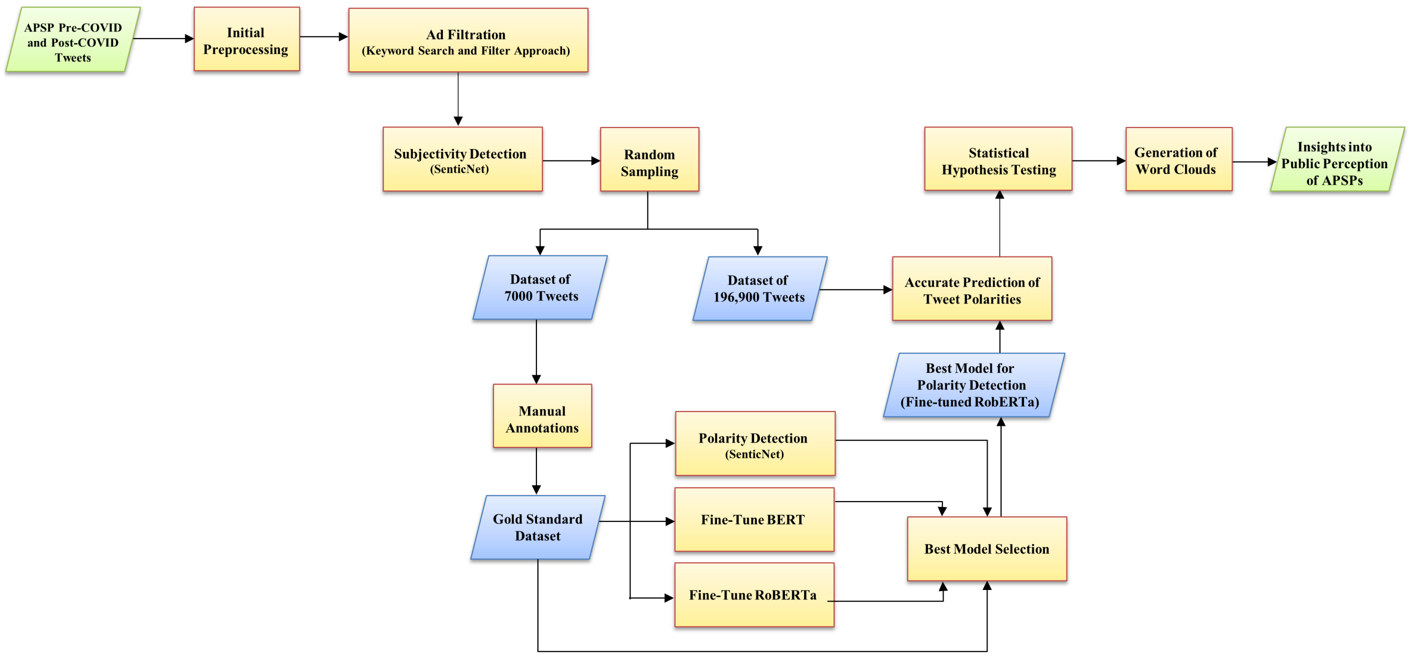

Fig. 1. Flowchart of our end-to-end system for the sentiment analysis of APSP tweets

TABLE I
KEYWORDS USED FOR SCRAPING APSP TWEETS

Rhinoplasty	Facelift	Blepharoplasty	Liposuction	Breast Augmentation	Abdominoplasty	Botox
rhinoplasty	facelift surgery	blepharoplasty	liposuction	breast augmentation	abdominoplasty	botox
nose job	rhytidectomy	eyelid surgery	lipo surgery	breast aug	tummy tuck	botulinum toxin
nose reshaping	facial fillers	eyelid lift	lipoplasty	breast implant		
nose plastic surgery	surgical facelift		liposculpture	breast enlargement		
cosmetic nose surgery	facial rejuvenation surgery		lipectomy	boob job		
aesthetic nose surgery	facial cosmetic surgery			breast enhancement		

TABLE II
NUMBER OF SCRAPED APSP TWEETS IN PRE-COVID AND POST-COVID DATASETS

Aesthetic Plastic Surgery Procedure (APSP)	Number of Scraped Tweets	
	Pre-COVID Dataset	***Post-COVID Dataset***
Rhinoplasty	252438	271685
Facelift	20665	14235
Blepharoplasty	24914	19129
Liposuction	89145	71309
Breast Augmentation	228138	217491
Abdominoplasty	62451	59007
Botox	517072	602041

Furthermore, interestingly, Facelift and Blepharoplasty remain the least tweeted (in comparison to the other APSPs) about APSPs in both the pre-COVID and post-COVID durations. Since it was highly difficult to distinguish the 6 surgical procedures from their non-surgical counterparts (and hence, remove the corresponding non-surgical tweets) considering the high unstructuredness of tweets, the tweets corresponding to the non-surgical counterparts (if scraped) of the 6 surgical procedures were retained and considered for further analysis as well, e.g., fillers have been clubbed with surgical facelift.

B. Data Pre-Processing

1) STAGE 1: Firstly, all the '@' mentions and URLs were removed from the tweets since these were immaterial for the subsequent analysis. This was followed by the removal of all tweets with null values and duplicate tweets. Furthermore, we found that many of the tweets were actually advertising of the APSPs by doctors, clinics, hospitals, etc and hence, were required to be removed before proceeding for sentiment analysis. This was because these were essentially biased tweets which were solely aimed at soliciting business. However, considering that there were some testimonies by patients in some of the tweets by doctors/clinics/hospitals which were reflective of the public's sentiments, a simple filtering of tweets using usernames would not work [20], [21]. Thus, we used a **'keyword search and filter approach'** and used 15 keywords: *'call us today'*, *'contact us today'*, *'learn more'*, *'for appointments'*, *'book consult'*, *'schedule consult'*, *'free consult'*, *'complimentary consult'*, *'on sale'*, *'special offer'*, *'visit website'*, *'offering \$ <any valid number> off'*, *'WhatsApp'*, *'for more information'*, *'text at'* and accounted for any common variations in them by making use of regular expressions [22]. Additionally, tweets containing more hashtags than words (another hallmark of advertising tweets) were removed.

All these aforementioned pre-processing steps resulted in a great reduction in number of the APSP tweets: the total number of tweets in the pre-COVID dataset were brought down from 1,194,823 to 966,510 and the total number of tweets in the post-COVID dataset were brought down from 1,254,897 to 1,102,496.

2) STAGE 2: For the purpose of further pre-processing, random sampling of the APSP tweets was done so as to obtain a maximum of 20,000 tweets per procedure (random sampling was only performed in case the number of tweets corresponding to an APSP was $>$ 20,000).

In order to affect the removal of a majority of the remaining advertisement tweets which contained neutral and factual information regarding the APSPs, we used SenticNet [23], an unsupervised, interpretable, and explainable framework for sentiment analysis. It follows a hybrid AI approach in that it utilizes subsymbolic AI techniques (e.g., deep learning-based auto-regressive models such as kernel-based methods) to create a symbolic representation (e.g., commonsense based knowledge graphs) that effectively converts text into a sort of protolanguage for natural language understanding [24]. In particular, input text is first broken down into concepts using Sentic Parser [25]; then, subjectivity detection is applied to filter out neutral or ambivalent content [26]; finally, semantic multidimensional scaling techniques are applied to extract polarity and emotion labels [27].

After processing the tweets using SenticNet, only the subjective tweets were retained for further analysis; all the ambivalent and neutral tweets were removed. While the motivation behind the removal of the objective tweets is obvious, ambivalent tweets were removed because these (majorly) represented the advertisements; a major chunk of the advertising tweets mostly contained open questions such as *'thinking about <name of the surgery> surgery?', 'want to look and feel better', 'only some slots left', 'get amazing <some kind of result>'* followed by some kind of description of the surgery, etc. leading to some kind of uncertainty and ambiguity in them (a sign of ambivalence). With the advertisement tweets being majorly filtered out, the other (general) remaining neutral tweets were dealt with later, i.e., during Polarity Detection.

C. Preparation of a Gold Standard Dataset

From the pool of subjective APSP tweets, we further performed random sampling so as to obtain 500 tweets per procedure. The resulting total of 7000 tweets (= 500 tweets $\times$ 14 datasets) were then classified by the authors plus 2 experienced plastic surgeons into the following 3 categories: Positive, Negative and Neutral keeping in mind the corresponding APSP as the context for classification.

The Fleiss' Kappa Score was found out to be 0.9275; well above the threshold of 0.80, indicating a very good inter-annotator agreement and further implying the high reliability of our annotations. This annotation exercise thus, resulted in the generation of a gold standard dataset of 7000 tweets.

D. Polarity Detection

In order to decipher the public's perception about APSPs, Polarity Detection was done. We experimented with 2 state-of-the-art Deep Learning models based on the powerful Transformer architecture [28]: BERT (Bidirectional Encoder Representations from Transformers) [29] and RoBERTa (A Robustly Optimized BERT Pretraining Approach) [30]. The pretrained versions of BERT Base and RoBERTa Base for sequence classification as available in the HuggingFace Transformers library [31] were fine-tuned on our gold standard dataset of 7000 tweets using Google Colaboratory (GPU [Tesla T4] enabled) with an 90:10 Train/Test Split using stratified sampling for 4 epochs using AdamW [32] as the optimizer with the values of the hyperparameters as follows: learning rate = 3e-5 and weight decay rate = 0.01 (in line with the recommendations of the authors of the BERT [29] and RoBERTa papers [30] for fine-tuning). Note that we used the uncased version of BERT Base and converted the tweet text into lowercase before using RoBERTa Base. This was done because the high unstructuredness of the tweet data implied its case insensitivity as regards the task of Polarity Classification into the following 3 categories: Positive, Negative and Neutral. An evaluation of the performance of the fine-tuned models can be found in the Results section.

E. Data Visualization and Statistical Hypothesis Testing

In order to gain useful insights into the sentiments of the public regarding APSPs, it is necessary to visualize the obtained results and draw statistically significant conclusions from the observations thereof. Category-specific word clouds were constructed for each of the 7 APSPs and for all the APSPs as a whole.

Word clouds are highly useful tools for visualization since they provide cues about relevant trends by displaying the most frequently occurring words in a corpus of text with the size of a word being proportional to its frequency of occurrence in the corpus. While generating word clouds, two important techniques were used. Firstly, due to the presence of a lot of profanity in the tweet text (which was unnecessary for the purpose of analysis), we used the profanity wordlist of the 'better_profanity' Python library [33] with some modifications: specifically, the word 'boob' (and spelling variants thereof) and the word 'breast' were removed from the profanity wordlist since for the present case, they were a part of the keywords corresponding to the Breast Augmentation procedure and were, hence important for the purpose of sentiment analysis. The modified profanity wordlist was added to the list of stopwords which were removed from the tokenized and lemmatized tweet text before the generation of the word clouds. Secondly, for a correct and meaningful analysis, the word 'not' was removed from the list of default stopwords in the NLTK library [15] and words preceded by 'not' were either replaced by their antonyms if the word's antonyms were present in WordNet [34] or else, joined by an underscore with the word 'not'.

This was a crucial step, often ignored by many, since, for example, 'good' and 'not good' would otherwise be both treated as the word 'good': something which would be detrimental to the very purpose of this analysis. Additionally, in order to extract insightful and possibly contrasting trends about each of the APSPs from the word clouds, hypothesis tests (to determine whether the pre-COVID and post-COVID polarity count differences were actually significant or not) were conducted and the results of the hypothesis tests (if significant) were correlated with the words (and their respective sizes) in the corresponding Word Clouds.

IV. Results

The performance of the fine-tuned BERT and RoBERTa models on the Test Set (as shown in Table III) was evaluated by making use of the following metrics[2]:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{1}$$

$$\mu - F1\ Score = \frac{2 \times \mu - Precision \times \mu - Recall}{(\mu - Precision + \mu - Recall)} \tag{2}$$

where:

$$\mu - Precision = \frac{\sum_{i=1}^{N}(TP_i)}{\sum_{i=1}^{N}(TP_i + FP_i)} \tag{3}$$

$$\mu - Recall = \frac{\sum_{i=1}^{N}(TP_i)}{\sum_{i=1}^{N}(TP_i + FN_i)} \tag{4}$$

and *TP* = True Positives, *FP* = False Positives, *TN* = True Negatives and *FN* = False Negatives.

TABLE III
RESULTS OF TRAINING DEEP LEARNING MODELS USING A 90:10 TRAIN/TEST SPLIT

Transformer	Test Set				
	F1 Score			*μ-F1 Score*	*Accuracy*
	Positive	*Negative*	*Neutral*		
BERT	0.76	0.55	0.55	67.244	65.857
RoBERTa	0.77	0.55	0.66	**69.904**	**67.428**

From these results, it was clear that RoBERTa outperformed[3] BERT in all the metrics. The reasons for RoBERTa's better performance than BERT is majorly owed to the former's robust techniques of Dynamic Masking of training data and Byte-Level Byte Pair Encoding of text and most importantly the former's huge amount of training data (160 GB) as compared to the latter's techniques of Static Masking of training data and Character-Level Byte Pair Encoding of text and the latter's relatively meagre amount of training data (16 GB) respectively [29], [30].

[2]The micro (μ) averaging method was used for computing the overall F1 Score across all *N* (= 3) classes since it weighs all the samples of each of the classes equally and in turn, treats no class as superior to the other

[3]matched as in the case with the F1 Score on Negative Tweets

V. Case Study

A total of 15,000 tweets corresponding to each of the top 7 APSPs for both pre-COVID and post-COVID durations (separately) were randomly sampled from the pre-COVID and post-COVID datasets of the subjective APSP tweets (Note that random sampling was only performed in case the number of tweets corresponding to an APSP was found to be greater than 15000). In total, all these tweets amounted to a large dataset of 196,900 APSP tweets. The fine-tuned RoBERTa model was then used for predicting the polarities of these 196,900 APSP tweets. Furthermore, word clouds were generated for the positive, negative and neutral tweets corresponding to each and every one of the 7 APSPs and statistical hypothesis tests were then conducted based on the polarity counts of the APSP tweets.

Since there was a significant overlap in the words corresponding to positive, negative and neutral tweets for both the post-COVID and pre-COVID categories, the category specific word clouds generated by taking into account both the corresponding pre-COVID and post-COVID tweets have been presented here (Fig. 2). For the purpose of analysis, however, category specific word clouds specific to the pre-COVID and post-COVID durations were considered. Following were the results (all significant results have a p-value < 0.001 while all highly significant results have a p-value $\ll 0.001$):

- **Rhinoplasty** – Although the majority of tweets in both pre-COVID and post-COVID durations were positive, there was a highly significant increase in the number of negative tweets ($\approx$ 3.3%) and a highly significant decrease in the number of positive tweets ($\approx$ 3.4%) in the post-COVID duration as compared to the pre-COVID duration. Changes in the number of neutral tweets in the 2 durations were insignificant. Negative words such as 'botch', 'bad', 'fake', 'hate' and 'money' feature with a significantly large size in Fig. 2(d) which leads to several important conclusions. Firstly, there has been increased media publicity of botched nose jobs post-COVID, for example, the alleged botched nose job of the American singer Summer Walker, which had been in the news in 2020 [35]('summer walker' is visible in Fig. 2(d) as well). The media has also reported on a significant increase in the number of follow up nose jobs (revision rhinoplasties) [36] which has led to people getting trapped in an unending cycle of facial dysmorphia. This has influenced a lot of people into thinking about rhinoplasty as being artificial and bad. Secondly, the appearance of the word 'money' suggests that the money crunch resulting from COVID has made people perceive rhinoplasty as an expensive venture; a waste of money. Thirdly, an increased number of sarcastic tweets and negative rhetoric questions has also been observed post-COVID as can be discerned from the appearance of some positive words intermingled with negative words in Fig. 2(d).

(a) Rhinoplasty (Positive) (b) Facelift (Positive) (c) Blepharoplasty (Positive) (d) Rhinoplasty (Negative)

(e) Facelift (Negative) (f) Blepharoplasty (Negative) (g) Liposuction (Positive) (h) Breast Augmentation (Positive)

(i) Abdominoplasty (Positive) (j) Liposuction (Negative) (k) Breast Augmentation (Negative) (l) Abdominoplasty (Negative)

(m) Botox (Positive) (n) Overall (Positive) (o) Botox (Negative) (p) Overall (Negative)

Fig. 2. Category-Specific Word Clouds for APSP Tweets [category indicated in brackets]

- **Facelift** – While the dominant tweet polarity shifted from positive in the pre-COVID duration to neutral in the post-COVID duration, there was a highly significant increase in the number of negative tweets ($\approx$ 12.7%) and a highly significant decrease in the number of positive tweets ($\approx$ 11.3%) in the post-COVID duration as compared to the pre-COVID duration. Changes in the number of neutral tweets in the 2 durations were insignificant. The main reason for the rise in negative sentiments post-COVID appears to be the reported adverse effects (localized swelling/inflammation of the face and/or lips) of the Moderna COVID vaccine as seen in people with facial fillers [37]. This is clearly evidenced by the large size of the following unigrams and bigrams: 'COVID vaccine', 'filler', 'moderna COVID', 'cause swell', 'inflammation patients', etc. in Fig. 2(e). Another reason lies hidden in the word 'expression' in Fig. 2(e): an excessive number of fillers/repeated facelift surgeries have rendered people's faces expressionless (the word 'expression' appears in Fig. 2(e) instead since different phrases such as 'no expression', 'devoid of any expression', etc. contain the word 'expression' in common though they ultimately imply the word 'expressionless'). An oft quoted example (both in pre-COVID and post-COVID) by people is that of Katie Price, an English celebrity (visible in Fig. 2(e) as well) who has been termed as looking artificial and unrecognizable after multiple facelift surgeries. Apart from celebrities, people have also taken to twitter to demean American politicians. They have used a heavy dose of sarcasm and negative rhetoric and more often than not have alleged that the Trumps have been using the tax payers' money for funding their 'kidney surgeries' (which are alleged to be facelifts, etc.).
- **Blepharoplasty** – Although the majority of tweets in both pre-COVID and post-COVID durations were neutral, there was a highly significant increase ($\approx$ 4.1%) in the number of negative tweets and a significant decrease in the number of neutral tweets ($\approx$ 2.3%) in the post-COVID duration as compared to the pre-COVID duration. Changes in the number of positive tweets in the 2 durations were insignificant. It turns out that an increased number of Asians (in particular, Koreans) have been opting for blepharoplasties; a phenomenon which has been labelled by some people as an emulation of western standards of beauty: the occidentalization of the orient and a manifestation of racism towards those termed as the non-Whites. Words such as 'korean', 'asian', 'white people' and 'beauty standard' are indeed quite large in Fig. 2(f) together with negative terms such as 'evil', 'scary', 'weird', 'bad', 'fake', 'die', 'pain', 'botch', etc.

- **Liposuction** – The majority of tweets in both pre-COVID and post-COVID durations were positive. Additionally, there was a highly significant increase ($\approx$ 3.9%) in the number of positive tweets, a significant increase in the number of negative tweets ($\approx$ 1.4%) and a highly significant decrease in the number of neutral tweets ($\approx$ 5.3%) in the post-COVID duration as compared to the pre-COVID duration. Since the increase in the positive tweets is more than twice the increase in negative tweets post-COVID, it can be concluded that the overall trend has majorly shifted in the positive direction.
  With reference to Fig. 2(g), the words 'need' and 'want' can be correlated with the terms: 'diet', 'exercise' and 'weight loss' in order to imply that liposuction is being increasingly viewed by many people as an easier way to get in shape as compared to dieting and exercising. This trend is in line with a study [38] which suggests that liposuction has contributed to a positive impact on the quality of life of the concerned patients. It can also be observed that liposuction is being increasingly opted for as a body contouring procedure by people in conjunction with a tummy tuck, boob job and butt lift.
- **Breast Augmentation** – The majority of tweets in both pre-COVID and post-COVID durations were positive. Additionally, there was a highly significant increase ($\approx$ 4.5%) in the number of positive tweets and a highly significant decrease in the number of neutral tweets ($\approx$ 3.9%) in the post-COVID duration as compared to the pre-COVID duration. Changes in the number of negative tweets in the 2 durations were insignificant.
  The high increase in positivity in this case can be attributed to the fact that most women desire an enhancement (a bigger size) of their breasts. A previous study [39] has also suggested a positive impact of breast augmentation on the psychosocial and sexual well-being of women. A plethora of positive words such as 'want', 'need', 'love', 'bigger', 'beautiful', 'dream', 'amaze', 'nice', 'happy', 'perfect' together with 'girl' and 'women' in Fig. 2(h) are a reflection of this fact.
- **Abdominoplasty** – The majority of tweets in both pre-COVID and post-COVID durations were positive. Additionally, there was a highly significant increase ($\approx$ 6.1%) in the number of positive tweets, a highly significant increase in the number of negative tweets ($\approx$ 2.6%) and a highly significant decrease in the number of neutral tweets ($\approx$ 8.7%) in the post-COVID duration as compared to the pre-COVID duration. Since the increase in the positive tweets is more than twice the increase in negative tweets post-COVID, it can be concluded that the overall trend has majorly shifted in the positive direction.
  With reference to Fig. 2(i), the words 'want', 'think', 'need', 'love', 'better' and 'great' together with the terms: 'workout', 'weight loss', 'excess skin' are suggestive of the fact that people are increasingly viewing abdominoplasty as a better way to get rid of excess skin and body fat and hence, get back in shape as compared to doing workouts in the gym. This observation is in line with a study [40] which suggests that abdominoplasty improves a patient's satisfaction with their own body and increases their self-esteem. Abdominoplasty is also being increasingly thought of as a 'feel-good factor' in conjunction with a liposuction, butt lift and boob job.
- **Botox** – The majority of tweets in both pre-COVID and post-COVID durations were negative. Additionally, there was a highly significant increase ($\approx$ 4.2%) in the number of negative tweets and a highly significant decrease in the number of neutral tweets ($\approx$ 3.2%) in the post-COVID duration as compared to the pre-COVID duration. Changes in the number of positive tweets in the 2 durations were insignificant. As visible in Fig. 2(o), the most often quoted phrase in the negative tweets is '[too] much botox'. Apparently, botox has become an American election propaganda tool as people are ridiculing American politicians such as Nancy Pelosi (termed as 'botox barbie' and/or 'botox queen' in the tweets), Biden and Melania Trump in order to gain one-upmanship in the race for political supremacy. Consequently, phrases such as 'botox has gone to her brain' are quite common and together with words such as 'evil', 'die', 'freeze [frozen faces]', 'hate', 'bad', 'crazy', 'cry', etc. are creating a whirlpool of negativity around botox.

VI. Conclusion, Discussion and Future Work

In this paper, we have successfully developed an end-to-end system for the sentiment analysis of tweets pertaining to APSPs with a baseline accuracy of 69.904% as measured on the Test Set. To the best of our knowledge, this is the largest of such studies published in the literature. The system was used for generating predictions of the polarities of 196,900 pre-COVID and post-COVID APSP tweets. A detailed sentiment analysis backed by statistical hypothesis testing was then conducted to figure out the public's perception of aesthetic plastic surgery and whether the COVID-19 pandemic had led to a significant change in the same. Overall, it was found that the sentiment of the people was largely positive, considering that the majority of tweets ($\approx$ 52%) in both pre-COVID and post-COVID durations were positive. This shows that people retained great positive interest in aesthetic plastic surgery despite the pandemic. However, a worrying trend is that there has been a significant increase ($\approx$ 3.9%) in the number of negative tweets post-COVID as compared to pre-COVID. While it is true that there has been a 'Zoom Boom' post-COVID and that consequently, aesthetic plastic surgeons are doing more work, that does not always equate with their standing and image in the society. The fact that aesthetic plastic surgery patients can be an object of ridicule and sarcasm does not behove well for any clinical speciality and clearly shows gaps in the communications between the surgeons and the public. Our study also highlights that even amongst the APSPs, body and breast procedures (liposuction, abdominoplasty and breast augmentation) have a very good acceptance among the public at large since they are perceived as ‘functional corrective procedures’ and fit into the overarching public desire to look fit and in shape. When it comes to facial appearance (facelift and blepharoplasty), however, cases of people trying to alter their looks so as to conform to a certain race or a society creates a lot of negative public perception and renders them as looking artificial/unnatural. Effectively, a desire to conform to western standards of beauty may not behove well for the confidence of other ethnicities as every ethnic race has its own strengths and culture and diversity is the basis of this world. Our study also shows the power of the media in affecting the perception of the public: sensationalism of botched up cosmetic surgeries, especially of some celebrities (as in the case of Rhinoplasty) serves a big blow to the reputation of this surgical speciality and creates a negative buzz around it. It also serves as a testimony to the use of Botox as a tool for political propaganda (as in the case of the 2020 US Presidential Elections) so as to gain one-upmanship in the run up to the centre stage of power. Future prospects include the incorporation of other fine-tuned deep learning models, implementation of an extra layer of sarcasm detection with regards to APSP outcomes in particular and image sentiment analysis using convolutional neural networks since in aesthetic plastic surgery, a lot of tweets also contain pictures which are *really* worth a thousand words; outcomes of undergoing a particular APSP can be deciphered from the picture and together with the corresponding tweet’s text can be mapped to the implied sentiments. A real-time display that reveals the sentiments of the public regarding APSPs and provides visualizations for comparisons between sets of tweets corresponding to different events could also be very helpful to the aesthetic plastic surgery community at large by providing them insights to cultivate a positive atmosphere around their surgical speciality.

ACKNOWLEDGMENTS

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funding Scheme (Project #A18A2b0046).

REFERENCES

[1] M. Grassi, E. Cambria, A. Hussain, and F. Piazza, “Sentic Web: A New Paradigm for Managing Social Media Affective Information,” Cognitive Computation, vol. 3, no. 3, pp. 480–489, 2011

[2] L. Oneto, F. Bisio, E. Cambria, and D. Anguita, “Statistical Learning Theory and ELM for Big Social Data Analysis," IEEE Computational Intelligence Magazine, vol. 11, no. 3, pp. 45–55, 2016

[3] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, “Statistical Approaches to Concept-Level Sentiment Analysis," IEEE Intelligent Systems, vol. 28, no. 3, pp. 6–9, 2013

[4] J. Edgar, “The Cosmetic Surgery “Zoom Boom” Is Real - But There’s More To The Story,” Refinery29, Accessed: Aug. 2, 2022. Available: https://www.refinery29.com/en-us/2021/06/10481534/zoom-boom-plastic-cosmetic-surgery-pandemic

[5] S. M. Rice, E. Graber, and A. S. Kourosh, “A Pandemic of Dysmorphia: “Zooming” into the Perception of Our Appearance,” Facial Plastic Surgery and Aesthetic Medicine, vol. 22, no. 6, pp. 401–402, 2020

[6] C. de Blacam, D. Kilmartin, C. Mc Dermott, and J. Kelly, “Public Perception of Plastic Surgery,” Journal of Plastic, Reconstructive and Aesthetic Surgery (JPRAS), vol. 2, no. 2, pp. 197–204, 2015

[7] A. Kalandar, S. Al-Youha, and B. Al-Habibi, “What Does the Public Think? Examining Plastic Surgery Perceptions through the Twitter-verse,” Plastic and Reconstructive Surgery (PRS), vol. 142, no. 1, pp. 265–274, 2018

[8] A. J. Reid and P. S. C. Malone, “Plastic Surgery in the Press,” Journal of Plastic, Reconstructive and Aesthetic Surgery (JPRAS), vol. 61, no. 8, pp. 866–869, 2008

[9] A. Maisel et al, “Self-reported Patient Motivations for Seeking Cosmetic Procedures,” Journal of the American Medical Association (JAMA) Dermatology, vol. 154, no. 10, pp. 1167–1174, 2018

[10] T. J. Lu, A. X. Nguyen, X. Trinh, and A. Y. Wu, “Sentiment Analysis Surrounding Blepharoplasty in Online Health Forums,” Plastic and Reconstructive Surgery - Global Open, vol. 10, no. 3, 2022

[11] M. Chopan, L. Sayadi, E. M. Clark, and K. Maguire, “Plastic Surgery and Social Media: Examining Perceptions,” Plastic and Reconstructive Surgery (PRS), vol. 143, no. 4, pp. 1259–1265, 2019

[12] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth, “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter,” Public Library of Science (PLOS) ONE, vol. 6, no. 12, pp. e26752, 2011

[13] L. L. Blankensteijn and S. J. Lin, “Discussion: Plastic Surgery and Social Media: Examining Perceptions,” Plastic and Reconstructive Surgery (PRS), vol. 143, no. 4, pp. 1266–1267, 2019

[14] W. Liu et al, “Public Interest in Cosmetic Surgical and Minimally Invasive Plastic Procedures During the COVID-19 Pandemic: Infodemiology Study of Twitter Data,” Journal of Medical Internet Research (JMIR), vol. 23, no. 3, pp. e23970, 2021

[15] S. Bird, E. Loper, and E. Klein, Natural Language Processing with Python, O’Reilly Media Inc, 2009

[16] C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text,” in ICWSM, pp. 216–225, 2014

[17] “2021 ISAPS International Survey on Aesthetic/Cosmetic Procedures performed in 2020,” International Society of Aesthetic Plastic Surgery (ISAPS), Accessed: Aug. 2, 2022. Available: https://www.isaps.org/wp-content/uploads/2022/01/ISAPS-Global-Survey_2020.pdf

[18] “2020 Plastic Surgery Statistics Report,” American Society of Plastic Surgeons (ASPS), Accessed: Aug. 2, 2022. Available: https://www.plasticsurgery.org/documents/News/Statistics/2020/plastic-surgery-statistics-full-report-2020.pdf

[19] “GitHub – JustAnotherArchivist/snscrape: snscrape: A social networking service scraper in Python,” GitHub, 2022, Accessed: Jan. 30, 2022. Available: https://github.com/JustAnotherArchivist/snscrape

[20] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, “Sentic Computing for Patient Centered Applications," Proceedings of ICSP, pp. 1279–1282, 2010

[21] E. Cambria, T. Benson, C. Eckl, and A. Hussain, “Sentic PROMs: Application of Sentic Computing to the Development of a Novel Unified Framework for Measuring Health-Care Quality," Expert Systems with Applications, vol. 39, no. 12, pp. 10533–10543, 2012

[22] G. van Rossum and the Python Development Team, “The Python Library Reference,” Release 3.9.7, Python Software Foundation, 2021

[23] E. Cambria, Q. Liu, S. Decherchi, F. Xing, and K. Kwok, “SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis,” in LREC, pp. 3829–3839, 2022

[24] Y. Susanto, A. Livingstone, B.C. Ng, and E. Cambria, “The Hourglass Model Revisited," IEEE Intelligent Systems, vol. 35, no. 5, pp. 96–102, 2020

[25] E. Cambria, R. Mao, S. Han, and Q. Liu, “Sentic Parser: A Graph-Based Approach to Concept Extraction for Sentiment Analysis," Proceedings of ICDM Workshops, 2022

[26] A. Valdivia, M. V. Luzon, E. Cambria, and F. Herrera, “Consensus Vote Models for Detecting and Filtering Neutrality in Sentiment Analysis,” Information Fusion, vol. 44, pp. 126–135, 2018

[27] E. Cambria, Y. Song, H. Wang, and N. Howard, “Semantic Multi-Dimensional Scaling for Open-Domain Sentiment Analysis," IEEE Intelligent Systems, vol. 29, no. 2, pp. 44–51, 2014

[28] A. Vaswani et al, “Attention is All you Need,” in NIPS, pp. 6000–6010, 2017

[29] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in NAACL-HLT, pp. 4171–4186, 2019

[30] Y. Liu et al, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv:1907.11692 [cs.CL], 2019

[31] T. Wolf et al, "Transformers: State-of-the-Art Natural Language Processing," in EMNLP, pp. 38–45, 2020

[32] I. Loshshilov and F. Hutter, "Decoupled Weight Decay Regularization," in ICLR, 2019

[33] S. N. Thanh, "better_profanity," Python Package Index (PyPI) 2022, Accessed: Jun. 1, 2022. Available: https://pypi.org/project/better-profanity/

[34] C. Fellbaum, "WordNet: An Electronic Lexical Database," Cambridge, MA, USA: MIT Press, 1998

[35] "Summer Walker Trolled Over Her Apparent Botched Nose Job," AceShowbiz, Accessed: Aug. 2, 2022. Available: https://www.aceshowbiz.com/news/view/00153124.html

[36] M. Kekatos, "Why are SO many Women getting Follow-Up Nose Jobs? Rise in Botched Ops and the Falling Trend of 'Pinched Tips' are driving Millions Patients back into the Operating Theater," DailyMail, Accessed: Aug. 2, 2022. Available: https://www.dailymail.co.uk/health/article-6680143/amp/Rise-botched-ops-falling-trend-pinched-tips-driving-women-follow-nose-jobs.html

[37] B. Buntz, "Moderna COVID-19 Vaccine might cause Facial Swelling for People with Dermal Fillers," Drug Discovery and Development, Accessed: Aug. 2, 2022. Available: https://www.drugdiscoverytrends.com/moderna-COVID-19-vaccine-might-cause-facial-swelling-for-people-with-dermal-fillers/

[38] N. A. Papadopolus et al, " Quality of Life following Aesthetic Liposuction: A Prospective Outcome Study," Journal of Plastic, Reconstructive and Aesthetic Surgery (JPRAS), vol. 72, no. 8, pp. 1363–1372, 2019

[39] H. Noorizadeh and B. K. Bari," The Effect of Breast Augmentation Surgery on Quality of Life, Satisfaction, and Marital Life in Married Women using BREAST-Q as a Validation Tool," Journal of Family Medicine and Primary Care, vol. 9, no. 2, pp. 711–713, 2020

[40] K. M. M. Saariniemi, "Abdominoplasty Improves Quality of Life, Psychological Distress, and Eating Disorder Symptoms: A Prospective Study," Plastic Surgery International, vol. 2014, 2014