Evolving Systems (2021) 12:57–68
https://doi.org/10.1007/s12530-020-09340-7

ORIGINAL PAPER

Suicidal tendencies prediction in Greek poetry

Stefanos Nikiforos[1] · Alexandros-Dimitrios Zervopoulos[1] · Evangelos Geramanis[1] · Alexandros Toulakis[1] · Asterios Papamichail[1] · Dimitrios Triantafylloy[1] · Theofanis Tasoulas[1] · Katia-Lida Kermanidis[1]

Received: 30 September 2019 / Accepted: 22 March 2020 / Published online: 4 April 2020

Abstract
Natural language processing (NLP) has been successfully used to predict a writer's tendency of committing suicide, using various text types: suicide notes, micro-blog posts, lyrics and poems. This paper is an extended version of earlier work. We extend our previous work on text mining in Greek Poetry by employing more sophisticated approaches. More specifically we have applied (i) Deep Neural Networks (DNN), (ii) additional morphosyntactic and semantic features based on writers' emotions and Big Five personality traits and (iii) feature selection, for suicide prediction in Greek poetry. We extend previous research to Greek, i.e. a language that has not been tackled before in this setting, using both language-dependent (but easily portable across languages) and language-independent linguistic features in order to represent the poems of 13 Greek poets of the twentieth century. Our results differ significantly from previous literature. In general, our proposed DNN model offers promising results for suicide prediction, despite the fact that this task poses multiple difficulties, especially for a language with limited related research support.

Keywords Natural language processing (NLP) · Suicide · Poetry · Deep learning · Big five · Machine learning

1 Introduction

Text analysis has been applied so far in many research approaches, aiming at identifying mental health illnesses. Such type of analysis has been used to predict various psychological states (Pennebaker and King 1999). Providing the ability to detect such psychological states in early stages, through the means of Natural Language Processing (NLP) techniques is of vital importance, before they lead to an unfortunate conclusion for the individual.

Furthermore, previous research has focused on detecting suicidal tendencies through the writer's language use, as suicide is a leading cause of death all over the world (Bertolote and Fleischmann 2012). More specifically, there have been multiple recent attempts focusing on suicide notes.

1.1 Literature review

Pestian et al. (2010) concluded that NLP could be used to differentiate between genuine and elicited suicide notes, by achieving a higher classification rate than mental health professionals. Many have tried to tackle the problem of emotion detection in suicide notes utilizing various methods, including entropy classification (Wicentowski and Sydes 2012) and latent sequence models (Cherry et al. 2012) in the pursuit of better understanding the suicidal mind.

✉ Stefanos Nikiforos
c13niki@ionio.gr

Alexandros-Dimitrios Zervopoulos
p15zerv@ionio.gr

Evangelos Geramanis
p15gera@ionio.gr

Alexandros Toulakis
p15toul1@ionio.gr

Asterios Papamichail
p15papa1@ionio.gr

Dimitrios Triantafylloy
p15tria@ionio.gr

Theofanis Tasoulas
p15taso@ionio.gr

Katia-Lida Kermanidis
kerman@ionio.gr

[1] Humanistic and Social Informatics Laboratory (HILab), Department of Informatics, Ionian University, 7 Tsirigoti square, 49100 Corfu, Greece

Additionally, through the widespread use of the Internet emerged the phenomenon of micro blogging and social media. This has resulted in an increasingly large proportion of people who use these means daily to express their feelings and opinions. Thus, there has been major interest in examining whether these can be useful in predicting suicidal ideation. Zhang et al. (2015) deduced that such a task is realizable using the Chinese version of the Linguistic Inquiry and Word Count (LIWC), as well as Latent Dirichlet Allocation (LDA), on Chinese micro-blog users' data. Litvinova et al. (2017) developed a mathematical model to predict suicidal tendencies based on Russian Internet texts, employing numerical rather than linguistic features. Burnap et al. (2015) aimed to identify suicide related topics from posts on Twitter by training baseline classifiers, then improving upon them with an ensemble classifier.

It has been proven that suicide rates among artists are significantly higher than rates pertaining to the general population (Jamison 1997; Raeburn 1999; Stack 1997). Lightman et al. (2007) deployed Coh-Metrix and LIWC to contrast textual features of suicidal and non-suicidal songwriters. Mulholland and Quinn (2013) composed a corpus (development, training and test set) consisting of songs from various English lyricists. This corpus was then used to derive lexical, syntactic, semantic class and n-gram features. Then these were input into the Waikato Environment for Knowledge Analysis (Weka) to compare the performance of multiple machine learning algorithms in classifying whether or not the song was written by a suicidal lyricist. SimpleCart was the algorithm with the highest classification rate, with an overall accuracy of 70.6%. Pająk and Trzebiński (2014) used LIWC to analyze Polish poems from six separate poets. Afterwards, ANOVA and logistic regression were used to extract the most prevalent features in identifying suicidal predisposition. They drew conclusions similar to those of previous research.

Perhaps one of the most influential contributions in the text analysis of poets is that of Stirman and Pennebaker (2001), considering their results have been the basis of many different studies, such as Lightman et al. (2007); Mulholland and Quinn (2013); Pająk and Trzebiński (2014). Their methodology consisted of the collection of 300 poems and the study, using LIWC, of linguistic characteristics that could be distinguished between suicidal and non-suicidal poets. Furthermore, they investigated how such characteristics accord with the two most dominant suicide models, namely Durkheim's model (Durkheim 2005), where suicide rate is linked to a society's integration level, and the hopelessness model (Petrie and Brook 1992), where an individual is overcome with negative emotions, such as hopelessness and helplessness, which ultimately drive them to suicide. These characteristics were then examined as to how they varied in different stages of the poets' careers. They concluded that suicide can be predicted by the language use, finding stronger support for the social integration suicide model, and that there's no significant variation over time.

As a basis for comparison, the results of previous studies are presented in this paragraph. Pestian et al. (2010) managed to accurately distinguish between elicited and genuine suicide notes 78% of the time. Litvinova et al. (2017) reached a 71.5% classification rate of deciding whether Russian internet texts were suicidal or not. Mulholland and Quinn (2013) achieved a performance of 70.6% in identifying suicidal tendencies of songwriters through their lyrics. Coppersmith et al. (2018) used Deep Learning to detect suicide risk on social media data, achieving 0.70–0.85 true positive (TP) rate. Nobles et al. (2018) presented a DNN architecture aiming at suicidality identification among young adults using text messages and achieved 70% accuracy and 81% recall. Du et al. (2018) used Convolutional Neural Networks (CNN) on Twitter data, for the detection of psychiatric stressors, as one major cause of suicide, achieving 74% accuracy, 78% precision and F-1 measure of 83%. Shing et al. (2018) created a dataset for studying the assessment of suicide risk via online postings in Reddit and used CNN at F-1 score of 0.42 on their baseline experimentation. Morales et al. (2019) used Deep Learning for suicide risk assessment of social media posts, achieving 52% accuracy and F-1 score of 0.57.

1.2 Aims and scope

The purpose of this study is to train a machine learning model, in order to classify a poem by the poet's tendency for suicide, based on textual and semantic features. In contrast to previous research, the study focuses on Greek poetry of the twentieth century and aims to examine whether preceding results, derived mostly from English language works, can be verified. These include higher use of the first-person active and passive verbs, compared to the second and third person use of such verbs, more frequent use of death and sexual-related words, overall positive emotion words (Stirman and Pennebaker 2001). Language analysis has been used as a tool to understand emotion and thoughts of a person. Tausczik and Pennebaker (2010) have proposed a computerized text analysis method using LIWC program aiming at the detection of the words which provide information about the emotional state and the motivation of a writer. Shing et al. (2018) identified thoughts and feelings (such as lack of hope, sense of agitation or impulsivity, mixed depressive state) as risk factors for suicide. They used empath and depression-based lexical categories and emotion features (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust), as well as syntactic features (proportion of active and passive verbs) in order to predict suicide risk. Morales et al. (2019) used personality features based on Big Five personality traits

(Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism—OCEAN) and tone features including emotion (anger, disgust, fear, joy, sadness) and social (openness, conscientiousness, extroversion, agreeableness, emotional range vs neuroticism) aiming at the assessment of suicide risk with Deep Learning systems.

Based on these studies we annotated the verses containing emotional words or words revealing personality traits and used them as features for suicide prediction in Greek poetry (Goldberg 1990). Overall, our methodology differs from previous efforts, given the lack of available tools for Greek, which resulted in the investigation of verb suffixes as features, as well as measuring in a novel manner the emotion of a poem's lyrics set in the range of [0,2], in hopes of increasing the overall reliability of the corpus' annotation. The algorithms used were for the most part examined in previous research, and have been shown to perform relatively well in these kinds of tasks.

The work described herein is an extension of earlier work (Zervopoulos et al. 2019). In detail, the novel aspects of the current work compared to the previous version include

- the use of deep learning structures, i.e. DNN, in contrast to the shallow learning schemata employed earlier, leading to a significant increase in predictive performance
- the extension of the feature set by including additional morphosyntactic and semantic features based on the writers' emotions and the Big Five Personality Trait model, enabling thus comparative evaluation of the predictive power of the various feature groups and
- the application of feature selection filters, for identifying the most appropriate feature subset for suicide prediction in Greek poetry.

The main purpose of this study was to apply NLP methods and machine learning techniques to Greek language text. Greek language poses certain difficulties, as it is a low resourced language, with lack of NLP tools. Applying sophisticated approaches (DNNs), as well as comparison between morphosyntactic (suffixes) and semantic (Big-5 personality trait model) features were also major challenges. In particular, these methods were applied aiming at suicide tendency detection in Greek poetry. To the authors' knowledge, this is the first time such a study is applied to Greek language poetry.

The rest of this paper is organized as follows. Section 2 describes the collected data the process and the criteria for gathering it, as well as it presents the feature vector that represents each poem, and why the respective values were selected. In Sect. 3, the classification experiments are described in detail, including feature selection, the algorithms used and their results. Section 4 illustrates some of the difficulties faced during the process of classifying a poet's suicidal ideation by their poem, especially in a language with a scarcity of processing tools. Finally, Sect. 5 presents future improvements that could be made and a conclusion of what was accomplished in this study.

2 Resources and proposed methodology

2.1 Data collection

A corpus of 90 poems was constructed, consisting of poems from 7 poets who committed suicide and 6 who did not. The number of poems is equally distributed between the two groups. The number of poems from each poet ranges from 5 to 9. The vast majority of poets are male (with the exception of two female poets in the suicidal group), who lived in approximately the same period of time, i.e. in the early to mid 1900s. The rationale behind this requirement was for the poets to belong to the same phase of the Greek language history, as the writing style of Greek changed significantly during the second half of the twentieth century. Specifically, Katharevousa was abolished as the official Greek language in 1976 and was replaced by Modern Greek.

It was of significant importance that the poets in the suicide group undoubtedly took their own life. Similarly, special attention was paid so that poets in the non-suicide group did not have a history of self-harm or suicide attempts (Table 1). The size of the poems varies between 80 and 300 words, with an average size of 155.35. For each poet, all poems were randomly selected, with no distinction as to the nature of their content, e.g. a poem particularly high in negative emotion was not specifically selected for a poet who committed suicide. The above

Table 1 Composition of corpus

Poet	Poems	Year of death	Suicide
Giorgos Makris	7	1968	Yes
Kostas Karyotakis	5	1928	Yes
Alexis Traianos	5	1980	Yes
Minas Dimakis	5	1980	Yes
Katerina Gogou	8	1993	Yes
Napoleon Lapathiotis	7	1944	Yes
KoraliaAndreadis	8	1976	Yes
George Seferis	9	1971	No
TellosAgras	7	1944	No
Constantine P. Cavafy	7	1933	No
Kostis Palamas	8	1943	No
Nikolaos Kavvadias	7	1975	No
Odysseus Elytis	7	1996	No

criteria, combined with the overall low suicide rates of Greek poets, significantly restricted the size of the corpus.

A Python script, implemented by the authors of this paper, was used in order to remove all the punctuation marks and accentuation from the poems. This script checks each character of a text whether it is accented (has a stress mark) or it is a punctuation mark. Accented characters are replaced (with the corresponding non-accented character), while punctuation marks are removed. For example the string: "ν᾽ ἀφήσουν ἄδεια σώματα κεῖ ποὺ οἱ ψυχὲς δὲν ἄντεχαν ἐκεῖ ποὺ ὁ νοῦς δὲν πρόφταινε καὶ λύγιζαν τὰ γόνατα." was converted to: "ν αφησουν αδεια σωματα κει που οι ψυχες δεν αντεχαν εκει που ο νους δεν προφταινε και λυγιζαν τα γονατα".

2.2 Feature description

2.2.1 Initial feature set

In the initial feature set (Zervopoulos et al. 2019) every poem was represented as a feature-value learning vector. Some features were based on the work of Kao and Jurafsky (2012), as well as Stirman and Pennebaker (2001), whereas others were novel, which were mostly applicable to the Greek language. A large number of features selected represented sums of occurrence of linguistic characteristics of interest throughout the entire poem, which were normalized by the poem's length. These sum values for each poem were then normalized across the entire corpus using the well known normalization transformation of Eq. 1, where v is the initial feature value, v_{min} and v_{max} are the minimum and maximum values of the feature across all poems, and v_{norm} is the resulting normalized value

$$v_{norm} = \frac{v - v_{\min}}{v_{\max} - v_{\min}}. \tag{1}$$

Overall, the number of features in the initial feature set that were explored reached 37 in total.

2.2.1.1 Vocabulary features The vocabulary feature that was chosen was the type to token ratio (TTR), which is used as an indicator of the poem's richness in vocabulary (Kao and Jurafsky 2012).

2.2.1.2 Morphosyntactic features The social integration model suggests that individuals who commit suicide have failed to integrate with society, so they are expected to be more self-centered, which seems to be manifested through the use of more first-person singular words, as opposed to first-person plural words (Stirman and Pennebaker 2001). Based on that observation, two morphosyntactic features were selected: the count of occurrences of (i) first-person singular and (ii) first-person plural verbs (Zervopoulos et al. 2019).

Due to the lack of reliable morphological analysis tools for Greek, especially for the historical phase of the language targeted herein, these counts were obtained using a set of predefined verb suffixes, that constitute person and number morphemes. These bases were also used as features, to examine whether a subset of them were more prevalent in the process of the classification. All in all, these constituted 32 of the 37 features that were examined, 30 of which were the aforementioned suffixes.

2.2.1.3 Semantic class features Suicidal poets are expected to deal more with negative emotion than with positive feelings, according to the hopelessness suicide model (Petrie and Brook 1992). To test this, each verse was assigned a number ranging from 0 to 2 for the positive emotion it expressed, and 0 to 2 for the negative emotion. The range is overall lower than ones used in previous studies, as [− 5, 5] has been used for example in Nielsen (2011). This was done to increase the overall reliability of the manual annotation. These numbers were then summed up for all verses of a poem, resulting in two features: one sum reflecting the positive, and one reflecting the negative emotion.

Sexual and death-related references were also included as features, as they are considered to contribute to identifying suicidal tendencies (Stirman and Pennebaker 2001). For these features, the number of verses for each poem that were deemed to contain sexual or death references were summed up for each poem.

2.2.2 Extended feature set

In addition to the earlier work (Zervopoulos et al. 2019), 21 new attributes were used: 4 morphosyntactic features and 17 semantic class features. The morphosyntactic features were: (1) 1st person active verbs, (2) 2nd or 3rd person active verbs, (3) 1st person passive verbs and (4) 2nd or 3rd person passive verbs. These features were used in previous research aiming at the detection of suicidal tendencies (Shing et al. 2018). These features were represented in sums of occurrence throughout the entire poem, which were normalized by the poem's total number of verbs. These sum values for each poem were then normalized across the entire corpus using the normalization transformation of Eq. 1 described above.

Furthermore, 17 semantic features regarding poets' feelings and personality traits were included. Some of these features (extraversion-introversion-optimism–pessimism-trust-denial- instability-joy-anger-fear-desperation) refer to the Big Five (OCEAN) personality traits model (Goldberg 1990; Morales et al. 2019), while others refer to the emotional state range of the poets (escapism-sadness-disgust-complaint-love-hate). Extraversion and introversion provide

Table 2 Annotation examples

Semantic features	Examples (verses)
Introversion	I put a strong lock in my mouth!
Extraversion	Come on in and drink
Instability	White love doves or the hyenas of the jungle
Desperation	Will I find a pillow to cry on?
Optimism	You shine new Spring, you shine new miracle
Pessimism	Like a dream, immersed in the sea of forgetfulness
Denial	Against the orgasm of consumption
Trust	Since I exist and you exist, despair will end soon
Escapism	Climb up your Calvary and crucify yourself
Joy	Somewhere there is a joy waiting for us…
Sadness	Oh desolate, sad elements of mine
Anger	My secret desire, has its beak ready, to search for blood
Fear	Scared sparrows their eyes
Disgust	(dying) from plague to horrible disintegration
Complaint	Our gods have fooled us
Love	Palace of my love
Hate	(our loneliness) is an ax in our hands turning over your heads

information regarding the openness of the poets, while trust and denial refer to their social attitude. Optimism and pessimism, along with emotional instability (joy-sadness, love-hate), fear, anger, complaint, disgust and desperation are emotional marks of a lack of hope for things to get better, which is associated to suicidal tendencies (Morales et. al. 2019; Zirikly et al. 2019). Escapism is used as a means of relieve persistent feelings and it is associated with depression (Cronkite et al. 1998). Selection of these emotion features aimed at the determination of the poets' psychological state. Poets' feelings revealing lack of hope, depressive sate, isolation could provide tone measures for suicide risk assessment (Morales et al. 2019; Shing et al. 2018; Zirikly et al. 2019). Annotation examples are presented in Table 2. Features were represented in sums of occurrence in each verse throughout the entire poem, which were normalized by the poem's total number of verses. These sum values for each poem were then normalized across the entire corpus using the normalization transformation of Eq. 1 described above.

Each poem was annotated with the aforementioned semantic information by two Greek language native speakers. When a verse led to contrasting annotators' decisions, the feature value was decided upon by the majority vote of the authors of the present work. All features used in this research are presented in Table 3. Morphosyntactic features are language-dependent related to the specific syntax of Greek language. Semantic features are language-independent, not related to the morphosyntactic features.

Various experiments were run using the features described in Table 2. Feature selection was applied aiming at results optimization. As a result, 7 feature sets were created (Table 4).

2.3 Model implementation

Each poem was represented using the aforementioned features. The RapidMiner Studio data science software platform (Hofmann and Klinkenberg 2013) was used for running prediction experiments pertaining to poets' suicidal tendencies. Since the corpus was not divided into different sets for training and testing, k-fold cross validation was employed for evaluating the learning schemata. A variety of tree and rule-based algorithms were tested in our previous work (Zervopoulos et al. 2019). In this extended work, the Deep Learning and the Generalized linear model (GLM) algorithms, both using the implementation of H2O 3.8.2.6 (Candel et al. 2018), were compared as to their performance at accomplishing this task. These are well-established algorithms in dealing with such tasks: GLM was used in Passos et al. (2016), while Deep Learning was used in Coppersmith et al. (2018); Du et al. (2018); Nobles et al. (2018); Shing et al. (2018); Zirikly et al. (2019). The parameters used, which achieved the best results when running these tests, are described below.

Deep Learning H2O is based on a multi-layer feed-forward artificial neural network that is trained with stochastic gradient descent using back-propagation. The network contained two (2) hidden layers consisting of 100 and 50 neurons respectively with the Exprectifier (Exponential Rectifier Linear Unit) activation function. Advanced features, such as adaptive learning rate, rate annealing, momentum training, dropout and L1 or L2 regularization enable high predictive accuracy. Each computed node trains a copy of the global model parameters on its local data with multi-threading (asynchronously), and contributes periodically to the global

Table 3 List of features by type

	Vocabulary	Feature types	
		Morphosyntactic	Semantic
Feature names	TTR	(30) suffixes	Positive emotion
		1st person singular words	Negative emotion
		1st person plural words	Death references
		1st person active verbs	Sexual references
		2nd-3rd person active verbs	Introversion
		1st person passive verbs	Extraversion
		2nd-3rd person passive verbs	Instability
			Desperation
			Optimism
			Pessimism
			Denial
			Trust
			Escapism
			Joy
			Sadness
			Anger
			Fear
			Disgust
			Complaint
			Love
			Hate
Total	1	36	21

Table 4 Description of feature sets

Feature set	No. of features	Selected features
A	7	Positive emotion, negative emotion, death references, sexual references, 1st person singular words, 1st person plural words, TTR
B	30	Suffixes
C	21	1st person active verbs, 2nd-3rd person active verbs, 1st person passive verbs, 2nd-3rd person passive verbs, introversion, extraversion, instability, desperation, optimism, pessimism, denial, trust, escapism, joy, sadness, anger, fear, disgust, complaint, love, hate
D	28	A + C feature sets
E	24	C feature set + positive emotion, negative emotion, TTR
F	26	C feature set + positive emotion, negative emotion, TTR, death references, sexual references
G	58	A + B + C feature sets

model via model averaging across the network. The operator initiates a 1-node local H2O cluster and runs the algorithm on it. Although it uses one node, the execution is parallel.

The Deep Learning H_2O-1 algorithm was run with the following settings: Leave one out cross validation with 2 epochs and the adaptive rate option activated. The implemented adaptive learning rate algorithm (ADADELTA) automatically combines the benefits of learning rate annealing and momentum training to avoid slow convergence. The specification of only two parameters (rho and epsilon) simplifies hyper parameter search. Values used in our experiments were rho = 0.99 and epsilon = 1.0E−8, as also used in previous research (Kamminga et al. 2017; Kim et al. 2018; Achmad et al. 2019). The standardization mode was set to automatically standardize the data, using the cross entropy loss function and the Bernouli distribution function. Deep Learning H2O-2 had the same settings as mentioned above with tenfold cross validation and stratified sampling, containing 2 hidden layers with 200 neurons each. The distribution function for the training data was set to multinomial with Quadratic loss function and 20 epochs.

Deep Learning-1 algorithm used tenfold cross validation with stratified sampling and use of a local random seed. Cross entropy was the selected loss function with

Adam updater, having 100 epochs. The Xavier uniform weight initialization was selected with a stochastic gradient descent optimization method. The network type consisted of a simple neural net with its input dimension set equal to the amount of the initial features, having 10 epochs per log. Neural network also included 3 hidden fully connected layers with 256, 128 and 2 neurons respectively, using the ReLU (Rectified Linear Unit) activation function. Regarding the output layer, the number of neurons was set to 2, because we had a binary classification task. As for the first two layers, the number of neurons was selected according to the trial and error rule (Sheela and Deepa 2013). Deep Learning-2 had the same settings as Deep Learning-1, with a different optimization method (Line Gradient Descent).

GLMs are an extension of traditional linear models. This algorithm fits generalized linear models to the data by maximizing the log-likelihood. The elastic net penalty is used for parameter regularization. The model fitting computation is parallel, extremely fast, and scales extremely well for models with a limited number of predictors with non-zero coefficients. The operator initiates a 1-node local H2O cluster, and runs the algorithm on it. Although it uses one node, the execution is parallel.

The GLM-1 algorithm was used with the following settings: tenfold cross validation with stratified sampling, use of a local random seed and L_BFGS solver. Standardize and use of regularization with lamda search were also selected. GLM-2 had the same settings as above mentioned with 20-fold cross validation. Despite 20-fold cross validation might be indeed too much for such a case, we used this setting in order to present comparative results with the other algorithms.

Table 5 Results on the initial dataset

Algorithm	Confusion matrix			Evaluation parameter	
		Correct labels		Precision	Recall
		FALSE	TRUE		
Deep Learning H2O-1	FALSE	36	10	0.783	0.800
	TRUE	9	35	0.796	0.778
Deep Learning H2O-2	FALSE	28	12	0.700	0.622
	TRUE	17	33	0.660	0.733
GLM-1	FALSE	38	6	0.864	0.845
	TRUE	7	39	0.848	0.867
GLM-2	FALSE	28	11	0.718	0.622
	TRUE	17	34	0.667	0.756
Deep Learning-1	FALSE	39	10	0.796	0.867
	TRUE	6	35	0.854	0.778
Deep Learning-2	FALSE	34	5	0.872	0.756
	TRUE	11	40	0.784	0.889

3 Results

At the beginning of the experiments the new algorithms were applied on the dataset (referred to as Dataset 1 hereinafter) described in Sect. 2.2.1. The results are presented in Table 5.

In our previous work (Zervopoulos et al. 2019), C4.5 was the algorithm that achieved the highest classification rate, reaching 84.5% (F:0.818). Deep Learning-1 reached at 82.2% (F:0.830), while GLM-1 outperformed them reaching at 85.56% (F:0.854).

As it has also been mentioned in (Zervopoulos et al. 2019), suffixes did not seem to form a clear pattern for suicidal tendency recognition. In order to confirm this assessment, we applied the algorithms on feature set A and feature B separately (Table 6).

Our previous assessment is confirmed, as the results from feature set B (suffixes) are very low, while the results for feature set A are remarkably higher (Table 6).

The algorithm that eventually achieved the highest classification rate was Deep Learning H2O-2 with feature set E, reaching 84.4% (F:0.854). Various statistics are presented in Table 7 and Fig. 1 detailing some of the tests done.

The resulting models indicate that the overall classification result is largely based on the semantic type features.

Deep Learning classifiers perform fairly well and result in simple models, which is in part attributed to the manual nature of the annotation process, as a result of the lack of reliable processing tools for the Greek language, as well as to the nature of the semantic features, which are of high-level linguistic knowledge. They have proven to perform fairly well in similar research conducted in the past as well. Our results are significantly better than previous research (mentioned above in Sect. 1.1), as our best algorithm achieved 84.4% accuracy and F-1 score of 0.854. DNN architecture contributes in better predictive results. Compared to previous work (Zervopoulos et al. 2019), recent results are also improved. Despite the fact that the accuracy score is almost the same (84.5%-84.4%), the F-1 score is improved (0.841–0.854).

4 Discussion

The main motivation for our study was to apply suicide tendency detection in Greek language text. Tackling the task of identifying suicidal tendencies in poetry, particularly for a language where no previous research has been conducted, is certain to pose many difficulties. Construction of the corpus was largely difficult due to the strict criteria described in Sect. 2.1. It was hard to determine whether someone did not

Table 6 Results for feature sets A and B

Algorithm	Feature set	Confusion matrix			Evaluation parameter	
			Correct labels		Precision	Recall
			FALSE	TRUE		
Deep Learning-1	A	FALSE	38	9	0.809	0.844
		TRUE	7	36	0.837	0.800
Deep Learning-1	B	FALSE	27	32	0.458	0.600
		TRUE	18	13	0.419	0.289
Deep Learning-2	A	FALSE	38	9	0.809	0.844
		TRUE	7	36	0.837	0.800
Deep Learning-2	B	FALSE	22	23	0.489	0.489
		TRUE	23	22	0.489	0.489
GLM-1	A	FALSE	36	7	0.837	0.800
		TRUE	9	38	0.809	0.844
GLM-1	B	FALSE	24	22	0.522	0.533
		TRUE	21	23	0.523	0.511
GLM-2	A	FALSE	35	6	0.854	0.778
		TRUE	10	39	0.796	0.867
GLM-2	B	FALSE	24	17	0.571	0.622
		TRUE	21	28	0.585	0.533
Deep Learning H2O-1	A	FALSE	36	11	0.766	0.800
		TRUE	9	34	0.791	0.756
Deep Learning H2O-1	B	FALSE	17	20	0.460	0.378
		TRUE	28	25	0.472	0.556
Deep Learning H2O-2	A	FALSE	36	10	0.783	0.800
		TRUE	9	35	0.796	0.778
Deep Learning H2O-2	B	FALSE	17	14	0.548	0.378
		TRUE	28	31	0.525	0.689

at least attempt to commit suicide in their lifetime, as it may not have necessarily become known to people outside their close social circle. Furthermore, in the case where it was ambiguous whether the cause of death was actually suicide, as is the case with Maria Polydouri, the poet was left out, which further shortened the number of candidate poems. Perhaps, one solution to this would have been to focus more on the few poets that have written a lot of poems. This does, however, introduce the risk of the classifier learning the patterns of those particular poets and being unable to accurately classify poems by others.

Additionally, and perhaps most importantly, the process of annotating the corpus was especially demanding, since the most prevalent tools used in previous research were not available for Greek. Such tools include LIWC, the UAM CorpusTool (O'Donnell 2008), various word-lists, such as AFINN and Affective Norms for English Words (ANEW) (Nielsen 2011), used in sentiment analysis tasks, as well as corpora used in previous studies, which could have been used for comparison. This resulted in the manual annotation of a significant portion of the selected features. It was also a major restriction in the selection of the features, as the more sophisticated a feature is, the more rigorous the process of identifying it in the text needs to be. Not adhering to the appropriate level of rigorousness introduces further potential bias into the data, and given the annotators' lack of a professional literary background, this would have been an increasingly precarious task.

Experiments confirmed that the suffixes do not form some clear pattern which could help identify suicidal ideation. Even in the case of the combination with the semantic class features, suffixes affect negatively the predictive results. Semantic class (E and F) feature sets perform better, regardless of the type of the algorithms. In contrast, semantic type features combined with Deep Neural Networks offer the ability for suicidal prediction in poetry. Positive and negative emotion words and TTR contribute in the improvement of the results. On the other hand, 1st person singular and 1st person plural words cannot be considered as essential. Introducing additional features, regarding the range of the emotional state and the personality traits of the poets proved to offer classification improvement.

In conclusion, semantic type (language-independent) features perform better, compared to the language type

Table 7 Performance for both classes for each algorithm

Algorithm	Feature set	Confusion matrix			Evaluation parameter	
			Correct labels		Precision	Recall
			FALSE	TRUE		
Deep Learning-1	C	FALSE	31	13	0.705	0.689
		TRUE	14	32	0.696	0.711
Deep Learning-1	D	FALSE	36	18	0.667	0.800
		TRUE	9	27	0.750	0.600
Deep Learning-1	E	FALSE	35	14	0.714	0.778
		TRUE	10	31	0.756	0.689
Deep Learning-1	F	FALSE	33	13	0.717	0.733
		TRUE	12	32	0.727	0.711
Deep Learning-1	G	FALSE	38	33	0.535	0.844
		TRUE	7	12	0.632	0.267
Deep Learning-1	C	FALSE	32	15	0.681	0.711
		TRUE	13	30	0.698	0.667
Deep Learning-1	D	FALSE	33	12	0.733	0.733
		TRUE	12	33	0.733	0.733
Deep Learning-1	E	FALSE	32	11	0.744	0.711
		TRUE	13	34	0.723	0.756
Deep Learning-1	F	FALSE	34	12	0.739	0.756
		TRUE	11	33	0.750	0.733
Deep Learning-1	G	FALSE	45	45	0.500	1.000
		TRUE	0	0	0.000	0.000
GLM-1	C	FALSE	31	16	0.660	0.689
		TRUE	14	29	0.674	0.644
GLM-1	D	FALSE	37	10	0.787	0.822
		TRUE	8	35	0.814	0.778
GLM-1	E	FALSE	35	13	0.729	0.778
		TRUE	10	32	0.762	0.711
GLM-1	F	FALSE	38	10	0.792	0.844
		TRUE	7	35	0.833	0.778
GLM-1	G	FALSE	30	9	0.769	0.667
		TRUE	15	36	0.706	0.800
GLM-2	C	FALSE	31	16	0.660	0.689
		TRUE	14	29	0.674	0.644
GLM-2	D	FALSE	36	9	0.800	0.800
		TRUE	9	36	0.800	0.800
GLM-2	E	FALSE	37	9	0.804	0.822
		TRUE	8	36	0.818	0.800
GLM-2	F	FALSE	39	10	0.796	0.867
		TRUE	6	35	0.854	0.778
GLM-2	G	FALSE	32	9	0.781	0.711
		TRUE	13	36	0.735	0.800
Deep Learning H2O-1	C	FALSE	25	16	0.610	0.556
		TRUE	20	29	0.592	0.644
Deep Learning H2O-1	D	FALSE	35	13	0.729	0.778
		TRUE	10	32	0.762	0.711
Deep Learning H2O-1	E	FALSE	38	13	0.745	0.844
		TRUE	7	32	0.821	0.711
Deep Learning H2O-1	F	FALSE	33	12	0.733	0.733
		TRUE	12	33	0.733	0.733

Table 7 (continued)

Algorithm	Feature set	Confusion matrix			Evaluation parameter	
			Correct labels		Precision	Recall
			FALSE	TRUE		
Deep Learning H2O-1	G	FALSE	32	14	0.696	0.711
		TRUE	13	31	0.705	0.689
Deep Learning H2O-2	C	FALSE	27	12	0.692	0.600
		TRUE	18	33	0.647	0.733
Deep Learning H2O-2	D	FALSE	40	12	0.769	0.889
		TRUE	5	33	0.868	0.733
Deep Learning H2O-2	E	FALSE	41	10	0.804	0.911
		TRUE	4	35	0.897	0.778
Deep Learning H2O-2	F	FALSE	38	12	0.760	0.844
		TRUE	7	33	0.825	0.733
Deep Learning H2O-2	G	FALSE	32	19	0.628	0.711
		TRUE	13	26	0.667	0.578

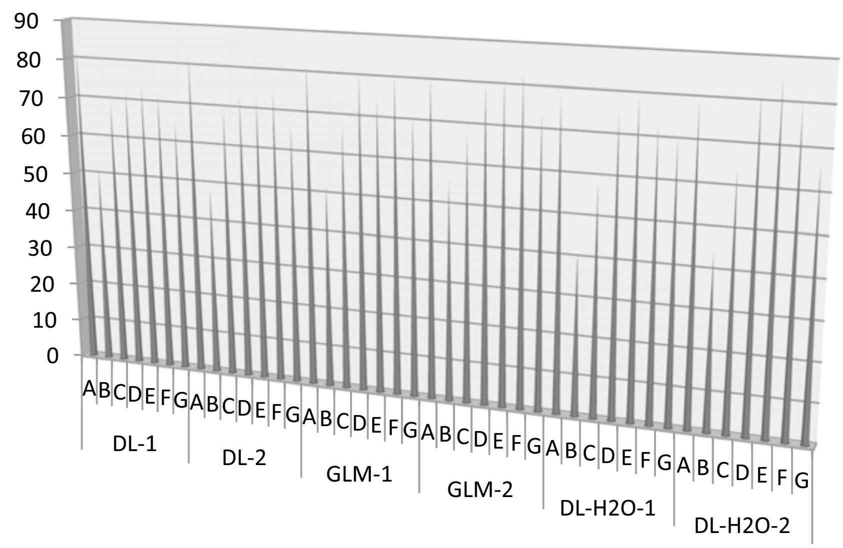

Fig. 1 F-1 scores per feature set for each algorithm

(language dependent) features. Consequently, the suggested model is language-independent. It can be also be applied to other languages, disregarding their specific syntax. For the same reason, it does not depend either on the generation of the poets, or on their writing style.

Algorithms tested achieved high accuracy and strong precision and recall, especially for the TRUE (suicide) class which is the sensitive one. The use of new attributes regarding the emotional state and the personality traits of the poets', based on the Big Five (OCEAN) model, contributed to the creation of a more robust prediction model. This is in line with the work of (Coppersmith et al. 2018). Results also confirm that deep learning models can outperform more traditional machine learning systems for suicide risk assessment (Morales et al. 2019).

5 Conclusions and implications

In this paper we extend our previous work on suicide prediction in Greek poetry (Zervopoulos et al. 2019) by using (i) DNN (ii) additional morphosyntactic and semantic features based on writers' emotions and the Big Five Personality Traits model and (siii) applying feature selection, for accurate suicide tendency prediction in Greek poetry.

In conclusion, a corpus of poems was composed to identify suicidal ideation in Greek poetry of the twentieth century. This proved to be a challenging task, since there has not been any previous work done on the language that was selected. Nonetheless, the resulting classifier, Deep Learning H2O-2, reached an accuracy of 84.5% (F:0.854). Results, compared to previous research have been improved. The features explored were mainly semantic in nature, while also utilizing morphosyntactic features, such as verb suffixes which are language specific and have not been investigated before.

As suicide rates among artists are significantly higher than rates pertaining to the general population, the results of this research cannot be applied to the general population and they cannot be used in psychology in general. They are suitable for poets. Nevertheless, the results are overall promising and the features selected are easily portable across different strategies, which should allow future studies to confirm how successful our methodology was.

The construction of a properly annotated corpus proved to be the most challenging part of this task. Therefore, the development of such a corpus would be worth looking into, as it would greatly aid future efforts in studying NLP related topics for Greek texts. It is of critical importance, as has been showcased by the attempts in this study, to pay special attention when manually annotating data, and to adhere to best practices which have been studied extensively before, for example by Wiebe, Wilson and Cardie (2005).

There's also a scarcity of available tools, which are often required in NLP related tasks, due to the complex structure of the Greek language. For example, tools for lemmatization and stemming are somewhat difficult to implement, partly due to punctuation and grammar rules. This indicates it would be meaningful to spend time developing such tools, before further progress into more intensive NLP tasks is made. Considering the above, any future advancements in the area will be interesting to behold, after this first attempt has been made, to keep track of how such difficulties are handled.

References

Achmad KA, Nugroho LE, Djunaedi AW (2019) Context-aware based restaurant recommender system: a prescriptive analytics. J Eng Sci Technol 14(5):2847–2864

Bertolote JM, Fleischmann A (2012) A global perspective on the magnitude of suicide mortality. In: Oxford University Press (Ed.), Oxford textbook of suicidology and suicide prevention: A global perspective, pp 92–98. 10.1093/med/9780198570059.003.0014

Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J (2015) Machine classification and analysis of suicide-related communication on twitter. In: ACM (Ed.), Proceedings of the 26th ACM conference on hypertext & social media. ACM, pp 75–84

Candel A, Parmar V, LeDell E, Arora A (2018) Deep learning with H_2O. Retrieved from https://h2o.ai/resources

Cherry C, Mohammad SM, De Bruijn B (2012) Binary classifiers and latent sequence models for emotion detection in suicide notes. Biomed Inform Insights. https://doi.org/10.4137/bii.s8933**(5s1, BII.S8933)**

Coppersmith G, Leary R, Crutchley P, Fine A (2018) Natural language processing of social media as screening for suicide risk. Biomed Inform Insights. https://doi.org/10.1177/1178222618792860

Cronkite RC, Moos RH, Twohey J, Cohen C, Swindle R (1998) Life circumstances and personal resources as predictors of the ten-year course of depression. Am J Community Psychol 26(2):255–280. https://doi.org/10.1023/A:1022180603266

Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, Xu H (2018) Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med Inform Decis Mak. https://doi.org/10.1186/s12911-018-0632-8

Durkheim É (2005) Suicide: a study in sociology (Routledge, ed.). 10.4324/9780203994320

Goldberg LR (1990) An alternative 'Description of Personality': the big-five factor structure. J Personal Soc Psychol 59(6):1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Hofmann M, Klinkenberg R (eds) (2013) RapidMiner: data mining use cases and business analytics applications. CRC Press, Boca Raton

Jamison KR (1997) Manic-depressive illness and creativity. Sci Am 276:44–52

Kamminga JW, Bisby HC, Le DV, Meratnia N, Havinga PJ (2017) Generic online animal activity recognition on collar tags. In: UbiComp/ISWC 2017—Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 597–606. https://doi.org/10.1145/3123024.3124407

Kao J, Jurafsky D (2012) A computational analysis of style, affect, and imageryin contemporary poetry. In: 1st workshop on computational linguistics for literature (CLfL 2012), 8–17.

Kim S, Shim S, Kim J, Jung J, Kim SA (2018) A Study on performance prediction in early design stage of apartment housing using machine learning. Int J Archit Environ Eng 11(9):1343–1351

Lightman EJ, McCarthy PM, Dufty DF, McNamara DS (2007) Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal song writers. In: McNamara JG, Trafton DS (Ed.) Proceedings of the annual meeting of the cognitive science society. Cognitive Science Society, Austin, 1217–1222

Litvinova TA, Seredin PV, Litvinova OA, Romanchenko OV (2017) Identification of suicidal tendencies of individuals based on the quantitative analysis of their internet texts. Computación y Sistemas 21(2):243–252. https://doi.org/10.13053/CyS-21-2-2721

Morales M, Dey P, Theisen T, Belitz D, Chernova N (2019) An Investigation of Deep Learning Systems for Suicide Risk Assessment. In: Association for computational linguistics (ACL) (Ed.), Proceedings of the sixth workshop on computational linguistics and clinical psychology, pp 177–181. https://doi.org/10.18653/v1/w19-3023

Mulholland M, Quinn J (2013) Suicidal tendencies: the automatic classification of suicidal and non-suicidal lyricists using NLP. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP), 680–684

Nielsen FÅ (2011) A new ANEW: evaluation of a word list for sentimentanalysis in microblogs. CEUR Workshop Proc 718:93–98

Nobles AL, Glenn JJ, Kowsari K, Teachman BA, Barnes LE (2018) Identification of imminent suicide risk among young adults using text messages. In: ACM (Ed.), Proceedings of the 2018 CHI conference on human factors in computing systems (p. Paper 413, 11 pages). https://doi.org/10.1145/3173574.3173987

O'Donnell, M. (2008). Demonstration of the UAM CorpusTool for text and image annotation. In: Proceedings of the ACL-08: HLT Demo Session, 13–16. https://doi.org/10.3115/1564144.1564148

Pajak K, Trzebiński J (2014) Escaping the world: linguistic indicators of suicide attempts in poets. J Loss Trauma 19(5):389–402. https://doi.org/10.1080/15325024.2013.794663

Passos IC, Mwangi B, Cao B, Hamilton JE, Wu MJ, Zhang XY, Soares JC (2016) Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. J Affect Disord 193:109–116. https://doi.org/10.1016/j.jad.2015.12.066

Pennebaker JW, King LA (1999) Linguistic styles: language use as anindividual difference. J Personal Soc Psychol 77(6):1296. https://doi.org/10.1037/0022-3514.77.6.1296

Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A (2010) Suicide note classification using natural language processing: a content analysis. Biomed Inform Insights. https://doi.org/10.4137/bii.s4706**(3, BII.S4706)**

Petrie K, Brook R (1992) Sense of coherence, self-esteem, depression and hopelessness as correlates of reattempting suicide. Br J Clin Psychol 31(3):293–300. https://doi.org/10.1111/j.2044-8260.1992.tb00996.x

Raeburn SD (1999) Psychological issues and treatment strategies in popular musicians: a review, part i. Med Probl Perform Artist 14:171–179

Sheela KG, Deepa SN (2013) Review on methods to fix number of hidden neurons in neural networks. Math Probl Eng. https://doi.org/10.1155/2013/425740

Shing H-C, Nair S, Zirikly A, Friedenberg M, Daumé III H, Resnik P (2018) Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Association for Computational Linguistics (ACL) (Ed.), proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic, pp 25–36. https://doi.org/10.18653/v1/w18-0603

Stack S (1997) Suicide among artists. The Journal of Social Psychology 137(1):129–130. https://doi.org/10.1080/00224549709595421

Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 29(1):24–54. https://doi.org/10.1177/0261927X09351676

Wicentowski R, Sydes MR (2012) Emotion detection in suicide notes using maximum entropy classification. Biomed Inform Insights. https://doi.org/10.4137/bii.s8972**(5s1, BII–S8972)**

Wiebe J, Wilson T, Cardie C (2005) Annotating expressions of opinions andemotions in language. Lang Resour Eval 39(2–3):165–210. https://doi.org/10.1007/s10579-005-7880-9

Wiltsey Stirman S, Pennebaker JW (2001) Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 63(4):517–522. https://doi.org/10.1097/00006842-200107000-00001

Zervopoulos AD, Geramanis E, Toulakis A, Papamichail A, Triantafylloy D, Tasoulas T, Kermanidis K (2019) Language processing for predicting suicidal tendencies: a case study in Greek poetry. IFIP Adv Inf Commun Technol 560:173–183. https://doi.org/10.1007/978-3-030-19909-8_15

Zhang L, Huang X, Liu T, Li A, Chen Z, Zhu T (2015) Using linguistic features to estimate suicide probability of chinese microblog users. In: Springer Verlag (Ed.), Lecture notes in computer science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 8944, pp 549–559. https://doi.org/10.1007/978-3-319-15554-8_45

Zirikly A, Resnik P, Uzuner Ö, Hollingshead K (2019) CLPsych 2019 shared task: predicting the degree of suicide risk in reddit posts. In: Association for Computational Linguistics (Ed.), proceedings of the sixth workshop on computational linguistics and clinical psychology, pp 24–33. https://doi.org/10.18653/v1/W19-3003