International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

December 2022	Issue 53	Volume 14	Number 4	Pages 211-218

SENTIMENT RETRIEVAL OF HEALTH RECORDS USING NLP-BASED ALGORITHM

A.M. Taha [1,2] S.F. Jabbar [3] A.H. Alwan [3]

1. Department of Business Information Technology, University of Information Technology and Communication, Baghdad, Iraq, ahmed.majid.taha@gmail.com
2. Soft Computing and Data Miming Center, University Tun Hussein Onn, Batu Pahat, Johor, Malaysia
3. College of Education, Ibn Rushed for Human Science, University of Baghdad, Baghdad, Iraq
saadya.fahad@ircoedu.uobaghdad.edu.iq, asmaa.hussien@ircoedu.uobaghdad.edu.iq

Abstract- This paper presents a sentiment retrieval based on natural language processing NLP-based Word2Vec method for health records in medical institutions of Iraq especially for Covid-19 patients. Sentiment retrieval of medical records has gained significant attention worldwide to understand the behaviors of both clinicians and patients. However, Sentiment retrieval of medical notes still not provides a clear picture of information retrieving from these summaries. Covid-19 Pandemic urges researchers in the field of medical records and AI modelling to establish a sentiment analysis based on discharge summary notes. The study is performed on 10000 medical notes from general hospitals with total of 8500 patients and a 15000 medical notes from general clinics with total of 12000 patients. The study is conducted in Iraq during May 2021 to May 2022. The main intensity of measured sentiment is captured with positive or negative in the health records. The SentiWordNet platform is used to standardize a gold sentiment dataset and the performance is evaluated using Word2Vec method. The Welch's t-test is used to validate the significance of the obtained results. It has been found that the statistical significance between Covid-19 health records reaches to 94.6% with p-value of 0.054.

Keywords: Covid-19, NLP, Sentiment Retrieval, SentiWordNet, Word2Vec.

1. INTRODUCTION

Generally, the using of health records known as discharge summary notes in coded data have established a widely investigation on the benefits toward pharmacovigilance to risk stratification [1]. This brings a great challenge in terms of decision-making and health institutions status due to the huge data that being coded. In addition, these health records such as discharge summary notes are not analyzed based on sentiments method. Basically, health records/notes are captured as unstructured data, which makes the analysis of the provided data easy to decision-making. However, as those records are captured as description notes, it is challenging to quantify it. Which reduces the overall efficiency of the data coded in the health systems. Another issue occurred based on the text reflections in medical records on the physicians or the patients, which triggers several researchers to study the effects of text reflections in health records on a specific medical subject. For instance, the word "positive" in medical report is usually referred to an infected disease, and word "negative" is referred to non-infected disease. Therefore, quantization methods and algorithms are proposed for the validation of the text reflections and feeling impact on medical text documents and refers it as sentiment [2]. These methods have enabled a significant investigation on the efficiency of the health care records such as health care contentment [3], tobacco impact on health [4], several approaches on cancer treatments [5], and wellness of health on social media [6, 7].

Recently, the Covid-19 pandemic has increased the amount of medical records exceptionally as the information obtained from medical institution is developed in several domains and applications. In order to understand the Covid-19 patients before and after treatment, medical process or opinions have a major area of interest. The data collected from the Covid-19 patients records in different clinics and hospitals have been widely used to help physicians in giving an accurate opinion about the patient attitude after fully treatment and recovery. Many researchers have implemented several Covid-19 records in order to anticipate recovered patients and retrieve their discharge Covid-19 records [8]. However, due to the huge data size and quantity challenges, the path toward data mining is proposed. Data mining techniques have helped in modelling several algorithms, techniques, and methods for handling big size data. In addition, data mining methods have the benefit of providing the suitable tools for increasing insightful learning [8, 9]. Moreover, using data mining method such as sentiment analysis in study the Covid-19 patient behavior has a key role in providing an efficient review for clinicians and physicians [9]. In that matter, the information from Covid-19 records can be analyzed by using sentiment retrieval method.

Such a method could help the physician's reviews and opinions to be accurately specified patients' behavior and utilized the sentiment analysis of Covid-19 records. Figure 1 shows the process of analyzing Covid-19 health notes on obtained dataset [9].

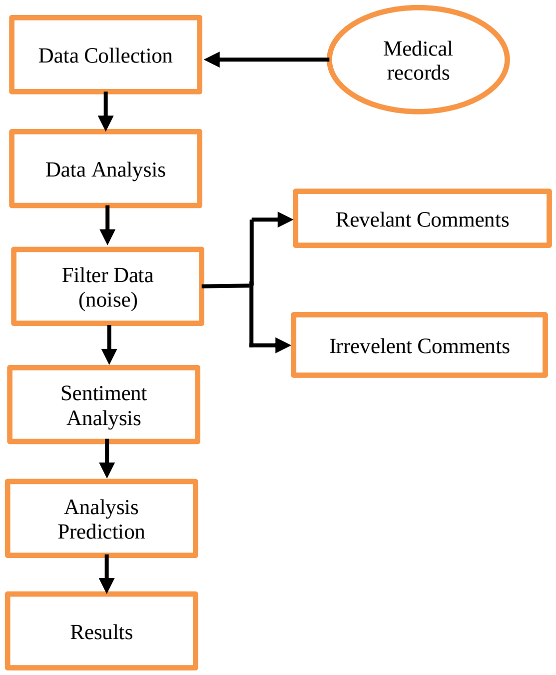

Figure 1. Health records analysis process [9]

Hence, this paper highlighted that using sentiment retrieval analysis on Covid-19 health record that obtained from physicians in Iraqi's medical hospitals and clinics can be measured as a data mining notes. This helps to better understanding of both physicians and patient's behaviors toward Covid-19 medical records. Specifically, the opportunity of these Covid-19 records to have a predictive validity and utilized the Covid-19 records as data coded sources. Therefore, this paper offers a sentiment retrieval analysis by applying opinion mining algorithm for quantifying the Covid-19 health records. The main contributions of this work are summarized as follows:

- A gold sentiment dataset is established for Covid-19 health records based on NLP-based SentiWordNet for Iraq's health database.
- Applying Word2Vec tool in sentiment analysis (unsupervised) of a set of medical Health records focusing on Covid-19 recovery cases.
- Evaluation the performance of Word2Vec in identifying sentiment similarity of the discharge summary notes.

2. RELATED WORKS

Typically, health records are sentiment analyzed by lexical and dictionary resources. This analysis includes MPQA, WordNetAffect, and SentiWordNet [10-12]. As these algorithms and tools are functioning in a reliable way of identify sentiment in highly reviewed text. However, for the Covid-19 health record and to the author knowledge, these methods are still not yet implemented to date [13].

In addition, the sentiment annotation and high cost of data mining of health/Covid-19 records are the major challenges facing sentiment retrieval analysis. In the other hand, using supervised learning algorithms in health records have caused an impracticable and incorrect data setting due to the wrongly accessing of training labelled data. Conversely, in sentiment analysis of health records, the unsupervised learning algorithm is the most suitable candidate for retrieval of health/Covid-19 records [14-17].

The authors in [18] presents a Sentiment analysis of medical tweets based on supervised learning method. The authors used sentiment specific word embedding (SSWE) is a part request of embedded word representation in outclassed models of neural network including Word2Vec. This approach tried to categorize the sentiment to differentiate between words and its opposite sentiment in the same arise of meaning. However, this method performed a poorly accuracy around 10 % for the measured lexicons [18]. As Covid-19 strokes the world in very fast time and cases, the urgent to apply sentiment analysis and using artificial intelligent to achieve better outcomes is urgently needed. The number of the cases reaches to nearly 77.7 million patients till this June 2022 [19]. Deep learning and AI methods aim to help to give better understanding and standardization for the battle of Covid-19. Hence, various researchers proposed using artificial intelligent algorithms to be implemented in sentiment retrial from data collecting through treatment to recovery stages [20-25]. Hence, this paper investigates two approaches; the first approach is to study the Covid-19 health records with low sentiment terms than other. The second one is to analysis sentiment retrieval by applying unsupervised learning algorithm on Covid-19 datasets.

3. PROPOSED METHODOLOGY

Figure 2 shows the research methodology which includes two specific phases [26]. The first phase is collecting data and the second phase is the sentiment analysis based on data mining for Covid-19 health records. The research proposed methodology is shown in Figure 3 which presents the flow chart of data collection, classifications, and applying unsupervised learning algorithm on health record text. It starts with data collection of Covid-19 health records. Following the utilization of Covid-19 health records in text and views. This stage is important for analyzing the results effectively. Then, the data mining algorithm is applied in the classification level with neutral entity denotes as (0), positive entity denotes as (1), and negative entity denotes as (-1). Finally, the unsupervised learning algorithm based on Word2Vec is applied for data analysis and modelling in such way it performed as opinion data mining.

Data is collected in the period of May 2021 to May 2022 using provided and electronically Covid-19 health notes from local server. In order to get access to the information stacked at the institution server, the i2b2 software for server access is used to gain access of health records data in the major medical institutions in Iraq [26, 27]. The present research only discusses the Covid-19 health records based on recovery cases and limited to the patients in Iraq.

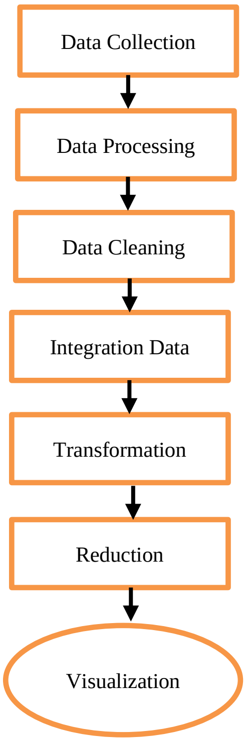

Figure 2. Data collection process from discharge summary notes [26]

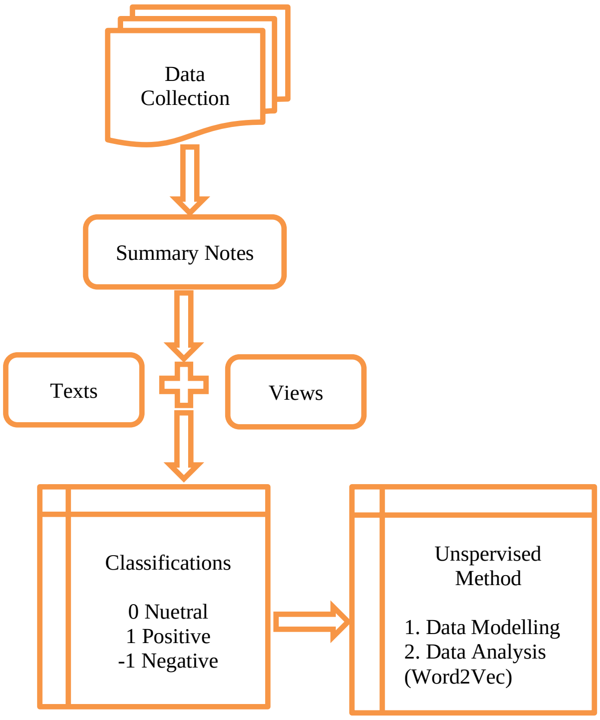

Figure 3. The proposed flow chart of the methodology

4. DATA COLLECTION AND PRE-PROCESSING

Two dissimilar cohorts are pulled out from the local server of specific patients to analyze the features of their Covid-19 health notes. The patients that have been admitted to psychiatric with fifty beds from May 2021 to May 2022 were assigned to the first cohort. The second cohort covers the patients that admitted in the hospital with same period of time. Patients above 19 years old are only considered in this research. The fascination outcomes with first cohort are caused by clinical readmission, and following the checking time index of Covid-19 records for determining admissions subsequent. In another hand, the second cohort time readmission is examined with the same period of all caused mortality. The number of data collection consists of 25000 Covid-19 health records that have been written by physicians and clinicians.

An annotation is defined for each set which identifies each Covid-19 health record with specific number of accompany diseases. Therefore, each annotations of existence record are classified into four main classes: Absent, Current, Uncertain, and Unknown. Using these classes, the Covid-19 health records are divided into sub-class that associates with particular disease. Following, the separation of each class is applied by determination of each sub-class to the main class Current. This means the Covid-19 health records is classed to specific sub-class when its main class of the disease is linked. The Covid-19 health records that linked to each sub-class disease is listed in Table 1. As we can be noticed from Table 1, several diseases such as Covid-19 with GERD, Covid-19 with Hypertriglyceridemia, and Covid-19 with Gout have few discharge notes with that disease being linked. This will perform an incorrect computational and statistical sentiment analysis associated with them. Therefore, the data set with best matching linked to specific five diseases are selected for this study which are: Covid-19 with Hypertension, Covid-19 with Diabetes, Covid-19 with Heart failure, Covid-19 with Gallstones, and Covid-19 with Depression.

Table 1. Distribution of disease classifications

Annotated Disease	Covid-19 Health Notes
Covid-19 with Hypertension	6500
Covid-19 with Diabetes	4500
Covid-19 with Heart failure	4130
Covid-19 with Gallstones	4000
Covid-19 with Depression	3125
Covid-19 with GERD	985
Covid-19 with Hypertriglyceridemia	915
Covid-19 with Gout	845

4.1. Processing Tool of Natural Language

Several methodologies are developed to characterize elements of written text in a high throughput manner [26, 27]. Generally, there are two conceptually different procedures exist: the first one is based on machine learning where a model of word or maybe phrase usage is applied to documents of sentimentality well understood. Then, it will apply to the whole document. The second method is based on using a curated lexicon of sentimentality and subjectivity to mark words in the whole document. However, for the first method, it is found that there is no gold standard for medical data notes in sentiment form. Therefore, the second method is preferred to be used in this study. Particularly, Lemmatization SentiWordNet method is used, a wide-open source implementation of opinion mining method. This method works on matching phrases and words to build-in lexicon with 3000 polarity words such as positive +1 and negative -1, intensity modifier 0.5 to 2x, and reverse polarity (negation) [26].

This method has demonstrated accuracy of 85% compared to a gold standard [26]. To the author knowledge, this method has been applied for movie reviews [26], yet no works stated or examined its performance on medical notes. Python panda's library is an open source data processing tool which is used in extraction of retrained information of the implemented method. By using this library, the Health records are extracted to its corresponding dataset. The process starts with cross reference data frames with identification (ID) of discharge summary notes. Then the ID links with its annotation's frames. After linking all the IDs, the sub-categories are organized based on their descending frequency. The retrained data is performed using Python with the assistance of the Natural Language processing toolkit (NLPTK).

This procedure starts with removing all the special characters and formatting text from all the discharge notes. Then, the English stop words are also removed from summaries by NLPTK tool. The remains Health records are standardized for retrained process before choosing whether lemmatization or stemming to be selected for further analysis of data. The next step is to determine which process is chosen for testing the results of sentiment; stemming or lemmatization. SentiWordNet method is performed to test the data of sentiment analysis by using Porter Stemming Algorithm [27] then lemmatization independently. The results then compared with the original obtained data. Table 2 reports the comparison results between the original data with stemming and lemmatization results. The results showed a best term matching with original data by using lemmatization with an average of 38.44% compared to the stemming terms matching of 27.14%. Therefore, this work will choose the lemmatization for analysis of the dataset.

Table 2. The percentage of matching of lemmatization and stemming on original data

Diseases	Original Data	Lemmatization	Stemming
Covid-19 with Hypertension	0.4574	0.43500	0.3540
Covid-19 with Diabetes	0.4120	0.4005	0.2844
Covid-19 with Heart failure	0.4245	0.4120	0.3100
Covid-19 with Gallstones	0.3469	0.3200	0.2100
Covid-19 with Depression	0.3602	0.3547	0.2005
Average percentage	0.4001	0.3844	0.2714

4.2. Analysis of Method

Two characteristics are chosen as a primary measure for sentiment analysis; which are objective '0' and subjective '1'. Then it associated with product of two scales of positive '+1' or negative '-1'. Therefore, for each discharge note two scores result in consistent range of 0-1 as positively subjective and negatively subjective respectively. Hence, for a discharge note scored 1 means entirely positively subjective, a note scored -1 is entirely negatively subjective and if note scored 0 it means an entirely neutral. All the notes with these two positive and negative scores are separately analyzed to permit the chance that several notes may have very high amounts of each. Relations between the variables of negative or positive sentiment subjectivity and clinical sociodemographic characteristics are analyzed using utilizing effects models. That is clearly observed for the account for attendance of several grouped opinions (i.e., several hospitalizations) for each patient. SentiWordNet is a sentiment method based on lexical resource as part from opinion mining formed in [27-29].

It is built on WordNet, another lexical resource and database for English language. It classified words with their corresponding sets of synonyms. SentiWordNet is a top level of WordNet which works by allocating each sentiment scores; Positive, Negative, and Objective scores. The subjective serving of the word is made up by the positive and negative scores, while objective score can be obtained by minus the summation of positive and negative score with number one. To analysis the overall sentiment scores of each sub-category, SentiWordNet is implemented to determine each summary linked to the sub-category. Then, the determination of sentiment ratios in terms of positive and negative scores for each sub-category is carried out. Word2Vec method is used for constructing the summaries data. In Word2Vec method, the similarities between each sub-category will be calculated against other sub-categories. Referring to Table 2, each of the five sub-categories are evaluated by SentiWordNet. Then, the positive, negative, and objective scores are calculated and the results are listed in Table 3.

Table 3. The results of sentiment scores of Covid-19 health notes

Diseases	Positive	Negative	Objective	Overall
Covid-19 with Hypertension	0.0521	0.0698	0.8781	0.0178
Covid-19 with Diabetes	0.0507	0.0717	0.8773	0.0208
Covid-19 with Heart Failure	0.0506	0.0719	0.8775	0.0213
Covid-19 with Gallstones	0.0507	0.0720	0.8785	0.0225
Covid-19 with Depression	0.0484	0.0736	0.8787	0.0245

Figure 4 illustrates the performance of positive and negative percentage scores of each sub-category in sentiment analysis. It can be observed from mentioned graph that there are particular sub-categories have a higher negative score than positive ones. This might be revealing of a negative partiality in relative to the other sub-categories. In dissimilarity, the reverse behavior can also be perceived. This might be revealing of a positive partiality in relative to the other sub-categories. In addition, the maximum negative overall score happens when the alteration between the negative and the positive scores is high. Correspondingly, the maximum positive overall score arises when that alteration is small. Therefore, it can be realized that the overall relation sentiment score of each sub-categories can be found by the difference between the positive and negative sentiments.

The overall score sentiment of each sub-category is found by computing the differences between the positive and negative sentiments for each sub-category as presented in Figure 5. As can be seen from this graph, the Covid-19 with Gallstones and Covid-19 with heart failure sub-category have the maximum negative overall score, with the sub-category Covid-19 with gallstones having the second maximum negative score, and the sub-category Covid-19 with heart failure having the third maximum negative. In dissimilarity, the Covid-19 with hypertension sub-category has the maximum positive overall score.

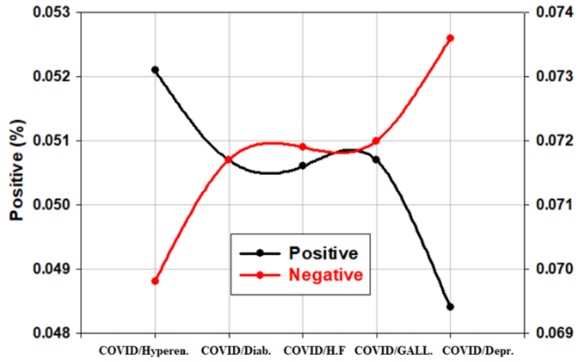

Figure 4. SentiWordNet percentages scores of Table 3

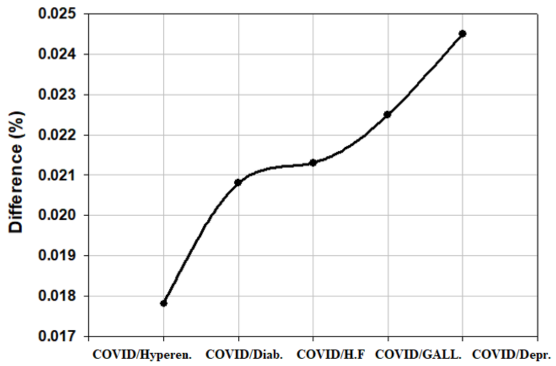

Figure 5. The difference of the negative and positive scores each discharge summary sub-category notes

5. RESULTS AND DISCUSSIONS

Word2Vec is collection of unsupervised narrow two-layers neural network models realized in [23]. Word inserting are the numeric demonstration of words in the form of vectors. Word2Vec provides word inserting based on the relative semantics of words in a text. Words with similar semantic contexts are mathematically assembled together in a vector planetary, which conserves the semantic connection between words. Word2Vec can then use these words inserting to provide estimations on a word's meaning. Additionally, each sub-category will associate with single ID as listed in Table 4. Table 5 shows the results terms between each sub-category similarity. It can be clearly noticed that all the terms similarity of the scores are quite similar to each other with maximum values of 0.99. That means the terms similarity in each sub-category with respect to all others is very high. The highest term similarity with dataset is sub-category 1 (Covid-19 with Hypertension) and sub-category 3 (Covid-19 with Diabetes) with similarity term of 0.9997. The lowest term similarity with dataset is sub-category 4 (Covid-19 with Gallstones) and sub-category 5 (Covid-19 with Depression) with similarity term of 0.9722. For more clarification, the highest term similarity is highlighted in blue color and the lowest term similarity is highlighted in red color.

Table 4. ID of the datasets

Data name	ID
Covid-19 with Hypertension	1
Covid-19 with Diabetes	2
Covid-19 with Heart failure	3
Covid-19 with Gallstones	4
Covid-19 with Depression	5

Table 5. Similarity between subsets by using Word2Vec method

IDs	1	2	3	4	5
1	0.9997	0.9899	0.9955	0.9966	0.9840
2	0.9877	0.9788	0.9655	0.9750	0.9700
3	0.9920	0.9850	0.9740	0.9744	0.9788
4	0.9980	0.9890	0.9870	0.9770	0.9780
5	0.9822	0.9722	0.9863	0.9772	0.9821

Then, to validate the Word2Vec sentiment scores results, a statistical analysis is performed in order to calculate whether the results are statistically significant or not. Welch's t-test on the SentiWordNet sentiment scores is implemented with two independent samples between each pair of sub-categories. Generally, Welch's t-test is performed as a deviation of the Student's t-test. Welch's t-test has better performance when it uses in comparing an unequal sample sizes and variances. This is a good approach to use Welch's t-test since all the five sub-categories are not equal in sample size. Hence, the Welch's t-test has been performed for each pair of sub-categories with positive, negative, and overall sentiment scores.

The Welch's t-test performance is measured with its p-value. Each value of tested p-value is represented the finding probability of the experimental results if the hypothesis (Null) is True. It means that the p-value represents the null hypothesis being rejected in the smallest level of significance. The SciPy statistics module is used in the Welch's t-test function. Table 6 lists the results of the positive sentiment Welch's t-test with its p-value. The minimum value for each test is highlighted in Red Color. Sub-category 2 (Covid-19 with Diabetes) and Sub-category 5 (Covid-19 with Depression) have the minimal p-value of 0.282. This means that p-value is high, which indicates that 28.2% of the time the hypothesis (null) is accepted or 71.8% is statistically significant. Table 7 lists the results of the negative sentiment Welch's t-test with its p-value. The minimum value for each test is highlighted in Red Color. Sub-category 2 (Covid-19 with Diabetes) and Sub-category 3 (Covid-19 with Heart failure) have the minimal p-value of 0.180. This means that p-value is high, which indicates that 18% of the time the hypothesis (null) is accepted or 82% is statistically significant.

Table 6. The results of the positive sentiment Welch's t-test with its p-value

ID	1	2	3	4	5
1	0.954	0.458	0.685	0.996	0.491
2	0.657	0.765	0.965	0.475	0.282
3	0.758	0.358	0.774	0.574	0.778
4	0.942	0.924	0.887	0.777	0.378
5	0.741	0.282	0.686	0.877	0.7821

Table 7. The results of the negative sentiment Welch's t-test with its p-value

ID	1	2	3	4	5
1	0.965	0.458	0.685	0.458	0.996
2	0.774	0.765	0.180	0.765	0.475
3	0.887	0.180	0.574	0.358	0.774
4	0.887	0.924	0.777	0.924	0.887
5	0.686	0.947	0.877	0.741	0.686

Table 8 lists the results of the overall sentiment Welch's t-test with its p-value. The minimum value for each test is highlighted in Red Color. Sub-category 4 (Covid-19 with Gallstones) and Sub-category 5 (Covid-19 with Depression) have the minimal p-value of 0.054. This means that p-value is quite high, which indicates that 5.4 % of the time the hypothesis (null) is accepted or 94.6% is statistically significant. From these above results, it can be concluded that the sentiment analysis using Word2Vec is significantly efficient with accuracy more than 90%.

Table 8. The results of the overall sentiment Welch's t-test with its p-value

ID	1	2	3	4	5
1	0.565	0.158	0.685	0.258	0.796
2	0.274	0.365	0.470	0.365	0.275
3	0.487	0.270	0.174	0.758	0.974
4	0.857	0.824	0.377	0.824	0.054
5	0.636	0.647	0.677	0.054	0.386

A comparison is done with other similar works in terms of p-value and accuracy. Figure 6a shows a good result for this proposed method compared to others with high p-value of 0.054. This is lead to significant high accuracy of more than 90% compared to other method used as shown in Figure 6b. Hence, using Word2Vec method in retrieving medical records is a good approach for analyzing and making opinion based on health electronic record. Table 9 shows the comparison of this work with other related researches in the same classifications.

6. CONCLUSIONS

In this paper, the Covid-19 health records sentiment retrieval analysis based on unsupervised learning NLP Word2Vec method is presented. The study is performed over 25000 Health records collected from medical institutions located in Iraq during May 2021 to May 2022. Word2Vec NLP method is a suitable tool to distinguish sentiment scores such as positive and negative dataset terms. Five different sub-classes associated with specific diseases are analyzed statistically using SentiWordNet base on Word2Vec model for Covid-19 health records. Welch's t-test is performed to evaluate the statistical significance of the sentiment scores results. The sentiment scores results showed that the Covid-19 with Hypertension sub-category has the highest positive sentiment score and Covid-19 with Gallstones sub-category is the lowest negative sentiment score. The p-value obtained from Welch's t-test showed a highest value of 0.054, which indicates that our sub-category is statistically significant by 94.6%.

Table 9. Compression of this work with related state of art approaches

Study	Methods	Approaches	Sample size	P-value	Accuracy
[26]	Hybrid	Opinion, AFINN	MIMIC database	0.033, 0.412	62%, 68%
[27]	semi-supervised	Random Fields (CRF)	500	0.1	60%
[28]	Unsupervised	CNN (SVM)	5000	0.37	58%
[29]	Unsupervised	Word2Vec	35000	0.46	75%
[30]	semi-supervised	ML algorithms	3500	0.080	69%
This work	Unsupervised	Word2Vec	25000	0.054	93%

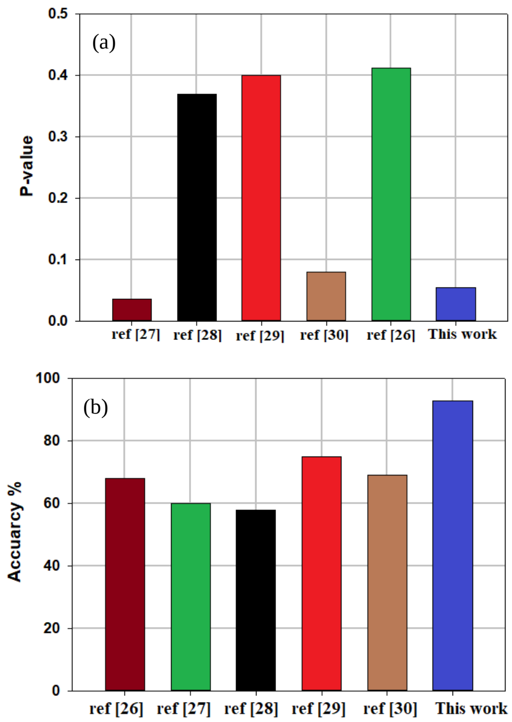

Figure 6. Comparison in terms of obtained results with other related works. (a) P-value, (b) Accuracy of the method

ACKNOWLEDGEMENTS

The authors would like to express their grateful for University of Information Technology and Communication and University of Baghdad, Baghdad, Iraq for supporting of this research work.

REFERENCES

[1] S. Gohil, S. Vuik, A. Darzi, "Sentiment Analysis of Health Care Tweets: Review of the Methods Used", JMIR Public Health Surveill, Vol. 4, No. 2, pp. 43-50, April 2018.

[2] L. Lizhen, S. Wei, W. Hanshi, L. Chuchu, L. Jingli, "A Novel Feature-Based Method for Sentiment Analysis of Chinese Product Reviews", China Communications, Vol. 11, Issue 3, pp. 154-164, March 2016.

[3] M. Azam Zia, Z. Zhang, L. Chen, M. Hashim, S. Su, "Exploration of Influential People for Viral Marketing", China Communications, Vol. 15, Issue 5, pp. 138-148, May 2018.

[4] V. Raghupathi, Y. Zhou, W. Raghupathi, "Exploring Big Data Analytic Approaches to Cancer Blog Text

Analysis", International Journal of Healthcare Information Systems and Informatics, Vol. 14, pp. 20-32, April 2019.
[5] L. Zhang, B. Liu, "Sentiment Analysis and Opinion Mining", Encyclopedia of Machine Learning and Data Mining, Vol. 1, pp. 1152-1161, April 2017.
[6] S. Park, S.H. Hong, "Identification of Primary Medication Concerns Regarding Thyroid Hormone Replacement Therapy from Online Patient Medication Reviews: Text Mining of Social Network Data", Journal of Medical Internet Research, Vol. 20, Issue 10, pp. 85-95, October 2018.
[7] Y. Lu, Y. Wu, J. Liu, J. Li, P. Zhang, "Understanding Health Care Social Media Use from Different Stakeholder Perspectives: A Content Analysis of an Online Health Community", Journal of Medical Internet Research, Vol. 19, Issue 4, pp. 87-97, April 2017.
[8] W3C Web Services Activity (2007) Retrieved March 3rd, 2007 from W3C Architectural Domain: www.w3.org/2002/ws/.
[9] Informatics for Integrating Biology and the Bedside (2007) Retrieved March 15th, 2007 from: www.i2b2.org.
[10] S.N. Murphy, M.E. Mendis, D.A. Berkowitz, I. Kohane, H.C. Chueh, "Integration of Clinical and Genetic Data in the i2b2 Architecture", AMIA Annu Symp Proc, 2006.
[11] E. Holderness, P. Cawkwell, K. Bolton, "Distinguishing Clinical Sentiment: The Importance of Domain Adaptation in Psychiatric Patient Health Records", The 2nd Clinical Natural Language Processing Workshop, pp. 117-123, Minnesota, USA, June 2019.
[12] K. Sugathadasa, B. Ayesha, N. de Silva, A. Shehan Perera, "Synergistic Union of Word2Vec and Lexicon for Domain Specific Semantic Similarity", arXiv, Cornell University, Vol. 2, June 2017.
[13] I.E.R. Waudby, N. Tran, J.A. Dubin, J. Lee, P. van Bogaert, "Sentiment in Nursing Notes as an Indicator of Out-of-Hospital Mortality in Intensive Care Patients", PLOS One, Vol. 13, Issue 6, pp. 1661-1671, June 2018.
[14] S. Sarabi, M. Asadnejad, S.A. Tabatabaei Hosseini, S. Rajebi, "Using Artificial Intelligence for Detection of Lymphatic Disease and Investigation on Various Methods of its Classifications", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 43, Vol. 12, No. 2, pp. 58-65, June 2020.
[15] A.H. Odeh, M.A. Odeh, "Increasing the Efficiency of Online Healthcare Services Software and Mobile Applications using Artificial Intelligence Technology", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 44, Vol. 12, No. 3, pp. 16-22, September 2020.
[16] Y. Haddi, A. Moumen, A. Kharchaf, "Study of A Mobile Robot's Obstacle Avoidance Behavior in A Radioactive Environment with A High Level of Autonomy", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 50, Vol. 14, No. 1, pp. 34-41, March 2022.
[17] A. Yadollahi, A.G. Shahraki, O.R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining", ACM Computing Surveys (CSUR), Vol. 50, Issue 2, pp. 25-35, February 2018.
[18] A. Rghioui, "Managing Patient Medical Record using Blockchain in Developing Countries: Challenges and Security Issues", The IEEE International Conference of Moroccan Geomatics (Morgeo), pp. 1-6, Casablanca, Morocco, May 2020.
[19] V. Chamola, V. Hassija, V. Gupta, M. Guizani, "A Comprehensive Review of the Covid-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact", The IEEE Access, Vol. 8, pp. 90225-90265, May 2020.
[20] F. Gao, K. Deng, C. Hu, "Construction of TCM Health Management Model for Patients with Convalescence of Coronavirus Disease Based on Artificial Intelligence", International Conference on Big Data and Informatization Education (ICBDIE), pp. 417-420, Zhangjiajie, China, November 2020.
[21] Q. Pham, D.C. Nguyen, T. Huynh The, W. Hwang, P. N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (Covid-19) Pandemic: A Survey on the State-of-the-Arts", IEEE Access, Vol. 8, pp. 130820-130839, July 2020.
[22] F. Shi, et al., "Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for Covid-19", The IEEE Reviews in Biomedical Engineering, Vol. 14, No. 4, pp. 4-15, January 2021.
[23] S. Allen, "Artificial Intelligence and the Future of Psychiatry", The IEEE Pulse, Vol. 11, No. 3, pp. 2-6, May-June 2020.
[24] A. Zunic, P. Corcoran, I. Spasic, "Sentiment Analysis in Health and Well-Being: Systematic Review", The JMIR Medical Informatics, Vol. 28, No. 8, pp. 23-30, January 2020.
[25] S. Yang, C. Li, T. Huang, "Exponential Stabilization and Synchronization for Fuzzy Model of Memristive Neural Networks by Periodically Intermittent Control", Neural Networks, Vol. 75, pp. 162-172, Chongqing, China, March 2016.
[26] G.E. Weissman, L.H. Ungar, M.O. Harhay, K.R. Courtright, S.D. Halpern, "Construct Validity of Six Sentiment Analysis Methods in the Text of Encounter Notes of Patients with Critical Illness", Journal of Biomedical Informatic, Vol. 89, pp. 114-121, June 2019.
[27] K.P. Chodey, G. Hu, "Clinical Text Analysis Using Machine Learning Methods", The IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), pp. 1-6, Okayama, Japan, June 2016.
[28] Y. Wang, S. Sohn, S. Liu, F. Shen, L. Wang, J. Atkinson, S. Amin, H. Liu, "A Clinical Text Classification Paradigm Using Weak Supervision and Deep Representation", BMC Medical Informatics and Decision Making, Vol. 19, No. 1, pp. 12-27, January 2018.
[29] J. Conrad, J. Chris, "Machine Learning in Medicine: A Practical Introduction to Natural Language Processing", BMC Medical Research Methodology, Vol. 21, No. 1, pp. 47-58, July 2021.
[30] G. Trivedi, "Clinical Text Analysis Using Interactive Natural Language Processing", The 20th International Conference on Intelligent User Interfaces Companion, pp. 113-116, New York, USA, March 2015.

BIOGRAPHIES

Ahmed M. Taha was born in Baghdad, Iraq, on March 2, 1984. He received the B.Sc. degree in Software Engineering from Al-Rafidian University, Baghdad, Iraq in 2007 and the Master degree in Computer Science from National University of Malaysia, Bangi, Malaysia in 2011 and the Ph.D. degree in Artificial Intelligence from University of Tenaga Nasional, Kajang, Malaysia in 2014. Currently, he is a senior lecturer at Businesses Informatics College, University of Information Technology and Communication, Baghdad, Iraq. He is also member of Soft Computing and Data Mining Center, University Tun Hussein Onn Malaysia. His research interest includes data mining, metaheuristics, algorithms and machine learning.

Saadya F. Jabbar was born in Baghdad, Iraq, on October 1, 1971. She received the B.Sc. degree in Computer Science from University of Baghdad, Baghdad, Iraq in 2003 and the Master degree in Computer Science from University of Al-Mustansiriyah, Baghdad, Iraq in 2015. Currently, she is a senior lecturer at Collage of Education, University of Baghdad. Her research interests are in data optimization, machine learning and natural language processing.

Asmaa H. Alwan was born in Baghdad, Iraq, on October 10, 1967. She received the B.Sc. degree in Computer Science from Al-Rafidian University, Baghdad, Iraq in 1994 and the Master degree in Operation Research Science from University of Baghdad, Baghdad, Iraq in 2003. Currently, she is a senior lecturer at Collage of Education, University of Baghdad. Her research interests are in AI technology, machine learning in data mining, image processing, security algorithms for key encryption, and natural language processing.