"Rough Day … Need a Hug": Learning Challenges and Experiences of the Alzheimer's Disease and Related Dementia Caregivers on Reddit

Congning Ni[1], **Bradley Malin**[1,2,3], **Lijun Song**[4], **Angela Jefferson**[5,6], **Patricia Commiskey**[5], **Zhijun Yin**[1,2]

[1]Department of Computer Science, Vanderbilt University
[2]Department of Biomedical Informatics, Vanderbilt University Medical Center
[3]Department of Biostatistics, Vanderbilt University Medical Center
[4]Department of Sociology, Vanderbilt University
[5]Department of Neurology, Vanderbilt University Medical Center
[6]Vanderbilt Memory and Alzheimer's Center, Vanderbilt University Medical Center
{congning.ni, lijun.song}@vanderbilt.edu, {b.malin, angela.jefferson, patricia.commiskey, zhijun.yin}@vumc.org

Abstract

Alzheimer's disease and related dementia (ADRD) is a collection of disorders involving mental deterioration, which is often quite distressing for the individuals and those caring for them. As online social media become prevalent, many people share their ADRD care challenges and experiences in online environments. Despite encouraging findings in the literature regarding online support among ADRD caregivers, studies to date have focused only on a single online community about ADRD, which leads to an incomplete picture of the needs of ADRD caregivers. Additionally, the large volume of data from online communities makes it challenging for both researchers and caregivers to discover discussions about ADRD care efficiently. In this paper, we focus on Reddit, an online rating and discussion platform that consists of many communities, or *subreddits*, and aim to analyze the topic difference regarding ADRD care between ADRD and non-ADRD subreddits. To do so, we develop a two-stage classification framework to extract posts about ADRD care from the entire Reddit. We then apply structured topic modeling to investigate what is discussed regarding ADRD care and the prevalence of such discussions in different types of subreddits. Our results show that non-ADRD subreddits contribute 68.5% of the submissions about ADRD care - more than twice as many as ADRD subreddits. Moreover, non-ADRD subreddits are more likely to disclose legal and financial issues, mental health, and negative relationships, while ADRD subreddits are more likely to talk about memory loss, issues about sleep and diet, the disease, and visits to healthcare providers. Our findings suggest that research in this area should consider discussions that take place beyond ADRD-specific communities to gain a comprehensive understanding of ADRD care experiences and challenges.

1 Introduction

Dementia is a clinical syndrome that impairs memory, thinking, and social abilities severely enough to interfere with daily life (Chertkow, Feldman et al. 2013). Alzheimer's disease is the most common cause of dementia and affects over 5 million people in the U.S. and more than 47 million people worldwide. It is a high-stress situation for the individuals diagnosed with the disease, as well as for their informal caregivers like family members, friends, or other unpaid caregivers (Bryden 2015). In 2020, more than 11 million informal caregivers in the U.S. provided approximately 15.3 billion hours of care to people with ADRD, which was valued at nearly $256.7 billion (Alzheimer's Association et al. 2021). The rapidly increasing number of Americans living with ADRD is overloading these caregivers with substantial physical, financial, and mental stress (Manzini and Do Vale 2020; Sajwani 2020), which makes them known as the *invisible second patients* (Brodaty and Donkin 2009). This has only worsened during the COVID-19 pandemic, which has further increased care challenges and needs for people providing care to those with ADRD (Lo, Ramic et al. 2020).

To improve the quality of life for people living with ADRD and their informal caregivers, it is critical for healthcare providers, researchers, or practitioners to understand and respond to the needs of these individuals (Wennberg et al. 2015). Various studies have summarized and evaluated the problems faced by ADRD caregivers through traditional interactions, such as surveys and interviews (Gibson, Walsh, and Brown 2018). Although these studies provide some insights into ADRD caregiving challenges, they are often biased toward predefined questionnaires and can provide an incomplete and inaccurate picture of the scenario due to response bias (Gove and Geerken 1977). In addition, the ways in which caregivers express their needs through face-to-face or access local support services may be restricted due to privacy concerns (Rose, Coop Gordon et al. 2021) and social stigma (Maxfield and Greenberg 2020).

ADRD caregivers have increasingly sought support and shared their experiences through various general-purpose online social platforms, such as Twitter (Berry, Lobban et al. 2017), Facebook (Pagán-Ortiz, Cortés et al. 2014), as well as online health communities like ALZConnected (Du, Paiva et al. 2020). While it has been shown that participating in online communities has certain benefits (e.g., a reduction in depressive symptoms) (Wilkerson, Brady et al. 2018), the collection of studies to date are limited in that they focused only on a single ADRD community, such as *r/Alzheimers* on Reddit (Wang, Zou et al. 2021), which does not provide a complete picture of the challenges that ADRD caregivers confront. To better understand the mechanism of online support for ADRD caregivers, it is necessary for researchers first to know *where* and ADRD caregivers communicate and *what* they talk about.

However, due to the free-form nature of online platforms, ADRD care posts can be published in any communities that their authors think proper. Given the large amount of online data generated in numerous communities, it would be exorbitantly expensive, in both time and cost, for researchers to manually extract and examine such information. Moreover, ADRD caregivers have limited time and often suffer from highly dynamic caregiving journeys due to the increasingly impaired ADRD patients over time (Schulz, Eden et al. 2016). Struggles to search for desired information in a massive quantity of online data only exacerbates stress that, in turn, may make it more likely that they fail to pursue support in an online setting.

In the research communicated in this paper, we aim to collect and understand the challenges and needs in ADRD caring-related submissions on an online social media platform, Reddit. Reddit is an American social news aggregation, web content rating and discussion website[1]. The platform contains millions of communities called *subreddits* that discuss specific topics (e.g., *r/dementia* and *r/politics*). A subreddit contains many *discussion threads* where the initial posts are called *submissions* and all of the following responses are called *comments*. More specifically, our work is guided by two hypotheses:

- **H1:** We can build efficient machine learning classification models to identify online submissions on Reddit that are related to ADRD care challenges and experiences.
- **H2:** ADRD care information disclosed in the non-ADRD communities differs from that in the ADRD communities.

In this paper, we apply computational methods to understand the challenges and experiences of ADRD caregivers on a large number of Reddit submissions. Figure 1 illustrates our research workflow, as well as how each component is organized in the following sections to address the proposed hypotheses. The primary contributions of our research are as follows:

- We introduce an efficient two-stage machine learning framework to identify submissions about ADRD care from Reddit on a large scale. We build and compare both traditional and deep learning-based models on common machine learning metrics. Our best-performing model achieves an AUC of 0.94 (± 0.01).
- We conduct a structured topic modeling to identify the main ADRD care challenges and experiences disclosed on Reddit. In addition to the common issues that are found in previous studies, we gain further insights into topics about detailed daily matters, including *routine recommendations* and *complex kinship relationships*.
- We show that there are meaningful differences in the topics learned from ADRD and non-ADRD subreddits, which explains why it is important to extend investigations about online ADRD caregivers into non-ADRD communities. We finally discuss the implications of our work to ADRD caregivers and researchers, as well as to text classification to solve real-world problems.

[1]https://www.reddit.com

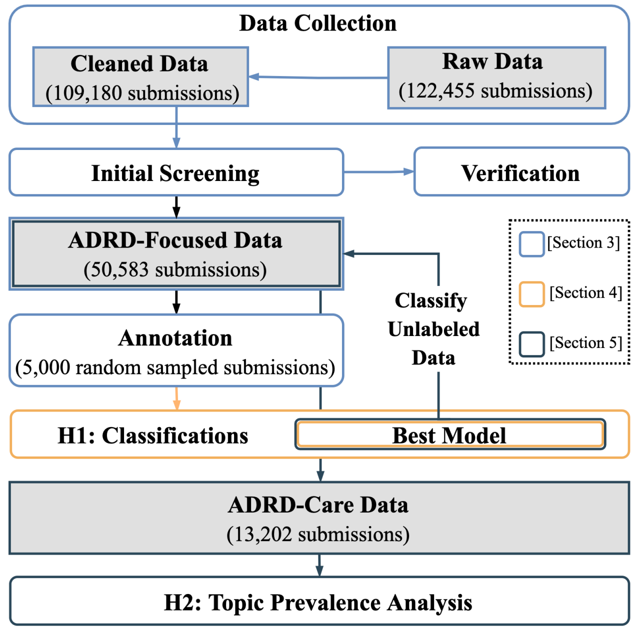

Figure 1: A depiction of the research workflow, where grey blocks represent data, non-grey blocks represent the processes associated with a particular experiment.

2 Related Work

Challenges in Caring for People Living With ADRD

Various investigations have relied upon both qualitative and quantitative methods to study the difficulties faced by ADRD caregivers. For example, it has been shown that caring for patients with ADRD is often overwhelming. ADRD caregivers experience more serious emotional, financial, and physical difficulties than those who provide care to patients without ADRD (Alzheimer's Association et al. 2021). Notably, Sikder et al. showed that family caregivers were often under high-intensity stress and were very likely to suffer from depression (Sikder, Yang et al. 2019). Given that people living with ADRD experience memory loss, decreased self-care ability, mood disorders, and other complications over time, their caregivers often suffer from stress due to the extended time needed to provide care and associated challenges, which are more common in individuals at late-stage ADRD (Han, Chi et al. 2019).

In addition, the physical health of caregivers is affected by daily caring responsibilities. High-intensity stress and workload increase caregivers' susceptibility of developing health complications (Riedel, Klotsche et al. 2016). Also, ADRD caregivers experience exacerbated hardships due to economic issues, lack of knowledge about clinical care, and insufficient access to medical resources (Stone 2015; Borson, Boustani et al. 2016; Alzheimer's Association et al. 2021). Furthermore, due to the continuity and incurability of ADRD, caregivers tend to experience a more fragile family structure. For example, in the face of natural disasters (Gibson, Walsh, and Brown 2018), or epidemics such as COVID-19 (Williamson, McCarthy et al. 2020), ADRD families are more vulnerable than other families. As a consequence, ADRD caregivers require more social attention and support. In this paper, we show that the information available in online communities provides a comprehensive characterization of ADRD care challenges and experiences.

Social Support for ADRD Caregivers

Offline support aids ADRD caregivers through local peer support groups and professional programs (National Institute of Aging 2019). For example, in a recent retrospective cohort study (Zechner, Lundquist et al. 2020), it was shown that peer supporters had the potential to address the practical, physical, social, and emotional needs of the surveyed caregivers. A scoping review of interventions found that a wide range of interventions were designed to improve ADRD family caregiver outcomes in countries with various income levels in Asia (Hinton et al. 2019). However, offline social support, from both peers and professionals, remains limited due to geographical constraints, the characteristics of caregivers (Wilkerson, Brady et al. 2018), and, more recently, the negative impact of the COVID-19 pandemic (Lo, Ramic et al. 2020).

Online social media platforms have proven to be valuable to supplement local peer support groups and professional support services for ADRD caregivers. For example, it was observed that information exchange about how to care for ADRD family members in an online community increased a sense of mastery and reduced symptoms of depression for caregivers from ethnic minority groups (Pagán-Ortiz, Cortés et al. 2014). More recently, a study showed that ALZConnected, an online Alzheimer's community, provided useful resources for caregivers to seek informational and emotional support while caring for patients with ADRD (Du, Paiva et al. 2020). Still, though such studies have demonstrated the significance of social media for supporting ADRD caregivers, they have been limited to specific online communities and groups. In this study, we focus on a substantially larger scale of data, ranging across many communities on Reddit, to characterize where and what online ADRD caregivers discuss regarding their challenges and experiences.

Machine Learning for Online Health Data

Online communities generate huge amounts of data, which requires machine learning to extract health-related information about personal and public health (Yin, Sulieman, and Malin 2019). Machine learning applications in this domain include, but are not limited to, opioid addiction detection (Fan et al. 2017), flu trend analysis (Alessa and Faezipour 2019), depression prediction (Aldarwish and Ahmad 2017), sentiment analysis of tweets on COVID-19 (Rustam, Khalid et al. 2021), and ADRD care experience (Al-Bahrani, Danilovich et al. 2017). Recently, Wang and colleagues proposed a method that combined domain knowledge and machine learning to detect different types of information exchange in *r/Alzheimers* (Wang, Zou et al. 2021). Our research is essentially different in that we focus on detecting all of the potential ADRD care submissions from Reddit. Another interesting study (Rajadesingan, Budak, and Resnick 2021) proposed a similar framework to study the toxicity of posts in political/non-political subreddits. However, they aimed to estimate the prevalence of political discussion in each subreddit, while we aim to predict whether a submission with ADRD keywords is talking about ADRD care. While both studies make comparisons between two types of subreddits that are complementary to each other, our work is fundamentally different from theirs in terms of methods, domain, and implications.

3 Data Preparation

Reddit is a popular social media platform with approximately 52 million daily active users and 430 million monthly users[2]. All of the Reddit posts, including submissions and their comments, are publicly accessible. In this paper, we focus on submissions only to understand the potential challenges and experiences faced by ADRD caregivers and defer the analysis of comments to future work. This study is exempt from human subjects research by the Institutional Review Board at our university. All of the submission quotes presented in this paper have been rephrased for privacy consideration and demonstration purpose (Proferes et al. 2021).

Data Collection

We apply two methods to collect the data through the pushshift.io API[3]. First, we collect all of the submissions within the four main subreddits of ADRD: *r/Alzheimer*, *r/Alzheimers*, *r/dementia*, and *r/AlzheimersGroup*. Second, we extract all of the submissions from across Reddit that contain at least one of the following pre-defined keywords: *alzheimer, alzheimers, dementia, dementias*. Third, we combine these two datasets and remove duplicate submissions. For convenience, in this paper, we refer to the aforementioned four ADRD-focused subreddits as *ADRD subreddits* and all of the other subreddits as *non-ADRD subreddits*. Table 1 provides a summary of the data used in this study.

ADRD Subreddits	Subreddits	4
	Submissions	15,675
	Users	9,051
Non-ADRD Subreddits	Subreddits	13,473
	Submissions	93,505
	Users	67,392
Time period	11/30/2005 - 12/04/2020	

Table 1: Summary statistics for the cleaned dataset.

[2]https://www.businessofapps.com/data/reddit-statistics/

[3]https://pushshift.io/

We remove the stop-words and apply the Ekphrasis Python package (version 0.5.1)[4], a text processing tool for social tokenization and spell correction, to correct misspellings. We also replace unstructured expressions, such as time and URLs into predefined notations (e.g., "☺" to "*smile*", "05/01/2021" to "*date*") using Ekphrasis. We convert the extracted tokens in each submission into lower case.

Initial Screening of ADRD Submissions

While our data collection methods ensure that a large number of submissions about ADRD care are collected, it also introduces many submissions that are not related to ADRD (e.g., "*He must be crazy, I think he got dementia.*"). Sampling posts from such data would make annotators read many obviously unrelated submissions, which would be a waste of time and effort. To narrow the search scope, we perform an *initial screening* by building a binary classification model to remove data that do not focus on ADRD topics.

This is accomplished by generating positive and negative classes of the same size. The positive class is generated by all of the submissions from *ADRD subreddits*. Specifically, we randomly select the same number of the most recent submissions from the following "Today's Top Growing Communities[5]" (on 4th Dec, 2020, the day we stopped data collection) to form a negative class: *r/politics, r/Games, r/Pets, r/travel, r/cyberpunkgame, r/science, r/gtaoline, r/explainlikemfive, r/lifeProTipcs*. Ideally, any combination of common subreddits that are highly unrelated to ADRD can serve this purpose. To further verify if there exist any posts in the negative class that are in fact related to ADRD (and would, thus, be false negatives), we conduct two additional checks. First, we search all of the posts in the negative class for any post that contained one of the ADRD keywords that we apply to search on Reddit. This search turns up 43 posts. After manually reading these posts, we find that none of them are related to ADRD. Second, among the posts that lack an ADRD keyword in the negative class, we randomly select 100 posts and, by manual examination, we find that all were unrelated to the ADRD topic. As such, we believe it is unlikely that the negative class we create has posts with ADRD topics.

We use the words in each submission as features and apply term frequency (TF) - inverse document frequency (IDF) to generate feature values. TF-IDF determines the relative frequency of words in a specific document compared to the inverse proportion of that word over the entire document corpus. We use *TfidfVectorizer* method in the python *scikit-learn* package (version 1.0) to generate the word features. We use these features to train a binary logistic regression model to predict whether a submission is related to ADRD. We utilize 10-fold cross-validation to evaluate the model performance. The fitted model results in an area under the ROC curve (AUC) score of 0.994 (± 0.012), suggesting a very well-defined separation between positive and negative classes.

[4]https://github.com/cbaziotis/ekphrasis

[5]https://www.reddit.com/subreddits/leaderboard/

Since our ultimate goal is to identify submissions related to ADRD care, we adjust the threshold of the decision boundary of the fitted classifier to ensure that it can recognize all of the potential submissions talking about ADRD. In other words, the model should have a recall of 1.0. This leads to a threshold of 0.8, which corresponds to a recall of 1.0 and a precision of 0.3. After applying this screening classifier to all of the collected data, we obtain 50,583 ADRD submissions.

Data Annotation

We construct a gold-standard dataset that indicated whether a submission is related to ADRD care by annotating 5,000 randomly selected submissions from the ADRD-focused data. The annotated data are then relied upon to train classification models to efficiently identify ADRD care submissions on a large scale **H1**.

Four annotators are recruited and trained on an annotation codebook. The codebook is created by the authors based on our review of submission samples and the literature. The annotators are all Reddit users and have general knowledge of ADRD. The following provides an example of is a submission labeled positive for ADRD: "*Hi!! my name is *name* and I'm *age*. I have a ton of anxiety (and comorbid depression, yay) that I'm in counseling for; Currently, I take care of my grandma with Dementia. I play bass and smoke a lot of weed, and I'm just looking for people to talk to*".

We use Labelbox[6], an online platform for data labeling, as the annotation tool. Each submission is annotated by two annotators independently. If a submission receives contradicting labels, they are labeled by a third annotator to break the tie. After the first round of annotation, 69% of submissions receive the same labels. The remaining submissions are sent to the second round of annotation to break a tie. This process produce 5,000 labeled submissions with 38.3% (1,914) of them related to ADRD care.

4 Classification Models

Now that we have the annotated dataset, we present the process to train and compare traditional and deep learning-based classification models to identify whether a submission is about ADRD care (**H1**). The best-performing model will be applied to all of the unlabeled submissions in the screened dataset for further analysis.

Traditional Classifiers

We apply four classical machine learning models for the binary classification: logistic regression (LR), random forest (RF), k-nearest neighbors (KNN), and support vector machine with a linear kernel (LinearSVC). We construct both sparse and dense features to train each model.

Sparse Features. A sparse feature is composed mostly of values equal to zero. We generate sparse features through a bag-of-word representation, where each submission is coded into a vector with each cell representative of a word in the vocabulary defined by the dataset. Within this representation, we use two types of feature values: 1) Word Count and

[6]https://labelbox.com/

2) Word TF-IDF. Similarly, we generate a bag of character n-grams ($n \in [2, 5]$) with TF-IDF as feature values (N-Grams).

Dense Features. Sparse features can potentially limit model generalizability. As such, we generate three additional types of dense features by applying 1) linguistic inquiry and word count (LIWC) (Pennebaker, Francis, and Booth 2001), 2) Word2Vec (Mikolov, Chen et al. 2013), and 3) pretrained Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018):

LIWC is utilized to map the words of a submission into linguistic categories (Chung and Pennebaker 2012). In this study, we include all of the 93 categories proposed in LIWC, including, but are not limited to, positive or negative emotions (e.g., happy and cried), biological process (e.g., blood and pain), and cognitive process (e.g., cause and know).

Word2Vec is used to generate a vector representation such that the semantic similarity of two given words can be measured by calculating the cosine distance of their according vectors. For this type of feature, we use the Google pretrained model *word2vec-google-news-300* to obtain word embedding for each word in the vocabulary. We generate a submission embedding by averaging the word vectors within the submission weighted by TF-IDF values.

BERT is a language model based on the encoder part of Transformer, which is designed to learn bidirectional representations from a very large amount of unlabeled text. In this task, we apply the Google pretrained BERT model, *bert-base-uncased* (12-layer, 768-hidden, 12-heads, 110M parameters), to obtain the vector representation of each submission. Each submission is limited to a maximum length of 256 words, and the submissions with less (more) words than the maximum are padded (truncated) accordingly. While the final hidden states (the transformer outputs) of the input tokens can be concatenated to get the encoded representation of a sentence, as is a standard convention, we use the hidden states of the [CLS] token of the last layer of the BERT pretrained model as the dense feature of submission for the further classification tasks.

Deep Learning Classifiers

We apply four deep learning classifiers based on a long short-term memory network (LSTM) (Greff, Srivastava et al. 2016) and BERT. LSTM is a type of recurrent neural network that is capable of learning long-term dependencies and has been widely applied in natural language processing (Zhou, Wan, and Xiao 2016). We implement these models using PyTorch (version 1.4.0) and transformers (version 3.5.1).

LSTM with one-hot word encoding (LSTM-One-Hot). We use one-hot encoding to represent each word in a submission. We truncate or pad the sequence of encoding in a submission to be of size 256. The sequences of encodings are then fed to an LSTM for model training.

LSTM with Word2Vec embeddings (LSTM-Word2Vec). Since word2vec is designed to represent semantic meaning, we introduce a similar model with LSTM-One-Hot that uses the output of the Google pretrained Word2Vec model in the embedding layer. We anticipate this model will outperform the previous model.

LSTM with attention (LSTM-Attention). We append a layer of single-head attention (Vaswani, Shazeer et al. 2017) to the Bidirectional-LSTM (Bi-LSTM) model (with Google pretrained Word2Vec as the input embedding layer) as an extended classifier. The attention mechanism is used to focus attention on the information output from the hidden layer of Bi-LSTM, which is expected to further boost the model performance.

BERT fine-tuning (BERT-Fine-Tuning). Finally, we fine-tune BERT for ADRD care submission identification. Fine-tuning is a relatively inexpensive process whereby we continue to train all of the parameters of a pretrained BERT model together with the downstream machine learning task (Devlin et al. 2018), which, in our case, is a binary classification. By adapting the task-specific context, this model is expected to outperform all of the other models.

Model Training and Evaluation

We adopt a nested 10-fold cross-validation to train and evaluate the performance of each classifier. Specifically, we first apply 10-fold stratified shuffle splits with a testing size of 0.3 to ensure the training and testing data in each sample set preserve the proportion of submissions for each class. Next, for each sample set, we utilize grid search with 10-fold cross-validation within training data to optimize the hyperparameters based on F1. Then, we test the model performance using the test data in each sample set. We report on the mean and standard deviation (SD) of accuracy, precision, recall, F1, and AUC for each model, and finally compare model performance using a Student's t-test at a significant level of 0.05.

Classification Results

Table 2 presents the performance of the traditional classifiers, where the best score of each metric is shown in bold. There are several notable findings to highlight. First, the models based on Word2Vec features (LR and SVM) achieve the best accuracy, recall, F1 and AUC ($p < 0.05$). By contrast, the features generated from the pretrained BERT model do not result in any of the best-performing models. This suggests that, without adapting the task-specific context, the complex BERT model may not be as efficient as the much simpler word2vec model in capturing effective submission embeddings. Second, RFC with word TF-IDF features achieves the best precision and accuracy. It is not surprising to observe that word TF-IDF features are more efficient than word count features for classification, since TF-IDF can help identify the most informative words of a submission. Third, RFC with LIWC features also achieves the best accuracy. Upon inspection of the top informative LIWC features, we observe it is notable that ADRD care submissions are more likely to use linguistic categories like *family, female, shehe, money, focuspresent, home, drives* and *health*, while submissions not about ADRD care are more likely to use linguistic categories like *focuspast, informal, leisure, body, bio* and *swear*. However, all of the traditional models exhibit recall under 0.70, which may not be sufficiently high to extract as many ADRD care submissions as needed.

Model	Metric	Sparse vector				Dense vector	
		Word Count	Word TF-IDF	N-Grams	LIWC	Word2Vec	BERT pretrained
LR	Accuracy	0.74 (0.01)	0.75 (0.01)	0.76 (0.01)	0.74 (0.01)	**0.77 (0.01)** *	0.73 (0.01)
	Precision	0.68 (0.02)	0.70 (0.03)	0.71 (0.01)	0.67 (0.02)	0.72 (0.01)	0.67 (0.02)
	Recall	0.58 (0.03)	0.61 (0.03)	0.65 (0.02)	0.60 (0.02)	0.67 (0.02)	0.59 (0.02)
	F1	0.63 (0.02)	0.66 (0.02)	0.68 (0.01)	0.64 (0.01)	**0.69 (0.01)**	0.62 (0.01)
	AUC	0.78 (0.01)	0.82 (0.01)	0.83 (0.01)	0.80 (0.01)	**0.85 (0.01)** *	0.80 (0.01)
RFC	Accuracy	0.76 (0.01)	**0.77 (0.01)** *	0.76 (0.01)	**0.77 (0.01)** *	0.76 (0.01)	0.72 (0.01)
	Precision	0.76 (0.02)	**0.77 (0.01)**	0.73 (0.02)	0.73 (0.02)	0.74 (0.02)	0.68 (0.03)
	Recall	0.56 (0.03)	0.56 (0.03)	0.58 (0.02)	0.62 (0.02)	0.58 (0.02)	0.50 (0.02)
	F1	0.64 (0.02)	0.65 (0.02)	0.65 (0.01)	0.67 (0.02)	0.65 (0.02)	0.58 (0.02)
	AUC	0.83 (0.01)	0.84 (0.01)	0.83 (0.01)	0.84 (0.01)	0.84 (0.01)	0.78 (0.02)
KNN	Accuracy	0.63 (0.01)	0.68 (0.01)	0.68 (0.01)	0.66 (0.02)	0.72 (0.01)	0.68 (0.02)
	Precision	0.54 (0.02)	0.63 (0.03)	0.64 (0.03)	0.55 (0.02)	0.69 (0.03)	0.57 (0.02)
	Recall	0.24 (0.02)	0.39 (0.02)	0.35 (0.02)	0.57 (0.03)	0.49 (0.01)	0.61 (0.03)
	F1	0.33 (0.02)	0.48 (0.01)	0.45 (0.02)	0.56 (0.02)	0.57 (0.02)	0.59 (0.02)
	AUC	0.59 (0.01)	0.70 (0.02)	0.69 (0.01)	0.70 (0.01)	0.77 (0.01)	0.72 (0.02)
SVM	Accuracy	0.70 (0.01)	0.74 (0.01)	0.76 (0.01)	0.62 (0.10)	**0.77 (0.01)** *	0.71 (0.01)
	Precision	0.61 (0.02)	0.67 (0.02)	0.70 (0.02)	0.58 (0.13)	0.71 (0.01)	0.64 (0.03)
	Recall	0.60 (0.04)	0.62 (0.03)	0.65 (0.02)	0.64 (0.33)	**0.68 (0.01)**	0.56 (0.05)
	F1	0.61 (0.02)	0.64 (0.02)	0.68 (0.01)	0.52 (0.16)	**0.69 (0.01)** *	0.60 (0.02)
	AUC	0.73 (0.01)	0.80 (0.01)	0.83 (0.01)	0.73 (0.07)	0.84 (0.01)	0.77 (0.01)

Table 2: The average (SD) accuracy, precision, recall, F1 and AUC of traditional machine learning classifiers. Bold font indicates the best-performing score for each metric. *indicates $p < 0.05$ on a Student's t-test that assessed whether the performance of the model in question is statistically significantly different than the second-best performing model.

Table 3 shows the mean and SD of each metric for the deep learning classifiers. The best model is BERT-Fine-Tuning ($p < 0.001$), followed by LSTM-Attention, LSTM-Word2vec, and LSTM-One-Hot. This confirms the general practice in natural language processing, as explained in the description of each proposed model, that each model has advantages over the next one in the list. It should be noted that the BERT-Fine-Tuning model outperforms all of the traditional machine learning models as well ($p < 0.001$). By contrast, the other three deep learning-based models are not competitive with the traditional models in any metric. The BERT-Fine-Tuning model is applied to classify all of the unlabeled data in the screened dataset.

Error Analysis of Misclassified Submissions

To gain insight into where the BERT-Fine-Tuning model makes mistakes, we extract the prediction results from one of the test sets to perform a qualitative analysis to the misclassified submissions.

Among the 1,500 test set submissions that we extract, the number of true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) submissions is 418 (27.7%), 878 (58.5%), 110 (7.3%) and 94 (6.3%), respectively. We conduct the error analysis on the misclassified submissions from two perspectives.

First, we perform a t-test to compare the length of TP submissions and the length of FN submissions. We find that TP submissions are significantly shorter than FN submissions (357 words vs. 533 words on average, $p = 0.003$). This may be due to the fact that the BERT-Fine-Tuning model truncates the posts during training and prediction. However, we do not observe a significant difference in submission length between TN and FP submissions under a t-test (568 words vs. 467 words on average, $p = 0.100$). This suggests that a BERT-Fine-Tuning model with larger segmentation may help improve model performance but may need more training data as well.

Second, we sample some of the FP and FN posts to examine the content manually. We find that many FP submissions mentioned ADRD patients but discuss negative life experiences that are not related to ADRD care. One example of such a submission is *"I really need vent/talk/sympathy, I'm really not used to crying around most people. I just found out that my dad is going to be hospitalized for four days next week for a test to see if he is eligible for a liver transplant. The maternal grandfather I grew up adoring was teetering under his growing dementia, now what? ...it's too much, but even when I cry, I don't feel like I can do anything."* On the other hand, FN submissions seem to communicate the care story with a positive tone. A representative example is *"Dealing with dementia reaching out in love - *number* years ago, my wife was diagnosed with a genetic disorder that causes dementia, uncontrollable body movements, and ultimately death... I created this blog to help our family, and hopefully all of the caregivers... I pray this blog is a blessing to you."* This suggests that Reddit submissions that describe multi-topic stories may bring challenges for classification. The issue may be mitigated with more training examples or multi-task learning.

Model	Metric	Score
	Accuracy	0.66 (0.02)
	Precision	0.56 (0.04)
LSTM-One-Hot	Recall	0.55 (0.08)
	F1	0.55 (0.04)
	AUC	0.71 (0.02)
	Accuracy	0.69 (0.02)
	Precision	0.60 (0.04)
LSTM-Word2Vec	Recall	0.60 (0.10)
	F1	0.60 (0.03)
	AUC	0.75 (0.02)
	Accuracy	0.73 (0.01)
	Precision	0.66 (0.03)
LSTM-Attention	Recall	0.62 (0.08)
	F1	0.63 (0.03)
	AUC	0.80 (0.01)
	Accuracy	**0.87 (0.01)** ***
	Precision	**0.79 (0.02)** ***
BERT-Fine-Tuning	Recall	**0.84 (0.02)** ***
	F1	**0.81 (0.01)** ***
	AUC	**0.94 (0.01)** ***

Table 3: Performance of the deep learning classifiers. Bold font indicates the best score for each metric. ***indicates $p < 0.001$ on a Student's t-test that assessed whether the performance of the model in question is statistically significantly different than the second-best performing model.

5 Topic Prevalence Analysis

In this section, we focus on analyzing discussion disparities between communities (**H2**) by studying topics difference in ADRD care submissions published in *ADRD subreddits* and *non-ADRD subreddits*.

Classifying Unlabeled Submissions

We fine-tune the BERT model with all 5,000 annotated samples to generate a classifier. Using all of the available labeled data points to refit the selected model is a common strategy that is widely adopted in Kaggle machine learning competition to maximize its performance. We apply this refitted model to all of the unlabeled data within the 50,583 screened submissions (see Figure 1) to predict whether each of them is related to ADRD care. We obtain 13,202 positive submissions (26.1% of the screened submissions, combining the labeled and the predicted data) that are from 1,197 subreddits.

Figure 2 presents the top 20 subreddits with the largest number of ADRD care submissions in a barplot and the top 50 subreddits in a word cloud. The font size of each subreddit in the word cloud is proportional to the number of ADRD care (positive) submissions in this subreddit. It can be seen that submissions related to ADRD care unsurprisingly appear in ADRD subreddits (e.g., r/dementia: 2,794 submissions and r/Alzheimers: 1,339 submissions). However, there is also a substantial sizable number of positive submissions from *non-ADRD subreddits* that are dedicated

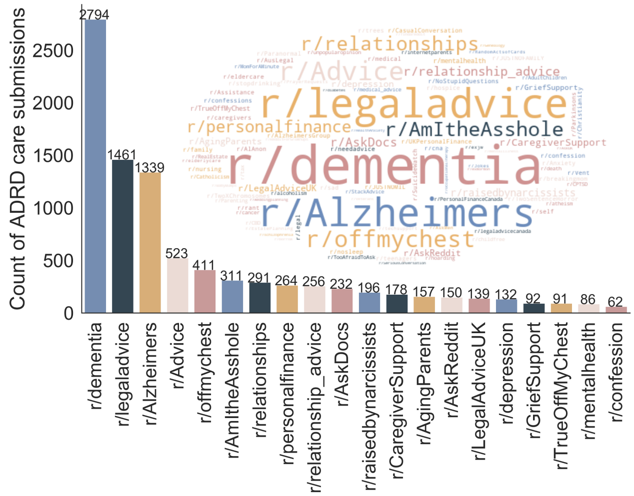

Figure 2: The bar chart shows the top 20 subreddits with the highest number of submissions related to ADRD care distributed in barplot. The word cloud is composed of the top 50 subreddits.

to many other topics such as consultation and suggestions (e.g., r/legaladvise: 1,461 submissions; r/Advice: 523 submissions). Together, these *non-ADRD subreddits* contribute 68.5% of all of the ADRD care submissions, more than twice as many as *ADRD subreddits*.

To better understand what has been communicated and the differences in topics between *ADRD subreddits* and *non-ADRD subreddits* (**H2**), we conduct the following topic prevalence analysis on the 13,202 positive submissions.

Structural Topic Modeling

In this task, we apply structural topic modeling (STM) (Roberts, Stewart, and Tingley 2019), as implemented in stm R package (v1.3.6), to investigate topic prevalence or proportion. In comparison to standard topic modeling, STM allows us to incorporate document-level metadata (e.g., authorship and creation time) into the topic modeling to investigate topic prevalence regarding the metadata. STM has been shown to improve inference and qualitative interpretability and is widely adopted in computational social science (Reber 2019; Schatto-Eckrodt, Janzik et al. 2020). In this paper, we introduce a *boolean* metadata variable to indicate whether a submission is posted within *ADRD subreddits* or *non-ADRD subreddits*, and investigate how the topics are different in these two types of subreddits.

To prepare the input of STM, we remove words with a frequency less than 3. Since STM is an unsupervised learning method, where the number of topics has to be set before running the algorithm, we rely on *exclusivity* and *semantic coherence* to select the optimal topic number K^* from a topic number candidate list of $\{5, 10, 15, 20, 25, 30\}$. Exclusivity refers to the distinctiveness of the words with the highest frequencies in the topic (Bischof and Airoldi 2012), while semantic coherence quantifies how the words in a topic frequently co-occur together in general contexts (Mimno et al.

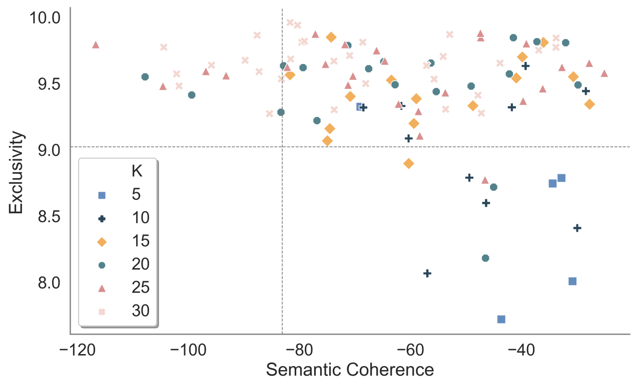

Figure 3: Exclusivity and semantic coherence between different candidate topic numbers. Each dashed line corresponds to one SD smaller than the average value of the metric, respectively. The model with the best K should have the largest number of topic markers residing on the top right region of the figure to achieve balanced exclusivity and semantic coherence scores.

2011). Although high exclusivity and semantic coherence scores are preferred, in practice, a higher exclusivity will lead to a lower semantic coherence score (vice versa). Therefore, an optimal topic number K^* should reach a balance between the two metrics for each topic. Figure 3 shows the values of the two metrics for each topic number candidate. Based on the aforementioned criterion, we prefer to select a K that has more topic markers appear on the top right region of Figure 3. Among the six predefined topic number candidates, $K = 15$ has relatively higher scores for both metrics, and no topics appear in areas where exclusivity or semantic coherence is small (e.g., left or bottom regions). Hence, we chose $K^* = 15$ and re-run STM until it converges.

Care Topics

Figure 4 shows the 15 resulting topics, along with their corresponding one-phase summaries, top words and topic proportions presented in a decreasing order. These one-phase summaries are generated by examining the submissions with the highest topic proportion for each topic by authors. The top words are the most representative words in each topic, which are ranked according to a score defined by $\beta_{w,k}(\log \beta_{w,k} - 1/K^* \sum_{k'} \log \beta_{w,k'})$, where $\beta_{w,k}$ is the probability of seeing word w conditioned on topic k and K^* is selected as 15 in this paper. If a word w has a higher probability in topic k than in other topics, it will have a larger score value based on the formula. Similar to TF-IDF, this score puts a higher rank on words that are more likely to differentiate a topic from others.

Specifically, Topic #14 [*Feelings*] (*feel, just, realli*) exhibits the highest proportion among the 15 topics. For example, one of the most representative submissions for Topic #14 is "*My grandpa has dementia and is getting worse ... I cannot stop crying and worrying, I do not know what to do. I feel so hopeless and stuck.*", which suggests that there are many caregivers who share their emotions about their care experiences in the online setting. Moreover, both positive and negative sentiments are included in these topics. For instance, Topic #3 [*Mental Issues*] contains the terms *depress, anxieti, suicid*, while Topic #1 [*Positive experience*] contains the terms *smile, dream, love*. Also, there exist some topics related to care experiences, like #10 [*Home care*] (*home, nurs, live*), #7 [*Legal and Financial Issues*] (*money, attorney, estat*) and #9 [*Communication*] (*call, told, phone*), which describe the problems and challenges that caregivers face when caring for a person with ADRD.

Additionally, caregivers discuss clinical issues regarding ADRD patients. For example, Topic #8 [*Recommendation*] (*app, phone, devic*) talks about seeking suggestions for caring, especially daily care (e.g., "*my dad has dementia ... had trouble operating our standard smart TV ... which TV to pick for ease of use?*"), and Topic #5 [*Disease and Symptoms*] (*hospit, doctor, activ*) summarizes information about the hospitals, doctors, and medication (e.g., "*My papa has dementia, he has to visit hospital a lot. Hospitals right now are full and with so much work in hands, so difficult now*").

Lastly, as expected, it is observed that many ADRD caregivers are relatives and family members, as shown in Topic #6 and #11 [*Family Member*]. Topic #4 [*Negative Relationship*] (**curse word*, sister, relationship*) suggests the emotional tense among family members during caring for ADRD patients, like "*she's in early stages of dementia, and she's extremely toxic towards me ... she's constantly belittling everyone and I am just sick of her *curse words**".

Topic Prevalence Across Subreddits

Figure 5 shows the effects of the topic prevalence contrast between *ADRD subreddits* and *non-ADRD subreddits*. The effect value for each topic is estimated by regressing a topic proportion in a post on the binary meta variable indicating whether the post is published in *ADRD subreddit*. This process is done by a repeat of sampling topic proportions from the posteriors estimated within STM. A positive (negative) effect suggests that the corresponding topic is more likely to be discussed in *ADRD (non-ADRD) subreddits*. There are several notable observations from this portion of the investigation.

First, in comparison to *non-ADRD subreddits*, the most prevalent topic in *ADRD subreddits* is Topic #2 [*ADRD*] (*alzheim, memori, diseas*), which is directly related to the disease. For example, "*#Diagnosed with early-onset Alzheimer's disease at *age*# My mom is *age* and her doctor gave her the diagnosis of early-onset.*". While this does not mean that the ADRD topic is not discussed in *non-ADRD subreddits*, an effect of 0.049 (± 0.001, $p < 0.001$) suggests that, on average, the ADRD topic prevalence will increase 0.049 if the online environment changes from *non-ADRD subreddits* to *ADRD subreddits*.

Second, the topics that are more disclosed in *ADRD subreddits* also include #13 [*Memory Loss*] (*orget, rememb, convers*) (e.g. "*Is it appropriate to tell her over and over again things that might upset her when she forgets things so quickly?*") and #12 [*Sleep and Eating Issues*] (*sleep, door, night*) (e.g. "*Are there any tips for dealing with severe nocturnal incontinence??*"). Both of these are problems caused

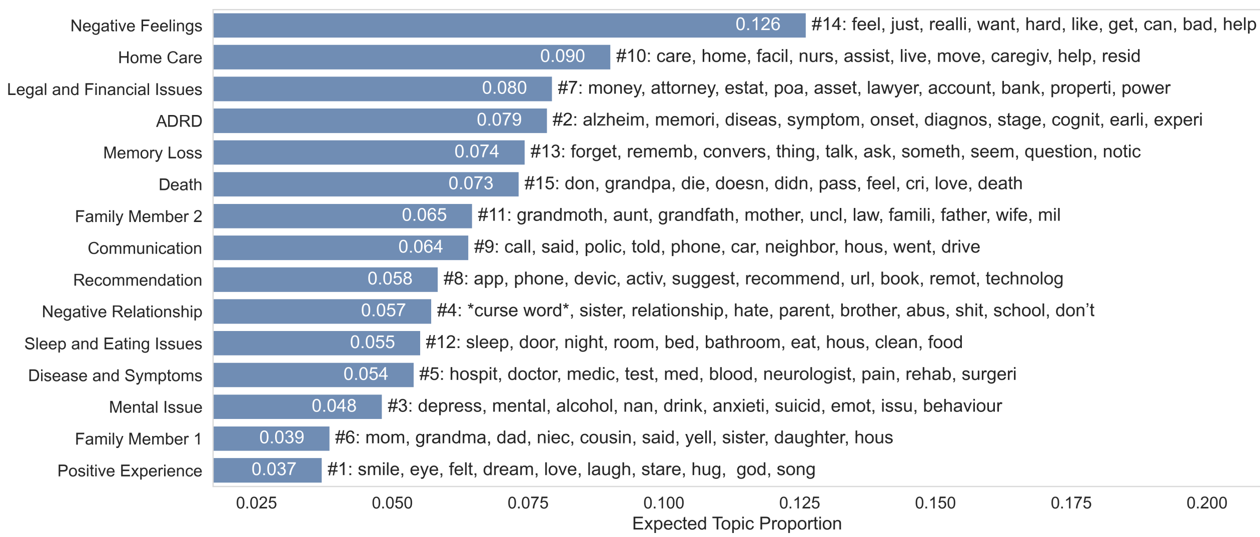

Figure 4: Topics generated by STM, sorted in decreasing order of expected topic proportions. The proportion score of each topic is shown to the right of each bar, while the 10 most representative words in the topic are shown to the right.

by the symptoms of ADRD disease.

Third, the most prevalent topic in *non-ADRD subreddits* is Topic #7 [*Legal and Financial Issues*] (*money, attorney, estat*), which highlights the legal and financial problems for ADRD caregivers due to the loss of memory for ADRD patients. For example, "*...Given that there is no dispute over the will, debts and relatively few assets, how does an estate attorney require?*". The caregivers facing this problem usually use more specific subreddits (for example, *r/bestoflegaladvice* and *r/personalfinance*) to seek support, instead of discussing in *ADRD subreddits*. Similarly, an effect of -0.079 (± 0.002, $p < 0.001$) suggests that, on average, the #7 [*Legal and Financial Issues*] topic prevalence will decrease 0.079 if we change online environment from *non-ADRD subreddits* to *ADRD subreddits*. However, considering that the proportion of this topic is only 0.080, it further suggests that most of the related posts are published in *non-ADRD subreddits* (e.g., *r/legaladvice*).

Fourth, Topic #4 [*Negative Relationships*] (**curse word*, sister, relationship*) and #6 [*Family Member 1*] (*mom, grandma, dad*), #11 [*Family Member 2*] (*grandmoth, aunt, grandfath*) are topics related to relationships, which are also more frequently discussed in *non-ADRD* rather than in *ADRD subreddits*.

Lastly, there is no significant difference on Topics #1 [*Positive Experience*] ($p = 0.137$), #9 [*Communications*] ($p = 0.206$), #10 [*Home Care*] ($p = 0.272$) between two types of subreddits. This means that both types of communities talk about the positive experience but neither widely mention it (e.g., Topic #1 has the smallest proportion), suggesting that challenges are still the major topics in these subreddits. In addition, it makes sense for Topics #9 and #10 to be insignificant since they are general words regarding caregiving and communications.

These differences in topic prevalence suggest that online ADRD caregivers go to different subreddits for seeking support and sharing experiences based on the subreddit topics. Research focusing on merely ADRD subreddits might not be able to gain a complete picture of the care challenges and experience.

6 Discussion

The objective of this study is to investigate what aspects of ADRD care are discussed on Reddit, as well as to characterize the differences in ADRD care topics within and outside of *ADRD subreddits*. Here, we discuss the implications of our research findings as follows.

ADRD Care Topics on Reddit

We identify many topics from submissions about ADRD care on Reddit, which can be summarized as *home care*, *mental health*, *memory loss*, *negative feelings* and *sleep and eating issues*. These topics are well-aligned with the findings in the previous investigations on the major challenges faced by ADRD caregivers (Van Den Kieboom et al. 2020; Gaugler, Zmora et al. 2019; Polzer, Nearing et al. 2021). This suggests that when an ADRD caregiver has a question or is looking for social support, it is likely that what they are seeking has already been discussed in an online community(Song and Chang 2012). If such discussion threads can be easily found, it could save the caregiver a substantial amount of time and effort. In turn, this will make online support a valuable, supplemental means to the local support groups or services to mitigate their certain limitations on privacy, social stigma, as well as limited face-to-face contact during the COVID-19 pandemic. In addition to the common challenges found in previous studies, Reddit contains many detailed descriptions of daily matters, including *clinical visits*, *negative family relationships* and *routine advice*. These are the topics that caregivers are more

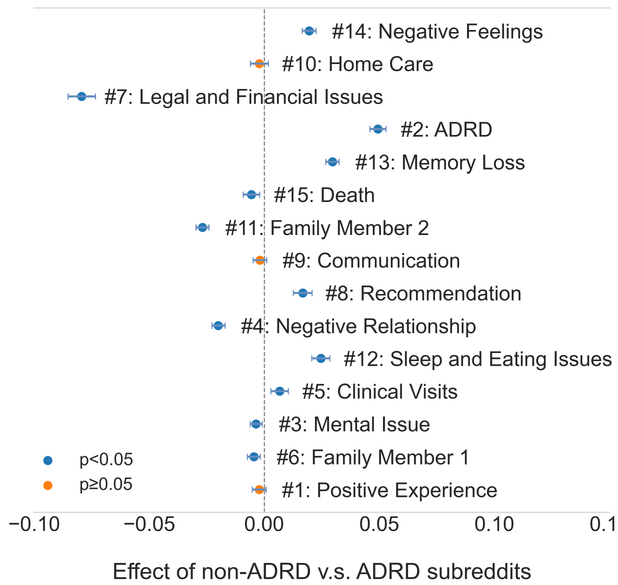

Figure 5: Topic prevalence contrast between *non-ADRD subreddits* and *ADRD-subreddits* with their 95% confidence intervals. A positive (negative) value of the X-axis indicates that the related topic is more prevalent in ADRD subreddits (non-ADRD subreddits). The topics are ordered by decreasing expected topic proportion value.

likely to seek suggestions about, as well as support from, online peers, which has been shown potentials to improve mental health (Gillard 2019). It should be noted that caregivers seeking online support may have to be cautious about the misinformation from open, public-accessible online environments (Bautista, Zhang, and Gwizdka 2021). However, dealing with, or evaluating the impact of misinformation is beyond our investigation.

Topic Prevalence Across Communities

Our experiments show that 68.5% of submissions about ADRD care are published in *non-ADRD subreddits* - far more than those in *ADRD subreddits*. For researchers, this suggests that focusing only on *ADRD subreddits* is insufficient to understand the needs of online caregivers. This further suggests relying solely on ADRD subreddits will not be sufficient to provide effective coping strategies under certain circumstances. Therefore, if decision support for ADRD caregivers relies on online environments, it is critical to ensure that they integrate discussion threads from within, as well as beyond, ADRD-specific communities. At the same time, for ADRD caregivers, this indicates that they may want to surf beyond *ADRD subreddits* to find out their desired information or support (e.g., regarding family relationships, legal or financial issues, see Figure 5). However, it might not be easy for ADRD caregivers to locate these submissions because they could be published anywhere outside *ADRD subreddits*, which may become a barrier for using online support to assist ADRD caregivers. An efficient tool or service to screen and search ADRD care posts from massive online data may benefit caregivers who have challenges in using technology in searching for relevant information.

More broadly, our topic and prevalence analysis suggest that it would be necessary and important to categorize and distinguish communities on social media into two types: identity-driven or demand-driven. Some online communities are created to meet the general needs of people with the same social or group identities (e.g., caregivers, LGBTQ people and teachers), while others are created to address a specific type of demand (e.g., legal advice, relationship, and stock). These two types of communities provide different forms of support (Song, Son, and Lin 2011; Song and Zhang Forthcoming). In our study, the *ADRD subreddits* can be considered as identity-driven communities where most of them are informal ADRD caregivers, while *non-ADRD subreddits* are more relevant to specific demands on topics such as legal advice, mental health, and relationships that caregivers encounter when caring for an ADRD patient. It is equally important to focus on demand-driven communities where caregivers with specific demands tend to go, and identity-driven communities in which how to take care of ADRD patients are often discussed. The partition may be motivated by empathy and identity, but further investigations are needed to investigate this conjecture.

Classification of ADRD Care Submissions

The best-performed classification we build, the BERT-Fine-Tuning model, achieves a high AUC of 0.94. This investigation provides additional evidence that pretrained deep learning-based language models can efficiently capture the context of online disclosure and improve classification performance. At the same time, we find that traditional methods have similar or better performance than other RNN-based deep learning models. For example, the logistic regression model that incorporates Word2Vec as features (with an AUC of 0.85) outperforms LSTM-attention classifier (with an AUC of 0.80). This suggests that when BERT pretrained models are not available for a certain domain (e.g., clinical care), traditional models can serve as reasonable surrogates to save training time and resources, while obtaining acceptable prediction performance and maintaining good interpretability.

Health-related and non-health-related communities often coexist in online environments. For example, COVID-19 infection can be discussed in *r/COVID-19Positive* or any other subreddits. Therefore, our approach to identifying and analyzing relevant information can be applied to studies that use online data to understand patients or their caregivers in general.

Limitations and Future Work

There are certain limitations of our study that can serve as a basis for future research. First, our findings are limited to a sample of contributors to Reddit. Although this platform is the seventh most-visited website in the U.S.[7], our data are biased in terms of language usage (mainly in English), geographic region (mainly in the United States and Canada),

[7] https://foundationinc.co/lab/reddit-statistics/

and the fact that not all caregivers are online users. As such, future research should consider other online platforms, and particularly characterize the special needs of minority ethnic, non-English-speaking caregivers (Nielsen, Nielsen, and Waldemar 2021). In addition, our investigation may benefit from supplemental surveys targeting caregivers who do not use social media. Second, while the BERT-Fine-Tuning model results in the best model performance, several strategies could be taken for boosting the performance further, like training models with larger segmentation and more annotated posts, performing multi-task learning, or constructing combinations of multiple classification models. Third, it will be meaningful to investigate how other online caregivers support or harm the ADRD caregivers who ask questions in the initial posts (or submissions on Reddit) through the following comments. Another notable research direction is to consider the extent to which the relationships between people suffering from ADRD and their caregivers (e.g., spouses, children, or friends) impact the ADRD care experiences through using online social media data. In conclusion, studying the challenges experienced by people with ADRD would be a valuable future endeavor.

7 Conclusions

In this study, we have applied structured topic modeling to analyze what are communicated in ADRD care submissions on Reddit, and how such topics vary between *ADRD subreddits* and *non-ADRD subreddits*. We find that while Reddit covers the challenges and experiences that are commonly perceived by ADRD caregivers, a large amount of such information is actually posted in *non-ADRD subreddits*, which have significantly different topic prevalence from *ADRD subreddits*. To facilitate the topic analysis, we introduce a two-stage machine learning framework to help identify submissions about ADRD care challenges and experiences. The BERT-Fine-Tuning model built upon the 5,000 annotated submissions achieves an AUC of 0.94, which significantly outperforms traditional machine learning or other deep learning-based models. Despite the merits of this investigation, future research should consider how online caregivers support each other in a discussion thread as well as in a community over time.

References

Al-Bahrani, R.; Danilovich, M. K.; et al. 2017. Analyzing informal caregiving expression in social media. In *IEEE International Conference on Data Mining Workshops (ICDMW)*, 342–349.

Aldarwish, M.; and Ahmad, F. 2017. Predicting depression levels using social media posts. In *2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS)*, 277–280.

Alessa, A.; and Faezipour, M. 2019. Flu outbreak prediction using twitter posts classification and linear regression with historical centers for disease control and prevention reports: prediction framework study. *JMIR Public Health and Surveillance* 5(2): e12383.

Alzheimer's Association; et al. 2021. 2021 Alzheimer's disease facts and figures. *Alzheimer's & dementia* 17(3).

Bautista, J. R.; Zhang, Y.; and Gwizdka, J. 2021. Healthcare professionals' acts of correcting health misinformation on social media. *International Journal of Medical Informatics* 148: 104375.

Berry, N.; Lobban, F.; et al. 2017. # WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. *Journal of medical Internet research* 19(4): e107.

Bischof, J.; and Airoldi, E. M. 2012. Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th International Conference on Machine Learning*, 201–208.

Borson, S.; Boustani, M. A.; et al. 2016. Report on milestones for care and support under the U.S. National Plan to Address Alzheimer's Disease. *Alzheimer's and Dementia* 12(3): 334–69.

Brodaty, H.; and Donkin, M. 2009. Family caregivers of people with dementia. *Dialogues in clinical neuroscience* 11(2): 217.

Bryden, C. 2015. *Nothing about us, without us!: 20 years of dementia advocacy*. Jessica Kingsley Publishers.

Chertkow, H.; Feldman, H. H.; et al. 2013. Definitions of dementia and predementia states in Alzheimer's disease and vascular cognitive impairment: consensus from the Canadian conference on diagnosis of dementia. *Alzheimer's research & therapy* 5(1): 1–8.

Chung, C.; and Pennebaker, J. 2012. *Linguistic Inquiry and Word Count (LIWC): pronounced "Luke",... and other useful facts.*, 206–229.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Du, Y.; Paiva, K.; et al. 2020. Diabetes-related topics in an online forum for caregivers of individuals living with Alzheimer disease and related dementias: Qualitative inquiry. *Journal of Medical Internet Research* 22(7): e17851.

Fan, Y.; Zhang, Y.; Ye, Y.; li, X.; and Zheng, W. 2017. *Social Media for Opioid Addiction Epidemiology: Automatic Detection of Opioid Addicts from Twitter and Case Studies*, 1259–1267.

Gaugler, J. E.; Zmora, R.; et al. 2019. Six-month effectiveness of remote activity monitoring for persons living with dementia and their family caregivers: an experimental mixed methods study. *The Gerontologist* 59(1): 78–89.

Gibson, A.; Walsh, J.; and Brown, L. M. 2018. A perfect storm: Challenges encountered by family caregivers of persons with Alzheimer's disease during natural disasters. *Journal of gerontological social work* 61(7): 775–789.

Gillard, S. 2019. Peer support in mental health services: where is the research taking us, and do we want to go there? *J Ment Health* 28(4): 341–344.

Gove, W. R.; and Geerken, M. R. 1977. Response bias in surveys of mental health: An empirical investigation. *American journal of Sociology* 82(6): 1289–1317.

Greff, K.; Srivastava, R. K.; et al. 2016. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28(10): 2222–2232.

Han, S.; Chi, N.-C.; et al. 2019. Adapting the resilience framework for family caregivers of hospice patients with dementia. *American Journal of ADRD* 34(6): 399–411.

Hinton, L.; Tran, D.; Nguyen, T.-N.; Ho, J.; and Gitlin, L. 2019. Interventions to support family caregivers of people living with dementia in high, middle and low-income countries in Asia: a scoping review. *BMJ Global Health* 4(6): e001830.

Lo, C. H.; Ramic, A.; et al. 2020. Home-based care for people with alzheimer's disease and related dementias (Adrd) during covid-19 pandemic: From challenges to solutions. *International journal of environmental research and public health* 17(24): 9303.

Manzini, C. S. S.; and Do Vale, F. A. C. 2020. Emotional disorders evidenced by family caregivers of older people with Alzheimer's disease. *Dementia e Neuropsychologia* 14(1): 56–61.

Maxfield, M.; and Greenberg, J. 2020. Anticipated stigma and dementia-related anxiety in middle-aged and older adults. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry* 34(1): 13–22.

Mikolov, T.; Chen, K.; et al. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

Mimno, D.; Wallach, H.; Talley, E.; Leenders, M.; and McCallum, A. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, 262–272.

National Institute of Aging. 2019. Getting Help with Alzheimer's Caregiving. Technical report, NIH National Institute on Aging (NIA). URL https://www.nia.nih.gov/health/getting-help-alzheimers-caregiving. Accessed March 2022.

Nielsen, T. R.; Nielsen, D. S.; and Waldemar, G. 2021. Barriers in access to dementia care in minority ethnic groups in Denmark: a qualitative study. *Aging & Mental Health* 25(8): 1424–1432.

Pagán-Ortiz, M. E.; Cortés, D. E.; et al. 2014. Use of an online community to provide support to caregivers of people with dementia. *Journal of gerontological social work* 57(6-7): 694–709.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71(2001): 2001.

Polzer, E.; Nearing, K.; et al. 2021. Firearm access in dementia: legal and logistic challenges for caregivers. *International review of psychiatry* 1–9.

Proferes, N.; Jones, N.; Gilbert, S.; Fiesler, C.; and Zimmer, M. 2021. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media+ Society* 7(2): 20563051211019004.

Rajadesingan, A.; Budak, C.; and Resnick, P. 2021. Political Discussion is Abundant in Non-political Subreddits (and Less Toxic) 15: 525–536.

Reber, U. 2019. Overcoming language barriers: Assessing the potential of machine translation and topic modeling for the comparative analysis of multilingual text corpora. *Communication methods and measures* 13(2): 102–125.

Riedel, O.; Klotsche, J.; et al. 2016. Overlooking informal dementia caregivers' burden. *Research in Gerontological Nursing* 9(4): 167–74.

Roberts, M. E.; Stewart, B. M.; and Tingley, D. 2019. stm: An R Package for Structural Topic Models. *Journal of Statistical Software, Articles* 91(2): 1–40.

Rose, K. M.; Coop Gordon, K.; et al. 2021. Smarthealth technology study protocol to improve relationships between older adults with dementia and family caregivers. *Journal of Advanced Nursing* 77(5): 2519–2529.

Rustam, F.; Khalid, M.; et al. 2021. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *Plos one* 16(2): e0245909.

Sajwani, Z. 2020. Challenges of caregivers of patients with Alzheimer's disease. *Innovation in Aging* 4(Suppl 1): 149–149.

Schatto-Eckrodt, T.; Janzik, R.; et al. 2020. A computational approach to analyzing the Twitter debate on gaming disorder. *Media and Communication* 8(3): 205.

Schulz, R.; Eden, J.; et al. 2016. Family caregiving roles and impacts. *Families caring for an aging America* 73 – 113.

Sikder, A. T.; Yang, F. C.; et al. 2019. Mentalizing imagery therapy mobile app to enhance the mood of family dementia caregivers: feasibility and limited efficacy testing. *JMIR aging* 2(1): e12850.

Song, L.; and Chang, T.-Y. 2012. Do Resources of Network Members Help in Help Seeking? Social Capital and Health Information Search. *Social Networks* 34: 658–69.

Song, L.; Son, J.; and Lin, N. 2011. *Social Support*, 116–128. *The Sage Handbook of Social Network Analysis*. London: SAGE.

Song, L.; and Zhang, Z. Forthcoming. *Social Support*. *The Sage Handbook of Social Network Analysis*. London: SAGE.

Stone, R. I. 2015. Factors affecting the future of family caregiving in the U.S. In *Family caregiving in the new normal*, 57–77.

Van Den Kieboom, R. C.; Snaphaan, L. J.; Mark, R. E.; and Bongers, I. 2020. The trajectory of Caregiver burden and risk factors in dementia progression: a systematic review. *Journal of Alzheimer's Disease* (Preprint): 1–9.

Vaswani, A.; Shazeer, N.; et al. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008.

Wang, Z.; Zou, N.; et al. 2021. Characterizing dementia caregivers' information exchange on social media: exploring an expert-machine co-development process. In *Diversity, Divergence, Dialogue: 16th International Conference, iConference 2021*, 47–67.

Wennberg, A.; Dye, C.; Streetman-Loy, B.; and Pham, H. 2015. Alzheimer's patient familial caregivers: A review of burden and interventions. *Health & Social Work* 40(4): e162–e169.

Wilkerson, D. A.; Brady, E.; et al. 2018. Friendsourcing peer support for Alzheimer's caregivers using Facebook social media. *Journal of Technology in Human Services* 36(2-3): 105–124.

Williamson, H. J.; McCarthy, M. J.; et al. 2020. Addressing the needs of rural caregivers of individuals with Alzheimer's disease and related dementias during and beyond coronavirus disease 2019 (COVID-19). *Public Policy & Aging Report* 30(4): 178–180.

Yin, Z.; Sulieman, L. M.; and Malin, B. A. 2019. A systematic literature review of machine learning in online personal health data. *Journal of the American Medical Informatics Association* 26(6): 561–576.

Zechner, M.; Lundquist, M. C.; et al. 2020. Care2Caregivers: A retrospective examination of needs addressed and services pro-vided by a peer support helpline. *Journal of Psychosocial Nursing and Mental Health Services* 58(10): 13–16.

Zhou, X.; Wan, X.; and Xiao, J. 2016. Attention-based LSTM network for cross-lingual sentiment classification 247–256.