

**OPEN ACCESS**

Full open access to this and thousands of other papers at <http://www.la-press.com>.

## Suicide Note Classification Using Natural Language Processing: A Content Analysis

John Pestian<sup>1</sup>, Henry Nasrallah<sup>2</sup>, Pawel Matykiewicz<sup>1</sup>, Aurora Bennett<sup>2</sup> and Antoon Leenaars<sup>3</sup>

<sup>1</sup>Department of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, <sup>2</sup>University of Cincinnati, College of Medicine, Cincinnati, OH 45229, USA. <sup>3</sup>Windsor ON, Canada. Email: <http://pestianlab.cchmc.org>; [john.pestian@cchmc.org](mailto:john.pestian@cchmc.org)

**Abstract:** Suicide is the second leading cause of death among 25–34 year olds and the third leading cause of death among 15–25 year olds in the United States. In the Emergency Department, where suicidal patients often present, estimating the risk of repeated attempts is generally left to clinical judgment. This paper presents our second attempt to determine the role of computational algorithms in understanding a suicidal patient's thoughts, as represented by suicide notes. We focus on developing methods of natural language processing that distinguish between genuine and elicited suicide notes. We hypothesize that machine learning algorithms can categorize suicide notes as well as mental health professionals and psychiatric physician trainees do. The data used are comprised of suicide notes from 33 suicide completers and matched to 33 elicited notes from healthy control group members. Eleven mental health professionals and 31 psychiatric trainees were asked to decide if a note was genuine or elicited. Their decisions were compared to nine different machine-learning algorithms. The results indicate that trainees accurately classified notes 49% of the time, mental health professionals accurately classified notes 63% of the time, and the best machine learning algorithm accurately classified the notes 78% of the time. This is an important step in developing an evidence-based predictor of repeated suicide attempts because it shows that natural language processing can aid in distinguishing between classes of suicidal notes.

**Keywords:** suicide, suicide prediction, suicide notes, machine learning

*Biomedical Informatics Insights* 2010:3 19–28

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



It is estimated that each year 800,000 die by suicide worldwide.<sup>1</sup> In the United States, suicide ranks second as the leading cause of death among 25–34-year olds and the third leading cause of death among 15–25-year olds.<sup>1</sup> The challenge in a clinical setting is to predict the likelihood of a serious repeated attempt.<sup>2</sup> This challenge is exacerbated by the heterogeneity of patients and clinical judgment. Two evidence-based, risk assessment tools that have shown conceptual success, but we know of none has been translated into standard medical practice.<sup>3,4</sup> The long-term goal of our research is to develop and implement an evidence-based tool for measuring the likelihood of repeated suicide attempts.

To gain insight into the suicidal frame of mind, researchers have suggested analyzing national mortality statistics, psychological autopsies, nonfatal suicide attempts and documents such as suicide notes.<sup>5</sup> Early research<sup>6,7</sup> on suicide notes usually used an anecdotal approach incorporating descriptive information.<sup>8</sup> Subsequent methods, based on Frederick's analytical approach have used content, classification, and theoretical-conceptual analysis. Content classification extracts explicit information from a suicide note, e.g. length of the message, words, and parts of speech. On the other hand, classification schemes use data such as age, sex, marital status, educational level, employment status and mental disorder.<sup>9,10,11–13</sup> It has been suggested that simple classification analysis has its limitations,<sup>14</sup> but comparison of note-writers with non-note-writers has consistently found no differences.<sup>15</sup>

Only a very few studies have used Theoretical-Conceptual Analysis<sup>8</sup> despite the assertion in the first formal study of suicide notes that such an analysis has much promise.<sup>5</sup> To address this paucity, Leenaars introduced a method that permits a theoretical analysis of suicide notes, increases the effectiveness of controls, and fosters development of some theoretical insights into problem of suicide.<sup>13,16–18</sup> He developed a cross-cultural model that consists of intrapsychic and interpersonal cluster themes. The intrapsychic cluster includes unbearable psychological pain (UP); cognitive constriction (CC) indirect expressions (IE), e.g. ambivalence, unconscious processes; inability to adjust (IA), or psychopathology interpersonal grouping that include: disturbed interpersonal relations (IR), rejection-aggression (RA); and identification-egression

(IEG) or escape.<sup>19</sup> Subsequent research on suicide notes have supported the utility of such research indicating that both content and psychological processes are critical to prediction.<sup>20,21</sup>

Using computational methods to study suicide notes is not new,<sup>14</sup> but applying advanced algorithms to clinical care of suicidal patients is. Recent computer analysis compared structural characteristics (average sentence length, parts of speech) with content variables (length of communication, instructions, active state, explanation provided, locus of control) in their predictive values.<sup>22</sup> Another approach focused on semantic content of words used in suicide notes by grouping words into linguistic variables (e.g. positive, negative emotions, hearing, references to people, time, religion).<sup>23</sup>

Content, classification and theoretical-conceptual analyzes have discovered many features that can be used to assess suicide risk. Yet few features are consistent among research protocols. Most of them overlap in meaning, and some are contradictory. These earlier studies, however, were unable to take advantage of current machine learning methods for feature extraction. Our preliminary studies show that it is possible to create machine-learning models that mix content and theoretical-conceptual features and classify suicide notes with higher accuracy than mental health professionals.<sup>24</sup> In this study we research if this trend is consistent when psychiatric physician trainees participate.

## Methods

This section describe the study's methods. It has the following components: experimental design, data, feature selection, expert classification, word mending, annotation, and machine learning.

## Experimental design and hypothesis

This is a cross-sectional design to test the hypothesis that machine learning algorithms can classify suicide notes as well as or better than practicing mental health professionals and psychiatry physician trainees. This study is approved by our Institutional Review Board (#2009-0664).

## Data

Completer and elicited notes were transcribed from Clues to Suicide.<sup>5</sup> The transcribed data were then



reviewed for errors and omissions. Sixty-six notes were divided into two groups: 33 completers and 33 elicitors. To create an elicitor note, Shneidman, asked individuals to write a note as if they were going to commit suicide. The groups were matched by gender (male), race (white), religion (Protestant), nationality (United States citizens); ages ranged between ages 25 and 59 years. Anyone suspected of the having a personality disorder or a tendency toward morbid thoughts was asked to write about the happiest day of their life. These notes were then discarded and the individual was not enrolled in the study.

### Expert classification

Each completer note was paired with its elicited counterpart. The paired notes were then randomly ordered and presented to 11 mental health professionals (psychiatrists, emergency room physicians with mental health training, and psychiatric social workers) and 30 psychiatry trainees who had between one and four years of post-MD training and one psychiatric fellow (five years of training) who were asked to classify the notes as either genuine or and elicited notes. There results were compared to the learning models described below.

### Feature selection

Feature selection, also called *variable selection* is a data reduction technique for selecting the most relevant features of a learning models. As irrelevant and redundant features are removed the model's accuracy increases. Multiple methods for feature selection were tested: bag-of-words, latent semantic analysis and heterogeneous selection. Ultimately, heterogeneous selection was used. To reduce co-linearity, highly correlated features were removed. To increase the certainty that a feature was

not randomly selected, that feature had to appear in at least 10% of the documents.<sup>25</sup> Finally, after preparing each bootstrap sample, only 66 features with highest information gain were selected. Information gain can intuitively be interpreted as measuring the reduction of uncertainty.<sup>26,27</sup> Table 1 shows the feature selection and reduction processes. An initial feature space of 1,063 variables was reduced to 66. Thus, the final matrix contains 66 documents and 66 features.

### Parts of speech

Tokenization is the first step in Natural Language Processing (NLP) analysis. It identifies those basic units which need not be decomposed in subsequent analysis and prepares for analysis like word checking, ambiguity checking and disambiguation.<sup>2</sup> This was done using an internally developed Perl program. Next, using the Penn-Treebank tag set. Using The Lingua-EN-Tagger-0.13, 2004 module, 18 part of speech tags were added to the feature space. This tagging is necessary to establish the relationship of a particular word to a particular concept

### Readability

The Flesch and Kincaid readability scores produced a high information gain and were included in the feature space. These scores are designed to indicate comprehension difficulty. They include an ease of reading and text-grade level calculation.<sup>28,29</sup> Computation of the Flesch and Kincaid indexes was completed by adding the Lingua::EN::Fathom module to our Perl program.

### Suicidal emotions

Each suicide note was annotated with emotional concepts. Developing an ontology to organize these concepts required both the Pubmed queries and expert literature reviews. Using the Pubmed

**Table 1.** Feature selection process and results.

	Initial features	Removal of high correlations	Removal of rare features	Removal of low information gain features*
Words	993	993	153	42.7 (0.13)
P.O.S.	36	32	24	18.47 (0.11)
Concepts	32	32	8	2.82 (0.10)
Read. Score	2	2	2	2 (0)
Total	1063	1053	186	66 (0)

**Note:** \*Done on training set only.

queries, a frequency analysis of the keywords in 2,000 suicide related manuscripts was conducted. Expert review of those keywords yielded 166 suicide related manuscripts that contained suicide emotional concepts. These emotional concepts were allocated to 19 different classes. Three mental health professionals then reviewed each of the 66 notes and assigned the emotional concepts found in those notes to the appropriate classes. For example, the emotional concept of *guilt* was assigned to the class of *emotional states*.

## Machine learning

There are multiple general types of machine learning: unsupervised, semi-supervised and supervised. Semi-supervised methods use both labeled and unlabeled data and is efficient when labeling data is expensive, which leads to small data sets. For this research the semi-supervised approach was selected mainly because the labeled data are small. Using the Waikato Environment for Knowledge Analysis (Weka) collections of data mining algorithms, we compared a number of machine learning methods.<sup>30</sup> Those germane to this research were organized into five categories:

- Decision trees: J48/C4.5, Logistic Model Trees, Decision Stump and M5P
- Classification Rules: JRIP, Repeated incremental Pruning to Produce Error Reduction (RIPPER), M5Rules, OneR, and PART;
- Function models: Sequential minimal optimization (SMO, a variant of SVM), PolyKernel, Puk, RBF Kernel, Logistic, and Linear Regression;
- Lazy Learners or Instance-based learner: IBk and LBR;
- Meta learners: AdaBoostM1, Bagging, Logit Boost, Multi Boost AB and Stacking.

## Machine categorization

Some features have been used in previous studies.<sup>22,23,31</sup> This study extends the previous work by creating a heterogeneous, multidimensional feature space. To do so, the following algorithms were used to extract and quantify the relevant content features:

1. Structure: number of paragraphs.<sup>32</sup>
2. Spelling: number of misspellings (perl module Text::SpellChecker).

3. Tagging: number of tokens, number of words, number of non-word characters, number of sentences, mean frequency of a word, standard deviation of frequency of a word, maximal frequency of a word, mean length of a sentence, standard deviation of length of a sentence, maximal length of a sentence, frequency of 32 parts of speech (perl module Lingua::EN::Tagger).<sup>33</sup>
4. Readability: Flesch-Kincaid grade level, Flesch reading ease (perl module Lingua::EN::Fathom).<sup>28</sup>
5. Parsing: mean depth of a sentence, standard deviation of depth of a sentence, maximal depth of a sentence (perl module Lingua::CollinsParser).<sup>34</sup>

Features came from different sources which led to their numeric values being in different ranges. To remedy this, feature values were normalized to a maximum value of one. In the end, a matrix with 66 documents and 49 features whose values ranged 0 and 1 was created. Since there are fewer features than documents additional features selection was not applicable.

## Algorithm classification

### Decision trees

A decision tree classifier is represented as a tree. Every node of the tree is represented by a list of possible decisions. The decision about the next branch is based on a single feature response. Leaves of the tree are represented by the decisions about which class should be assigned to a single document. For decision tree analysis the following algorithms:

- J48 generates unpruned or pruned C4.5 revision for 8 decision trees.
- LMT implements “Logistic Model Trees”.
- Decision Stump implements trees with a single split only (i.e. one-level-decision trees), which are frequently used as base learners for meta learners such as Boosting.

### Classification rules

The classifier algorithm is represented by a set of logical implications. If a condition for a document is true, then a class is assigned. Conditions are composed of a set of feature responses to OR-ed or AND-ed together. These rules can also be viewed as a simplified representation of a decision tree. For classification analysis the following algorithms are used:



- JR implements a fast propositional rule learner, “Repeated Incremental Pruning to Produce Error Reduction” (RIPPER).
- OneR builds a simple 1R classifier; it is a set of rules that test a response of only one attribute.
- PART generates a set of simplified rules from a C4.5 decision tree.

### Function models

Classifiers can be written down as mathematical equations. Decision trees and rules cannot. There are two classifiers in this category. For functional models the following algorithms are used:

- SMO implements a sequential minimal optimization algorithm for training a support vector classifier using linear kernel. Note: SMO is a variant of the popular SVM algorithm. It breaks the data into two-dimensional sub-problems that may be solved analytically, rather than numerical optimization.
- Logistic builds multinomial logistic regression models based on ridge estimation.

### Instance-based learning or Lazy learners

Classifiers in this category do no real work until classification time. It is done by reviewing every instance in the training set separately. Only one algorithm is used in this category:

- IBk provides a k-nearest neighbors classifier, which uses Euclidean metric as a distance measure.

### Bayesian classifiers

Classifiers use Bayes theorem and the assumption of independence of features. Only one algorithm is used in this category:

- NB implements the probabilistic Naive Bayes classifier.

### Estimation

Bootstrapping is used to estimate classifier performance.<sup>35</sup> Bootstrapping has been shown to provide stable estimates.<sup>36</sup> It is the practice of estimating properties of a classifier by measuring them when sampling from an approximating distribution. The advantage of bootstrapping over other analytical methods is its simplicity. Derivation of standard

error estimates and confidence intervals for complex estimators of complex parameters is straightforward. The disadvantage of bootstrapping is that while (under some conditions) it is asymptotically consistent, it does not provide general finite sample guarantees and it has a tendency to be overly optimistic. The apparent simplicity may conceal the fact that important assumptions are being made when undertaking the bootstrap analysis, e.g. independence of samples whereas these would be more formally stated in other approaches.<sup>37</sup>

In the case of mental health professionals and psychiatry trainees ratings the result is a simple weighted average and so bootstrapping is not necessary. In the case of machine-categorization estimation, a random sample of 66 documents with replacements is drawn to create a training set. After that the model is tested against all 66 documents. The 632+ method is used for bias correction.<sup>35</sup> Each bootstrap estimate had 100 samples redrawn. The procedure was repeated 25 times to calculate the stability of the estimate.<sup>36</sup> Thus, there was total  $25 \times 100$  samples with replacement drawn.

Previous computer analysis of suicide notes have used t-test, chi-square or ANOVA statistics to show the best features that can discriminate between two categories. In our case, the Kolmogorov-Smirnov-test showed that not all features are normally distributed. Hence, a Wilcoxon-test is used to calculate the difference in distribution shift of a feature in elicited and in genuine notes.

### Results

Overall features most often selected for the machine algorithms included:

**Words:** am, and, are, Betty, but, could, did, do, everything, for, good, goodbye, had, have, he, her, I, in, is, it, Jones, leave, life, longer, love, Mary, more, mother, my, n't, now, Smith, so, that, the, things, this, to, Tom, was, with, and you. While it is reasonable to suggest that the anonymized proper nouns like Jones, Mary, Tom and Smith should not be part of the feature space, they were included because they can act as proxies for individual names;

**Part of speech tagging:** Cardinal number (CD), Determiner (DET), preposition or subordinating conjunction (IN), Adjective (JJ), Adjective superlative (JJS), Modal (MD), Noun, singular or mass (NN),



Proper noun plural (NNP), Noun plural (NNS), Personal pronoun (PP), Prepositional phrase (PP), Personal pronoun (PRP), Possessive pronoun (PRPS), Adverb (RB), Verb, base form (VB), Verb, past participle (VBN), Verb, non-3rd person singular present (VBP), and Verb, 3rd person singular present (VBZ);

**Readability:** the Flesch Reading Ease score is a 100-point scale, with higher scores easier to read. The Flesch-Kincaid Grade Level is a number that corresponds with grade level; and

**Emotions:** giving things away, hopeless, regret and sorrow.

- Suicide
  - Emotions
  - Methods
  - Notes
    - Reasons in notes
      - Situations
        - Unbearableness
        - Perturbation
        - Dread
        - Pain
      - Relations
        - Total rejection
        - Lack of successful relationship
        - Broken relationship
        - Unrealistic relationship
        - Abandonment
        - Substance abuse
      - Emotional states
        - Hostility
        - Self vengefulness
        - Self aggression
        - Self punishment
        - Depression
        - Guilt
        - Anger
        - Worthlessness
        - Hate
        - Shame
        - Fear
        - Despair
        - Hopelessness
        - Helplessness
        - Love
      - Cognitive states
        - Suppressed belief that leads to deadly consequence
        - Unaware of own distortion
        - Vague objective reasoning
        - Use of cognitive constriction
        - Confused concept of self
        - Denigration of person place and things
        - Motivation by unconscious dynamic

**Figure 1.** Partial suicide ontology.

**Table 2.** Genuine and elicited notes descriptive statistics.

Description	Genuine notes mean (SD)	Elicited notes mean (SD)	P-value
Number of sentences	9.242 (8.058)	4.848 (3.308)	0.001
Maximal frequency of a word	7.909 (6.090)	4.212 (3.525)	0.002
Number of non-word characters	13.182 (11.078)	7.212 (7.039)	0.004
Standard deviation of word	1.102 (0.809)	0.632 (0.519)	0.005
Flesch-Kincaid grade level	4.719 (2.142)	6.517 (2.994)	0.008
Cardinal numbers per note	1.121 (1.867)	0.182 (0.465)	0.015
Verb, past tense per note	4.848 (9.398)	1.152 (1.787)	0.018
Personal pronoun per note	16.030 (16.318)	8.879 (8.112)	0.019
Flesch reading ease per note	81.776 (7.566)	75.949 (11.201)	0.024
Number of paragraphs per note	1.909 (1.508)	1.333 (0.777)	0.027
Number of tokens per note	122.000 (108.096)	73.364 (63.465)	0.031
Number of words per note	108.818 (97.944)	66.152 (56.801)	0.041
Adjective, superlative per note	0.485 (0.795)	0.182 (0.584)	0.042
Verb, non-3rd person singular present per note	4.727 (3.476)	3.394 (3.297)	0.072

**Abbreviation:** SD, standard deviation.

The human raters relied on the ontology shown in Figure 1: *Suicide Ontology*. This ontology is much more extensive than the four emotions aliquoted by the information gain function.

Feature selection and data reduction are listed in Table 1: *Feature Selection Process and Results*. Information gain was calculated only for the training data in each bootstrapped sample. From the initial 1063 possible features 66 were ultimately selected based on information gain and frequency. They included words, parts of speech, concepts and reading scores.

Table 2: *Genuine and Elicited Notes Descriptive Statistics* provides mean and standard deviation of a number of note characteristics. It shows fifteen features with the smallest p-values in two sample Wilcoxon-tests. Hypothesis testing as a feature selection is only one of many methods and may not always describe the data accurately.<sup>38</sup> Some machine learning algorithms, like LMT, have feature selection embedded. It is worth looking if different feature selection algorithms give same results. This is possible only for some simple problems.

Table 3: *Human & Machine Raters in after 25 × bootstraps* shows that mental health providers perform better than psychiatry trainees, but not as well as the best machine learning algorithms. For the psychiatry trainees, their overall categorization was roughly equal to the flip of a coin. Mental health providers were significantly better than trainees. They accurately classified notes about 63% of time.

Table 3 compares the machine classification algorithms with psychiatry trainees and mental health providers. On average, the best machine algorithm (Logistic Model Trees) performed significantly better than the mental health providers. All the algorithms did significantly better than the psychiatry trainees. Nine of ten machine algorithms performed significantly better than mental health providers.

Performance of different machine learning algorithms is complementary. J48/PART (0.640/0.645) suggests that a tree representation is mediocre for the data. On the other hand Linear SMO (SVM)/LMT (0.705/0.744) suggests that there is some linear

**Table 3.** Human & Machine Raters in after 25 × bootstraps.

	Mean accuracy (SD)
Human Raters	
Psychiatric Physician Trainees	0.510 (0.002)
Mental Health Professionals	0.609* (0.001)
Machine Model	
LMT	0.744** (0.009)
LinSMO	0.705** (0.012)
Decision	0.667** (0.007)
JRip	0.661** (0.015)
NB	0.645** (0.005)
PART	0.645** (0.009)
J48	0.640** (0.005)
Logistic	0.633** (0.014)
IB3	0.623** (0.010)
OneR	0.605 (0.014)

**Notes:** \*Significant between students and professionals  $\leq 0.001$ .

\*\*Significant between professionals and machine  $\leq 0.001$ .

**Table 4.** Logistic Model Tree when all features and all suicide notes are used for training.

Genuine note equation	Elicited note equation
$f = 0.04 +$ “Flesch-Kincaid grade level”* $-4.4 +$ “maximal depth of a sentence”* $1.31 +$ “mean depth of a sentence”* $-0.82 +$ “mean length of a sentence”* $-0.45 +$ “standard deviation of a length of a sentence”* $0.6 +$ “comma frequency”* $-2.35 +$ “other punctuation frequency”* $0.87 +$ “cardinal number frequency”* $1.17 +$ “proper noun, singular frequency”* $9.61 +$ “symbol frequency”* $2.83 +$ “verb, gerund or present participle frequency”* $-2.02 +$ “verb, past participle frequency”* $-2.03 +$ “wh-pronoun frequency”* $0.74 +$ “maximal frequency of a word”* $1.5$	$f = -0.04 +$ “Flesch-Kincaid grade level”* $4.4 +$ “maximal length of a sentence”* $-1.19 +$ “mean depth of a sentence”* $0.82 +$ “mean length of a sentence”* $0.45 +$ “standard deviation of a length of a sentence”* $-0.6 +$ “comma frequency”* $2.35 +$ “other punctuation frequency”* $-0.87 +$ “cardinal number frequency”* $-1.17 +$ “proper noun, singular frequency”* $-9.61 +$ “symbol frequency”* $-2.83 +$ “verb, gerund or present participle frequency”* $2.02 +$ “verb, past participle frequency”* $2.03 +$ “wh-pronoun frequency”* $-0.74 +$ “maximal frequency of a word”* $-1.5$

**Note:** \*Coefficient's absolute value represent strength of feature.

separability of the two categories. In addition, a logistic regression outperformed linear support vector machines, i.e. LMT (0.744).

Table 4: *Logistic Model Tree*. When all features and all suicide notes are used for training and LMT is trained on the entire data set, there was only one leaf with two linear functions that categorize suicide notes. There are only three features shared by the Wilcoxon-test and LMT (maximal frequency of a word, Flesch-Kincaid grade level, cardinal number frequency). Two of the equations in Table 4 misclassified only four documents. Features selected by LMT describe sentences (number of words, depth of the parsed tree), whereas hypotheses testing selected features that describe different aspects of the notes.

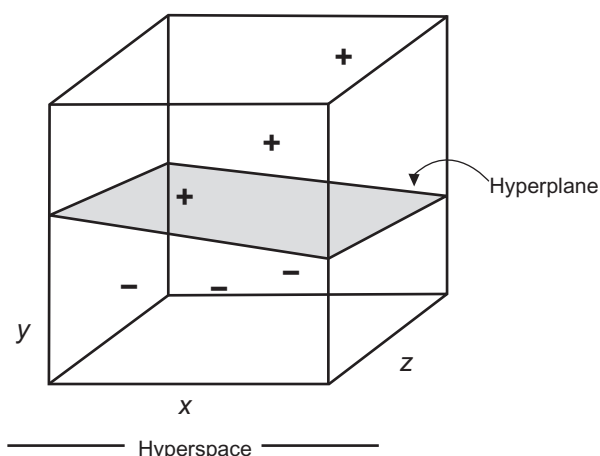
**Figure 2.** Hyperspace example.

Table 4 can be difficult to read and so we offer the following example as an explanation. Figure 2: *Hyperspace Definition* shows a three dimensional cube, or a hyperspace with three features. Axis  $z$  represent the Flesch-Kincaid reading score, Axis  $y$  represents the MLS method, and axis  $x$  represents the MDS method. In this case, the difference each methods computes creates a hyperplane. This hyperplane is shown in the center of the defined hypercube. Those features above the hyperplane are labeled with a “+”. In our case this represents genuine notes. Those features below the hyperspace are labeled “-”. In our case, this represents elicited notes.

## Discussion

The purpose of this research was to understand how well different machine learning algorithms performed compared to humans who were asked to distinguish between elicited and genuine suicide notes. We confirmed that, at least in part, machine algorithms could do as well as humans. We speculate here on the possible reasons.

One possible explanation can be found in psychological phenomenology. Psychological phenomenology focuses on the *experience* of the subject. True, the term *experience* is a complex concept, but in principal an *experience* is not directly observable by an external observer:<sup>39</sup> a mental health providers cannot truly observe the internal pain of a suicidal patient. What then gives insight into how the mental



health, psychiatry trainees and machine algorithms experience the act of classifying genuine and elicited suicide notes? We propose that it is the features. Figure 2 show what features the humans used for classification. There are four classes and 40 emotional concepts. The machine algorithms include four emotional concepts, 42 specific words (none emotional), and readability scores. Considering these selections it is reasonable that the human raters focused on content, while the machine algorithms focused on structure.

The results of this research has a number of potential applications. One potential application is that using machine algorithms to discriminating between genuine and elicited suicidal notes has important clinical and forensic implications, especially as it relates to advanced decision support. The findings also suggest that algorithms such as the one used in this study may have applications for the prospective clinical assessment of psychiatric patients suffering not only from suicidal ideation or intent, but also homicidal impulses which are vital to predict. Finally, this study can have relevant applications for distinguishing malingerers who feign psychiatric illness for ulterior motives.

Finally, addressing one item would enhance the strength of the study's generalizability; that is, the sample size. We understand that 66 notes is low. To our knowledge, however, this is the only data set that lends itself to this type of research.

## Acknowledgements

We acknowledge E.S. Shneidman for access to the data and for comment and guidance. We acknowledge the divisions of Biomedical Informatics, Emergency Medicine and Psychiatry at Cincinnati Children's Hospital Medical Center, University of Cincinnati and the Ohio Third Frontier program for their generous support of this work. This work is funded from the Ohio Third Frontier program.

## Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

## References

1. Sussman MP, Jones SE, Wilson TW, Kann L. The Youth Risk Behavior Surveillance System: updating policy and program applications. *J Sch Health*. 2002 Jan;72(1):13–7.
2. Freedenthal S. Challenges in assessing intent to die: can suicide attempters be trusted? *Omega (Westport)*. 2007;55(1):57–70.
3. Erdman HP, Greist JH, Gustafson DH, Taves JE, Klein MH. Suicide risk prediction by computer interview: a prospective study. *J Clin Psychiatry*. Dec 1987;48(12):464–7.
4. Levine S, Ancill RJ, Roberts AP. Assessment of suicide risk by computer-delivered self-rating questionnaire: preliminary findings. *Acta Psychiatr Scand*. 1989 Sep;80(3):216–20.
5. Shneidman E, Farberow N. *Clues to Suicide*: McGraw Hill Paperbacks; 1957.
6. Boismont B. De l'influence de la civilisation sur le developpement du suicide. Paris: Germer Bailliere; 1856.
7. Wolff H. Suicide notes. *American Mercury*. 1931;24(193):264–72.
8. Frederick CM. Suicide notes: A survey and evaluation. *Bulletin of Suicidology*. 1969;1:27–32.
9. Chavez-Hernandez A-M, Paramo D, Leenaars AA, Leenaars L. Suicide notes in Mexico: what do they tell us? *Suicide Life Threat Behav*. 2006; 36(6):709–15.
10. Ho T, Yip P, Chiu C, Halliday P. Suicide notes: what do they tell us? *Acta Psychiatrica Scandinavica*. 1998;98(6):467–73.
11. Leenaars A. *Suicide notes predictive clues and patters*: Human Science Press; 1988.
12. Leenaars AA, Lester D, Wenckstern S, Rudzinski D, Breward A. A comparison of suicide notes written by suicide notes and parasuicide notes. *Death Studies*. 1992;16.
13. Leenaars A, Balance W. A predictive approach to the study of manifest content in suicide notes. *Journal of Clinical Psychology*. 1981;37(1):50.
14. Ogilvie D, Stone P, Shneidman E. Some characteristics of genuine versus simulated suicide notes. *National Institute of Mental Health Bulletin of Suicidology*. 1966:27–32.
15. Foster T. Suicide note themes and suicide prevention. *International Journal of Psychiatry in Medicine*. 2003/2004;33(4):323–30.
16. Leenaars A, Cantor C, Connolly J, et al. Controlling the environment to prevent suicide: International perspectives. *Canadian Journal of Psychiatry*. 2000;45(7):639–44.
17. Leenaars A. *Suicide notes predictive clues and patterns*: Human Science Press; 1988.
18. Leenaars A, Balance W. A predictive approach to freud's formulations regarding suicide. *Suicide and Life-Threatening Behavior*. 1984;14(4): 275–83.
19. Connor R, Leenaars A. A thematic comparison of suicide notes drawn from Northern Ireland and the United States. *Current Psychology*. 2004;22(4):339–47.
20. O'Connor R, Sheehy N, O'Connor D. A thematic analysis of suicide notes. *Crisis: the Journal of Crisis Intervention and Suicide Prevention*. 1999; 20(3):37–41.
21. Barak A, Miron O. Writing characteristics of suicidal people on the Internet: a psychological investigation of emerging social environments. *Suicide Life Threat Behav*. 2005;35(5):507–24.
22. Jones N, Bennell C. The development and validation of statistical prediction rules for discriminating between genuine and simulated suicide notes. *Archives of suicide research: official journal of the International Academy for Suicide Research*. 2007;11(2):219.
23. Handelman LD, Lester D. The content of suicide notes from attempters and completers. *Crisis*. 2007;28(2):102–4.
24. Pestian J, Matykiewicz P, Kowatch R, Grupp-Phelan J, Leenaars AA. Suicide notes classification using natural language processing: a phenomenological perspective. *Suicide an Life-Threatening Behavior*. 2008;Under Review.
25. Manning C, Scheutze H. *Foundations of statistical natural language processing*: MIT Press; 1999.
26. Breiman L, Freidman JJ, Olshen RA, Stone CJ. *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group; 1984.



27. Quinlan JR. *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufman; 1993.
28. Flesch R. A new readability yardstick. *Journal of Applied Psychology*. 1948;32:221–33.
29. Kincaid J, Fishburne R, Rogers R, Chissom B. *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Memphis, TN: Naval Technical Training, US Naval Air Station; 1975.
30. Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. 2nd ed: Morgan Kaufman; 2005.
31. Edelman A, Renshaw S. Genuine versus simulated suicide notes: an issue revisited through discourse analysis. *Suicide and life-threatening behavior*. 1982;12(2):103.
32. Wall L, Christiansen T, Randal L. *Programming Perl*. New York: O'Reilly; 1996.
33. Ceglowski M, Coburn A, Cuadrado J. An Automated Management Tool for Unstructured Data. Paper presented at: WI '03: Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence 2003; Washington, DC, USA.
34. Collins AM, Quillan M. Retrieval Time for Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*. 1996;8:2969.
35. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. London: Chapman and Hall; 1993.
36. Molinaro N, Carreiras M. Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biol Psychol*. 2009 Dec 22.
37. Chernick MR. *Bootstrap Methods, A practitioner's guide*.: Wiley Series in Probability and Statistics; 1999.
38. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*. 2007;23(19):2507.
39. Snygg D. The need for a phenomenological system of psychology. *Psychological Review*. 1941;48:404–28.

**Publish with Libertas Academica and every scientist working in your field can read your article**

*"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."*

*"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."*

*"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."*

**Your paper will be:**

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

**<http://www.la-press.com>**