HIR

Healthcare Informatics Research

Concerns of Thalassemia Patients, Carriers, and their Caregivers in Malaysia: Text Mining Information Shared on Social Media

Yuen Chi Phang[1], Azleena Mohd Kassim[1], Ernest Mangantig[2]

[1]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia
[2]Regenerative Medicine Cluster, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Penang, Malaysia

Objectives: The main aim of this study was to use text mining on social media to analyze information and gain insight into the health-related concerns of thalassemia patients, thalassemia carriers, and their caregivers. **Methods:** Posts from two Facebook groups whose members consisted of thalassemia patients, thalassemia carriers, and caregivers in Malaysia were extracted using the Data Miner tool. In this study, a new framework known as Malay-English social media text pre-processing was proposed for performing the steps of pre-processing the noisy mixed language (Malay-English language) of social media posts. Topic modeling was used to identify hidden topics within posts shared among members. Three different topic models—latent Dirichlet allocation (LDA) in GenSim, LDA in MALLET, and latent semantic analysis—were applied to the dataset with and without stemming using Python. **Results:** LDA in MALLET without stemming was found to be the best topic model for this dataset. Eight topics were identified within the posts shared by members. Of those eight topics, four were newly discovered by this study, and four others corresponded to the findings of previous studies that used an interview approach. **Conclusions:** Topic 2 (the challenges faced by thalassemia patients) was found to be the topic with the highest attention and engagement. Healthcare practitioners and other concerned parties should make an effort to build a stronger support system related to this issue for those affected by thalassemia.

Keywords: Data Mining, Thalassemia, Natural Language Processing, Social Media, Data Science

I. Introduction

Thalassemia is an inherited disorder characterized by the inability or diminished ability to produce hemoglobin, which affects the oxygen-carrying capacity of red blood cells [1]. Approximately 20% of the world's population are alpha-thalassemia carriers and 5.2% are significant variant carriers for beta-thalassemia [2]. Severe thalassemia patients require lifelong regular blood transfusions and expensive iron chelation therapy to survive. The 2018 Malaysian Thalassemia Registry Report showed a significant rise in the total number of thalassemia patients, increasing from 6,805 living patients in 2014 to 7,984 living patients in 2018, with 697 patient deaths reported in November 2018 [3]. As the prevalence

of thalassemia in Malaysia increases, there will be greater demand for resources such as healthcare facilities, medical personnel, various mediations, and counseling services.

Many thalassemia-related studies in Malaysia have focused on its molecular characterization and identification of genetic mutations, and there has been a lack of studies that explore the quality of life and experiences of those affected by thalassemia in Malaysia. One qualitative study examined the concerns of the thalassemia patients, carriers, and caregivers in Malaysia, including the patients' beliefs related to thalassemia, by conducting face-to-face interview sessions with patients and their parents [4]. The results found that patients and parents were concerned about education, self-image and body image, employment, marriage, medical financing, relationships, social integration, and self-esteem [4]. In addition, while the majority of thalassemia patients understood that thalassemia is a genetic disease and believed modern treatment to be effective, some thalassemia patients did not seek treatment due to their fears of the side effects [5]. The first quality of life study among Malaysian thalassemia patients examined children aged 3 to 18 years old with transfusion-dependent thalassemia and used the PedsQL (Pediatric Quality of Life Inventory) Generic Core Scales to assess the impact of thalassemia on patients' quality of life [6].

There were several studies from other countries that explored the concerns of thalassemia patients. A study in Iran that examined the burden of caregivers of thalassemia patients found that there was insufficient social support for caregivers despite the high burden of care [7]. Another study in Italy found that thalassemia patients coped better with their condition when they had social support and a proactive personality [8]. An additional survey was conducted by another study to examine the burden of thalassemia patients and their caregivers in Italy, the United Kingdom, and the United States using a smartphone application [9]. The survey results indicated that patients and caregivers suffered from burdens on time management, fatigue, pain, and impaired quality of life.

The insights provided by these qualitative studies, however, are limited to a small number of participants who were willing to take part in interview sessions or digital surveys. The findings might not be generalizable to patients with different ethnicities and socioeconomic backgrounds in Malaysia. In addition, it can be time-consuming to carry out interview sessions while still yielding limited results.

As social media becomes increasingly ubiquitous, people are becoming more comfortable sharing their thoughts and experiences openly, even for health-related issues, and it is important to study the information that can be extracted from this medium [10]. One survey suggested that medical treatment plans that integrate social support networks into treatment could reduce the mental burden of thalassemia patients [11]. Some patients and caregivers use online social networks to seek support and share their experiences, which suggests that collecting and examining data from social networking platforms could help to identify and mitigate issues related to those affected by thalassemia.

Hence, in this study, an alternative method of text mining was used to explore topics and information frequently shared on social media by those affected by thalassemia to gain insight into their health-related concerns. Previous studies that used text mining to identify patterns on social media were explored. For example, topic modeling was used in one study to identify different recurring topics and concerns related to breast cancer discussed in a public Facebook group and a public health forum [12]. In addition, topic modeling has been used to categorize user-generated content from Twitter and Reddit [13]. To the best of our knowledge, this is the first study to use text mining on social media to gain insight into the concerns of thalassemia patients, carriers, and caregivers for enhancing understanding of thalassemia health-related concerns and providing recommendations for improving the quality of life of those affected by thalassemia. In addition, a new framework, Malay-English social media text pre-processing (MESMTPP), is proposed for pre-processing noisy, mixed-language text in social media posts.

Therefore, this study aimed to apply text mining to a social media context to gather information and develop a better understanding of the concerns that affect the quality of life of thalassemia patients, thalassemia carriers, and their caregivers. A comparison between previous works and this study is presented in Table 1, with previous work categorized based on their methods and limitations.

II. Methods

1. Data Extraction
The data were collected from Facebook, which is a social media service through which users can share and exchange information. Specifically, data came from two different Facebook groups: "For ALL Thalassemia MALAYSIA" and "Kelab Thalassemia Malaysia." These Facebook groups were created for sharing thalassemia-related information. The members of these groups are those affected by thalassemia, including caregivers, in Malaysia. The data were extracted using the Data Miner tool (https://data-miner.io/).

Table 1. Comparison between previous studies and this study

Previous studies		This study[a]
Methods	Limitation	Contribution
Interview sessions with thalassemia patients and caregivers [4–8]	It was time-consuming and costly to organize interview sessions, as travel to different places was often required to carry out face-to-face interviews.	Time and costs were saved due to not having to organize interview sessions or prepare survey questions.
	Patients were reluctant to attend interviews or were sensitive when discussing their concerns.	Patients and caregivers are free to post anything on social media.
	The findings were generalizable to all patients from different backgrounds, as information was only obtained from small numbers of patients and caregivers who attend interviews.	More information could be obtained from a larger group of people with different backgrounds.
Digital surveys of thalassemia patients and caregivers [9]	Some patients and caregivers may not have understood all the questions and possibly gave wrong answers.	Patients and caregivers were free to post issues related to thalassemia on social media, from their levels of understanding and points of view. Text mining was used to process and understand their social media posts.
Text mining approach on social media for breast cancer patients [12]	Text data was in one language only (French).	A detailed workflow to pre-process social media texts was presented. A pre-processing method known as Malay-English social media text pre-processing (MESMTPP) was introduced.
Text mining approach on social media for general posts [13]	Basic text data cleaning. Text data was in one language only (English).	The text data included a mixture of two languages (English and Malay).

[a]The study method is text mining of social media posts by thalassemia patients, thalassemia carriers, and caregivers.

2. Data Exploration

All posts from January 2015 to April 2020 were extracted. The total number of posts in both Facebook groups and descriptions of their attributes and types of data are shown in Table 2. Data from the two groups were combined, and 1,045 posts were collected in total. After the preliminary data cleaning, there were 922 posts, which comprised 73,553 words. Figure 1 is a visualization of the raw dataset showing the number of status posts, likes, and comments for each year and the distribution of the total number of words in each post. A clear increase in the number of posts each year can be seen. The total number of posts decreased in 2020 because data from only 4 months were extracted. This indicates that, over time, more people began to take advantage of social media to seek information related to thalassemia and became increasingly active on social media as more members participated in the groups. The results regarding word count indicate that the typical length of posts was short, usually ranging from one to 50 words per post.

A word cloud using a bag-of-words model was generated to identify the most frequently used words. Additionally, term frequency-inverse document frequency (TF-IDF) was calculated to determine the importance of individual words that appeared in posts. In addition, the hashtags used in posts were extracted for further analysis, as further discussed in Section III.

3. Pre-processing of Noisy Social Media Texts

Social media users usually do not strictly follow correct language conventions when making posts, which, for the purposes of text mining, results in a high proportion of noisy and formally incorrect vocabulary use and sentence structures, which subsequently influence the analysis. As a result, the MESMTPP framework was introduced to pre-process text from social media. Figure 2 shows the steps that were performed to clean and normalize noisy text from social media using MESMTPP.

The first few steps (steps 1–6) of social media text pre-processing involved removing some characters to direct the focus towards the essence of the posts. In step 7, all capitalized words were changed to lowercase to make them uniform, since social media users tend to freely mix word cases.

Table 2. Number of posts extracted in each group with attribute descriptions and data type

	For ALL Thalassemia MALAYSIA	Kelab Thalassemia Malaysia	Description
Social media platform	Facebook	Facebook	
Total posts	784	261	
Posts with text	768	154	
Attribute[a]			
Posts			Posts from Facebook group
Date			Date each post was made
Year			Year each post was made (2015–2020)
Number of likes			Number of likes received by each post
Number of comments			Number of comments given on each post
Group			The group ("For ALL Thalassemia MALAYSIA" or "Kelab Thalassemia Malaysia") in which each post was made

[a]Attribute data types can be text-based or numerical.
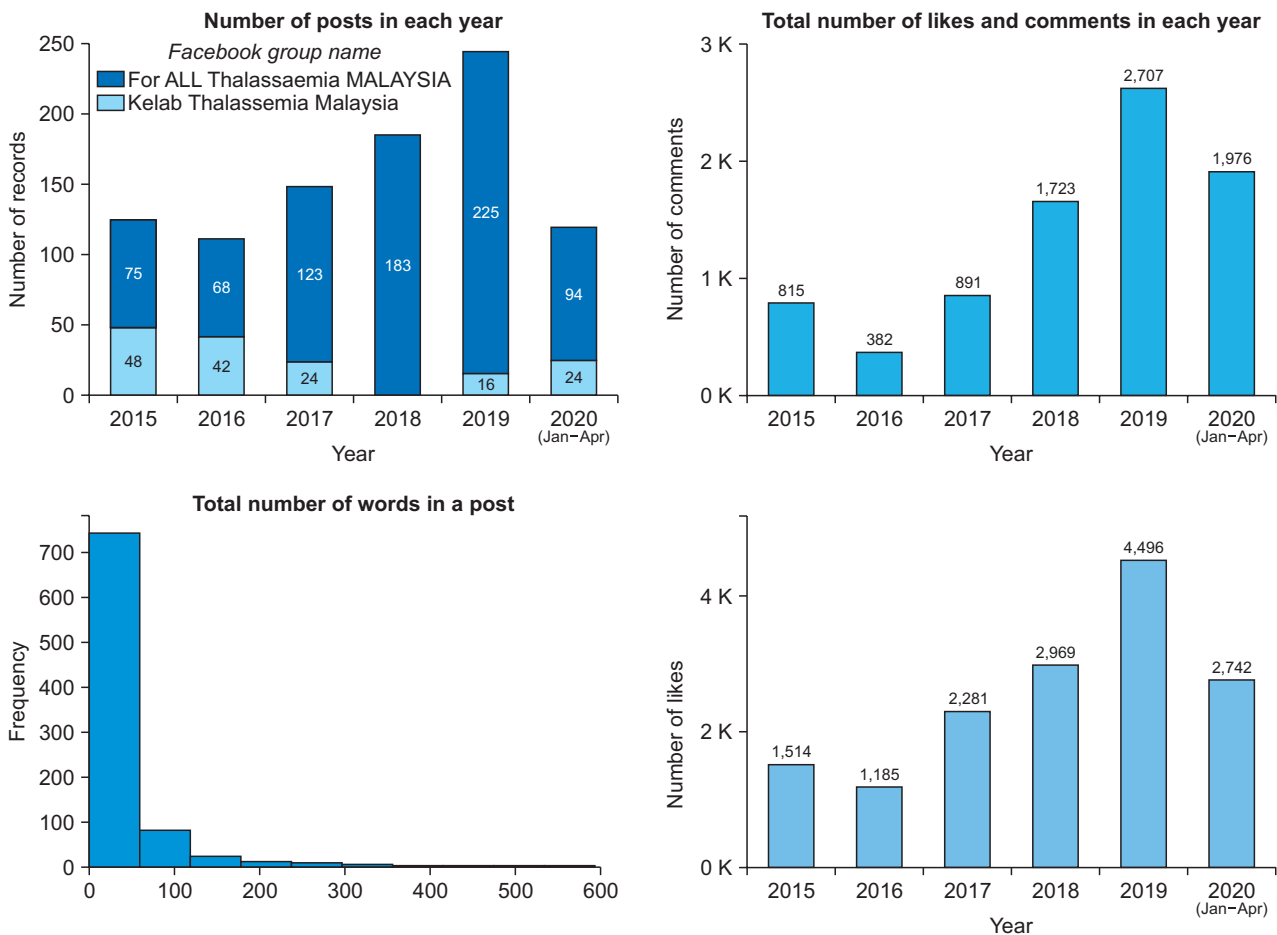

Figure 1. Number of posts, comments, and likes for each year, with the distribution of the total number of words per post.

In step 8, tokenization was carried out to split the given text into smaller parts, followed by step 9, at which point the names of people and places were removed. Abbreviations were processed in step 10 by identifying patterns of abbre-

viations, as suggested in studies that analyzed pre-processing tasks related to social media posts in Spanish [14] and constructed a Malay abbreviation corpus based on social media data [15]. In step 11, all English words were translated into
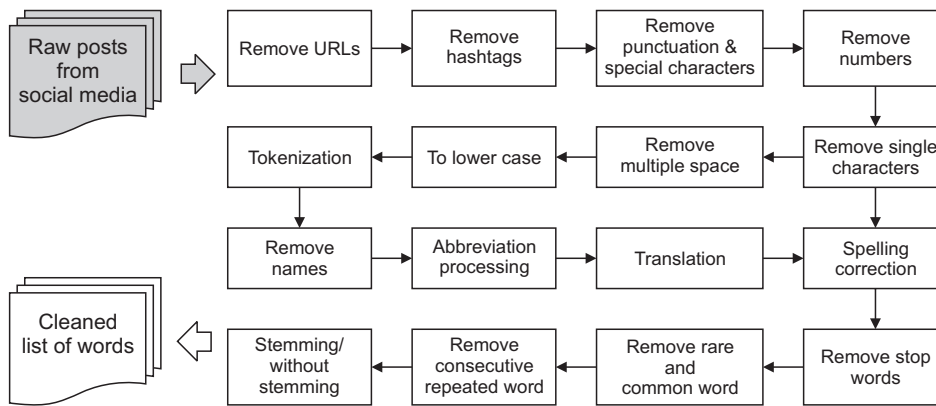
Figure 2. Malay-English social media text pre-processing (MESMTPP) framework: a procedure to pre-process text from social media posts.
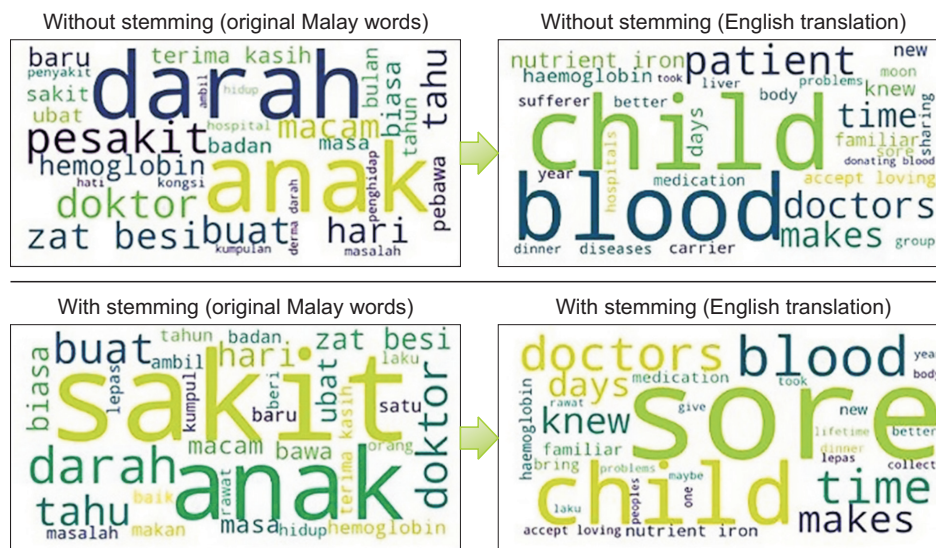

Figure 3. Word cloud: cleaned data without stemming and with stemming (with English translation).

Malay by consulting a Malay-English dictionary.

Next, spelling was checked in step 12 using the Malaya library [16], followed by step 13, during which the Malay stop words were removed. During this step, a custom list of stop words for the Malay language was created and added to the stop word list. Part of the Malay stop word list was adopted based on a paper that proposed a list of Malay stop words for novelty detection on Malay documents [17]. Rare words that appeared fewer than three times were removed at step 14. At step 15, words that contained three or more repeated letters were removed. In the final step, stemming was done to eliminate affixes of words to obtain the root term.

4. Topic Identification using Topic Modeling

After pre-processing, the data were arranged into a suitable format for topic modeling. Two topic modeling algorithms were used: latent Dirichlet allocation (LDA) [18] and latent semantic analysis (LSA) [19]. Both LDA in GenSim and MALLET were tested.

III. Results

1. Text Data Analysis Results

The results from the word clouds are shown in Figure 3: one without stemming and one with stemming. For each of the word clouds in Malay, a second word cloud was generated for the English translations of Malay words. From the word clouds, *darah* (blood) and *anak* (child) had the highest frequency, as shown in Figure 3. However, after stemming was performed, *sakit* (sore or sick) showed the highest frequency due to all *pesakit* (patient) words being stemmed to *sakit* (sore or sick).

In addition, the TF-IDF value of one of the posts is shown in Figure 4, with an English translation of the original Malay words. This figure indicates that the most important word in the post was *cuti* (leave or holiday), followed by "*mc*" (medical certificate). Thus, it was understood that the post was about taking leave from work. The content of the post could therefore be summarized using TF-IDF to understand what it was about.

Out of 933 posts, 163 posts contained hashtags, and the most common hashtags used were #Thalassemiamylifelongcompanion and #jomdermadarah, which appeared 24 and 20 times, respectively. A correlation matrix is shown in Figure 5 visualizing associations between hashtags. The hashtag #zerothalassemia was highly correlated with the hashtag #Thalassemiaawareness and #kempenkesedaranTalasemia. This

indicates that many organized awareness campaigns were undertaken to raise awareness of thalassemia and reduce its prevalence.

2. Topic Modeling Result

In a comparison of the topic modeling algorithms based on coherence score, LDA in MALLET showed the best results for both datasets with and without stemming. LDA in MALLET yielded better results with the dataset without stemming than LSA, but LSA yielded better results with stemming when there were fewer topics. Nonetheless, LDA in MALLET was better when the number of topics was higher. Therefore, the results obtained from the LDA model that used a dataset without stemming were chosen for analysis, since it produced the best coherence score overall, as shown in Table 3. The topic modeling identified eight topics in total. The word distribution per topic produced was evaluated based on human judgment. Table 4 shows the eight main topics discussed in Facebook groups by people affected by thalassemia, with their associated labels and keywords. English translations for the original keywords in Malay were added. In addition, examples of posts are shown in Malay, with the content described in English (Table 4).
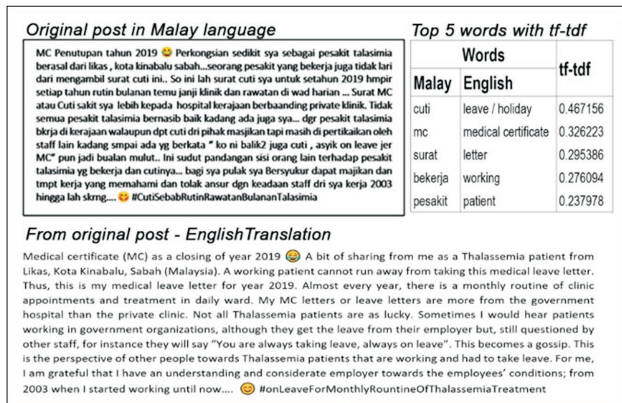

Figure 4. Top 5 words with the highest TF–IDF (term frequency–inverse document frequency) values from a sample post (with English translation).
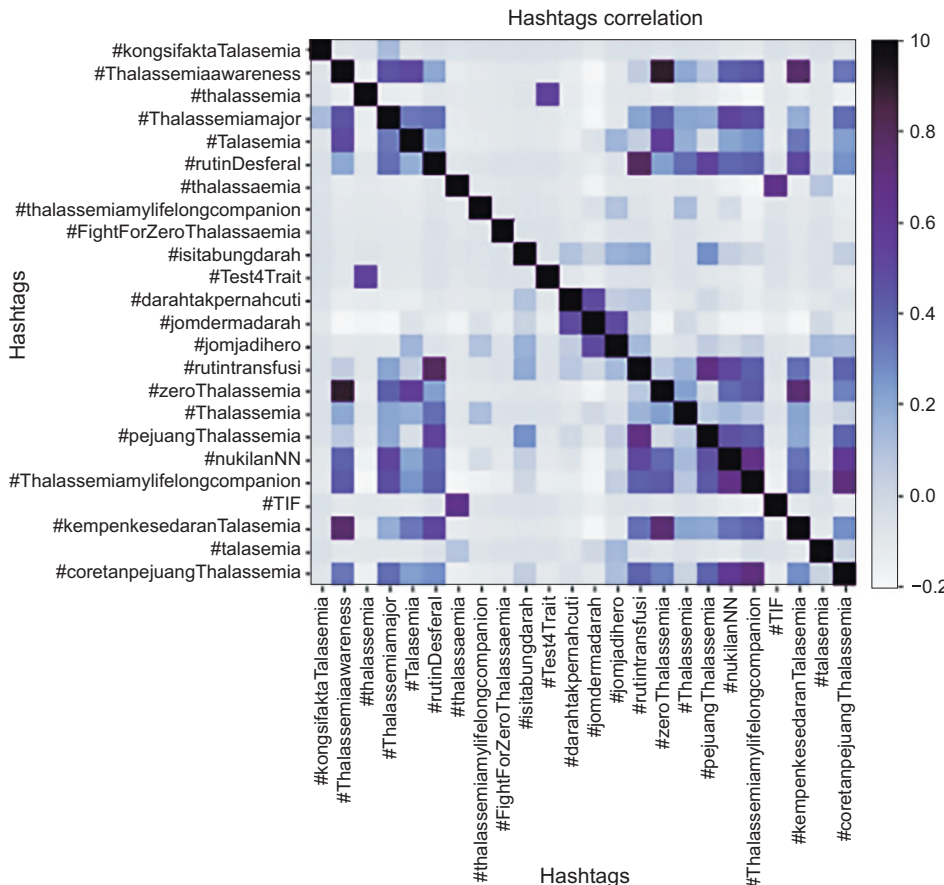

Figure 5. Hashtag correlation plot.

Table 3. Comparison of coherence scores across a range of topic numbers after applying LDA in GenSim, LDA in MALLET, and LSA to the dataset

| Topic number | Coherence score | | | | | |
| | Without stemming | | | With stemming | | |
	LDA in GenSim	LDA in MALLET	LSA	LDA in GenSim	LDA in MALLET	LSA
2	0.29	0.332	0.334	0.293	0.314	0.354
4	0.293	0.365	0.363	0.319	0.357	**0.411**
6	0.301	0.335	**0.387**	0.308	0.402	0.398
8	0.304	**0.426**	0.352	0.317	0.396	0.403
10	0.285	0.365	**0.387**	**0.346**	0.396	0.376
12	0.288	0.398	0.377	0.307	0.390	0.353
14	0.29	0.374	0.354	0.321	**0.406**	0.331
16	**0.329**	0.410	0.373	0.310	0.386	0.348
18	0.324	0.406	0.370	0.305	0.391	0.360

Bold text is the best coherence score on each method.
LDA: latent Dirichlet allocation, LSA: latent semantic analysis.

IV. Discussion

The raw data were very noisy, with many colloquial or vernacular words in addition to many posts being written in some mixture of Malay and English. MESMTPP was performed without eliminating the English words since a translation step (step 11) was included to translate all English words into Malay. In step 12, spelling was checked to ensure correctness using the Malaya library [16]. Some misspelled words were still detected when consulting the Malaya library. Therefore, a custom dictionary was created to correct spelling, in which the key was a misspelled word and the value was the correct spelling of the word. The misspelled word would then be automatically replaced with its correct form. After MESMTPP, the data were mostly cleaned and ready for modeling, although they were not 100% clean due to the constantly evolving nature of language on social media. The Malay text corpora were limited but publicly available. Hence, the dictionary of the Malay corpus should be updated routinely.

As a result of topic modeling (Table 4), half of the topics (topics 1, 3, 5, and 7) revealed several concerns of thalassemia patients and caregivers that have not been reported in previous qualitative studies, while topics 2, 4, 6, and 8 are consistent with findings from previous interview-based qualitative studies.

Topic 1 encompasses treatment for thalassemia, indicating the need for healthcare providers to offer education to strengthen patients' and caregivers' understanding of treatment and to ensure that they received updated information.

Topic 2 encompasses the challenges related to managing the illness at work. Thalassemia patients who work and the parents of children with thalassemia may face difficulties at work due to the regularity with which they need to take leave from work for blood transfusion treatments. This finding is consistent with previous studies that reported thalassemia patients and the parents of children with thalassemia, some of whom also suffer from the condition, had problems with their employers [4]. Similar studies found that thalassemia patients were exhausted by physical changes and treatment [9,20].

Topic 3, a new finding, encompasses thalassemia patients' concerns about diet and dietary supplements since they cannot consume foods that are high in iron. Topic 4 covers religious faith as it relates to coping with thalassemia and praying to God for strength to continue living, which was also reported by previous studies [5,21,22]. Topic 5—another new finding—shows that Facebook has become a valuable platform to ask questions and solicit opinions of those affected by thalassemia. Topic 6 covered patients' and caretakers' concerns about blood donation, either via posts asking for blood donation or expressing concern about blood insufficiency. Another study also found that patients raised the issue of possible blood shortage [21].

Topic 7 encompasses members of the group sharing information about their involvement in a thalassemia society and the society's activities. The society functions as a support group for thalassemia patients and parents, and organizes activities to spread awareness about thalassemia. Topic 8 addresses the genetic nature of thalassemia, which indicates a

Table 4. Eight main topics generated with keywords and examples of posts (with English translations and descriptions)

No.	Topic	Keywords		Example of original posts in Malay with English descriptions
		Malay	English	
1	Treatment for thalassemia (umbilical cord, iron chelation therapy, and bone marrow-related treatment)	pesakit	patient	"*Alhamdulillah… Semakin sihat… After bmt… Hampir 5 bulan dah x transfuse.*"
		doktor	doctor	**English description:**
		rawatan	treatment	This post is from a person who expressed gratitude to God as he/she got healthier after bone
		sihat	healthy	marrow transplantation and had not received transfusion for 5 months.
		pusat	center	
		ujian	test	
		tahun	year	
		hasil	outcome	
		hospital	hospital	
		bayi	baby	
2	Challenges faced by thalassemia patients (illness, work, and treatment side effects)	buat	make	"*sy bekeje kakitangan awam mmg keje yang byk keje kene anta surat, byk berjalan laa. kadang*
		bulan	month	*bila tak larat memang mc n cuti jer rehat. lagi kerap sakit skang nie n byk cuti dgn mc ambik…*"
		tahun	year	**English description:**
		sakit	sick or sore	This post is about the person's problems at work, where the person had to walk great distances
		baru	new	to deliver letters. When he/she could not cope with the strain, he/she would obtain a medical
		dekat	close	certificate and take leave from work. The frequency of the patient's illness increased, so he/
		kena	hit	she had to take leave more often.
		rasa	feel	
		lepas	last	
		kawan	friend	

Table 4. Continued 1

No.	Topic	Keywords		Example of original posts in Malay with English descriptions
		Malay	English	
3	Iron-rich foods and diet	besi	iron	"What should we eat in Thalassemia? Nutrition & Thalassemia. It is recommended that patients going through blood transfusion should opt for a low iron diet…" (**originally posted in English**)
		zat	nutrients	
		ubat	medicine	
		makan	eat	
		makanan	food	
		utama	main	
		hati	liver	
		tinggi	high	
		transfusi	transfusion	
		kadar	rate	
4	Praying for strength and peer support	anak	child	"Semoga kita semua dilindungi sentiasa dalam lindungan Allah swt.. Anak anak Thalassemia semoga sentiasa dalam lindungan Allah swt"
		penghidap	sufferer	**English description:**
		ibu	mother	This post describes someone's prayers that all thalassemia patients, including children, are always under the protection of Allah (God).
		diri	self	
		masa	time	
		Allah	Allah (God)	
		mampu	able	
		moga	hope	
		terus	continue	
		kuat	strong	

Table 4. Continued 2

No.	Topic	Keywords		Example of original posts in Malay with English descriptions
		Malay	English	
5	Asking questions/opinion about the information of disease and machine used for treatment.	*kumpulan*	group	*"Sebelum ni pernah kongsi tentang permohonan zakat untuk yang perlu beli kelengkapan rawatan Desferal."*
		hospital	hospital	
		ahli	members	**English description:**
		tempat	place	This post is from a person that mentioned that he/she had shared about the zakat (alms) application for those who need to buy Desferal equipment for treatment.
		maklumat	information	
		selamat	safe	
		kongsi	share	
		sejahtera	prosperous	
		sahabat	friend	
		darah	blood	
6	Blood donation (issues related to shortage of blood and side effects)	*derma*	donation	*"Masalah kekurangan darah bukan hanya di Malaysia sekarang, tapi di seluruh dunia juga akibat wabak. Jom wujudkan kesedaran untuk yang sihat…"*
		sel	cell	
		jenis	type	**English description:**
		badan	body	This post stated that blood shortage is not only a problem in Malaysia, but all over the world as a result of an epidemic. He/she reminded others to raise awareness among those without thalassemia.
		pemindahan	transplant	
		rendah	low	
		biasa	normal	
		tanda	sign	
		masalah	problem	

Table 4. Continued 3

| No.	Topic	Keywords		Example of original posts in Malay with English descriptions
		Malay	English	
7	Sharing experiences related to a thalassemia society and activities	*kesihatan*	health	"*Thalassemia dan diabetes. Ada yang nak kongsi pengalaman? Thalassemia dan diabetes.*"
		hidup	life	**English description:**
		kongsi	share	This post invites others to share their experiences of thalassemia and diabetes.
		keluarga	family	
		dunia	world	
		rakan	friend	
		pengalaman	experience	
		negara	country	
		penyakit	disease	
		aktiviti	activity	
8	Genetics of thalassemia and issues among carrier couples	*Thalassemia*	Thalassemia	"*Possible tak hala 1/3 dari anak?? Pembawa halassemia, tapi bila buat ujian darah ibu dan bapa clear, bkn pengidap mahupun pembawa…?*"
		pebawa	carrier	**English description:**
		anak	child	This post is about someone inquiring as to whether it is possible that one-third of children could be thalassemia carriers, but when blood tests were conducted for the parents, both parents are neither thalassemia patient nor carrier. He/she wondered whether the result is accurate or not.
		beta	beta	
		alfa	alpha	
		sifat	trait	
		hala	"*hala*" (short form for thalassemia)	
		baik	good	
		pakar	expert	

need for community awareness for pre-marital thalassemia screening and counseling for carrier couples to reduce the prevalence of thalassemia in Malaysia. Studies have shown that some parents of thalassemia patients are not aware of their carrier status prior to marriage [4]. In addition, it was found that married couples often had inadequate knowledge related to the genetic nature of thalassemia and did not undergo pre-marital screening [23].

Figure 6 shows the number of posts, likes, and comments on each topic. For the number of posts in each topic, topics 2 and 5 had the highest frequency, and there was more engagement with topic 2 than topic 5 in terms of the number of likes and comments. The frequency of topic 2 indicates that members were very concerned about physical changes, illness, and employment.

In conclusion, this study found that the most common topics related to thalassemia discussed on social media were the challenges of thalassemia patients and questions about treatment for thalassemia. Eight topics were discovered related to the concerns of thalassemia patients and their caregivers on social media. Topics 2, 4, 6, and 8 are consistent with the findings of past qualitative studies, while topics 1, 3, 5, and 7 are new discoveries resulting from the analysis of this study.

Apart from regular clinical care, thalassemia patients and caretakers should be provided with more resources for improving their quality of life, including offering more weekend thalassemia treatments to reduce work absenteeism.

Healthcare providers and other concerned parties, such as the government and non-governmental organizations, should also provide more health education and informational support to thalassemia patients, carriers, and caregivers to improve their understanding of the disease and their quality of life. Social media was used in this study to explore the health-related concerns of thalassemia patients, carriers, and caregivers. In addition, a new framework, known as MESMTPP, was applied to pre-process the noisy mixed-language (Malay and English) social media posts by normalizing and reducing the text of posts. Three topic models were tested, and the results showed that LDA in MALLET performed best according to the coherence score and the interpretation of the researchers. This study was limited to one social media platform only (Facebook). Thus, in the future, the MESMTPP framework can be applied to different social media platforms, as well as for other types of health issues to gather information and develop a better understanding of patients' health-related concerns.
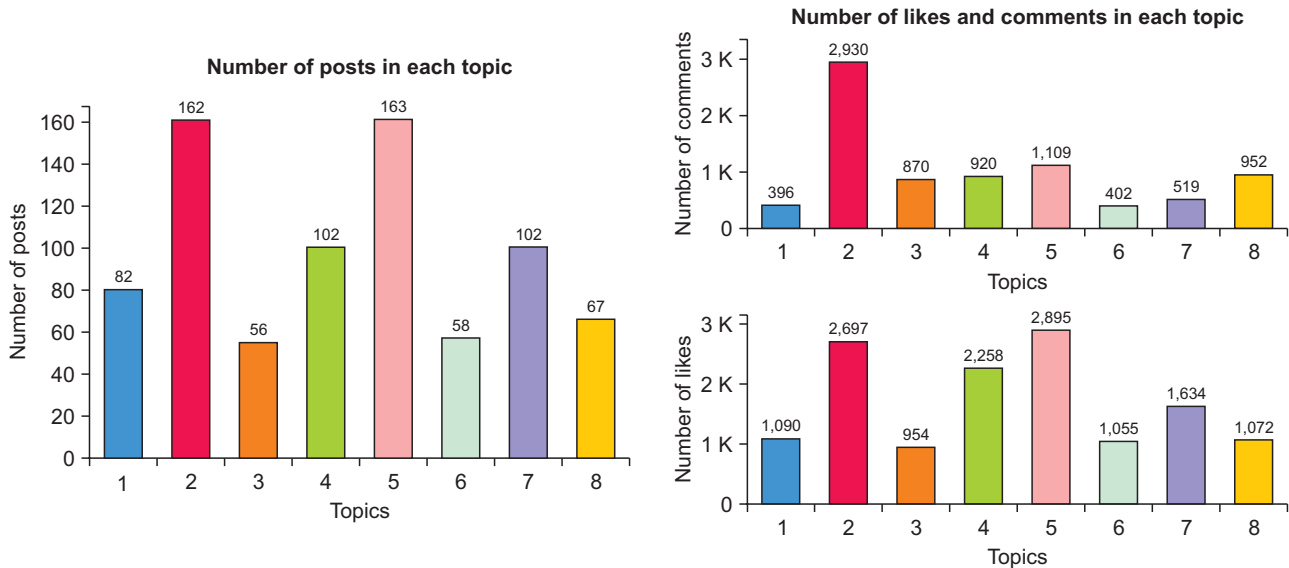
No.	Topic	No.	Topic
1	Treatment of thalassemia (umbilical cord and bone marrow-related treatment)	5	Asking questions/opinion about the information of the disease
2	Challenge of thalassemia patients (illness, work and life treatment side effect)	6	Blood donation (issues related to lack of blood and side effect)
3	Iron-rich foods and diet	7	Sharing experience thalassemia society and activities within the group
4	Praying for strength and peer support	8	Genetics of thalassemia and issues among carrier couples

Figure 6. Visualization of the number of posts, likes, and comments on each topic.

Conflict of Interest

Acknowledgments

ORCID

Yuen Chi Phang (https://orcid.org/0000-0002-4733-6930)
Azleena Mohd Kassim (https://orcid.org/0000-0002-7168-6535)
Ernest Mangantig (https://orcid.org/0000-0003-3355-3026)

References

1. Alnaami A, Wazqar D. Disease knowledge and treatment adherence among adult patients with thalassemia: a cross-sectional correlational study. Pielegniarstwo XXI wieku/Nurs 21st Century 2019;18(2):95-101.
2. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ 2008;86(6):480-7.
3. Mohd Ibrahim H. Malaysian thalassaemia registry report 2018. Putrajaya, Malaysia: Medical Development Division, Ministry of Health; 2019.
4. Wahab IA, Naznin M, Nora MZ, Suzanah AR, Zulaiho M, Faszrul AR, et al. Thalassaemia: a study on the perception of patients and family members. Med J Malaysia 2011;66(4):326-34.
5. Ismail WI, Hassali MA, Farooqui M, Saleem F, Aljadhey H. Perceptions of thalassemia and its treatment among Malaysian thalassemia patients: a qualitative study. Australas Med J (Online) 2016;9(5):103-10.
6. Shafie AA, Chhabra IK, Wong JH, Mohammed NS, Ibrahim HM, Alias H. Health-related quality of life among children with transfusion-dependent thalassemia: a cross-sectional study in Malaysia. Health Qual Life Outcomes 2020;18(1):141.
7. Mashayekhi F, Jozdani RH, Chamak MN, Mehni S. Caregiver burden and social support in mothers with β-thalassemia children. Glob J Health Sci 2016;8(12):206-12.
8. Platania S, Gruttadauria S, Citelli G, Giambrone L, Di Nuovo S. Associations of thalassemia major and satisfaction with quality of life: the mediating effect of social support. Health Psychol Open 2017;4(2):2055102917742054.
9. Paramore C, Levine L, Bagshaw E, Ouyang C, Kudlac A, Larkin M. Patient- and caregiver-reported burden of transfusion-dependent β-thalassemia measured using a digital application. Patient 2021;14:197-208.
10. Rocha HM, Savatt JM, Riggs ER, Wagner JK, Faucett WA, Martin CL. Incorporating social media into your support tool box: points to consider from genetics-based communities. J Genet Couns 2018;27(2):470-80.
11. Maheri A, Sadeghi R, Shojaeizadeh D, Tol A, Yaseri M, Rohban A. Depression, anxiety, and perceived social support among adults with beta-thalassemia major: cross-sectional study. Korean J Fam Med 2018;39(2):101-7.
12. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017;5(3):e23.
13. Curiskis SA, Drake B, Osborn TR, Kennedy PJ. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Inf Process Manag 2020;57(2):102034.
14. Tessore JP, Esnaola LM, Russo CC, Baldassarri S. Comparative analysis of preprocessing tasks over social media texts in Spanish. Proceedings of the XX International Conference on Human Computer Interaction; 2019 Jun 25-28; Donostia, Spain. p. 1-8.
15. Omar N, Hamsani AF, Abdullah NA, Abidin SZ. Construction of Malay abbreviation corpus based on social media data. J Eng Appl Sci 2017;12(3):468-74.
16. Husein Z. Malaya: Natural Language Toolkit for bahasa Malaysia [Internet]. GitHub Repository; 2018 [cited at 2021 Jun 29]. Available from: https://github.com/huseinzol05/malaya.
17. Kwee AT, Tsai FS, Tang W. Sentence-level novelty detection in English and Malay. In: Theeramunkong T, Kijsirikul B, Cercone N, Ho TB, editors. Advances in Knowledge Discovery and Data Mining. Heidelberg, Germany: Springer, 20009. p. 40-51.
18. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
19. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Discourse Process 1998;25(2-3):259-284.
20. Pouraboli B, Abedi HA, Abbaszadeh A, Kazemi M. The

burden of care: experiences of parents of children with thalassemia. J Nurs Care 2017;6(2):389.

21. Dadipoor S, Haghighi H, Madani A, Ghanbarnejad A, Shojaei F, Hesam A, et al. Investigating the mental health and coping strategies of parents with major thalassemic children in Bandar Abbas. J Educ Health Promot 2015;4:59.

22. Mufti GE, Towell T, Cartwright T. Pakistani children's experiences of growing up with beta-thalassemia major. Qual Health Res 2015;25(3):386-96.

23. Ishfaq K, Hashmi M, Naeem SB. Mothers' awareness and experiences of having a thalassemic child: a qualitative approach. Pak J Appl Soc Sci 2015;2(1):35-53.