

Struggle to Settle down! Examining the Voices of Migrants and Refugees on Twitter Platform

Aparup Khatua

L3S Research Center, Leibniz Universität Hannover,
Hannover, Germany
khatua@l3s.de

Wolfgang Nejdl

L3S Research Center, Leibniz Universität Hannover,
Hannover, Germany
nejdl@l3s.de

ABSTRACT

Prior studies, mostly from the social science domain, have observed that mental stress and struggles are high for refugees. Information science researchers have found that social media data can be insightful for probing psychological stress. However, none of the previous studies, to the best of our knowledge, investigated social media data to identify the voices of migrants and refugees and analyzed their concerns. We have collected 0.15 million tweets, but only 2% of these tweets are the voices of migrants and refugees. In addition to non-refugee and non-migrant voices, we have classified their voices into three themes as follows: their generic views, initial struggles, and subsequent settlement in the host country. We have employed deep learning and transformer-based models for identifying these themes. Our best-performing transformer-based model has reported an accuracy of 75.89%. We have also identified some exciting avenues for future research.

ACM Reference Format:

Aparup Khatua and Wolfgang Nejdl. 2021. Struggle to Settle down! Examining the Voices of Migrants and Refugees on Twitter Platform. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21 Companion)*, October 23–27, 2021, Virtual Event, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3462204.3481773>

1 INTRODUCTION

The presence of refugees and migrants can have various socio-economic consequences in host countries. A vast amount of existing literature from the social science domain [4, 9, 12, 25] explored health concerns of refugees [1, 5], their integration process [2], refugee-related policies and law [25], and the societal change due to refugees in the host country. In recent times, the information science domain is also probing social media platforms, such as Twitter, to understand the concerns related to migrants and refugees [3, 14, 18, 21, 35]. These studies employed sentiment analysis [23], explored the usage of hashtags [30, 31], or investigated the emergence of communication networks [28]. A few studies also probed Twitter data to analyze migration movement from one country to

another [35]. In brief, these studies mostly explored public opinions about refugees and migrants.

To the best of the knowledge, none of the prior Twitter-based studies has exclusively examined the voices and concerns of refugees and migrants. Extant literature from the social science stream has observed that refugees are constantly struggling due to anxiety [15], discrimination [22], hatred [16], and other stressful events before settling down in the host country. On the methodology front, these works have considered focused group discussion [22, 33], questionnaire-based survey [27, 29], and individual interviews [15]. These approaches generally employ structured or semi-structured questionnaires to investigate the concerns of migrants and refugees. On the contrary, the Twitter platform allows a user to express her opinions without any probing. Hence, Twitter deliberations might offer an exhaustive view of the struggles they face. Thus, this study attempts to contribute to the *Artificial Intelligence (AI) for social good* stream of the literature.

2 STRUGGLE TO SETTLE DOWN: A REVIEW

The online behavior of social media users can be an indicator of their mental well-being. For instance, [6] explored the online behavior of Facebook users after job loss. This study reveals that talking with close friends helps them to control their stress level and provide social supports; however, distant friends rarely provide these supports [6]. Covid19 crisis is one of the most stressful global events in the history of mankind. [34] explored the search pattern, on Microsoft's Bing platform, during the pandemic to investigate the shift in human needs across societal, economic, and psychosocial dimensions. Basic human needs went up drastically during this crisis, whereas higher-level aspirational needs experienced a downfall [34]. Additionally, temporal variations in human needs indicate instances of psychological and economic resilience [13, 34]. Similarly, childbirth can lead to postpartum depression in new mothers. [10] has probed the changes in emotion and behavior during the postpartum period and argued that postpartum depressions of new mothers could be predicted by analyzing their 'social engagement, emotion, ego-network, and linguistic styles' on the Twitter platform around childbirth.

A similar study in the Canadian context [5] has found that the propensity of having postpartum depression of refugee and asylum-seeking women is five times more than Canadian-born women. Not only postpartum depression, but also migrants experience loneliness due to separation from their own family members, face ethnic conflicts in the new country, face difficulties in getting an appropriate accommodation, struggle to learn a new language, and also face communication-related difficulties in the host country [22, 32]. These struggles to settle down create psychological stresses that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CSCW '21 Companion, October 23–27, 2021, Virtual Event, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8479-7/21/10...\$15.00

<https://doi.org/10.1145/3462204.3481773>

potentially lead to mental disorders [22]. Consequently, prior studies have found evidence of mental stress among refugees in the context of Australia, Canada, Denmark, Germany, Iraq, Netherlands, Switzerland, Syria, and in the Middle Eastern countries [5, 22, 26, 27, 29, 32, 33]. A meta-analysis of anxiety, depression, and post-traumatic stress disorder (PTSD) in refugees reveals that a significant portion of refugees has self-reported anxiety, depression, and PTSD [15]. [32] argues that exposure to traumatic experiences leads to substantial psychological impairment, and this study has also observed a strong correlation between the lack of social integration and mental health problem. Hence, [32] suggests fostering the social integration process for these traumatized refugees. Alarming, the prevalence of mental disorders of refugees from politically disturbed countries is substantially higher in comparison to non-refugee populations, and this pattern is similar for both child/adolescent refugees as well as adult refugees [15, 27]. For instance, [26] found a strong relationship between mental health, especially in young refugees, and traumatic events, such as arrest and imprisonment of family members, loss and separation from the family, street shooting, war, and bombing in the Middle East. The struggles of migrants and refugees to settle down are broadly similar irrespective of their origin and host countries.

Thankfully, the traumatic experiences of refugees and subsequent mental stress are not the end of the story. The resilience of migrants and refugees helps them to overcome these adverse situations (i.e., struggles) and keep them optimistic in achieving their target (i.e., settling down in the host countries). For example, [20] has observed that refugees from Arabs have displayed ‘resilience in successfully adapting to life in the USA’. Resilience can be defined as ‘the personal qualities that enables one to thrive in the face of adversity’ [13]. The resilience to come out of the adverse situation can be captured through qualities such as adaptability to a changed context, confidence to face a new challenge, ability to handle stress, don’t give up attitude, ability to handle unpleasant feelings, and tenacity to attain the desired goal [8, 13]. In brief, resilience enhances the propensity of successful adaptation and settling down in the host country after the initial phase of not-so-pleasant hostile experiences and adverse situations [26].

To sum up, our literature review reveals that social media data can be insightful for probing mental well-being (from the information science domain) and the level of mental stress is high for migrants and refugees (from the social science domain). However, none of the prior studies, to the best of our knowledge, probed the *first-person tweets from migrants and refugees* to understand their voices. We attempt to address this research gap in this paper.

3 DATA AND THEMES

Twitter data was widely used in the migration contexts [18, 21, 30, 35]. Hence, we also collect migrant and refugee-related tweets from May 2020 to March 2021 using the Twitter search API. For the initial crawling purpose, we have considered the following keywords: ‘*asylums*’, ‘*migrants*’, ‘*refugee*’, ‘*immigrants*’, etc. Our final corpus comprises 0.15 million tweets after discarding duplicate tweet-ids and tweets with similar text content from the initial corpus. A careful introspection reveals that a significant portion of our corpus comprises tweets by non-refugee or non-migrant users, but we

need to probe the refugee perspectives – their struggles, concerns, experiences, and resilience. We note that only a minuscule portion of our corpus captures these issues.

Tweets from migrants and refugees occasionally mention phrases such as “*I’m a/an refugee/migrant/immigrant*”, “*as a refugee/migrant I*”, and “*being a refugee/migrant*” (e.g., tweets #1 to 6). Some of the tweets from second-generation refugees and migrants also use phrases such as “*father/mother was a refugee/migrant*” or “*my parents were refugees/migrants*” (e.g., tweet # 4). Using the above phrases as search themes, we find that less than 3000 tweets are first-person tweets from migrants and refugees, i.e., a mere 2% of our corpus. This scant presence of first-person tweets elucidates the challenges and need to analyze the faint voices of migrants and refugees on Twitter platform. Table 1 reports a few representative tweets. Our tweets can be broadly divided into two classes: migrant and refugee voices and ‘*public views*’ (i.e., tweets by non-refugee or non-migrant users). We find public opinions are either supportive to refugees (e.g., tweet # 8) or abusive to them (e.g., tweet # 7).

Next, we have divided refugee voices into three classes: ‘*struggles*’ (i.e., tweets sharing their sufferings and mental stresses due to traumatic experiences), ‘*settlement*’ (i.e., successful adaptation in the host country or recovery through resilience), and ‘*generic views*’ (i.e., tweets about larger political or social issues, mostly related to refugee and migrant issues, but not about their personal experiences). Refugees use the Twitter platform to share their initial struggles and mental stress (e.g., tweet #1), such as traumatic detention experiences (e.g., tweet # 2) after arriving at the host country. These tweets capture the fear, anxiety, and challenges associated with accessing primary education and essential health facilities. Our next theme tells how their resilience helps some of them to overcome adverse situations to settle down in the host country (e.g., tweets # 3 & 4). Interestingly, a few tweets are generic, and these tweets are deliberating issues that are not specific to their personal journey (e.g., tweets # 5 & 6). We have analyzed the linguistic content of tweets and manually annotated 623 tweets into the above four categories for the final analysis. Some of the tweets are having multiple themes within it, but we considered the most salient theme/view for classifying these tweets.

4 METHODOLOGY AND ANALYSIS

Deep learning models, such as Bidirectional Long Short-Term Memory (Bi-LSTM) and Convolutional neural network (CNN), outperform traditional NLP models [7, 17]. Bi-LSTM models can efficiently capture long-term dependencies of very lengthy sequences; whereas the CNN models can capture the spatial information of words in low-dimensional vectors, and convolutional filters are applied to the word embedding matrix [19]. We have considered Bi-LSTM and CNN models with pre-trained embedding *fasttext* (i.e., wiki-news300d-1M of 16B tokens) for training. We have also considered transformer-based models, which evaluate the context of a given word from both left and right directions. Bidirectional Encoder Representations from Transformers (BERT) [11] and Robustly Optimized BERT Pretraining Approach (RoBERTa) [24] have considered broad-based large corpus for pre-training purpose, and an essential feature of these models is their multi-head self-attention to extract the semantic properties of input tokens for contextual

Table 1: Sample Tweets from our corpus

Sl.#	Representative Tweets	Themes
1.	Yes, it is so sad. <i>I am a refugee</i> with family about 7 years with horrible conditions. We are fighting days, nights. High level of depression, stress, uncertain future and mental; psychological problems cased (sic) him end his life. #UNHCR please do something.	Struggles of Migrants & Refugees (168 tweets)
2.	1178 days in a NZ jail. No crime, no charge, just for <i>being a refugee</i> .	
3.	<i>I am an immigrant</i> every year I pay my taxes, and nobody gives me anything.	Settlement in the host country (156 tweets)
4.	A century ago, my great-grandparents migrated for better economic opportunities. A few years later their daughter -my grandmother- left their new home country for love and family. I'm a proud descendant of migrants and today, <i>I am a migrant</i> myself.	
5.	<i>I'm an immigrant</i> rights activist, so you're wrong on every count. Every time you open your mouth, you're wrong. Did you get your right-wing feelings hurt? Poor thing. #JoeBidenInauguration	Generic Views of Migrants & Refugees (147 tweets)
6.	<i>I am a refugee</i> from Fox News who tuned in last month and now watch Newsmax. My bro and neighbor recommended we try and now we love your channel!	
7.	I demand an explanation from ____ why it takes so long to deport foreign criminals and illegal immigrants	Public views about Migrants & Refugees (152 tweets)
8.	Refugees deserve respect as much as the next person they are part of the solution you can be too (sic) listen to them join them support them on #worldrefugeeday and beyond	

Table 2: Accuracies in Identifying Refugee and Migrant Concerns

Batch Size	Bi-LSTM	CNN	Dropout	BERT	RoBERTa	Learning rate
16	61.61	67.86	0.4	75.00	75.00	2e-5
16	60.71	65.18	0.5	70.54	75.89	3e-5
32	56.25	66.07	0.4	68.75	74.11	2e-5
32	54.89	64.29	0.5	71.43	72.32	3e-5

representation. We consider pre-trained implementation from the HuggingFace library for our BERT and RoBERTa models [36]. We have considered 16 and 32 batch sizes and 'adam' optimizer for all these models.

Table 2 reports the classification accuracies. We have considered multiple hyperparameters for robustness. Accuracies of CNN-based models are higher (in the range of 64% to 67%) than Bi-LSTM models (55% to 60%) for identical hyperparameters. The best F1-Score for CNN is 67.86%. As expected, BERT and RoBERTa models have significantly outperformed Bi-LSTM and CNN models. Broadly, the accuracies of RoBERTa models are higher than BERT models. The best performing RoBERTa model has reported an F1-Score of 75.89% (for batch size 16 and learning rate 3e-5). Table 2 strongly indicates that our transformer-based models are efficient in identifying and analyzing the concerns of migrants and refugees. Hence, opinion mining of user-generated social media data can be insightful for regulators and policymakers.

5 CONCLUSION AND FUTURE WORK

Concerns of refugees were probed by using focused group discussion, questionnaire-based analysis or individual interviews [14, 22, 29, 33]. Recent studies have also employed NLP tools on social media data, such as Twitter, to gauge the public opinions about refugees and migrants [18, 21, 28]. Our literature review

suggests that none of the prior studies probed the voices and concerns of refugees and migrants on the Twitter platform. Hence, we have considered *first-person tweets* sharing the personal experiences of refugees and migrants. We have broadly categorized their voices into two categories: *struggle* and *settlement*. Additionally, we have also identified another class where these migrants or refugees are not sharing their success stories or personal struggles but talking about generic issues. Our last category captures public opinion. Some of these tweets are supportive, whereas some of them are abusive towards refugees and migrants. We have used deep learning (Bi-LSTM and CNN) and transformer-based models (BERT and RoBERTa) for our classification task. We note that transformer-based models efficiently identify and classify these voices and concerns, and RoBERTa has outperformed other models. However, still, there is a room for improvement in modeling to increase the accuracy and the corpus size can be extended in future work.

Our study has a few limitations which might offer exciting avenues for future research. First, a further fine-grained classification of these themes might offer a more nuanced view. For example, we have clubbed a wide range of challenges and constraints as struggles. It is worth noting that some of them are struggling due to their inability to access essential medical services. In contrast, others are worried about the future of their kids in the asylum because they are not getting the primary education. Some don't have legal

documents, and some are traumatized due to hostile behaviors or lack of a proper social integration process. These struggles are not uniform, and needs are different. So, a fine-grained classification will help organizations like United Nations Commission on Human Rights (UNCHR) to take appropriate policy measures. It is worth noting that we have collected data for a year, and we could gather around 3000 tweets. Intuitively, it can be argued that refugees in the asylums might not have the facility or internet connectivity to tweet about their sufferings. Hence, understanding their voices and concerns would be a challenging task.

Second, our paper has considered refugees, migrants, and immigrants synonymously. It is quite possible that xenophobic social media users can assume most refugees and migrants are illegally entering their country. However, UNCHR policies have clearly pointed out that these three categories are not the same. The mental stress of refugees coming from politically disturbed countries will be much higher than an immigrant. Our corpus reveals that the struggles of refugees are associated more with basic human needs like food or shelter. On the contrary, the struggles of an immigrant are primarily associated with a lack of proper social integration or discriminating behaviors from a specific section of the host country. Similarly, we also note that most support activities by charitable organizations are for distressed refugees. We also note that a significant portion of the success stories of settling down in the host country mostly comes from migrants and immigrants. Hence, future work should tease out the voices of refugees, migrants, and immigrants. A fine-grained analysis will enrich the contextual understandings about migrants and refugees.

ACKNOWLEDGMENTS

Funding for this paper was, in part, provided by the European Union's Horizon 2020 research and innovation program under Grant Agreement No: 832921.

REFERENCES

- [1] M. F. Al Qutob. 2018. Refugee healthcare: Refugees in crisis. *British Dental Journal* 225, 3: 201–201.
- [2] Amanda Alencar. 2018. Refugee integration and social media: a local and experiential perspective. *Information, Communication & Society* 21, 11: 1588–1603.
- [3] Cairtin Armstrong, Ate Poorthuis, Matthew Zook, Derek Ruths, and Thomas Soehl. 2021. Challenges when identifying migration from geo-located Twitter data. *EPJ Data Science* 10, 1: 1.
- [4] Sami Atallah and Dima Mahdi. Law and Politics of “Safe Zones” and Forced Return to Syria: Refugee Politics in Lebanon. *Beirut: The Lebanese Center for Policy Studies*.
- [5] Amy Brown-Bowers, Kelly McShane, Karline Wilson-Mitchell, and Maria Gurevich. 2015. Postpartum depression in refugee and asylum-seeking women in Canada: A critical health psychology perspective. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 19, 3: 318–335.
- [6] Moira Burke and Robert Kraut. 2013. Using facebook after losing a job: differential benefits of strong and weak ties. In *Proceedings of the 2013 conference on Computer supported cooperative work* 1419–1430.
- [7] Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann Lecun. 2017. Very Deep Convolutional Networks for Text Classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Association for Computational Linguistics, 1107–1116.
- [8] Kathryn M. Connor and Jonathan R.T. Davidson. 2003. Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety* 18, 2: 76–82.
- [9] Heaven Crawley. 2017. Migration: Refugee economics. *Nature* 544, 7648: 26–27.
- [10] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting post-partum changes in emotion and behavior via social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 3267–3276.
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*.
- [12] Christian Dustmann, Francesco Fasani, Tommaso Frattini, Luigi Minale, and Uta Schönberg. 2017. On the economics and politics of refugee migration. *Economic Policy* 32, 91: 497–550.
- [13] David Fletcher and Mustafa Sarkar. 2013. Psychological Resilience: A Review and Critique of Definitions, Concepts, and Theory. *European Psychologist* 18, 1: 12–23.
- [14] Mar Gallego, Estrella Gualda, and Carolina Rebollo. 2017. Women and Refugees in Twitter: Rhetorics on Abuse, Vulnerability and Violence from a Gender Perspective. *Journal of Mediterranean Knowledge* 2.
- [15] Jens-R. Henkelmann, Sanne de Best, Carla Deckers, et al. 2020. Anxiety, depression and post-traumatic stress disorder in refugees resettling in high-income countries: systematic review and meta-analysis. *BJPsych Open* 6, 4: e68.
- [16] Matouš Hrdina. 2016. Identity, Activism and Hatred: Hate Speech against Migrants on Facebook in the Czech Republic in 2015. *Náš společnost* 1, 14: 38.
- [17] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. *arXiv:1404.2188 [cs]*.
- [18] Aparup Khatua and Wolfgang Nejdl. 2021. Analyzing European Migrant-related Twitter Deliberations. *Companion Proceedings of the Web Conference 2021*, ACM, 166–170.
- [19] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. *arXiv:1408.5882 [cs]*.
- [20] Ibrahim Aref Kira, Mona M. Amer, and Nancy Howells Wrobel. 2014. Arab Refugees: Trauma, Resilience, and Recovery. In S.C. Nassar-McMillan, K.J. Ajrouch, and J. Hakim-Larson, eds., *Biopsychosocial Perspectives on Arab Americans*. Springer US, Boston, MA, 175–195.
- [21] Ramona Kreis. 2017. #refugeesnotwelcome: Anti-refugee discourse on Twitter. *Discourse & Communication* 11, 5: 498–514.
- [22] Cornelis J. Laban, Hajo B. P. E. Gernaat, Ivan H. Komproe, Ingeborg van der Tweel, and Joop T. V. M. De Jong. 2005. Postmigration Living Problems and Common Psychiatric Disorders in Iraqi Asylum Seekers in the Netherlands: *The Journal of Nervous and Mental Disease* 193, 12: 825–832.
- [23] Ju-Sung Lee and Adina Nerghe. 2018. Refugee or Migrant Crisis? Labels, Perceived Agency, and Sentiment Polarity in Online Discussions. *Social Media + Society* 4, 3: 205630511878563.
- [24] Yinhan Liu, Myle Ott, Naman Goyal, et al. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [cs]*.
- [25] Valsamis Mitsilegas. 2017. Humanizing solidarity in European refugee law: The promise of mutual recognition. *Maastricht Journal of European and Comparative Law* 24, 5: 721–739.
- [26] Edith Montgomery. 2010. Trauma and resilience in young refugees: A 9-year follow-up study. *Development and Psychopathology* 22, 2: 477–489.
- [27] Wadih J. Naja, Michaelangelo P. Aoun, Eliane L. El Khoury, Fabiola J. Bou Abdallah, and Ramzi S. Haddad. 2016. Prevalence of depression in Syrian refugees and the influence of religiosity. *Comprehensive Psychiatry* 68: 78–85.
- [28] Adina Nerghe and Ju-Sung Lee. 2018. The Refugee/Migrant Crisis Dichotomy on Twitter: A Network and Sentiment Perspective. *Proceedings of the 10th ACM Conference on Web Science*, ACM, 271–280.
- [29] Y. Nesterko, D. Jäckle, M. Friedrich, L. Holzapfel, and H. Glaesmer. 2020. Prevalence of post-traumatic stress disorder, depression and somatisation in recently arrived refugees in Germany: an epidemiological study. *Epidemiology and Psychiatric Sciences* 29: e40.
- [30] Mehmet Gökyaz Özerim and Juliette Toley. 2020. Discussing the populist features of anti-refugee discourses on social media: an anti-Syrian hashtag in Turkish Twitter. *Journal of Refugee Studies*: feaa022.
- [31] Nandita Adilfi Eti Riyadi and Maria Regina Widhiasti. 2020. Racism and Stereotyping of Refugees: The Use of the Hashtag #Aufschrei by Twitter Users in Germany. *Proceedings of the International University Symposium on Humanities and Arts (INUSHARTS 2019)*, Atlantis Press.
- [32] Matthias Schick, Andre Zumwald, Bina Knöpfli, et al. 2016. Challenging future, challenging past: the relationship of social integration and psychological impairment in traumatized refugees. *European Journal of Psychotraumatology* 7, 1: 28057.
- [33] Robert Schweitzer, Lisa Buckley, and Donata Rossi. 2002. The psychological treatment of refugees and asylum seekers: what does the literature tell us? *Mots Pluriels*, 21.
- [34] Jina Suh, Eric Horvitz, Ryen W. White, and Tim Althoff. 2021. Population-Scale Study of Human Needs During the COVID-19 Pandemic: Analysis and Implications. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, ACM, 4–12.
- [35] Stefanie Urchs, Lorenz Wendlinger, Jelena Mitrovic, and Michael Granitzer. 2019. MMoveT15: A Twitter Dataset for Extracting and Analysing Migration-Movement Data of the European Migration Crisis 2015. *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET-ICE)*, IEEE, 146–149.
- [36] Thomas Wolf, Lysandre Debut, Victor Sanh, et al. 2020. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv:1910.03771 [cs]*.