

Analysis of Parkinson's Disease using Deep Learning and Word Embedding Models

^{1,2}Feyza Çevik and ²Zeynep Hilal Kilimci*

¹IQBender Technology Company, Istanbul, Turkey

²Faculty of Engineering, Department of Computer Engineering Dogus University, Turkey

¹feYZa.cevik@iqbender.com, ²hkilimci@dogus.edu.tr

Abstract

Parkinson's disease is a common neurodegenerative neurological disorder, which affects the patient's quality of life, has significant social and economic effects, and is difficult to diagnose early due to the gradual appearance of symptoms. Examining the discussion of Parkinson's disease in social media platforms such as Twitter provides a platform where patients communicate each other in both diagnosis and treatment stage of the Parkinson's disease. The purpose of this work is to evaluate and compare the sentiment analysis of people about Parkinson's disease by using deep learning and word embedding models. To the best of our knowledge, this is the very first study to analyze Parkinson's disease from social media by using word embedding models and deep learning algorithms. In this study, Word2Vec, GloVe, and FastText are employed as word embedding models for the purpose of enriching tweets in terms of semantic, context, and syntax. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory Networks (LSTMs) are implemented for the classification task. This study demonstrates the efficiency of using word embedding models and deep learning algorithms to understand the needs of patients' and provide a valuable contribution to the treatment process by analyzing sentiments of them with 93.63% accuracy performance.

Key words: Parkinson's disease, sentiment analysis, deep learning, long short-term memory networks, word embedding models.

1. Introduction

With the advancement of internet and mobile technology in recent years, social media platforms have grown rapidly. It has become an important source for informing patients and their relatives about the disease process, sharing their opinions and experiences and obtaining detailed information about treatment options [1,2]. Twitter is the most popular social networking service known to share information as it is, and to connect with others in real time. With around 350 million active Twitter users per month, Twitter is a powerful tool for gathering information about the real world. Twitter is a powerful health data source and allow monitoring epidemics, understanding patients' experiences and sentiments [3-8].

The application of deep learning (DL) algorithms and word embedding (WE) models are very popular in recent years in different research fields such as image processing, natural language processing, speech recognition, and machine translation. DL methods and WE models are preferred by the researchers because of representing better predictions and results when compared with traditional machine learning algorithms. Deep learning models are mainly used to provide automatic feature extraction by training complex features with minimal external support to obtain meaningful representation of data through deep neural networks. Furthermore, deep learning methods are also employed for the purpose of classification tasks in many fields. Convolutional

*Corresponding author: Address: Faculty of Engineering, Department of Computer Engineering Dogus University, 34722, Istanbul, TURKEY. E-mail address: hkilimci@dogus.edu.tr, Phone: 4447997-1232

neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs) [15], deep belief networks (DBNs) as deep learning models and Word2Vec [11, 12], GloVe [13], FastText [14] as word embedding models are well-known architectures.

In this work, we propose to facilitate the development of a personal care plan for Parkinson's patients and provide a valuable medical resource for personal health information to a growing population of Parkinson's patients by using sentiment analysis of individual and organizational users. For this purpose, three different word embedding models namely, Word2Vec, GloVe, and FastText are utilized to enrich the meanings and context of about Parkinsons' disease related tweets. Furthermore, deep learning models are employed to interpret the tweets of individual and organizational users and understand perception on Parkinsons' disease by classifying tweets such as positive, negative, and neutral. To the best of our knowledge, this is the very first attempt to analyze sentiments for Parkinson's disease by using deep learning and word embedding based models. Experiment results demonstrate that the combination of RU+RH and LSTM outperforms others with 93.63% accuracy result. The rest of the article is organized as follows: Section 2 presents a summary of studies on Parkinson's disease. Section 3 presents materials and methods including word embedding models and deep learning algorithms. The experiment results, discussion, and conclusions are given in Sections 4, and 5.

2. Related Work

This section provides a brief summary of the literature review of studies on Parkinson's Disease (PD).

In [16], the Probabilistic Neural Network (PNN) classifier is employed for the diagnosis of Parkinson's disease and achieved a test success rate of 80.92%. In another study [17], patients and normal individuals are classified on the Parkinson's dataset which consists of Parkinson's patients (42 subjects) and contains approximately 200 voice recordings per patient. Sakar and Kursun [18] use the attributes picked up by the conventional bootstrapping or leave-one-out validation method and a mutually knowledge-based attribute selection method (maximum relevance minimum redundancy - MRMR). In another study [19], has applied four independent classification models namely, neural networks, data mining neural networking model (DMneural), regression and decision tree to differentiate Parkinson's patients. The neural network-based classifier outperforms other with 92.9% accuracy result. Caglar et al. [20] employ the linguistic strong adaptive nerve-fuzzy classifier (ANFC + LH) on Parkinson's disease dataset samples including human biomedical voice for both classification task and feature selection. They report that the utilization of ANFC + LH exhibits 94.72 classification performance.

In another study [21], classification of Parkinson's disease is proposed using feature weighting method on the basis of fuzzy C-means clustering. Luukka [22] achieves 85.03% classification performance through similarity classifier (SC) and fuzzy entropy (FE) feature selection methods by using four medical data sets, dermatology, Pima-Indian diabetes, breast cancer and Parkinson's data set. In [23], Kihel and Benyettou concentrate on the Parkinson's disease recognition using artificial immune system (AIS) and clone attribute selection approaches. In [24], Eskidere focuses on the comparison of feature selection methods for diagnosis of Parkinson's disease from vocal

measurements. In [25], authors propose novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning models by using logistic regression (OLR), support vector machine (SVM), Adaptive Boosting (AdaBoost), and Random Under sampling Boosting (RUSBoost) based classifiers. In [31], authors focus on the early detection of Parkinson's disease through patient questionnaire and predictive modelling.

Our work differs from the above mentioned literature studies in that this is the very first attempt to determine the perception and sentiment analysis of Parkinson's disease from social media employing deep learning algorithms and word embedding models. Unlike the literature studies, we propose to facilitate the development of a personal care plan for Parkinson's patients and provide a valuable medical resource for personal health information to a growing population of Parkinson's patients by employing textual data from social media.

3. Materials and Methods

A summary of the methods, materials, and proposed framework are presented in this section.

3.1. Data Collection and Proposed Framework

In this study, we concentrate on the detection of Parkinson's disease by evaluating sentiment analysis of Parkinson's disease related tweets. For this purpose, Selenium crawler [26] is employed to gather tweets related Parkinson's disease which are composed of these keywords: "Parkinson", "ParkinsonDisease", "ParkinsonsCure", "ParkinsonsTreatment", "ParkinsonDiagnosis", "ParkinsonSymptom". There are two main user accounts, namely individual and organizational in Twitter. Individual and organizational user comments, called tweets in Twitter, are collected to understand and interpret the opinions and experiences of Parkinson's patients and their relatives about Parkinson's disease. While individual accounts can be belonged to both Parkinson's patients and their relatives, organization accounts can include neurologists, news sources and non-patients. In this study, individual accounts with public tweets are analyzed because of the protected tweets of some individual accounts. Totally, 11,043 tweets are downloaded from both individual and organizational accounts. By using Selenium, we pick up as many tweets as we like, without worrying about the limit issue allowed by the Twitter API. In this work, we focus on the supervised machine learning strategy. Because of this, there is a need to label as positive, negative, or neutral each tweet posted from users to determine the attitude of each user to Parkinsons' disease by using TextBlob [27]. TextBlob employs naive Bayes classifier to determine the sentiment and generates the class probability as positive, negative, or neutral of each tweet. It exhibits approximately 79.13% average classification success for determining the sense of user comments. The raw dataset gathered from each user is fairly dirty in social media platforms. Because of this reason, there is need to apply different preprocessing techniques. In this work, stop-word elimination, removing hashtags, removing URLs, and stemming techniques are implemented.

It is a fact that users cannot express their feelings sufficiently due to Twitter's character limitation. To eliminate this issue, we concentrate on the word embedding models such as Word2Vec, GloVe, and FastText. In this way, each comment is enriched in terms of meaning, context and syntax by using word embedding models. Through these methods, the limit trouble in expressing ideas on

Twitter is not a problem for understanding user feelings. After that, instead of using conventional machine learning algorithms, three different deep learning architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) are employed for the classification purpose. The flowchart of proposed system is given in Figure 1.

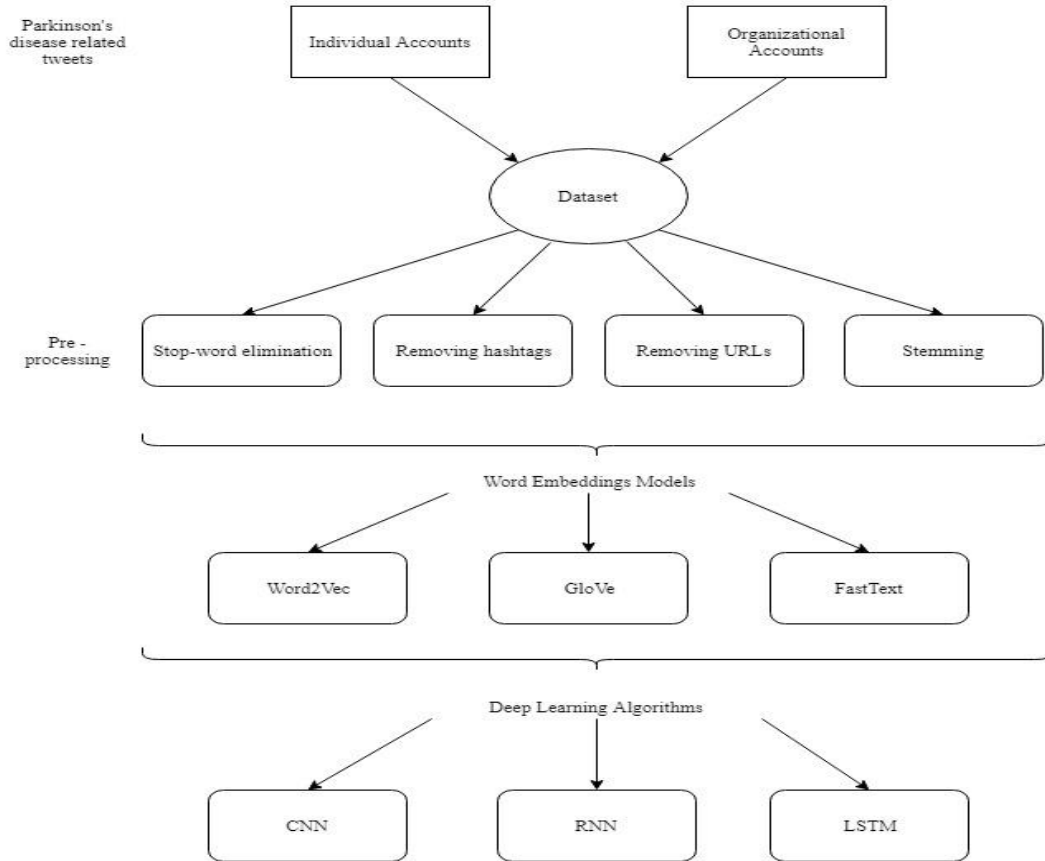


Figure 1. Flowchart of the proposed system

3.2. Word Embedding Models

Word2Vec is a two-tier neural network that processes text. Although Word2Vec is not a deep neural network, it converts text into a numeric form that deep networks can understand. Mikolov et al. [11,12] propose two different models namely, Skip-gram and the Continuous Bag of Word (CBOW) models. Skip-gram models are to guess the words around a word in a sentence or document [15]. On the other hand, CBOW model estimates the target word w_T , from the surrounding words. In other words, CBOW learns the target word w_t by training the words n surrounding the target word w_t .

GloVe, Global Vectors for Word Representation, is another word embedding model proposed in [14]. Statistics are based on unchecked algorithms. Models such as Skip-gram and CBOW capture

semantic information, but do not use statistics for interoperability. Although matrix decomposition methods use these statistics, they cannot capture semantic relationships. These models have no meaning. This word embedding model, proposed by Pennington et al. [14], aims to solve this problem by creating a new purpose function using probability statistics. GloVe method first contracts a word co-occurrence matrix X . Each element of X_{ij} shows how many times the word i appears in the context word j .

FastText is another method of word embedding model that is an extension of the Word2Vec model. Instead of learning vectors for direct words, FastText displays each word as n-gram characters. For example, if we select the word “artificial” with $n=3$; where FastText stages of this word <ar, art, rti, tif, ifi, fic, ici, ial, al>, beginning and end of the word.

3.3. Deep Learning Algorithms

In this work, we focus on the widely-used three deep learning algorithms namely, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM).

Convolutional neural networks (CNNs) are a type of multilayer perceptron (MLP). There are advantages with usage of CNN structure as mentioned below [32-35]:

- The layers go deeper and deeper.
- Calculation performance is improved (ReLU, dropout, batch normalization).
- As the connections between the network layers increase, the backpropagation algorithm is improved. Geoffrey Hinton is known to popularize the backpropagation algorithm in the 80s. This algorithm is currently used in almost all of the deep learning applications.

Recurrent neural network (RNN) is a type of neural network in which the output from the previous step is fed as input to the current step. RNN is proposed due to the requirement of remembering the words. This problem is solved with the help of hidden layer [15, 28, 29]. The most important feature of RNN is the secret state, which remembers some information about a sequence. RNN has a “memory” that remembers all information about what is calculated. RNN, unlike other neural networks, reduces the complexity of the parameters. Performs the same task on all inputs or hidden layers to produce the output. Using the same parameters for each input reduces the complexity of the parameters.

Long Short-Term Memory networks – usually just called “LSTMs” - are a special kind of RNN, capable of learning long-term dependencies. The starting point is to provide a solution to the problem of exponential growth of errors using the backpropagation algorithm while training deep neural networks. The main reason for this problem is that the values generated by the activation function are constantly in the range of -1, 1, so that these values are given to the backpropagation algorithm and multiplied by multiplying to zero. LSTM, which is developed in order to avoid this problem and to design better learning algorithms for complex structures, gives good results in problems where long-term dependencies and long-term information should be remembered. In this study, it is used to learn the meaning of the words used previously and to produce guesses based on these meanings [15, 36].

4. Results

In this study, the extensive experiments are carried out to analyze the sentiments of Parkinson's disease related user comments using word embedding models and deep learning algorithms. Accuracy as an evaluation metric is employed in the experiments in order to demonstrate the classification performances of each model and the contribution of our work. We perform experiments by varying the training set sizes as 1%, 5%, 10%, 30%, 50%, 80%, and 90% with repeated holdout method. The holdout is applied 10 times on the dataset. These percentages are displayed with "ts" prefix to prevent confusion with accuracy percentages. This approach is similar to the previous literature studies [37, 38] where they employ 80% of data for training data and 20% for the test. We use pre-trained vectors with the Word2Vec Skip-gram model [12, 15, 39] from Google News dataset to represent documents with 300 dimensions. Abbreviations are employed for the preprocessing methods, word embedding models, and deep learning algorithms as follows: SWE: Stop-word elimination, RH: Removing hashtags, RU: Removing URLs, STM: Stemming, AOT: All of them, CNN: Convolutional neural network, RNN: Recurrent neural network, LSTM: Long short-term memory network. The best accuracy results are acquired is indicated with boldface. At first, we analyze the classification performances of word embedding models and deep learning algorithms in terms of training set percentages to compare them with each other as observed in Table 1 and 2.

Table 1. The classification accuracies of word embedding models in terms of training set percentage.

Ts	Word2Vec	GloVe	FastText
90	75.20	71.49	71.90
80	76.64	73.46	74.62
50	78.56	75.42	75.33
30	63.52	50.72	51.12
10	58.99	47.89	39.73
5	57.83	42.83	35.87
1	52.61	34.32	19.12

In Table 1 and 2, the impact of training set sizes is investigated on the word embedding models and deep learning methodologies, respectively. In Table 1, it is clearly observed that the best classification accuracies are obtained at 50% training set percentage for all word embedding models covering Word2Vec, GloVe, and FastText. From ts50 to ts90 for all word embedding models, there is an approximately between 2% and 4% decrement in classification success. At the lower training set sizes, the decrease in classification performance is sharply observed compared to the interval of ts50 and ts90. For this reason, 50% training set percentage is set as a baseline training set size in the experiments. From a wider perspective, Word2Vec exhibits superior classification performance with 78.56% of accuracy compared to the others at ts50. It is followed by GloVe with 75.42% of accuracy and FastText with 75.33 accuracy result. It is important to emphasize that GloVe and FastText demonstrate close each other in terms of classification performance but they are competitive compared to Word2Vec. As a result of Table 1, the utilization of Word2Vec takes an advantage for the proposed system in terms of classification success.

In Table 2, the classification accuracies of deep learning algorithms are evaluated in terms of training set percentage. LSTM is best performing deep learning model at all training set levels.

CNN has a better success of 1% compared to the RNN, however both of them also demonstrate classification performances very close to each other. Because of slightly difference in classification success between CNN and RNN, it is clearly seen that one of them is not better than the other deep learning algorithm. Moreover, LSTM outperforms other deep learning models with a better accuracy of 3% - 4% at ts50. At the other training set percentages, the difference in accuracy between LSTM and RNN varies from nearly 2% to 10% while the LSTM exhibits minimum 1% and maximum 12% enhancement compared to CNN in terms of accuracy percentage. Similar to word embedding models, the highest accuracy rate is obtained at ts50 for all deep learning techniques. As a summary, the classification success of deep learning algorithms is ordered as: LSTM> RNN> CNN at all training set sizes. As a consequence of Table 1 and Table 2, 50% training set size is adjusted as a baseline in all experiments because of the observed superior classification performance at ts50.

Table 2. The classification accuracies of deep learning algorithms in terms of training set percentage.

Ts	CNN	RNN	LSTM
90	75.54	76.47	79.46
80	77.98	78.59	80.68
50	78.39	79.36	82.33
30	74.21	74.85	78.34
10	72.46	73.92	75.20
5	56.53	58.68	68.44
1	44.87	42.34	45.89

Secondly, after determining the best training/test split in terms of accuracy results, we investigate the impact of preprocessing methods on both word embedding models and deep learning algorithms as given in Table 3 and 4, respectively.

Table 3. The impact of preprocessing methods on word embedding models at 50% training set.

Preprocessing methods	Word2Vec	GloVe	FastText
SWE	81.23	80.87	78.98
RH	85.75	81.51	80.72
RU	87.67	82.25	81.86
STM	78.36	76.79	75.69
AOT	82.48	80.59	78.83

In Table 3, the effect of preprocessing methods on word embedding models at 50% training set is demonstrated as above. RU exhibits superior classification performance with 87.67% of accuracy compared to the RH, which is competitive with the 85.75% accuracy success, by boosting the success of the proposed system. It is followed by AOT with 82.48%, SWE with 81.23%, and STM with 78.36%. Among the word embedding models, Word2Vec remarkably demonstrates classification success with 87.67% of accuracy when the preprocessing method is adjusted as RU. Moreover, Word2Vec outperforms other word embedding models while FastText has the poorest classification performance with 81.86% among word embedding models. Furthermore, Word2Vec maintains approximately 6% improvement considering the success of FastText while GloVe provides nearly 1% enhancement compared to the classification performance of FastText.

In Table 4, RU and RH are generally the best two preprocessing models when all deep learning

algorithms are considered. The combination of LSTM and RU demonstrate remarkable classification success with 92.50% while the impact of RU on CNN is evaluated by 83.13% accuracy result. In addition, LSTM provide roughly 6% contribution in classification compared to the RNN. As a result, it is obviously observed that LSTM demonstrates the superior classification performance among deep learning classifiers for all preprocessing methods. The classification success for each deep learning model is ordered as: LSTM > RNN > CNN. RU with 92.50% of accuracy result boosts the classification success of the proposed system when LSTM is set as a deep learning model. The classification success of system is changeable when the combination of LSTM and preprocessing models. In addition to the superior success of RU, STM demonstrates the worst performance in classification. Surprisingly, AOT is the second-best preprocessing model and competitive with 85.56% when the combination of RNN and RU is considered. The outcome of Table 4, the sequence of classification accuracies of preprocessing models is generally as: RU > RH > AOT > SWE > STM. In this study, the consolidation of RU and RH is chosen as preprocessing models of the proposed system because of the outstanding impact to the classification.

Table 4. The impact of preprocessing methods on deep learning algorithms at 50% training set.

Preprocessing methods	CNN	RNN	LSTM
SWE	80.25	81.75	83.81
RH	81.56	84.05	90.57
RU	83.13	86.82	92.50
STM	78.31	78.75	79.64
AOT	80.83	85.56	86.39

As a third step, we analyze the best classification results by varying training set percentages from 1 to 90 and employing RH+RU as preprocessing methods as seen in Figure 2. The classification performances of each word embedding and deep learning model are investigated in Figure 2.

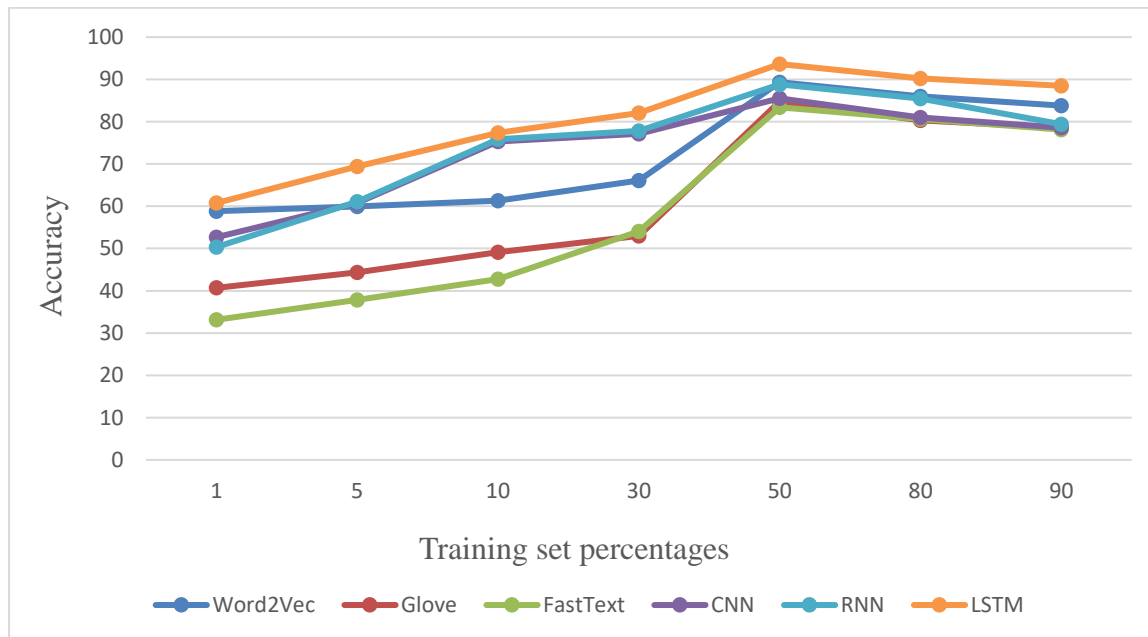


Figure 2. The classification performances of each word embedding and deep learning model in terms of training set percentages when RH+RU is set as a preprocessing method.

From ts1 to ts50, a continuous increase is observed for all word embedding and deep learning models. It is clearly observed that LSTM is the best model with 93.63% of accuracy for determining the sentiment analysis of Parkinson's disease related tweets while FastText exhibits the poor classification performance. At higher training set levels, the accuracy results of each model decreases. Surprisingly, the classification successes of RNN and CNN is very similar to each other in both ts90 and ts30. This means the growth of the training set cannot improve the classification performance in the same way as expected. As a result of Figure 2, the usage of combination of RU+RH at the preprocessing stage provides approximately 2% enhancement at ts50 compared to Table 1 and Table 2.

5. Discussion and Conclusions

In this study, we concentrate on the detection of Parkinson's disease by analyzing the sentiment of the users from social media platforms unlike the recent studies on Parkinson's disease. To understand and analyze of the opinions of the users on Parkinson's disease, deep learning approaches and word embedding models are evaluated. For this purpose, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) are employed as deep learning algorithms and Word2Vec, Glove and FastText are evaluated as word embedding models. Moreover, we also focus on the stop-word elimination (SWE), removing hashtags (RH), removing URLs (RU), stemming (STM) as preprocessing methods to boost the classification performance of the proposed model. In addition, the impact of training set sizes is also investigated to the classification performances. As a result, the use of RU+RH combination in the preprocessing stage by blending with LSTM performs the best classification success at ts50 to determine the sentiment of the users on Parkinson's disease.

To the best of our knowledge, this is the first study to analyze the perception of Parkinson's disease using word embedding models and deep learning algorithms. In this way, the sentiment analysis of users on Parkinson's disease can provide a new perspective for Parkinson's patients and their relatives in the process of both early diagnosis and treatment. Furthermore, this study demonstrates that the analysis of sentiments from tweets about Parkinson's disease represents a valuable source of health care that reveals what needs in the care process of Parkinson patients and how Parkinson's patients and their relatives communicate with each other about their health care problems. As a future work, we plan to improve a hybrid model includes both textual and vocal data recording data by empowering the early detection of Parkinson's disease.

References

- [1] I. De Martino, R. D'Apolito, A.S. McLawhorn, K.A. Fehring, P.K. Sculco, G. Gasparini, Social media for patients: benefits and drawbacks, *Curr. Rev. Musculoskelet. Med.* 10 (2017) 141–145.
- [2] P. Eckler, G. Worsowicz, J.W. Rayburn, Social media and health care: an overview, *Pm R.* 2 (2010) 1046–1050.
- [3] V.M. Prieto, S. Matos, M. Álvarez, F. CACHEDA, J.L. Oliveira, Twitter: a good place to detect

- health conditions, PloS One 9 (2014) e86191, , <https://doi.org/10.1371/journal.pone.0086191>.
- [4] L. Sinnenberg, A.M. Buitteheim, K. Padrez, C. Mancheno, L. Ungar, R.M. Merchant, Twitter as a tool for health research: a systematic review, *Am. J. Public Health* 107 (2017) e1–e8
- [5] B.L. Neiger, R. Thackeray, S.H. Burton, C.R. Thackeray, J.H. Reese, Use of twitter among local health departments: an analysis of information sharing, engagement, and action, *J. Med. Internet Res.* 15 (2013) e177, , <https://doi.org/10.2196/jmir.2775>.
- [6] C. Chew, G. Eysenbach, Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak, *PloS One* 5 (2010) e14118, , <https://doi.org/10.1371/journal.pone.0014118>.
- [7] A. Beykikhoshk, O. Arandjelović, D. Phung, S. Venkatesh, T. Caelli, Data-Mining Twitter and the Autism Spectrum Disorder: A Pilot Study, (2014), pp. 1–8.
- [8] A. Beykikhoshk, O. Arandjelović, D. Phung, S. Venkatesh, Overcoming Data Scarcity of Twitter: Using Tweets as Bootstrap with Application to Autism-Related Topic Content Analysis, ACM, New York, New York, USA, 2015, <https://doi.org/10.1145/2808797.2808908>.
- [9] S. Fahn, Description of Parkinson's disease as a clinical syndrome, *Ann N Y Acad Sci*, 991 (2003) 1-14.
- [10] M.C. de Rijk, C. Tzourio, M.M. Breteler, J.F. Dartigues, L. Amaducci, S. Lopez-Pousa, J.M. Manubens-Bertran, A. Alperovitch, W.A. Rocca, Prevalence of parkinsonism and Parkinson's disease in Europe: the EUROPARKINSON Collaborative Study. European Community Concerted Action on the Epidemiology of Parkinson's disease, *J Neurol Neurosurg Psychiatry*, 62 (1997) 10-15.
- [11] Q. V. Le and T. Mikolov, “Distributed Representations of Sentences and Documents,” vol. 32, 2014.
- [12] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- [13] J. Pennington, R. Socher, and C. Manning, “Glove: Global Vectors for Word Representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.
- [14] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of Tricks for Efficient Text Classification,” 2016.
- [15] Kilimci, Z. H., & Akyokus, S. (2018). Deep Learning-and Word Embedding-Based Heterogeneous Classifier Ensembles for Text Classification. *Complexity*, 2018.
- [16] Ene, Marius. "Neural network-based approach to discriminate healthy people from those with Parkinson's disease." *Annals of the University of Craiova-Mathematics and*

Computer Science Series 35 (2008): 112-116.

[17] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE transactions on bio-medical engineering*, 56(4), 1015.

[18] Sakar, C. Okan, and Olcay Kursun. "Telediagnosis of Parkinson's disease using measurements of dysphonia." *Journal of medical systems* 34.4 (2010): 591-599.

[19] Das, Resul. "A comparison of multiple classification methods for diagnosis of Parkinson disease." *Expert Systems with Applications* 37.2 (2010): 1568-1572.

[20] Caglar, Mehmet Fatih, Bayram Cetisli, and Inayet Burcu Toprak. "Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier." *Journal of Engineering Science and Design* 1.2 (2010): 59-64.

[21] Polat, Kemal. "Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering." *International Journal of Systems Science* 43.4 (2012): 597-609.

[22] Luukka, Pasi. "Feature selection using fuzzy entropy measures with similarity classifier." *Expert Systems with Applications* 38.4 (2011): 4600-4607.

[23] Kihel, Badra Khellat, and Mohamed Benyettou. "Parkinson's Disease Recognition Using Artificial Immune System." *JSEA* 4.7 (2011): 391-395.

[24] Eskidere, Ö. (2012). A Comparison Of Feature Selection Methods For Diagnosis Of Parkinson's Disease From Vocal Measurements. *Sigma*, 30, 402-414.

[25] Prashanth, R., & Roy, S. D. (2018). Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning. *Neurocomputing*, 305, 78-103.

[26] Bruns, A., Kornstadt, A., & Wichmann, D. (2009). Web application tests with selenium. *IEEE software*, 26(5), 88-91.

[27] Loria, S. (2018). textblob Documentation (pp. 1-73). Technical report.

[28] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," May 2015.

[29] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, 1990.

[30] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal*

Processing - Proceedings, 2013.

[31] Prashanth, R., & Roy, S. D. (2018). Early detection of Parkinson's disease through patient questionnaire and predictive modelling. *International journal of medical informatics*, 119, 75-87.

[32] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[33] G. H. Yann LeCun, Yoshua Bengio, "Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton," *Nature*, 2015.

[34] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Comput. Intell. Neurosci.*, 2018.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *ImageNet Classification with Deep Convolutional Neural Networks*, 2012.

[36] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.

[37] Schneider KM. On Word Frequency Information and Negative Evidence in Naive Bayes Text Classification. In: *4th International Conference on Advances in Natural Language Processing*; 2004; Alacant, Spain: pp. 474–485.

[38] Kilimci, Z. H., Selim Akyokus, and Sevinc Ilhan Omurca. "The effectiveness of homogenous ensemble classifiers for Turkish and English texts." In *INnovations in Intelligent SysTems and Applications (INISTA)*, 2016 International Symposium on, pp. 1-7. IEEE, 2016.

[39] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "English pretrained Word2vec model," <https://code.google.com/archive/p/word2vec/>.