Using of Natural Language Processing Techniques in Suicide Research

Azam Orooji [a*], Mostafa Langarizadeh [b]

[a] PhD Student in Medical Informatics, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

[b] Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Abstract

It is estimated that each year many people, most of whom are teenagers and young adults die by suicide worldwide. Suicide receives special attention with many countries developing national strategies for prevention. Since, more medical information is available in text, Preventing the growing trend of suicide in communities requires analyzing various textual resources, such as patient records, information on the web or questionnaires. For this purpose, this study systematically reviews recent studies related to the use of natural language processing techniques in the area of people's health who have completed suicide or are at risk. After electronically searching for the PubMed and ScienceDirect databases and studying articles by two reviewers, 21 articles matched the inclusion criteria. This study revealed that, if a suitable data set is available, natural language processing techniques are well suited for various types of suicide related research.

1- Introduction

Suicide accounts for a great many mortalities worldwide which has made it a serious health issue especially involved in mortalities among the young and teenagers [1]. In 2000, the World Health Organization (WHO) maintained a global suicide rate of one person per 40 seconds [2]. An annual rate of about one million mortalities has been reported on a global scale [3] and suicide is considered as the thirteenth most frequent causes of death worldwide and accounts for 5-6% of all mortalities [4]. The risk of suicide varies across countries as a function of different socio-demographic features [4]. The highest probability of suicide belongs to the young, teenagers and men [4]. It is noteworthy that the rate of suicide among men is four times as high as women and about one out of twelve attempt to suicide is estimated to cause death [5]. How the victims behave is complicated and affected by an array of social, contextual, ecological and psychological factors [5].

A number of risk factors vary according to age, ethnicity and sex and occasionally vary through time. They include depression, previous attempts of suicide, family history of mental disorders or substance abuse, young age, family history of suicide, presence of firearms at home, experience of imprisonment, domestic and sexual violence, genetic problems and severity of illness [5-8].

Suicidal patients are of three types: 1) an ideator, or one who is obsessed with the thought of suicide or according to the definition by WHO attaches no value to life, 2) an attempted, one who attempts to commit suicide or according to the definition put forth by WHO attempted to terminate one's life but did not succeed, and 3) a completer, one who has managed to commit suicide [2, 9]. According to statistics, although women think about suicide more than men, in practice they are not as successful in ending their lives as men. Moreover, research showed a higher rate of male-female disparity in high-income countries [4, 5].

Estimating the adverse effects of obsession with suicide in the long run is far from easy for the public, households

---

and community [4]. Moreover, it differs across communities and that is why a copious amount of attention has been drawn to this issue worldwide. There are usually great attempts made at a national level to prevent this disaster to occur or prevail [10].

In its first Mental Health Action Plan report in 2014 [11], WHO maintained that an improved state of data in hospital registries is the key way to cut down on the rate of suicide and its victims [4]. So far, adherence to old methods of monitoring public health has made it hard to investigate issues related to suicide [4]. Therefore, it is essential to use cost-effectiveness tools and strategies to collect and interpret the required data for suicide prevention [4]. Researchers believe that thinking about suicide mostly leads to the action itself and that is why it is essential to identify those who tend to do so [1]. Computational methods such as NLP can use Electronic Health Record (EHR) data to identify those prone to suicide [4]. The other key textual sources that help to predict suicide are notes, interviews and questionnaires filled by people who already committed suicide. Suicide notes help to recognize their motivations and thoughts [5]. With this concern, NLP-based tools can be applied too to spot those at the risk of suicide or under mental pressures and to prevent undesirable consequences [4].

NLP deals with a structured data extraction from free text (not following a particular format). Knowledge-based NLP algorithms were based on terminologies and rules developed by experts. In recent years, however, annotated clinical texts attracted researchers' attention to use ML algorithms which were more flexible and time-saving. However, knowledge-based NLP algorithms were more suitable for solving particular problems. A third method exists which has mixed the benefits of the first two methods and is accordingly known as the hybrid method [12, 13]. NLP data extraction methods are less costly and less time-consuming which made helped them to achieve much in different health-related domains [4, 13].

The present study aimed to systematically review recent investigations which used NLP techniques in issues related to suicide within the past ten years. The following of this paper is organized in four sections. The second section deals with the method used to search and select the articles. The third section presents the results and the fourth discusses the findings in the light of the related literature. The final section presents a summary of findings and also makes suggestions for further research.

2- Method

This paper is a systematic review to insure that search and retrieval process is accurate and comprehensive enough. The articles published in the last ten years in Pubmed and ScienceDirect electronic databases were reviewed with various combination of related keywords. Table 1 and 2 show the search query used in this study and inclusion/exclusion criteria, respectively.

Table 1. The search query used in this study

Database	Items found	Search Query
Pubmed	26	Suicide[Title/Abstract] AND (natural language processing[Title/Abstract] OR text mining[Title/Abstract] OR information extraction[Title/Abstract]) AND ("2007/06/14"[PDat] : "2017/06/10"[PDat])
SD	13	pub-date > 2006 and TITLE-ABSTR-KEY(suicide) and (TITLE-ABSTR-KEY(natural language processing) or TITLE-ABSTR-KEY(information extraction) or TITLE-ABSTR-KEY(text mining)).

Table 2. Inclusion and exclusion criteria

Exclusion Criteria	Inclusion Criteria
Studies when access to full-text articles was not available	Articles that have been used NLP to analysis texts
Newspapers, Review, letter to editor, workshops, posters, short report, book and thesis	Articles published in English
Articles written in non-English language	Articles published during the last ten years

A total of 39 papers were identified in the first stage using the search query (Table 1). All duplicated articles were removed automatically using Endnote and a manual revision was done for verification. Based on the criteria for inclusion and exclusion, two reviewers independently screened titles, abstracts and then full text of articles identified through search, with discrepancies resolved by consensus. Finally, 21 publications were included on the basis of eligibility criteria. The agreement between the two reviewers was calculated by using the κ statistic. The resulting k-statistic is equal to 0.81 (P<0.001). Figure 1 shows PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of selecting the studies.

3- Results

As previously mentioned in the introduction, the selected articles reviewed here were analysed from various aspects

including the purpose of research, data set, setting, methodology and the main results.

As already described, suicide notes were mostly affective in content and could be useful in reading victims' minds. Therefore, some studies addressed this issue [9, 14]. Those who attempted suicide and were taken to the hospital emergency unit are likely to commit suicide again. The notes written by those who commit suicide and those pretending to do so differ widely. The research [9] aimed to classify completer and simulated suicide notes. Therefore, three experts, according to an ontology, searched for all emotional concepts (anger, depression, affection, etc.) in the content and annotated them. Five experts then divided these annotations into two groups, real and pretended. Subsequently, these texts were used for machine-learning so as to distinguish between linguistic and emotional characteristics of the notes. The results revealed that Machine Learning (ML) algorithms were capable of distinguishing people pretending to commit suicide and those who really aim to do so. ML algorithms showed to be useful for a prospective clinical analysis of the mentally ill who not only intend to commit suicide but also think of murder. Also, the authors in the other paper(14) showed that how well different machine learning algorithms performed compared to humans (practicing
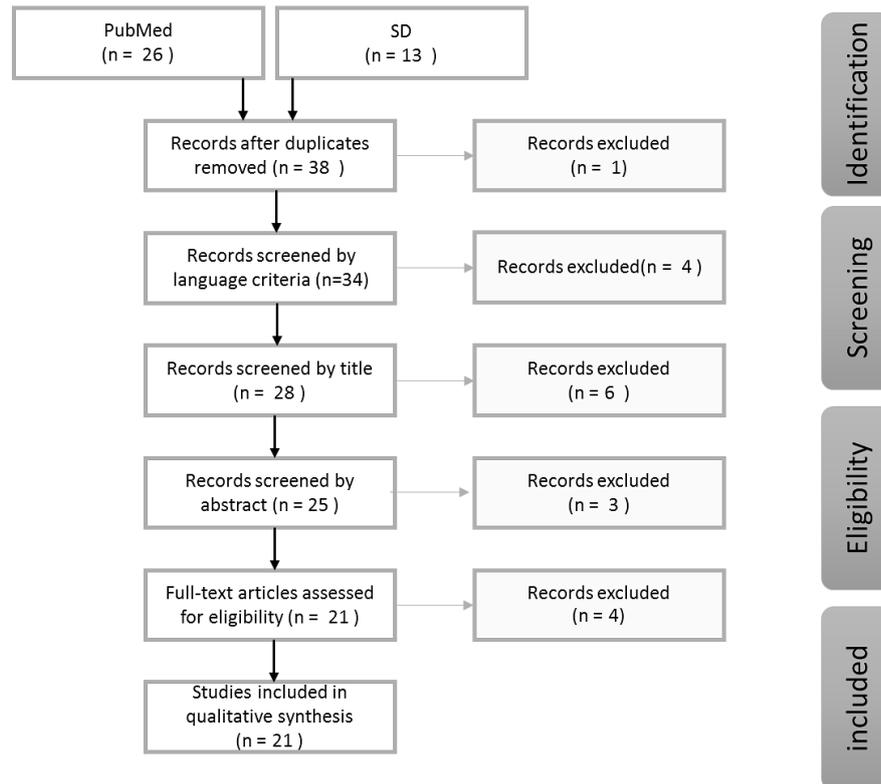

Figure 1. PRISMA flowchart of selecting the studies

Mental health professionals and psychiatry physician trainees) who were asked to distinguish between elicited and genuine suicide notes.

Sentiment Analysis (SA) has recently turned into a main area of research in computational linguistics. Emotion recognition is a type of SA used to detect the emotion fragments involved in free texts [15]. NLP challenge proposed by Informatics for Integrating Biology and Bedside (i2b2) in 2011 dealt with SA of suicide notes [16]. Many researchers participated in this challenge and the results were published in several papers [3, 5, 15, 17-23]. This challenge provided researchers with 900 notes made by those who already committed suicide and died as a result. The content of 600 notes were annotated according to 15 emotion classes such as fear, sin, anger and so on [16].

As biomedical relations are highly involved in biological processes, study of interactions within biology has attracted research interest. Many attempts have been made to extract the effective biological relationship such as protein-protein interactions or gene-disease associations. Knowing how genetic factors are involved in the occurrence of diseases can help to develop new techniques for prevention, diagnosis and treatment of diseases. According to the related literature, genetics is a risk factor of suicide. The majority of biomedical facts can be accessed in free texts such as articles. So, text mining techniques can be used to extract the required data. Compared to other diseases, finding gene-suicide associations through experience is hard. In different databases few samples contain information related to suicide and the relevant genes. That is why supervised methods are not possible. Therefore, Quan [10] used supervised and semi-supervised methods for finding the gene-suicide associations.

The depressed are usually prone to suicide and are often treated and kept in primary care unit. Therefore, a close monitoring of suicidal patients in primary care and how they behave can reduce the rate of suicide. Using the existing

data in EHR can contribute to this supervision; however, it hardly contains the frequency of one's attempts to commit suicide. In [24] the content of Patient Health Questionnaire [PHQ-9]) regularly completed by suicidal patients and history of present illness (HPI) notes field in the EHR was analysed and compared to diagnostic codes recorded in EHR.

A correct and comprehensive record of data in EHR helps to facilitate quite many investigations. The paper [2] explored whether it was possible to make more precise estimates of the national rate of suicide annually based on an automatic EHR analysis system rather than the manual? To this aim, initially all the records were coded via an automatic system based on UMLS (Unified Medical Language System). Then seven ML algorithms were used to classify the reasons (suicide vs. non-suicide) why people were taken to hospital emergence unit.

Symptoms of depression can range from temporary sadness to suicide. Shyness, embarrassment and the negative social aspect of suicide dissuade one from asking for help. Current forms of social media especially internet forums and web-logs create the chance of discussing feelings in a safe and confidential environment. However, the majority of members are not capable of diagnosing the severity of their depression and their need for cure. Online diagnosis, as the target of Kamren's study [25], is a strategy for finding those at risk, facilitating the right and timely intervention and improving public health. Similar investigations have been conducted in Korea which used web-based data. As an instance, a Korean study [1] explored the search queries on web-based social media that are commonly used with adolescents and young adults and not only analysed the risk of suicide but also its relation to the monthly employment rate, rental prices index, youth suicide rate, and number of bully victims. Woo et al. [26] analysed Twitter content to explore social behaviour in response to the sinking of Sewol ship. Disasters often influence mental health tremendously. However, most investigations only explore the effects of disasters on those affected e.g. victims or their families. A few studies have addressed the massive effect of disasters on public mental health. The data presented in social media can help to analyse public behaviour and estimate the rate of suicide after a certain disaster occurs.

As there is a high risk of suicide after discharge from hospital, a system was proposed by McCoy [27] to reduce the rate of post-discharge suicide. NLP was initially used in this system to extract the positive and negative (suicide-related) terms. Then, a regression analysis was performed to estimate suicide risk for each individual. The results were also compared to what was recorded in patients' follow-up records as the cause of death.

Another system was developed by Cook [4] to predict suicide and sever mental symptoms in adults just discharged from hospital emergency unit or a mental clinic. To this aim, 1453 young adults (>18 yrs.) who survived from a suicide attempt and were just discharged from an emergency unit or a mental clinic were sent, within a year, text messages containing a link to a structured questionnaire. The content was about the state of sleep, well-being and anger. There was also an unstructured question to be answered: "How do you feel today?" A regression analysis was used on the structured questionnaire while NLP helped to analyse the unstructured question. The results were then compared together.

Though the rate of suicide and self-harm is significantly reduced during pregnancy, women already suffering from mental disorders get more prone to suicide. However, a few studies have been done with this concern and the risk factors involved are unknown. In [8], NLP was used to extract information from the medical records of pregnant women with mental disorders. Such information was history of domestic, substances abuse, suicide ideation, location and method of self-harm. Then the statistical analysis particularly regression analysis was performed to explore the prevalence of these injuries, their locations and methods. The main question explored was whether women with self-harm during pregnancy differed from the others in terms of demographic and clinical features. A summary of the articles explored is presented in Table 3.

Table 3. Characteristics of selected articles

Author	Country	Purpose(s)	Dataset	Method (s)	Best results
Pestian [9]	US	Classification of completer and simulated suicide notes	Suicide notes from 33 completers and 33 simulators	Decision tree(J48, C4.5, LMT,Decision stump, M5P) ; Rule-based classifiers (JR, M5, OneR, PART); Function models(SMO, logistic builds, multinomial logistic regression, linear regression); lazy learners and meta learners	A significant difference was found between the linguistic (word count, verbs, prepositions and nouns) and emotional (completers gave away their possessions) characteristics of the notes. Classification accuracy: Mental health professionals = 71% The Best ML method (SMO)= 78%
Pestian [14]	US	Classification of completer and simulated suicide notes	Suicide notes from 33 completers and 33 simulators	Decision tree(J48, LMT,Decision stump) ; Rule-based classifiers (JR, OneR, PART); Function models(SMO, logistic regression); lazy	Classification accuracy: Mental health professionals = 63% psychiatric physician trainees = 49% The Best ML method (SMO)= 78%

				learners(IB3) and Naïve Bayes (NB)
Cherry [23]	Canada	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Latent sequence model and Support Vector Machine (SVM)	F-measure=55.22%
Desmet [18]	Belgium	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	SVM	F-measure=51.19%
Kovacevic [22]	UK	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Integrated rules and machine learning (NB)	F-measure (hybrid model) = 53.36% Recall (hybrid model) =50.47% Precision (using Rules)=67.17%
McCart [5]	US	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Integrated rules and machine learning (Decision tree, k-Nearest Neighbor (kNN),SVM)	F-measure(using rules, Descision tree and SVM)= 50.23% Recall (using rules, Descision tree and SVM) = 50.55% Precision (using Rules)=56.67%
Sohn [19]	US	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Integrated rules and machine learning (NB, RIPPER)	F-measure (hybrid model) =56.40 % Recall (hybrid model) = 55.74% Precision (ML-based model) = 61.12%
Spasic [21]	UK	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Integrated rules and machine learning (NB)	F-measure = 53% Recall =52% Precision =55%
Wicentowski [3]	US	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	An ensemble of supervised maximum entropy (ME) classifiers	F-measure=53.43% Recall=49.61% Precision =58.61% (different feature set was used) *
Yang [15]	UK	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Hybrid model (Integrated lexicon-based keyword spotting and different merging of ML algorithms including Conditional Random Field (CRF), SVM, NB and ME)	F-measure= 61.39% Precision =58.21% Recall (using different merging method) = 68.03%
Yeh [20]	US	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	Logistic Regression	F-measure =49.33% Recall =47.64% Precision =51.14%
Desmet [17]	Belgium	Automatic emotion classification by assigning right label to suicide notes with relevant emotions	A set of 900 suicide notes (600 notes for training and 300 notes for testing)	SVM	F-measure =53.87%
Quan [10]	Japan	Identifying gene–suicide associations by mining the amount of biomedical literature.	Dataset 1: Genetic Association Database (GAD, http:// geneticassociationdb .nih.gov/cgi-bin/index.cgi), abstracts of PubMed papers that describe suicide in GAD Dataset 2: GeneCards gene list (http://www.genecards.org/index.php?path =/Search /keyword/suicide /0/500/ score/desc) , abstracts from PubMed Central (PMC) Open Access based on the query of ''human+suicide' and GAD gene list	Polynomial kernel (PK) based pattern clustering method	Dataset 1 (using GAD gene list): F-measure= 82.05% Recall =90.06% Precision =75.35% Dataset 1 (using GeneCard gene list): F-measure= 79.01% Recall =94.66% Precision =67.80% Dataset 2 (using GAD gene list): F-measure= 41.03% Recall =67.12% Precision =29.55% Dataset 2 (using GeneCard gene list): F-measure= 47.20% Recall =81.12% Precision =33.28%
Anderson [24]	US	To estimate the use of diagnostic codes in EHRs to document suicidal	A number of 32385 patient record	Rule-based	Agreement between the ICD-9 code for suicidal ideation and suicidal ideation documented in the clinician notes field,

		ideation and attempt among patients seen in primary care.			item 9 of the 9-item Patient Health Questionnaire _PHQ-9 or logical disjunction of them was low (K-statistic= 0.036 , 0.068 and 0.04 respectively)
Karmen [25]	Germany	Detection of depression symptoms and their frequency using NLP methods	1304 posts extracted from an internet forum	Pattern matching	Average F-measure =79% Average recall =75% Average precision =84%
Song [1]	South Korea	Assessing the risk of suicide by monitoring online search activity of suicide-related words in South Korean adolescents through data mining of social media-Web sites	99,693 suicide-related documents posted by adolescents, ages 19 or younger	Regression	The largest total effect was observed in the grade pressure to depression to suicide risk. A lower employment rate, a higher rental prices index, and more victims of bullying, but not monthly youth suicide rate, were associated with increased online search of suicide-related words.
McCoy [27]	US	Improving prediction of Suicide and accidental death after discharge from general hospitals With NLP	845417 hospital discharges (including 458053 unique individuals)	Regression	Greater positive valence expressed in narrative discharge summaries was associated with substantially diminished risk for suicide.
Metzger [2]	France	Designing an expert system based on automated processing of EHRs to estimate of the annual rate of emergency department (ED) visits for suicide attempts	EHRs of 444 patients admitted to the ED in 2012 (98 suicide attempts, 48 cases of suicidal ideation, and 292 controls with no recorded non-fatal suicidal behavior)	Association rules, neural network, NB, random forest (RF), regression, decision tree and SVM	F-measure (RF,NB) =95.3% Positive Predictive Value (PPV) (NB) =96.9% Sensitivity (RF)= 95.9%
Woo [26]	South Korea	To explore how the public mood in Korea changed following the Sewol disaster using Twitter data	Daily Twitter posts (3 years before and 2 months after the Sewol disaster)	Time series analysis	In Korea, a human-made disaster can lead to an immediate increase in the suicidal preoccupation of the general public.
Taylor [8]	UK	Investigating the prevalence and correlates of self-harm in pregnant women with psychotic disorders	Of 420 women (33 women with and 387 women without (group 1,2 respectively) a recorded self-harm event in their index pregnancy)	Regression	Self-harm in pregnancy was associated with younger age, a history of child abuse or domestic violence, current (i.e. during pregnancy) domestic violence, a history of self-harm in the 2 years preceding pregnancy, substance misuse, smoking, non-affective disorder, acute admissions in the 2 years preceding pregnancy and stopping or switching rather than continuing a maintenance medication in the first trimester of pregnancy. The majority of self-harm events took place at home (73.1 %) and through overdoses (38.5%). There were no differences in age, ethnicity, diagnosis and admission rate between two groups
Cook [4]	Spain	Prediction of suicidal ideation and heightened psychiatric symptoms among adults recently discharged from psychiatric inpatient or emergency room	Two questionnaires: structured (such as General Health Questionnaire (GHQ-12) and WHO-5) and unstructured	Regression (for structured questionnaires) and NLP techniques (for unstructured questionnaire)	Suicide ideation (heightened psychiatric symptoms) prediction: 1. Regression PPV=73% (79%) Sensitivity =76% (79%) Specificity =62% (85%) 2. NLP PPV= 64% (64%) Sensitivity= 57% (31%) Specificity = 62% (79%)

4- Discussion

Though the majority of those committing suicide already visit doctors or specialists, it is hard to predict their suicide [27]. Those with a failed attempt of suicide are at a high risk of making a new suicide attempt (12-30%) or completing suicide (1-3%) within the next one year [2]. An analysis of suicide notes can partly help to delve into their mind [3, 5, 9, 14, 15, 18-23]. Even in an absence of any notes, questionnaires can be used to explore their emotions [4, 24]. Notes

and questionnaires are both considered as free texts with no particular format. Therefore, NLP techniques can be used for data extraction, which showed to be effective in doing so. Besides, online platforms are commonly used today to express and exchange feelings openly. Adolescents and young adults often create online content and publish it in social networks and websites. To prevent suicide, the risks need to be analysed in the early stages and with the help of reliable and supportive people who are referred to. As the existing data in social networks are large in amount, those willing to access and investigate them (e.g. suicide prevention agencies or health domain experts) cannot manually and consistently monitor them. There is a need for automatic systems to do so. IT-based solutions shave succeeded in identifying these behaviours and managed to enquire the data confidentially so as to ensure privacy matters(17). NLP-based models have proved effective in analysing large amounts of data produced in social media publically available (such as Facebook and Twitter) [1, 25, 26].

To act efficiently, NLP-based systems require a comprehensive and precise record of data in patients' records. As mentioned in the previous section, if the data are adequately comprehensive, epidemiologic research can be conducted [2] and certain target groups can be monitored too [8].

Unfortunately, due to the lingual, structural and typical diversity of the existing texts, it cannot be concluded which NLP method was the most efficient. However, developing standardized datasets such as that by i2b2 can create a basis for comparing different algorithms.

5- Conclusion

The present study was a systematic review of recent articles which used NLP in their investigation of suicide. It attempted to analyse the target articles from different aspects. The realm of medicine is replete with textual data. Considering the high rate of suicide worldwide, using systems capable of processing textual data can help to find trends, people at risk, the risk factors involved and can facilitate the right intervention taken to prevent or reduce the risk of suicide.

6- References

[1] Song J, Song TM, Seo D-C, Jin JH. Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents. Journal of Adolescent Health. 2016;59 (6):668-73.

[2] Metzger MH, Tvardik N, Gicquel Q, Bouvry C, Poulet E, Potinet-Pagliaroli V. Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. International journal of methods in psychiatric research. 2016.

[3] Wicentowski R, Sydes MR. Emotion Detection in Suicide Notes using Maximum Entropy Classification. Biomedical informatics insights. 2012;5(Suppl. 1):51-60.

[4] Cook, Benjamin L., Ana M. Progovac, Pei Chen, Brian Mullin, Sherry Hou, and Enrique Baca-Garcia. "Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid." Computational and mathematical methods in medicine 2016 (2016).

[5] McCart JA, Finch DK, Jarman J, Hickling E, Lind JD, Richardson MR, et al. Using ensemble models to classify the sentiment expressed in suicide notes. Biomedical informatics insights. 2012; 5(Suppl. 1):77-85.

[6] Mościcki EK. Epidemiology of completed and attempted suicide: toward a framework for prevention. Clinical Neuroscience Research. 2001;1 (5):310-23.

[7] Shekelle P, Bagley S, Munjas B. Strategies for Suicide Prevention in Veterans. 2009.

[8] Taylor CL, van Ravesteyn LM, van denBerg MP, Stewart RJ, Howard LM. The prevalence and correlates of self-harm in pregnant women with psychotic disorder and bipolar disorder. Archives of women's mental health. 2016;19 (5):909-15.

[9] Pestian JP, Matykiewicz P, Grupp-Phelan J, Lavanier SA, Combs J, Kowatch R. Using natural language processing to classify suicide notes. AMIA Annual Symposium proceedings AMIA Symposium. 2008:1091.

[10] Quan C, Wang M, Ren F. An unsupervised text mining method for relation extraction from biomedical literature. PloS one. 2014;9 (7):e102039.

[11] Organization WH. Preventing suicide: a global imperative: World Health Organization; 2014.

[12] Olive J, Christianson C, McCary J. Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation: Springer Science & Business Media; 2011.

[13] Tang B, Chen Q, Wang X, Wu Y, Zhang Y, Jiang M, et al., editors. Recognizing Disjoint Clinical Concepts in Clinical Text Using Machine Learning-based Methods. AMIA Annual Symposium Proceedings; 2015: American Medical Informatics Association.

[14] Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide Note Classification Using Natural Language Processing: A Content Analysis. Biomedical informatics insights. 2010;2010 (3):19-28.

[15] Yang H, Willis A, de Roeck A, Nuseibeh B. A hybrid model for automatic emotion recognition in suicide notes. Biomedical informatics insights. 2012;5 (Suppl. 1):17-30.

[16] Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, Wiebe J, et al. Sentiment analysis of suicide notes: A shared task. Biomedical informatics insights. 2012;5 (Suppl. 1):3.

[17] Desmet B, Hoste V. Emotion detection in suicide notes. Expert Systems with Applications. 2013;40 (16):6351-8.

[18] Desmet B, Hoste V. Combining Lexico-semantic Features for Emotion Classification in Suicide Notes. Biomedical informatics insights. 2012;5 (Suppl. 1):125-8.

[19] Sohn S, Torii M, Li D, Wagholikar K, Wu S, Liu H. A hybrid approach to sentiment sentence classification in suicide notes. Biomedical informatics insights. 2012;5 (Suppl. 1):43-50.

[20] Yeh E, Jarrold W, Jordan J. Leveraging psycholinguistic resources and emotional sequence models for suicide note emotion annotation. Biomedical informatics insights. 2012;5 (Suppl. 1):155-63.

[21] Spasic I, Burnap P, Greenwood M, Arribas-Ayllon M. A naive bayes approach to classifying topics in suicide notes. Biomedical informatics insights. 2012;5(Suppl. 1):87-97.

[22] Kovacevic A, Dehghan A, Keane JA, Nenadic G. Topic categorisation of statements in suicide notes with integrated rules and machine learning. Biomedical informatics insights. 2012;5 (Suppl. 1):115-24.

[23] Cherry C, Mohammad SM, de Bruijn B. Binary classifiers and latent sequence models for emotion detection in suicide notes. Biomedical informatics insights. 2012;5 (Suppl. 1):147-54.

[24] Anderson HD, Pace WD, Brandt E, Nielsen RD, Allen RR, Libby AM, et al. Monitoring suicidal patients in primary care using electronic health records. Journal of the American Board of Family Medicine: JABFM. 2015;28 (1):65-71.

[25] Karmen C, Hsiung RC, Wetter T. Screening Internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. Computer methods and programs in biomedicine. 2015;120 (1):27-36.

[26] Woo H, Cho Y, Shim E, Lee K, Song G. Public Trauma after the Sewol Ferry Disaster: The Role of Social Media in Understanding the Public Mood. International journal of environmental research and public health. 2015;12 (9):10974-83.

[27] McCoy TH, Jr., Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. JAMA psychiatry. 2016;73 (10):1064-71.