

# A Machine Learning Approach to Evaluating Illness-Induced Religious Struggle

Joshua Glauser<sup>1,2</sup>, Brian Connolly<sup>1</sup>, Paul Nash<sup>3</sup> and Daniel H Grossoehme<sup>2</sup>

<sup>1</sup>Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>2</sup>Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>3</sup>Chaplaincy Department, Birmingham Children's Hospital, Birmingham, UK.

Biomedical Informatics Insights  
1–9  
© The Author(s) 2017  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1178222616686067



**ABSTRACT:** Religious or spiritual struggles are clinically important to health care chaplains because they are related to poorer health outcomes, involving both mental and physical health problems. Identifying persons experiencing religious struggle poses a challenge for chaplains. One potentially underappreciated means of triaging chaplaincy effort are prayers written in chapel notebooks. We show that religious struggle can be identified in these notebooks through instances of negative religious coping, such as feeling anger or abandonment toward God. We built a data set of entries in chapel notebooks and classified them as showing religious struggle, or not. We show that natural language processing techniques can be used to automatically classify the entries with respect to whether or not they reflect religious struggle with as much accuracy as humans. The work has potential applications to triaging chapel notebook entries for further attention from pastoral care staff.

**KEYWORDS:** Prayer, chaplaincy, natural language processing, machine learning, spiritual struggle

**RECEIVED:** August 5, 2016. **ACCEPTED:** October 17, 2016.

**PEER REVIEW:** Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 1548 words, excluding any confidential comments to the academic editor.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Partial funding for this study was provided by the Summer Undergraduate Research Fellowship Program of the University of

Cincinnati and by the Divisions of Pulmonary Medicine and Biomedical Informatics at Cincinnati Children's Hospital Medical Center.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Daniel H Grossoehme, Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA. Email: daniel.grossoehme@cchmc.org

## Introduction

Spiritual and religious beliefs and practices often play a positive role in people's lives. Although some religious beliefs and practices can confer health benefits,<sup>1</sup> this is not always the case, and some ways of using faith to cope have been associated with poorer health outcomes. Spiritual struggle describes the conflicts, questions, or tensions that arise around religious or spiritual issues.<sup>2</sup> Such struggles fall into 3 main categories: intrapersonal, interpersonal, and divine struggles.<sup>3</sup> Intrapersonal religious struggles are those in which the tension is internal, such as questioning of formerly helpful beliefs that are no longer adequate for the stressor or struggle due to guilt. Interpersonal religious struggles are those in which the tension exists between the individual and one or more other important people, such as a feeling of abandonment by members of one's congregation. Divine struggles are those in which one's relationship with the Divine is impaired, whether by anger, feelings of abandonment, or questions. These distinctions, although useful for selecting an intervention approach, are not mutually exclusive in actual experience, and a person may experience one or all forms of religious struggle simultaneously. Spiritual struggles are frequently operationalized using the negative spiritual coping scale of the Brief RCOPE.<sup>4</sup>

Religious or spiritual struggles are clinically important to health care chaplains because they are related to poorer health outcomes, involving both mental<sup>5</sup> and physical<sup>6</sup> health problems. For example, spiritual struggle has been associated with depression (4% variance), anxiety, and negative illness adjustment in

155 patients newly diagnosed with breast cancer.<sup>7</sup> In a similar longitudinal study involving Orthodox Jews, spiritual struggle preceded and was a potential cause of future depression.<sup>8</sup> Finally, a study of 577 patients aged 55 years or older at Duke University Medical Center demonstrated that negative religious coping (reappraisals of demonic forces and punishing, spiritual discontent, and negative attitudes toward God and clergy) generated depressive symptoms and a poorer quality of life.<sup>9</sup>

Struggles occur across religious traditions and have been described in samples composed of Jews,<sup>10</sup> Muslims,<sup>11</sup> and Hindus.<sup>12</sup> Spiritual struggle has been referred to as a catalyst that can either lead to further growth or become chronic.<sup>13</sup> Chaplaincy care is a potential catalyst to promote resolution of religious struggles in ways that lead to posttraumatic growth, as this growth is often attributed to religious coping. For example, surveys conducted with trauma survivors found that spiritual or religious coping played an important role in long-term recovery.<sup>14</sup> In addition, spirituality was shown to have a positive relationship with posttraumatic growth in bereaved human immunodeficiency virus (HIV)/AIDS caregivers.<sup>15</sup>

Religious struggle can be identified through instances of negative religious coping, such as feeling anger or abandonment toward God. Measures of the prevalence of religious struggle vary considerably between studies, depending on the population sample and the way religious struggle was operationalized. A study of more than 5000 college students demonstrated that 25% reported religious struggles.<sup>16</sup> In a study of 48



adolescents with sickle cell disease (SCD) and some of their parents (N = 42), the adolescents more frequently used negative religious coping than did the parents, with reported rates of spiritual struggle ranging from 20% to 36% among adolescents with SCD and 4% to 12% among parents of adolescents with SCD.<sup>17</sup> More than half of women with breast cancer (53%) were found to have religious struggle in one hospital in the United Kingdom, which is particularly notable given the more secular environment in Europe compared with the United States.<sup>18</sup> Grosseohme and colleagues<sup>19</sup> found that 31% of parents of children (N = 22) with cystic fibrosis in one hospital setting had evidence of religious struggle. With religious struggle very relevant among patients with serious life-altering diseases, negative religious coping can affect everyday life, from treatment adherence to general mental health, creating the need for an effective way to identify this problem before it affects the patient in a long-term fashion.

Despite the prevalence of religious struggle and the emotional disease that is associated with it, identifying persons experiencing religious struggle poses a challenge for chaplains. In one study, only 31% of those with high spiritual needs and low spiritual resources requested any form of spiritual care.<sup>20</sup> Various methods have been proposed to identify persons who could potentially benefit from chaplaincy care. These include questionnaires, such as the Brief RCOPE<sup>4</sup> and others, as well as computer-delivered assessment modules, many of which have been discussed elsewhere.<sup>6</sup> However, each of these current methods contains drawbacks. The Brief RCOPE is not practical in the clinical setting as the questionnaire is lengthy and requires high participant burden. However, the assessment modules may lack construct validity and represent measures for depression rather than spiritual distress. In addition to these methods, chaplaincy has historically been an itinerant, reactive service that is driven by a combination of self-referrals and referrals from others.<sup>21–24</sup> Thus, a feasible, efficient, and reliable means of identifying persons with potential religious struggle is important and would represent a significant shift in clinical practice for many health care chaplains.

One potentially underappreciated way of triaging chaplaincy effort is through prayers written in chapel notebooks. The role that written prayers play in coping in hospitals and other health care institutions has been described in several contexts, in both the United States and the United Kingdom.<sup>25–27</sup> Prayer books in chapels have been compared with psalms of lament,<sup>25</sup> and it has been suggested that at least some people write prayers as a means of seeking not only divine aid but also the support of other people.<sup>25,26</sup> A framework for conceptualizing such prayers has been used by ap Siôn and Nash,<sup>27</sup> framing them in terms of their reference, intention, and objective and how those constructs relate to health and communication. These prayers may also help inform chaplains about who is using the prayer books as a resource to cope and who may be expressing religious struggle. As a result, the written prayers in

chapels provide a different religious perspective for the patient and the parents in their relationships with God.

An analogous situation of using written texts as a means of analysis is the classification of genuine and elicited suicide notes by mental health experts using natural language processing (NLP).<sup>28,29</sup> Natural language processing has been used to create constant, reliable means of prediction of future needs or behaviors based on written texts. Natural language processing is a field that stands at the crossroads of computer science and linguistics, where machine learning computer algorithms are used to decipher corpus of texts across different disciplines. Machine learning is a technique in which a computer learns from a data set to classify future data, and these techniques have been used for NLP in a variety of tasks involving the classification of free-text, including identifying differences in retrospective suicide notes, newsgroups, and social media<sup>30–32</sup>. Although there are many approaches to this type of classification, this study seeks to identify the optimal approach to classify the severity of religious struggle. Consequently, identifying and automating this approach provide decision support to chaplains for triaging. We follow the innovation process of design, prototype, pilot, and implement described by Provost and Hoppenjans.<sup>33</sup> To test the design and prototype phases, the hypothesis of this study was that machine learning algorithms can categorize written prayer texts of patients and their parents found in a pediatric academic medical center's chapel notebooks for expressing religious struggle or no religious struggle using NLP.

## Methods

### Data

This study was approved by the Cincinnati Children's Hospital Medical Center Institutional Review Board (2014-3384). Prayers were collected from chapels at 2 pediatric medical centers. Cincinnati Children's Hospital Medical Center (site 1) was a 575-bed academic pediatric medical center in the US Midwest. There were 3 chapels at this site, each with its own prayer notebook. Prayers were collected weekly from those notebooks. At the top of each page read the following statement: "You are welcome to write a prayer in this book. Your prayers may be read by others. For the confidentiality of our children and families, please use initials rather than names when referring to a patient." Birmingham Children's Hospital (site 2) was a 364-bed pediatric hospital in the West Midlands of England in the United Kingdom. There were 3 places of worship and reflection on this site (Christian chapel, Muslim prayer room, and multifaith quiet room), and all were open continuously. The prayers used in this study were taken solely from the Christian chapel. Although the chapel is clearly a Christian place of worship and prayer, patients and families of all faiths and beliefs may have used the chapel for quiet prayer and reflection. This prayer book was perpetually open and

replaced when it was full. It is explained at the beginning of each book that “These prayers will be prayed for at our weekly Holy Communion service and will also be reflected upon to inform and improve our support to patients, families and staff.”

### Procedure

Following the process by Pestian et al,<sup>28</sup> an ontology for this study was developed by a quasi-Delphi consensus-building process.<sup>31</sup> A panel of 8 pediatric chaplains from site 1, all of whom were board-certified, generated a list of expressions of spiritual struggle derived from clinical experience and items from the negative religious subscale of the Brief RCOPE.<sup>4</sup> Items were continuously reevaluated for expressing spiritual struggle, and a consensus list was developed of expressions of spiritual struggle. The final list became the ontology for use by the annotators in this study (Table 1).

First, each prayer was transcribed and anonymized. The transcription process involved copying written prayers into a separate text (.txt) document as a virtual catalog with all misspellings and grammar errors kept *verbatim*. For anonymization, patient names were replaced with “[N]” (other protected health information [PHI] was not encountered during transcription). In addition, a new line in the prayer would be represented as “NEWLINE,” any writings that could not be understood were replaced with “[unintelligible],” and any drawings were replaced with a double-bracketed description of the drawing (eg, a picture of a cross would become “[[cross]]”).

The next step was to annotate a set of prayers. Seven annotators were recruited for the study. Three annotators were assigned to annotate all of the US prayers, and the other study staff annotated subsets of the prayers within study time frame. The annotators consisted of a male board-certified chaplain, a male layperson, and 3 female board-certified chaplains from the United States, a male professional chaplain from England, and a female layperson from the United Kingdom doing graduate work on prayers. The male layperson and 2 female board-certified chaplains (all from United States) annotated the US prayers. The annotation process called for each annotator to individually read each prayer in the block and determine whether language indicative of spiritual struggle was present using the ontology. If at least one annotator considered the prayer to contain religious struggle, then the prayer was tagged as having “struggle.” Otherwise, a prayer was considered “not struggle.” The reason for this is because in the clinical setting, if one chaplain detects religious struggle in a patient, then that patient is considered in need of chaplain assistance. This label of “struggle” or “not struggle” would then be attributed to the prayers. This combination of a prayer with its corresponding label would become the training set for which machine learning techniques would be applied.

The goal of machine learning is to train a computational model from training data that can then generalize to unseen

(test) data. Machine learning can be roughly divided into 3 types: supervised learning, when the training data are already labeled; semi-supervised learning, when only part of the training set is labeled; and unsupervised learning, when the challenge is to learn patterns in unlabeled data. Here, a supervised learning support vector machine (SVM) approach was used<sup>35</sup>.

### Machine learning classifier

The task of building a machine learning model to classify text can be separated into 5 parts: (1) transcribing the text, (2) preprocessing of the transcribed text, (3) extraction of features (unique word or phrase found in any of the prayers), (4) selection of those features that most differentiate the text between 2 separate classes, and (5) optimization of the classifier. Before the features were extracted, the prayers were preprocessed such that punctuation and line spaces (paragraphs) were identified and removed from the final set for the classifier. Individual words (unigrams), word pairs (bigrams), 3-word phrases (trigrams), and 4-word phrases (quadgrams) were then extracted from each prayer and used as the backbone for the classifier. In addition, the number of words in each prayer was also considered in analysis.

Although data are frequently split into training, validation, and test sets, the sample size here was limited such that any performance measure from a “set aside” test would have unacceptable statistical uncertainty. The approach described by Kohavi and John<sup>36</sup> was therefore used, in which both the classification and feature selection were contained in the training set, allowing the validation set to act as a proxy for the test set. The prayers contained more than 5000 unique features. Most of those provided little or no discrimination between prayers with or without religious struggle. Most of the features were either common words, such as “the” and “a” which appeared with equal frequency in both types of prayers, or words or phrases that appeared only in 2 or fewer prayers. Therefore, to reduce “noise” in the classifier, only those features that provided the best discrimination (ie, those words that appeared with the most different frequencies in struggle and not-struggle prayers) were selected for input. This was important because the “noise,” or unnecessary words that did not distinguish whether prayers contained religious struggle or not, would cause the classifier to give a false result. Consequently, the frequency of type I errors and type II errors would increase. To reduce the “noise,” methods were fine-tuned in 2 different manners.

The first method was changing the type of feature selection test itself. One way the differences in the frequency distributions were quantified was with the Kolmogorov-Smirnov test (KS-test) *P*-value.<sup>33</sup> The KS-test was performed by first determining the largest difference in the cumulative distributions of 2 samples: struggle and not struggle. The KS-test *P*-value was then evaluated, which is the probability of obtaining a difference larger than the one observed. The *P*-values from all of

**Table 1.** Chaplain-generated ontology of expressions of religious struggle.

DOMAIN	ITEM
<b>Wonders where God is</b>	"Are you there?"
	"Are you listening?"
	"Why have you forsaken/left/abandoned me?"
	"Are you real?"
	"Are you there?"
<b>Expresses the feeling punished</b>	"Are you up there?"
	Uses the word "punish" or "punishment"
	"What have I done?"
	"I terminated a pregnancy . . ."
<b>Questions God's love</b>	"I gave up a child for adoption . . ."
	"I had an affair . . ."
	"If you love me/her/him . . ."
<b>Attributes the problem to the devil or Satan</b>	"Deliver her from . . ."
<b>Questions why this is happening</b>	"Cast out . . ."
	"Why?"
	"Why are you doing this?"
	"Why my child?"
<b>Doubts the adequacy of their faith</b>	"Why so ill?"
	"God doesn't give us more than we can handle, but . . ."
	"You don't lay on us more than we can bear, but . . ."
<b>Blames God for child's illness or other problem</b>	". . . you did this . . ."
	". . . you sent this . . ."
	". . . you laid this on/upon him/her . . ."
	". . . you gave him/her this . . ."
<b>Expresses guilt</b>	"I am sorry for . . ."
	"please forgive me for . . ."
	"I don't know what I am doing wrong . . ."
<b>Expresses anger toward God for letting this problem happen</b>	"How could you . . ."
<b>Asks God for child/person to remain alive</b>	"Don't take . . ."
	"Let him/her stay here . . ."
	"Let him/her live . . ."
	". . . help him/her to live"
	"I'm not ready to let him/her go . . ."
	"If she go I go . . ."
	"Let her get to see more days"
	"Please save my son, he's a good baby, I love him very much"
	"I wish you would let me stay around for a much longer time . . ."
	"We almost lost him as he feels no one in this world cares . . ."
<b>Expresses suicidal ideation</b>	"Please save my life before I end it"
	"God I feel like dying . . ."
<b>Expresses feeling spiritually exhausted</b>	"Can't take this no more"
	"never knew it could hurt this bad"
	"My faith is fading . . ."
	"Get through this most challenging and difficult time"
<b>Expresses fear</b>	"I'm awful scared . . ."

**Table 2.** AROC for different classifiers in determining whether written prayers contain religious struggle.

	AROC	OPTIMAL NUMBER OF FEATURES	TOP FEATURES
KS-test (wrapper)	0.76±0.04	128	“a,” “give,” “have,” “I,” “to,” “. I,” “her,” “is,” “please,” “me,” “my,” Number of Words in the Prayer
ANOVA (wrapper)	0.74±0.04	64	“me,” “lord please,” “i have,” “is,” “please heal,” “can,” “but,” “. i,” “forgive,” “give,” “in my,” “through this”
KS-test (IF)	0.73±0.04	12	“a,” “give,” “have,” “I,” “to,” “. I,” “her,” “is,” “please,” “me,” “my,” Number of Words in the Prayer
ANOVA (IF)	0.74±0.12	5425	“me,” “lord please,” “i have,” “is,” “please heal,” “can,” “but,” “. i,” “forgive,” “give,” “in my,” “through this”

Abbreviations: AROC, area under a receiving operating characteristic; ANOVA, analysis of variance; KS, Kolmogorov-Smirnov. Techniques include wrapper method and information foraging (IF).

the features were then ranked. The other test that was used was analysis of variance (ANOVA)<sup>38</sup>. The ANOVA test was performed in a similar manner to the KS-test. These tests were used because of their frequent occurrences in studies regarding machine learning and feature selection. For example, the KS-test has been used as a common method for classification algorithms<sup>39, 40</sup> and ANOVA has been applied to email spam classification<sup>41</sup>, as well as other machine learning classifications.

The other method to reduce noise was choosing the number of top features to include in the classification. The first way was by manually choosing the top number of features in logarithmic steps of 2 (2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024), called the wrapper method (Saey, 2007). The other way was by using an information foraging algorithm to decide how many of the top features should be used to optimize the classifier<sup>42</sup>. Both methods were used in different combinations in the SVM to decide which one produced the best results.

An SVM, a type of machine learning classifier, was used to determine whether prayers contained religious struggle or not. SVMs are based on a computational learning theory called structural risk minimization, whose goal is to find a hypothesis with the lowest true error<sup>43</sup>. The SVM constructs a hyperplane in a high-dimensional space, which can be used for classification, regression, or other tasks<sup>44</sup>. The SVM is appropriate for this study’s data because its connection to computational learning enables it to be a universal learner<sup>45</sup>. Consequently, it tends to be fairly robust to overfitting<sup>46</sup>. The performance of the classifier was based on the area under the receiver operating characteristic, which was estimated using leave-one-out by subject cross-validation<sup>47, 48</sup>.

The area under a receiving operating characteristic (AROC)<sup>49</sup> measures the probability that given a randomly selected prayer with religious struggle and a second randomly selected prayer without religious struggle, the SVM will give a higher probability that the prayer contains religious struggle compared with a prayer actually expressing religious struggle. The performance of the classifier is optimized using the AROC for 2 reasons. First, the AROC provides a single statistic that quantifies the spectrum of possible sensitivities given

the desired specificities for the classifier. Second, the AROC provides a measure of how accurately prayers are classified and thereby directly quantifies the increased efficiency of finding a struggling patient or family.

### Results

The main analysis was conducted on 243 American prayers, which were only annotated by American members in the study (2 board-certified chaplains and 1 layperson). From these prayers, 42 contained signs of religious struggle, whereas 201 did not indicate religious struggle. Overall, interrater reliability was .41 using the Krippendorff alpha scale, which corresponds to moderate agreement. In addition, the annotators produced an overall agreement of 83.3% (SE, 5.8%). The most successful classifier had an AROC of 0.73 (±4%) and used only 12 features with information foraging. Different methods were used to tune the classifier, and the entire results can be found in Table 2. The list of top features is detailed in Table 3. Figure 1 depicts the full receiving operating characteristic (ROC) curve.

A second analysis was performed, in which the total 528 prayers were annotated, 245 of which were obtained from site 1 and 283 of which were obtained from site 2. For this analysis, a combination of 1 to 4 annotators (both from the United States and United Kingdom) annotated each prayer. Overall, interrater reliability was .38 using the Krippendorff alpha scale, which corresponds to moderate agreement. In addition, the annotators achieved an overall agreement of 74.1% (SE, 2.8%). The purpose of this analysis was to test whether there was a significant difference in language that would require different classifiers for different cultures. The classifier was able to classify prayers containing religious struggle with an AROC score of 80% (±3%) using 256 features. Figure 2 depicts the full ROC curve.

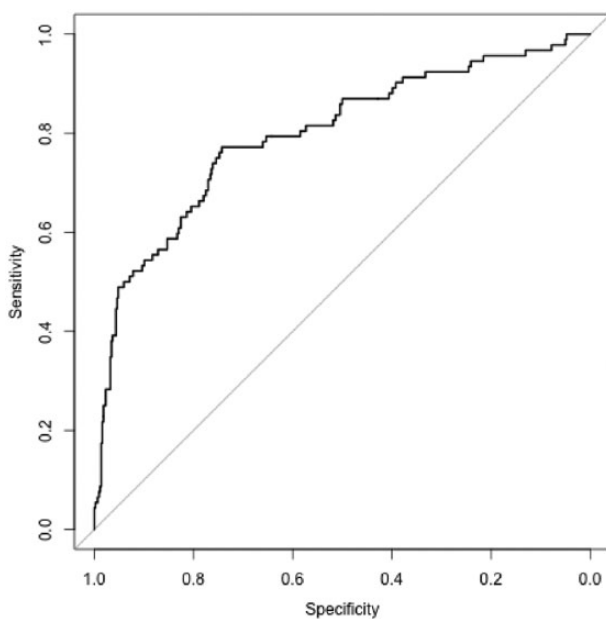
### Discussion

We present an exploratory attempt to use NLP to identify the spiritual struggle written in the prayer books of a pediatric academic medical center. The classifier’s performance (high AROC) supports our conclusion that an NLP classifier is a

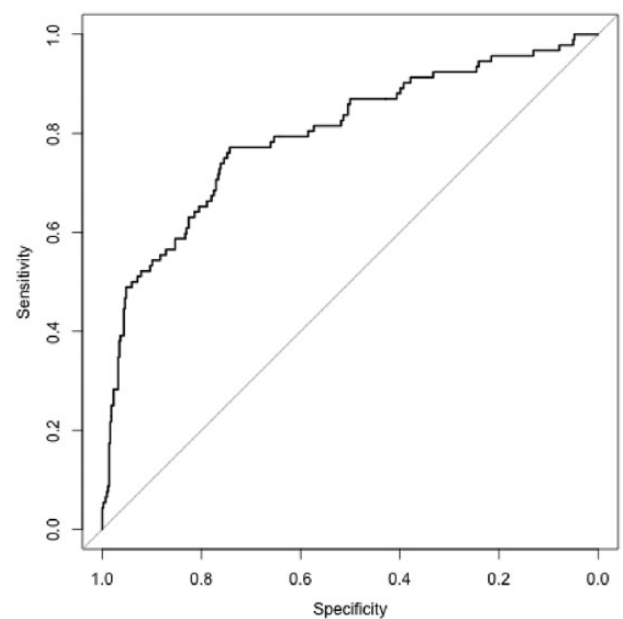
**Table 3.** Probabilities and *P*-values of top features occurring in a prayer of both classes.

FEATURE	“STRUGGLE”	“STRUGGLE” ERROR	“NONE”	“NONE” ERROR	<i>P</i> -VALUES
a	0.310	0.098	0.134	0.028	<.0001
give	0.333	0.103	0.119	0.026	<.0001
have	0.286	0.094	0.080	0.021	<.0001
i	0.524	0.138	0.313	0.045	<.0001
To	0.714	0.171	0.463	0.058	<.0001
. i	0.262	0.089	0.065	0.019	<.0001
her	0.357	0.107	0.169	0.031	<.0001
is	0.381	0.112	0.100	0.023	<.0001
please	0.524	0.138	0.259	0.040	<.0001
me	0.452	0.125	0.114	0.025	<.0001
my	0.500	0.134	0.224	0.037	<.0001

*P*-values based on the Welch *t*-test comparing prayers with struggle ( $n=42$ ) and prayers without struggle ( $n=201$ ).



**Figure 1.** Receiver operating characteristic for the classifier's discrimination of religious struggle within American prayers. The gray line is the area under a receiving operating characteristic curve for a baseline (random) classifier.



**Figure 2.** Receiver operating characteristic for the classifier's discrimination of religious struggle within British and American prayers. The gray line is the area under a receiving operating characteristic curve for a baseline (random) classifier.

useful objective tool that clinicians and others can use to determine religious struggle in parents and patients through written prayers. The classifier can identify religious struggle by examining patterns of words and phrases within the corpus of texts. In addition, the classification algorithm used can be applied across data sets with different prayers and cultures because the classifier with just American prayers and annotators performed comparably as the classifier with both British and American

prayers and annotators. Consequently, the classifier can provide additional decision support to chaplains in identifying persons with religious struggle.

The lack of significant differences in recognizing religious struggle by chaplains in the United States and the United Kingdom and by board-certified chaplains and nonchaplains supports the ontology and classifier's possibility for clinical utility. Prayers across 2 English-speaking cultures as having

expressions of religious struggle could be identified. This suggests that although there may be geographically unique expressions of religious struggle, there are also some expressions that are sufficiently common as to make the use of the classifier possible across regions. These results also suggest that it is feasible to train persons who are not board-certified chaplains to recognize expressions of religious struggle. It may therefore be possible to have prayers in chapel notebooks read in a cost-efficient manner by someone trained to use the ontology to recognize expressions of religious struggle and who is not a board-certified chaplain. Until the classifier has undergone further development and testing, this could be a viable way to triage chaplains' clinical efforts.

The use of the female pronoun "her" more frequently in prayers with religious struggle was statistically significant, but may not be clinically meaningful. To determine whether the context in which gendered pronouns occurred might provide a different perspective, a further review of 100 randomly selected prayers was undertaken by the senior author (DHG). Prayers containing "his" used the pronoun primarily in the context of God being asked to "look over" or "watch over." Prayers in which the pronoun "her" occurred were related to a girl or woman going through a surgery or procedure. This could reflect language of belief reflecting personal experience, an "ordinary theology,"<sup>50</sup> that is more troubled when a girl or woman is in need. It could also reflect a finding that is statistically significant and yet not clinically meaningful.

Another interesting aspect of the top features is the inclusion of "I," "me," and "my." These personal pronouns indicate a focus on the self rather than their community. A study focusing on personal pronouns in social communication indicated that too much attention to one's self is attributed to negative emotional states such as depression<sup>51</sup>. Because these self personal pronouns were demonstrated to be some of the most distinguishing features in the classifier, they might indicate the negative emotional state contained within prayers that have religious struggle, suggesting a need for further analysis.

The study has limitations that we acknowledge and accept for an exploratory study. The classifier was built around 528 prayers and 7 annotators from 2 different contexts. The restraints of time and the availability of the annotators meant that only 1 annotator annotated 100 British prayers in the supplementary analysis. As a result, more false-negatives and false-positives could have occurred than intended. However, with relatively stable interannotator agreement, much deviation was not expected and the chance of a type I error or type II error was low. Another potential limitation arises from the concern over the privacy of health information. This has led some institutions to discourage prayer writers from using surnames or patient room numbers, which would be helpful or necessary to correctly identify a prayer writer for follow-up. This concern need not be irremediable: Instructions at the top of each notebook page or on a bookstand could indicate that prayers are

being reviewed and that a member of a chaplaincy department may follow up. This is little different in principle for similar notations that prayers may be read aloud during public worship in the chapel space. The lack of need of oversight of this or similar studies stems from the very public nature of the prayer books, the voluntary surrender of some privacy that is made in choosing to write a prayer, and the degree of self-disclosure writers choose to make. One study has even noted that some prayer writers appear to write what is in effect a plea for others to read and lift up their person in prayer.<sup>26</sup> The anonymity of the prayer writers also precluded linking prayers with the patient's health status. However, it is the religious struggle of the prayer writer that is of primary clinical interest to the chaplain, which may or may not be directly related to the health status or religious struggle of the patient (or the person being prayed about). This is one difference, perhaps more common to pediatric than adult hospitals, in which the apparent proxy person is in fact the chaplain's "patient" as opposed to the identified patient. A limitation of the current classifier is this inability to distinguish between identified patient and the prayer writer/proxy. Another limitation is the inability to directly evaluate the individuals who wrote prayers for religious struggle by a clinically trained chaplain, rather than relying solely on chaplains' evaluation of the written prayer itself and the classifier. Prayers written in a pediatric hospital may differ in significant ways from those written in adult hospitals. For example, prayers relating to a child's dying may be both more prevalent in a pediatric hospital and express different sentiments and contents than those found in adult centers (especially those without a maternity unit). It is also impossible to definitively obtain demographic data about the prayer writers. More women than men attend religious services in the United States,<sup>52</sup> and it may well be that most of the prayer writers were women. Nevertheless, important conclusions can be drawn. The classifier can discriminate between prayers with or without religious struggle. This validates the classifier's design and prototype and its continued development to prepare for pilot testing as a tool to assist clinical chaplains in determining how to ration a limited resource, their time, by indicating prayer writers who are using faith to cope and those who are struggling and unlikely to self-refer for chaplaincy care.

Potential future research directions should focus on preparing for the next (pilot) phase of development. This translational research is needed to address the limitations mentioned above. These include, first, increasing the number of prayers included to strengthen the classifier, as well as give more evidence regarding the interannotator reliability across the cultures of the United States and the United Kingdom. Second is identifying and operationalizing a means of linking a prayer with an identifiable person while maintaining an appropriate level of privacy. For example, a header statement in such prayer books could invite the use of surnames or room numbers as an indication of a writer's desire for direct, follow-up contact based on

the prayer's content. After overcoming the translational barriers, the algorithm may be implemented by making it available to medical centers or publishable to be used conveniently and with regional data (if our current presented data show drastic contrast to the region). The issue of religious struggle occurring more frequently when a woman is the subject of prayer in a hospital notebook bears further exploration. Another intriguing use for the classifier and this methodology lies outside the medical scope: The current state of foreign affairs for the United States increasingly deals with potential persons who might become radicalized. The machine learning algorithm of religious struggle might be useful to identify potential persons at risk of radicalization at a faster or even more reliable rate, using social media outlets (eg, Facebook and Twitter).

### Acknowledgements

The authors acknowledge all the members of the Pestian Lab, especially Lesley Rohlf, Robert Faist, and Rachel McCourt, for their support and ongoing commitment. They also gratefully acknowledge the assistance of the following persons: Gail Pyne-Geithman, Diana Turov, Sofia Newaz, Sophia M Dimitriou, Joy Cheng, Sarah Lohbeck, McKenzie Roedig, Natasha Yanes, Alexis Teeters, Bill Scrivener, MaryAnn Hegner, Karen Behm, Lou Langford, Maggi Hunt, and Pawel Matykiewicz.

### Author Contributions

DHG conceived and designed the experiments. DHG and PN collected study data. JG, DHG, and BG analyzed the data. JG and PN provided annotations for data. JG wrote the first draft of the manuscript. JG, DHG, and BC contributed to the writing of the manuscript. PN reviewed the manuscript for intellectual content. All authors reviewed and approved the final manuscript.

### Disclosures and Ethics

As a requirement of publication, author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including, but not limited to, the following: authorship and contributorship, conflicts of interest, privacy and confidentiality, and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

### REFERENCES

- Oman D, Thoresen CE. Do spirituality and religion influence health? In: Paloutzian RF, Park CL, eds. *Handbook of the Psychology of Religion and Spirituality*. New York, NY: Guilford Press; 2005:435–459.
- Oemig-Dworsky CK. *Responding to Spiritual Struggles: Experiential Avoidance and Mindfulness in Adjustment*. Bowling Green, OH: Department of Psychology, Bowling Green State University; 2014.
- Oemig-Dworsky CK, Pargament KI, Reist Gibbel M, et al. Winding road: preliminary support for a spiritually integrated intervention addressing college students' spiritual struggles. In: Piedmont RL and Village A, eds. *Research in the Social Scientific Study of Religion*. Vol 24. Leiden, The Netherlands: Koninklijke Brill NV; 2013:309–339.
- Pargament KI, Koenig HG, Perez LM. The many methods of religious coping: development and initial validation of the RCOPE. *J Clin Psychol*. 2000;56:519–543.
- Smith TB, McCullough ME, Poll J. Religiousness and depression: evidence for a main effect and the moderating influence of stressful life events. *Psychol Bull*. 2003;129:614–636.
- Fitchett G. Selected resources for screening for spiritual risk. *Chaplain Today*. 1999;15:13–26.
- Thune-Bolye IC, Stygall J, Keshtgar MRS, Davidson TI, Newman SP. Religious/spiritual coping resources and their relationship with adjustment in patients newly diagnosed with breast cancer in the UK. *Psychooncology*. 2013;22:646–658.
- Pirutinsky S, Rosmarin DH, Pargament KI, Midlarsky E. Does negative religious coping accompany, precede, or follow depression among Orthodox Jews? *J Affect Disord*. 2011;132:401–405.
- Koenig HG, Pargament KI, Nielsen J. Religious coping and health status in medically ill hospitalized older adults. *J Nerv Ment Dis*. 1998;186:513–521.
- Rosmarin DH, Pargament KI, Flannelly KJ. Do spiritual struggles predict poorer physical/mental health among Jews? *Int J Psychol Religion*. 2009;19:244–258.
- Abu Raiya H, Pargament KI, Mahoney A, Stein C. A psychological measure of Islamic religiousness: development and evidence for reliability and validity. *Int J Psychol Religion*. 2008;18:291–315.
- Tarakeshwar N, Pargament KI, Mahoney A. Measures of Hindu pathways: development and preliminary evidence of reliability and validity. *Cultur Divers Ethnic Minor Psychol*. 2003;9:316–332.
- Pargament KI. Religious struggles: a fork in the road to growth or decline. Paper presented at: Spirituality, Health, and Human Flourishing: Meaning? Measurement? Implications?; 26 June 2008; Durham, NC.
- Wiechman Askay S, Magyar-Russell G. Post-traumatic growth and spirituality in burn recovery. *Int Rev Psychiatry (Abingdon, England)*. 2009;21:570–579.
- Cadell S, Regehr C, Hemsworth D. Factors contributing to posttraumatic growth: a proposed structural equation model. *Am J Orthopsychiatry*. 2003;73:279–287.
- Johnson CV, Hayes JA. Troubled spirits: prevalence and predictors of religious and spiritual concerns among university students and counseling center clients. *J Counseling Psychol*. 2003;50:409–419.
- Grossoehme DH, Cotton S. Religious/spiritual coping by parents of adolescents with sickle cell disease: a Pilot Study. Paper presented at: 1st Annual Meeting, Society for Spirituality, Theology and Health; June 26, 2008; Durham, NC.
- Thune-Bolye IC, Stygall J, Keshtgar MRS, Davidson TI, Newman SP. Religious coping strategies in patients diagnosed with breast cancer in the UK. *Psychooncology*. 2011;20:771–782.
- Grossoehme DH, Fitchett G. Testing the validity of a protocol to screen for spiritual struggle among parents of children with cystic fibrosis. *Res Soc Sci Study Relig*. 2013;24:281–308.
- Fitchett G, Meyer PM, Burton LA. Spiritual care in the hospital: who requests it? Who needs it? *J Pastoral Care*. 2000;54:173–186.
- Galek K, Flannelly KJ, Koenig HG, Fogg SL. Referrals to chaplains: the role of religion and spirituality in healthcare settings. *Ment Health Relig Cult*. 2007;10:363–377.
- Galek K, Vanderwerker LC, Flannelly KJ, et al. Topography of referrals to chaplains in the Metropolitan Chaplaincy Study. *J Pastoral Care Counsel*. 2009;63:1–13.
- Vanderwerker LC, Flannelly KJ, Galek K, et al. What do chaplains really do? III. Referrals in the New York Chaplaincy Study. *J Health Care Chaplain*. 2008;14:57–73.
- Chapman TR, Grossoehme DH. Adolescent patient and nurse referrals for pastoral care: a comparison of psychiatric vs. medical-surgical populations. *J Child Adolesc Psychiatr Nurs*. 2002;15:118–123.
- O'Reilly J. A community of strangers: finding solidarity through prayer request journals. *Caregiver J*. 1995;11:26–30.
- Grossoehme DH. Prayer reveals belief: images of God from hospital prayers. *J Pastoral Care*. 1996;50:33–39.
- ap Si6n T, Nash P. Coping through prayer: an empirical study in implicit religion concerning prayers for children in hospital. *Ment Health Relig Cult*. 2013;16:936–952.
- Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide note classification using natural language processing: a content analysis. *Biomed Inform Insights*. 2010;2010:19–28.



29. Pestian JP, Matykiewicz P, Grupp-Phelan J. Using natural language processing to classify suicide notes. Paper presented at: Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing; 19 June 2008 Columbus, OH.
30. Gómez, J. M. "Language Technologies for Suicide Prevention in Social Media." Processing in the 5th Information Systems Research Working Days (JISIC 2014): 21.
31. Huang, Y.-P., T. Goh and C. L. Liew (2007). *Hunting suicide notes in web 2.0—Preliminary findings*. Multimedia Workshops, 2007. ISMW'07.
32. Matykiewicz, P. and J. Pestian (2012). *Effect of small sample size on text categorization with support vector machines*. Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics.
33. Provost SM, Hoppenjans N. Evaluating innovation and improvement in the Collaborative Care Network. Paper presented at: Doing Research at the Front Line of Improving Health Care; April 25-26, 2013; Arlington, VA.
34. Stone Fish L, Busby DM. The Delphi method. In: Sprenkle DH, Moon SM, eds. *Research Methods in Family Therapy*. New York, NY: Guilford Press; 1996:469–484.
35. Schölkopf, B. and A. Smola (1998). "Support vector machines." Encyclopedia of Biostatistics. Hoboken, NJ, Wiley Publishing, ISBN 978-0471975762
36. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Int*. 1997;97:273–324.
37. Kolmogorov A. Sulla determinazione empirica di una legge di distribuzione [On the empirical determination of a distribution law]. *G Inst Ital Attuari*. 1933;4:83–91 (translated in English in Breakthroughs in Statistics, by Kotz and Johnson. Springer-Verlag; 1992).
38. Fisher, L., D. C. Ransom, H. E. Terry and S. Burge (1992). "The California Family Health Project: IV. Family structure/organization and adult health." *Fam Process* 31(4): 399-419.
39. Saeys, Y., I. Inza and P. Larranaga (2007). "A review of feature selection techniques in bioinformatics." *Bioinformatics* 23(19): 2507-2517.
40. Blachnik, M. and W. Duch (2011). "LVQ algorithm with instance weighting for generation of prototype-based rules." *Neural Netw* 24(8): 824-830.
41. Elssied, N. and O. Ibrahim (2014). "K-means clustering scheme for enhanced spam detection." *Research J. Applied Sci. Eng. Tech.* 7(10): 1940-1952.
42. Cohen, K. B., B. Glass, H. M. Greiner, K. Holland-Bouley, S. Standridge, R. Arya, R. Faist, D. Morita, F. Mangano, B. Connolly, T. Glauser and J. Pestian (2016). "Methodological Issues in Predicting Pediatric Epilepsy Surgery Candidates Through Natural Language Processing and Machine Learning." *Biomed Inform Insights* 8
43. Vapnik, V. N. (1999). "An overview of statistical learning theory." *IEEE Trans Neural Netw* 10(5): 988-999.
44. Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (2007). Section 16.5. Support Vector Machines. *Numerical recipes: the art of scientific computing*. New York, NY, Cambridge University Press.
45. Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *European Conference on Machine Learning*, Springer.
46. Sebastiani, F. (2002). "Machine learning in automated text categorization." *ACM Computing Surveys* 34(1): 1-47.
47. Efron, B. and R. Tibshirani (1991). "Statistical data analysis in the computer age." *Science* 253(5018): 390-395.
48. Molinaro, A. M., R. Simon and R. M. Pfeiffer (2005). "Prediction error estimation: a comparison of resampling methods." *Bioinformatics* 21(15): 3301-3307.
49. Hanley, J. A. and B. J. McNeil (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology* 143(1): 29-36.
50. Astley J. *Ordinary Theology*. Burlington, VT: Ashgate Publishing; 2002.
51. Chung, C. K. and J. W. Pennebaker (2008). "Variations in the spacing of expressive writing sessions." *Br J Health Psychol* 13(Pt 1): 15-21.
52. McClendon D. Gender gap in religious service attendance has narrowed in U.S. <http://pewrsr.ch/1TC4cPd>. Published 2016. Accessed October 13, 2016.