

# The impact of social media messages on Parkinson's disease treatment: detecting genuine sentiment in patient notes

El impacto de los mensajes de las redes sociales en el tratamiento de la enfermedad de Parkinson: detección de sentimientos genuinos en las notas de los pacientes

Hanane Grissette<sup>1</sup>, El Habib Nfaoui<sup>2</sup>

---

Grissette H.; Nfaoui E.H. The impact of social media messages on Parkinson's disease treatment: detecting genuine sentiment in patient notes. *Tecnología en marcha*. Vol. 35, special issue. November, 2022. International Work Conference on Bioinspired Intelligence. Pág. 5-15.

 <https://doi.org/10.18845/tm.v35i8.6441>

- 1 LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco.  
E-mail: [hanane.grissette@usmba.ac.ma](mailto:hanane.grissette@usmba.ac.ma)
- 2 LISAC Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco.  
E-mail: [elhabib.nfaoui@usmba.ac.ma](mailto:elhabib.nfaoui@usmba.ac.ma)

## Keywords

Distributed Biomedical representation; sentiment analysis; emotion modelling; patient narratives; social networks.

## Abstract

Parkinson's Disease (PD), one of the most serious neurodegenerative diseases that known huge controversy on social networks. Following medical lexicons, few approaches have been extended to leverage sentiment information that obviously reflects the patient's health status in terms of related-narratives observations. It is been crucial to analyze online narratives and detect sentiment in patients' self-reports. In this paper, we propose an automatic concept-level neural network method to distilling genuine sentiment in patients' notes as medical polar facts into true positives and true negatives. Towards building emotional Parkinsonism assisted method from Parkinson's Disease daily narratives di- gests, we characterize polar facts of defined medical configuration space through distributed biomedical representation at the concept-level as- sociated with real-world entities, which are operated to quantifying the emotional status of the speaker context. We conduct comparisons with state-of-art neural networks algorithms and biomedical distributed systems. Finally, as a result, we achieve an 85.3% accuracy performance, and the approach shows a well-understanding of medical natural language concepts.

## Palabras clave

Representación biomédica distribuida; análisis de sentimiento; modelado de emociones; Narrativas de pacientes; Redes sociales.

## Resumen

Enfermedad de Parkinson (EP), una de las enfermedades neurodegenerativas más graves que ha suscitado una gran polémica en las redes sociales. Siguiendo los léxicos médicos, se han extendido pocos enfoques para aprovechar la información del sentimiento que obviamente refleja el estado de salud del paciente en términos de observaciones narrativas relacionadas. Es crucial analizar las narrativas en línea y detectar el sentimiento en los autoinformes de los pacientes. En este artículo, proponemos un método automático de red neuronal a nivel de concepto para destilar el sentimiento genuino en las notas de los pacientes como hechos médicos polares en verdaderos positivos y verdaderos negativos. Hacia la construcción del método asistido por el parkinsonismo emocional a partir de los diálogos narrativos diarios de la enfermedad de Parkinson, caracterizamos los hechos polares del espacio de configuración médica definida a través de la representación biomédica distribuida a nivel de concepto asociada con entidades del mundo real, que se operan para cuantificar el estado emocional del contexto del hablante. Realizamos comparaciones con algoritmos de redes neuronales de última generación y sistemas distribuidos biomédicos. Finalmente, como resultado, logramos un rendimiento de precisión del 85,3%, y el enfoque muestra una buena comprensión de los conceptos del lenguaje natural médico.

## Introduction

Computational approaches are widely known surge application for alleviating Parkinson disease shared problems and experiences such as drug misuse or symptoms. Whereby, they used them for many cutting edge benefits such as mining shared experience to discover the risk of disease at an early stage whether a given patient has already developed or is expected to

develop Parkinson's disease in the future. Parkinson's Disease [1] Multiple examples of unsafe and incorrect treatment recommendations are shared every day. The challenge, however, to yield efficient, quick, consistent access to reliable insight. In the first sub-step, medical concept discrimination is proposed for data normalization [2], they automatically recognizing medical concepts mentioned in social media narratives enables several related-studies for enhancing health quality of people in a community, e.g. real-time monitoring of infectious diseases in the population, 1) we investigate a based-phrase transition between social media messages and formal description on medical anthologies such as MedLine.

Social Networks are frequently becoming popular as a platform for sharing personal health-related experiences [3] e.g. feedback, motivation, encouragement, emotional living, concrete support. Indeed, they support exchanged among those who go through emotional experiences too and usually, are influenced by the context surrounding the patient [4]. The advance of machine learning algorithms generally depends on data representation [5], and this information can be used for monitoring public health status. It hypothesized that these shared experiences might have distinct representations that can entangle and hide more or less the different descriptive factors of variation behind data and their properties. The advance of machine learning algorithms generally depends on data representation [5], neural networks based approaches such as Embeddings provide many advantages :(1) automatic feature extraction, (2) rich representation capabilities, (3) dimensionality reduction, (4) state-of-art performance for sentiment classification than tradition machine learning algorithms.

As many based neural networks approaches are becoming resolved many related medication text challenges, traditional based Lexicons methods fail requires an unaffordable huge labelled medical data or high qualified natural concept identification process. Due to such various factors as noised text errors and limited lexicons for related-medication text, the existing approaches cannot adequately make use of the sentiment information in the sentence for sentiment analysis. Many Distributions combined different pre-trained embeddings to enhance sentiment classification model performance, [6] proposed Sentiment Information Extractor based on Bi-directional Long Short Term Memory structure that applied to join the results of various sub-extractors. In particular scope, few approaches have been proposed for mining sentiment from related-medication text [7]. Patients notes on social networks and the use of informal medical language pose additional text challenges, including non-standard format, wrongly spelt, and abbreviation forms, as well as typos in social media messages. Nowadays, it is of crucial importance to providing an efficient solution to both mine medical components that extract aspects-based sentiment information. Most of the existing research, in this case, assumed at building a hand-annotated dictionary that is a time-consuming effort and may be subject to annotators/experts bias. However adopting an existing approach for sentiment analysis to mine related-medication texts may result in low classification accuracy, which employs less useful features sets and therefore lacks discriminatory potential. Many recent contributions aim at enhancing medical natural concepts processing by incorporating formal medical knowledge such as MedLine [8]. They developed a Neural network model, which further exploited N-gram-based convolutions for creating medical vocabulary scheme, which is dedicated mainly to featuring text under medical setting and clarifying related sentiment at the same level. In this paper, we provide an efficient method to both deploy medical concept-level conceptualization through generating real-world embeddings and define dynamic sentimental measure. Our main aims are as follow:

- Distill Parkinson's related-messages to treat what kind of PD's classes may uncover greater insights.
- Build an Emotion Parkinson's Assisted Tool to attempt a public concern about Parkinson's disease and their attributes.

- Propose an Inference component for evaluating note's positiveness/negativeness value, which measures the impact of given related-PD drugs or treatments.

The remainder of this paper is structured as follows. Section(2) provides a summary of the literature review concerning PD studies. Section(3) explains the offered methods. Section(4) offers large social network experimentation out- puts convolving large biomedical embedding. Section(5) presumes this paper and presents future aims.

## Literature review

The accuracy of traditionally and potentially proposed clinical approaches [9] for PD classification, and prediction models have known slow progress due to data inaccessibility and complexity, where shows apparent diversity adopting these techniques on similar online narratives. They jointly processed varied techniques included data aggregation, rebalancing imbalanced cohorts, and meanings correspondence of heterogeneous data elements. They have been mostly integrated solutions from multiple sources offers unrivalled opportunities to track disease progress, e.g. early stages of prevalent neurodegenerative manners, and promptly discover the efficacies of alternative treatments based on the clinical definition of Parkinson's disease.

However, recent studies [10] hypothesize that it is likely to develop a computing model, which experiments results in acceptable accuracy, significantly, when he/she is anticipated to develop the Parkinson's disease to overcome data properties difficulties and challenges. Where artificial neural networks significantly outperformed the other methods in the literature and achieved the highest classification results reported so far. Most of the existing studies and PD methodologies rely on voice disorders patterns analysis; few works have been involved in leveraging self-reported experiences shared on social networks. Nevertheless, the results of several machine-learning based classification methods involved mandatory additional tasks, including Parkinson's disease evaluation metrics such as consistent accuracy, sensitivity, and statistical n-fold cross-validation. Indeed, it requires patterns recognition preliminary steps and medical preprocessing or normalization.

According to [11], the huge use and communications share of many social plat- forms such as AskPatient, DailyStrength, and Twitter, draw colossal attention from many researchers, which the potential for monitoring health conditions (e.g. adverse drug reactions, infectious diseases) in particular communities. They employed the power of the neural network to construct dynamic and credible Parkinson's Disease Digital Biomarker Dataset Using Neural Network Construction (NNC) Methodology Discriminates Patient Motor Status [12]. Nevertheless, these online shared experiences pose additional related-medical challenges. Refers to [8], daily patients' notes on social media may be a pivotal role to re- veal crucial public health issues through distilling hidden shared information, which remained so far unconsidered. Embeddings widely propose efficient dense representations that enhance sentimental analysis performance, which mostly focuses on discovering grammar directions such as semantic direction or only rely on upon represent sentiment property of distinct words at word-level. Sentence- level SA approaches to leverage the sentiment through the ensemble of tokens. In contrast, most of them have not capable of discovering sentiment conveyed towards related drug entities, or determine aspects-based sentiment within patients narratives. where a given sentence may contain complicated descriptive and nontechnical medical entities such as disease symptoms, drug misuse, and adverse drug effects.

However, The use of informal language while patients note concerning their treatments are characterized by a combination of nontechnical medical language with specific terminology (e.g. adverse effects, drug names) with greater lexical diversity. Deep learning methods have been used to cover these limitations. Most existing studies adopt word-mapping techniques to

normalize and translate related medical concepts and their attributes. [11] adopt a combination of a phrase-based Machine Translation(MT) method and the similarity between dense representations of words for overlapping social media phrase to the formal medical concept in standard ontologies; where it implicitly and explicitly succeed health quality monitoring in real-time window [2]. Moreover, Indeed, Medical sentiment analysis and emotion recognition have known by aspects' diversity. The previous study focused on different aspects, such as drugs and doctors' aspects [13]. At this end, Bi-directional Long Short Term Memory (BiLSTM) approach widely used as state-of-the-art methods to retrieve and classify sentiment within unstructured text [6] where the ensemble of convolutional approaches applied for medical recognition and extraction tasks.

## Methods

The concept-level approaches are considered the key to common sense in neural network thinking. In this section, we seek to demonstrate the importance of establishing PD natural medical descriptive entities to the formal correspondence. Then, quantifying the emotional status of the speaker context. The way we can uncover the impact of social messages from shared Parkinson's Disease experiences through distilling genuine sentiment in patients' notes as polar medical facts into true positives and negatives narratives. In this section, we first present the rationale for this study case background. After, we address the proposed methods in detail.

### Aspect-based distributed conceptualization

The concept-level approach is the key to provide an appropriate medical abstraction that is termed implicitly in all fundamental aspects of relationships among words, concepts, and medical definitions. Aspect-based conceptualization offers an integrative theory based on biomedical distributed representation. To the aim to find the relevant correspond of natural medical concepts and entities. Following existing Medical Sentiment analysis approaches, most of them fail due to the inability to extract related-medication concept aspects from the text.

Medical components normalization, from generated narratives on social networks, frequently needs medical lexical comprehension at the features extraction level, which can be difficult without sufficient background knowledge of the formal medical contexts. These explain why Sentiment Analysis systems can perform sentiment analysis towards a given entity and services reasonably well but far on clarifying sentiment towards a set of words that may refer to the specific medical components such as drug misuse multi-word expression. At this end, we choose to build a feature-based mechanism in terms of embedding vectors that incorporate external biomedical definitions and universal words for higher generalization performance. The embeds related-drug concepts preserve to define unrelated-medical items/ concepts to the formal medical language used in the descriptions of medical concepts in standard ontologies. The efficient use of pre-trained language medical knowledge guarantees a contextualized and semantical overlapping of primary medical observations in text. In the experiments, we used a data set that consists of two corpora for mining drug-related knowledge from online data, as illustrated in table 1.

**Table 1.** Statistics of distributed model for mining drug related knowledge from online data.

Distributed model	Description	Statistics
Drug-related, word representations from Twitter [14]	Distributed models based on distributional semantics capturing sequential patterns	It consists of 267,215 Twitter with over 250 drug-related posts mention keywords using drug names as keywords, including their common misspelled forms
Biomedical Word embeddings with subword information and MeSH (named BioWordVec) [15]	An open set of biomedical word vectors/ embeddings that combines subword information from unlabeled biomedical text with a widely-used biomedical controlled vocabulary called Medical Subject Headings (MeSH)	Biomedical Word and sentence embeddings using PubMed and the clinical notes from MIMIC-III Clinical Database as described in Table 2.

**Table 2.** Statistics of Controlled medical vocabulary we used.

Sources	Documents	Sentences	Tokens
Pubmed	28,714,373	181,634,210	4,354,171,148
MIMIC III Clinical notes	2,083,180	41,674,775	539,006,967
The PubMed Central (PMC)	2,083,180	41,674,775	539,006,967

### Sentiment Language Modelling

Sentiment language modelling is the core of comprehending the genuine sentiment in patients' notes by quantifying the patient emotional status degree to state the social impact. It consists of an automatic online method that produces relevant aspects-based related-sentiment features, identifies and stores related- drug natural patterns from varied patient spaces, and then define the impactful class concerning medical aspects. So far, our sentimental analysis leverage differently concerning embed and contextual sentiment. It involved standard medical knowledge because it enables to encapsulate and parametrize new unobserved related-drug concepts throughout the downstream deep distributed medical components representations.

The relationship between emotion and language attributes to many challenges, especially, PD shared data has additional problems. Neural network methods have known massive distributions, where they mostly operated embeddings to get vectors of unknown raw data and define local semantics and contextual features to track related-sentiment value. Nevertheless, adopting the same manner on related-medication data results in slow accuracy than desired because of the model neglect medical entities' semantics. We propose to deploy both approximate modelling as follows :

- Generating biomedical Embedding vector built by adopting controlled vocabulary from medical ontologies and online version as described in Section 3.1.
- Re-calculate embeddings for unknown terms and medical relations.
- Computing statistical sentiment value regard medical aspects and targets defined Before.
- Constructing Our sentiment inference model based on Bidirectional LSTM, enhancing results by using an Attention mechanism.



The sentiment value is inferred regarding the relationships over words and concepts. we defined sentiment relationship by computing the pointwise mutual information (PMI) as quantified value that measure sentiment association within given words and/or concepts and the average concepts as follow:

$$PMI(x, y) = \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

By approximating the probabilities  $p(x)$  and  $p(x,y)$  respectively, we define sentiment association value with regard to biomedical embeddings. We present and discuss, in several different perspectives, the evaluation results of Machine learning models (LSTM, BiLSTM, SVM, simple RNN, linear regression) with based- associative biomedical embedding, and sentiment scale generation. The evaluation models have slightly different results, the most suitable algorithms that both leverage sentiment from semantics reasoning in varied contexts.

## Implementation and Output Result

Keep track of health issues such as infectious diseases through Sentiment Analysis(SA) is of critical importance. Parkinson's disease(PD) long-term degenerative disorder that affects an estimated seven to 10 million people and families worldwide. Reports pour daily into healthcare communities, micro-blogs at a staggering rate, e.g. PD patients tend to be affected by severe emotional disorders, especially, PD's n-polar emotional state. In this paper, we ingest Parkinson's related messages from Michael J. Fox Foundation<sup>1</sup> and Twitter Data, to attain a better view of distinct drug or treatment in real-time.

### Dataset and parameters

Twitter is the common popular micro-blogging that attracted millions of users to share and disseminate the most up-to-date information. In this study, we validated the proposed model by training our ensembles models using twitter data. We are also interested in evaluating the proposed method on extensive discussions from typical PD community with for evaluating how many methods comprehend varied particularities. Most of the PD communities are of public access, where PD patients and their families prefer to gather or share essential information about their minute treatments and problems from specific PD forums. In this paper, we ingest Parkinson's disease-related messages from the Michael J. Fox Foundation community to prove the efficiency of data on large forums that remained more details concerning drug and treatment vs disease. We collect data from this forum according to discussion topics and their attributes as described in table.4.

In our crawling system, data are collected using Twitter APIs. We used a list of related-PD Keywords Parkinson and dopamine. The table. 3 summarizes statistics of raw data in terms of the context of varied medical concepts and keywords. Parkinson communities have a particular structure; we choose categorized discussions into sub-forums, then into topics such as (1) Coronavirus (COVID-19) and Parkinson's for discussing the Corona virus pandemic, and how it coincides with Parkinson's disease. (2) Parkinson's Disease symptoms are uniquely challenging. For example, As PD symptoms are progressively damaged, their treatments description on social networks may be ambiguous and appear in the same manner in various discussions.

**Table 3.** Statistics of raw data in terms of the context of varied medical concepts and keywords.

Data set	Scheme and attributes	Keywords and statistics
Twitter Data	['id', 'created_at', 'source', 'original_text', 'retweet_count', 'original_author', 'hashtags', 'user_mentions', 'place', 'place_coord_boundaries']	15000 tweets collected regarding set of related-PD term and drugs e.g. <i>Tremor, bradykinesia, synuclein</i>
Parkinson's Disease(PD) forums	[Post_ids, topic, links, voices, replies, comments]	4807 original post by topics, subforum reference and their attributes.

Do, disambiguating related-entities for given related-PD discussions help to connect with others and access resources that can help explain, and cope with PD symptoms. As described, we streamed data regarding a set of keywords and topics. Then, raw data are stored to a set of \_les regarding drugs vs disease components and time axis; we focused on associated criteria of chronic conditions, PD symptoms and related-medication concepts. Moreover, these samples were passing by a set of normalization and processing steps to be able for our neural inference model. Indeed, every single tweet in our dataset consists of a separate document that saves the life of correlate information contained regarding medical and pharma objects.

**Table 4.** Statistics of raw data in terms of the context of varied medical concepts and keywords.

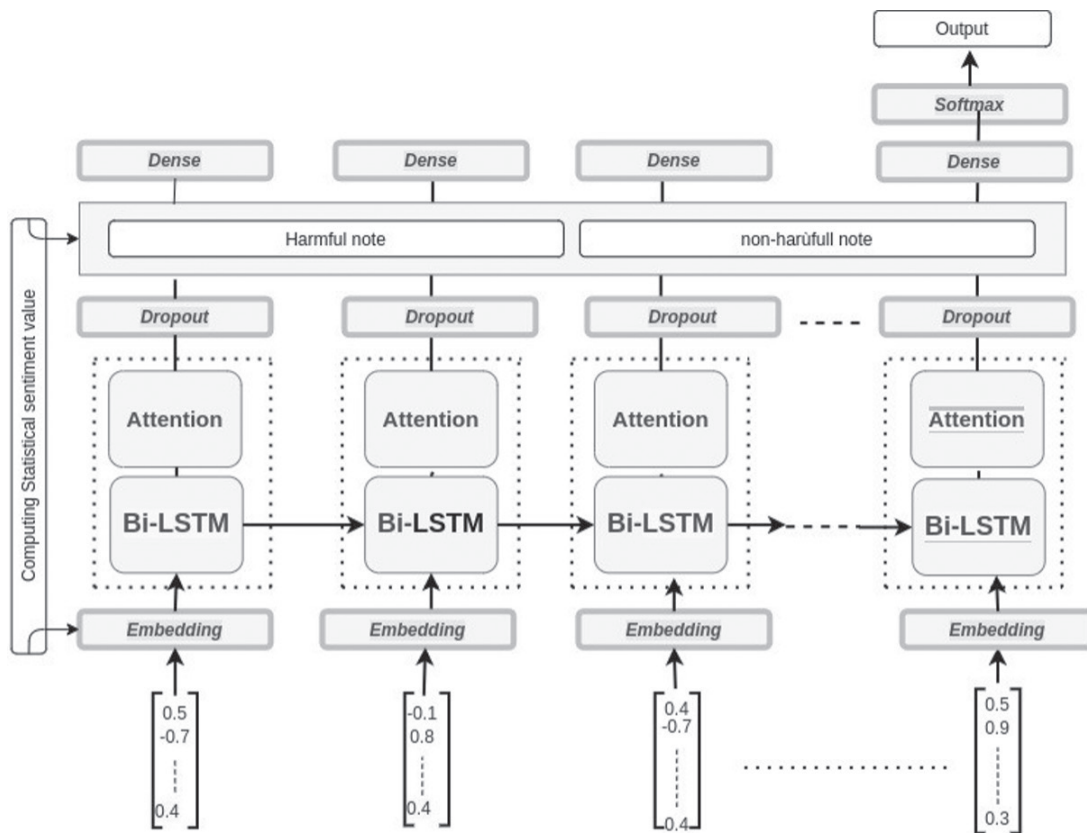
Raw data topic	Statistics
PD & Dystonia	180 posts
CBD Oil and Parkinsons	115 posts
Butyric Acid and Niacin	97 posts
Has Covid-19 changed your doctor appointments?	98 posts
No pharmaceutical drugs?	121 posts
CRISPR GENE THERAPY! What do you know? Is it hope?	77 posts
Magnesium-L-threonate and PD	28 posts
How has your diet changed since the diagnosis?	45 posts

### Implementation

A sample of 2807 PD-related posts, from the online healthcare community of Parkinson's disease, were collected and normalized to enrich the vocabulary. For twitter, we collect more than 15000 tweets that have been prepared to be input to the neural classifier for defining medical concepts and then re-define distributed representation for unrelated items of natural medical concepts cited in real-life patients narratives.



## Experimentation and Results



**Figure 1.** Sentimental inference model.

In this section, we present the results obtained by applying the proposed methods on both Twitter data and patients notes and messages on the forums. The experiments executed consider the language in which the messages are written, emoticons, mentions, and the length of texts analyzed.

The sentiment value is floated between  $[-1, +1]$  -1 extremely negative +1 extremely positive. we create a novel sentiment scale for interpretable sentiment analysis model, which is dedicated to labelling positiveness and negativeness by harmful and non-harmful notes of given posts regarding related-medical targets, the way we can state false positives and negatives from texts. As described in figure 1. Each sentiment word belongs to a harmful class has to be convolved with an extremely negative value and clear medical co-annotation.

We conducted various experimentation originated from the application of the proposed method based on hybrid medical corpora based-text and concepts; a minute conceptualization is released. Sentiment detection is assessed through experiments on Three online datasets from various observations. An extensive evaluation of different features, including medical corpora, neural network algorithms. The based stacked-LSTM and BiLSTM model consistently improved the sentiment classification performance but is efficient when we exploit proposed configuration on a big online dataset as illustrated in table 5. We conduct development vocabulary scheme to enhance the effectiveness and transferability of this study across various online related-medication posts (Twitter posts, Parkinson's disease forum's posts), which was significantly better than all other baselines.

**Table 5.** Experiments results overview on Both Twitter Data and forums discussions (discussed in Section 3.1.)

Raw data	Neural Network algorithm	Accuracy
PD Tweets	LSTM	77%
	BiLSTM	85.3%
	Stacked-LSTM	79%
PD forums posts	LSTM	79.5%
	BiLSTM	87%
	Stacked-LSTM	78%

## Discussion

Parkinson's disease (PD) is the second most crucial age-related disorder, with a prevalence ranging from 41 per 100,000 in the fourth decade of life to over 1900 per 100,000 in people over 80 years of age [16]. Refers to [12], the problem with Parkinson's disease (PD) patient monitoring fail is due to the limited resource. Where shared data suffered from inadequate, sporadic symptom monitoring, infrequent access, and sparsity, leading to poor medical decision making and sub-optimal patient health-related outcomes. Few Data Mining and NLP techniques have proposed in the PD context. As the existing tools are assisted for clinical decision-making, which is relying on the guidance of physicians to determine various measurement variables depends on diagnostic and treatments. In this study, we aim at conducting a demonstration of how can social media data can automatically leverage the social impact of PD patients and their families. Emotional dysregulation is an essential dimension that may occur in several psychiatric and neurologic disorders. We focused on clinical characteristics of emotional state variations in bipolar disorder and Parkinson's disease [17]. In both pathologies, the emotional intensity variability involves essential diagnostic and therapeutic issues for monitoring the emotional state of PD patients and track the impact of social media messages. Technically, instead of investing based-deep learning techniques to classify patients' self-reported messages on social media as positive or negative statements, we extend the sentiment inference model to distil sentiment aspects to distinguish negative facts degrees regarding Parkinson's disease-related drug-targets. Indeed, we are working on powered- Neural Network fact polar model to probe what kind of based-treatment target may result in improved Emotion Parkinson's model performance.

Likewise, harmful and non-harmful notes of given posts regarding related- medical targets noticeably may fail to retrieve the correct impact due to the inability to define complex medical components in text. Each post may refer to a drug reaction or/and misuse is categorized to harmful impact, where beneficial adverse reactions may also consider as harmful. As a solution, we propose to detect primarily drug reaction multi-expression in the text that should be considered as targets.

## Conclusion

In this research, we studied a based neural network approach through multiple Bi-LSTM components to build a dynamic configuration space from unsupervised medical concepts representations. Thus, the embeddings from this joint model are used to generate sentiment scale powered by additional statistical salient for medical concept-aspects sentiment Inference.

In the future, we aim at defining related-PD drug related-events such as drug effectiveness, drug reaction, or drug misuse in text, which is assured to distinguish credible harmful insights.

## References

- [1] Salama A. Mostafa, Aida Mustapha, Mazin Abed Mohammed, Raed Ibraheem Hamed, N. Arunkumar, Mohd Khanapi Abd Ghani, Mustafa Musa Jaber, and Shihab Hamad Khaleefah. Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease. *Cognitive Systems Research*, 2019.
- [2] Nut Limsopatham and Nigel Collier. Normalising medical concepts in social media texts by learning semantic representation. In *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 2016.
- [3] Azadeh Nikfarjam, Abeed Sarker, Karen O'Connor, Rachel Ginn, and Graciela Gonzalez. Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, 2015.
- [4] Rodrigues R.G., das Dores R.M., Camilo-Junior C.G., and Rosa T.C. SentiHealth- Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks. *International Journal of Medical Informatics*, 2016.
- [5] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.
- [6] Kai Shuang, Zhixuan Zhang, Hao Guo, and Jonathan Loo. A sentiment information Collector–Extractor architecture based neural network for sentiment analysis. *Information Sciences*, 2018.
- [7] Hanane Grissette, EL Habib Nfaoui, and Adil Bahir. Sentiment Analysis Tool for Pharmaceutical Industry & Healthcare. *Transactions on Machine Learning and Artificial Intelligence*, 2017.
- [8] Hanane Grissette and El Habib Nfaoui. Enhancing convolution-based sentiment extractor via dubbed N-gram embedding-related drug vocabulary. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1):42, 2020.
- [9] Ivo D. Dinov, Ben Heavner, Ming Tang, Gustavo Glusman, Kyle Chard, Mike Darcy, Ravi Madduri, Judy Pa, Cathie Spino, Carl Kesselman, Ian Foster, Eric W. Deutsch, Nathan D. Price, John D. Van Horn, Joseph Ames, Kristi Clark, Leroy Hood, Benjamin M. Hampstead, William Dauer, and Arthur W. Toga. Predictive big data analytics: A study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PLoS ONE*, 2016.
- [10] Niousha Karimi Dastjerd, Onur Can Sert, Tansel Ozyer, and Reda Alhajj. Fuzzy Classification Methods Based Diagnosis of Parkinson's disease from Speech Test Cases. *Current Aging Science*, 2019.
- [11] Nut Limsopatham and Nigel Collier. Adapting phrase-based machine translation to normalise medical terms in social media messages. In *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 2015.
- [12] Ioannis G. Tsoulos, Georgia Mitsi, Athanassios Stavrakoudis and Spyros Papapetropoulos. Application of machine learning in a Parkinson's disease digital biomarker dataset using Neural Network Construction (NNC) methodology discriminates patient motor status. *Frontiers in ICT*, 2019.
- [13] Salud María Jiménez-Zafra, M. Teresa Martín-Valdivia, M. Dolores Molina-González, and L. Alfonso Ureña-López. How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. *Artificial Intelligence in Medicine*, 2019.
- [14] Abeed Sarker and Graciela Gonzalez. A corpus for mining drug-related knowledge from Twitter chatter: Language models and their utilities. *Data in Brief*, 2017.
- [15] Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, and Zhiyong Lu. BioWord-Vec, improving biomedical word embeddings with subword information and MeSH. *Scientific Data*, 2019.
- [16] Ramon Cacabelos. *Parkinson's disease: From pathogenesis to pharmacogenomics*, 2017.
- [17] Max Little, Paul Wicks, Timothy Vaughan, and Alex Pentland. Quantifying short-term dynamics of parkinson's disease using self-reported symptom data from an internet social network. *Journal of Medical Internet Research*, 2013.