

Original Article

# Sentiment Classification in Medical Care with Psychometric Analysis Using Emotion Detection

Muddada Murali Krishna<sup>1</sup>, Vankara Jayavani<sup>2</sup>, Pooja Gotety<sup>3</sup>

<sup>1,2</sup>Assistant Professor, Department of Computer Science Engineering, Dr Lankapalli Bullayya College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

<sup>3</sup>Department of Electronics and Communication Engineering, Dr Lankapalli Bullayya College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

Received Date: 12 February 2020

Revised Date: 24 March 2020

Accepted Date: 27 March 2020

**Abstract** — The advancement of technology in the present era is accelerating exponentially, which leads to an increase in peer competition, mental tension and different mental problems like depression, schizophrenia, different disorders etc. So, a need for psychometric analysis is felt. Emotion recognition and sentiment analysis have gained a high level of popularity in research in social networking, but they have not been applied to the complicated problems of healthcare. But candidly speaking, both domains have great potential in solving some complex and interesting problems in medical science and engineering technology. This paper introduces a data science application, which acts as a psychometric analysis using the concept of Emotion recognition and sentiment analysis.

**Keyword** – Emotion Recognition, Sentiment Analysis, Schizophrenia, Natural Language Processing and Machine Learning.

## I. INTRODUCTION

Emotion Recognition and sentiment analysis is a trending research field that uses many machine learning algorithms on different types of data and its analysis. Different modalities of sentiment analysis are text [2] [3], facial recognition and speech, biosensors and thermal sensors. But the main drawback is with the accuracy of these different techniques. Nowadays, the advancements have reached a level where automated sentiment analysis software is designed for automated replies to tweets [4], complaints, customer reviews etc., but the accuracy rate is quite low and not much precise. In the era of big data, some of the major problems that computational linguistics are facing are time, accuracy etc. Handling such vast corpora is possible only if different algorithms are combined together efficiently. This paper proposes a model in the form of an algorithm, which combines the concept of Sentiment and Emotion analysis [5]. This application can

be used by psychologists and patients for solving many problems related to health on a reduced time-saving basis. This model (analyser) will use speech oriented emotional analysis [6] handwritten and typed text-based sentiment analysis. The proposed system in the form of a model is a result of a survey of current and previous technological and research advancements. The model works by the acquisition of data, followed by an analysis of typed mixed and handwritten data [7] and recorded voice in the form of a dataset. Then the sentiment analysis based on text and speech is performed, and finally, data is visualized. Finally, the analyzed psychological data is graphically represented [8].

## II. LITERATURE REVIEW

### A. Background

Already been stated before that the fields of Sentiment Analysis, handwritten/typed text analysis, Emotion Recognition NLP have deeply been researched. The literature of the same is presented in this section. M.A.M. Abhirami et al. explain the complete study of all the techniques of the sentiment analysis. Their research provides an efficient summary of all the methods available for analysis. P. Barlas et al.[7] has explained the various algorithms for text information extraction from the typed/mixed type and handwritten files. The preprocessing of voice signals based on neural networks is well studied by G.K. Berdibaeva [10]. S. Casale et al. [6] explained the concept of various techniques of Emotion Recognition using ML. K.Denecke et al. [3] studied the art of Emotion Recognition and Sentiment Analysis in the field of healthcare and medicines. K. Duretec et al. [11] studied an extraction and designed a tool that can be associated with various user-friendly devices. Forensic data analysis using a digital dataset is presented by G.Hales et al. [8]. The work presented by [8] presents a milestone in the area of forensic science. A proper algorithm for a visualization based on trigonometric components of splines constituting



the fundamentals of data visualization was proposed by F.Ibraheem et al. [12]. S.Kisan et al. [4] studied the sentiment analysis result of social media websites based on Stanford NLP libraries, thus designing software that can auto-reply the tweets. The concept of handwritten and typed mixed text analysis was explained by Louloudis et al. [13]. T.D. Nguyen et al. [5] proposed the sentiment analysis of a medical text. Their method used a Machine Learning algorithm. S. Niklander et al. has presented a combined approach of sentiment analysis and human interpretation for sentiment analysis. Prime research was done by A.Sarkar et al. [14] on polarity analysis of a medical document. K. Saeed et al. [15] worked on handwritten and typed text analysis.

P.H. Sahana et al. [9] studied and presented sentiment analysis based on various feature sets providing it with a new depth. Z.Shi et al. [16] proposed some of the best methods for text extraction from handwritten documents with an efficiency of 99%. R. Shikder et al. studied the emotion analysis based on keystrokes and mouse usage, which was quite a new concept. M. Soua et al. [17] proposed a real-time methodology of text extraction. Proper alignment text extraction with proper reality from various transcripts was proposed by N.Stamatopoulos et al. [18]. S. Tokuno et al. [19] have described the use of emotion analysis in the military and general health using the concept of voice recognition. A proper technique (algorithm) was proposed by P.Yerpude et al. in their article "Algorithm for Text to Graph Conversion and Summarizing Using NLP: A New Approach for Business Solutions". They implemented using Python Libraries. Anneketh Vij et al. [2] very nicely presented an analyser combining the concepts of Sentiment Analysis and Emotion Recognition.

### B. Problem Statement

Many a time, it happens that when a patient visits a doctor, he has to a doctor he has to repeat his problem again and again to the doctor. The case becomes very painful when the patient is suffering from a nervous breakdown etc. So this problem is solved by designing an automated kiosk using sentiment analysis and voice recognition, which helps the doctor in knowing the condition of the patient.

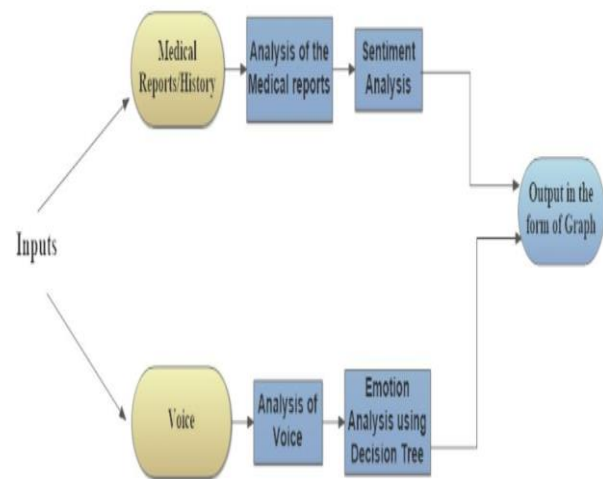


Fig. 1 Flow of the process

### III. PROPOSED MODEL

Our model (Psychometric Analysis) is an extension of already existing concepts of ML with little modifications. The aim of this study is to design an analyser (psychometric analyser) that results in an emotional and sentimental analysis of person health based on historical medical and recorded voice. Any context of sentiment analysis can be kept under one of the following components.

- Intensity Analysis (Extreme, High, Normal, Low).
- Emotion Analysis (Anger, Happy, Disgusting, Joy, Love, Hatred, etc.).
- Polarity Analysis (Good, Bad).
- Subjectivity Analysis (Opinion based Biased/ Unbiased).

The total modelling of the analyzer consists of three different steps, namely Data Acquisition, Data Analysis, and Data Visualization. Out of three, Data Visualization is the output phase.

#### A. Data Acquisition

Our model uses the concepts of Machine Learning. Hence, we have to train the system (using both supervised and unsupervised learning algorithms) by pre-feeding the data. The model uses the concepts of supervised learning (SVM or any classifier) and unsupervised learning (k-means clustering). The input data is collected through web mining (WebMD, Netmeds.com, drugs.com etc.) because it is easy and legal to get. Now, this data will be used to process the standard unseen data. But the system trained on these training data may not respond efficiently for acronym data, frequently used by a psychologist. For this reason, the acronym data is collected by an online survey form filled out by various psychologists. Some words creating ambiguity in the context of sense based on medical and English terms are also included by including medical and English dictionaries. It removes the ambiguity in the usage of the words. Apart from that, some grammatical data (noun, adjectives, adverbs etc.) are also fed into the system for intensity and emotion analysis. To

make the system robust, it is also fed with speech data (voice).

**B. Data Analysis**

This step is the most complicated step in designing the model. It takes the input in two different forms, namely medical history/medical reports and recorded voice. Finally, the output is generated in the form of a graph. This is clearly shown in figure1.

**a) Mixed Text Analysis (Typed text and Handwritten analysis)**

The first step of the analysis is to analyze the medical reports /history, prescription, reports of radiology which are in 80:20 ratios of handwritten and typed text. So, this analysis can be understood as a complex analysis [7]. If the analysis is done with the best possible algorithm, the final result will be more efficient. For this, we can perform text line segmentation by CUBS method or Hough transform [20] followed by segmentation of the word LSP-LWSeg-09 to get the best possible result. CUBS method is a technique of text line segmentation method based on line separation algorithm while the best possible result. CUBS method is a technique of text line segmentation method based on line separation algorithm while LSP-LWSeg-09 on Support Vector Machine. Using this method, the text is separated from the document, and the typed text is separated from written text and the white spaces. Now the white free space is masked, and after masking, the handwritten text is transformed into standard typed text according to an available dataset.

**C. Voice Analysis**

To visualize the analysis more accurately, we use voice data for emotion analysis. The voice will be recorded using the noise cancellation recorders, which results in faster sentiment analysis. The voice that is to be analyzed will be analyzed based on the normal speech frequency (fn), Normal Intensity (In) determined by using the K-means clustering algorithm. The different clusters formed will help to determine the overall state of emotions based on fear, joy, happiness, and stress using Decision Trees.

**D. Sentiment Analysis of Data in Text Format**

The text contains some emotions which need to be classified for sentiment analysis. Support Vector Machine is used for the sentiment analysis of the text. However, regression algorithms are helpful in finding the emotional intensity. Before applying any algorithm to the text, it is subjected to its preprocessing. In preprocessing phase, all the lowercase letters are changed to uppercase letters and unwanted punctuations are removed. Once the preprocessing phase is over, the N-GRAMS concept is taken into consideration. RANC (Relative Average Negative Count) is based on a number of negatives and their positions. The formula for RANC is given by

$$RANC_{it} = \sum_{n=1}^i \left( n g_{it} \times \frac{n}{i} \right) + t^{-1}$$

Here RANC is Relative Average Negative Count for document m. n is the negative term in  $i^{th}$  sentence, and t is the exact total number of sentences. The accurate result in RANC is obtained when a combination of unigrams and bigrams are considered. The medical terms are replaced by token M and their presence by P. Actually, and all medical terms are not replaced by M rather diseases by D and rest by na. Now the polarity analysis will be easier. For the determination of the intensity, the grammatical dataset is very useful if it is implemented using the technique of SVM. The intensity is determined by using the grammatical word data. If the degree of the adjective is first degree, then the intensity will be low when the degree of the adjective is second, then the intensity will be more, and if the adjective is of third-degree, then the intensity will be high. For example, worst is more negative than worse, and worse is more than b a d .

**E. Data Visualization**

Now, after the data processing phase, the data has to be converted in the form of graphs for the representation of the output. The graphical representation allows better visualization. Any programming language can be used for graphical representation. It can be plotted by using Python language using the class Pyplot(). Matlab can also be used for visualization purposes. It is one of the easiest steps.

**IV. RESULTS AFTER VISUALIZATION**

The results shown after the sentiment analysis is shown below. It clearly depicts the keyword, outcome polarity and outcome intensity. However, the final result is presented in the form of a graph.

Words Occurrence	Outcome Polarity	Outcome Intensity
Terminal Illness	Negative	10
Very Critical	Negative	10
Good Immunity	<b>Positive</b>	8
AIDS Negative	Positive	7
Sore Throat	Neutral	0
Fungal infection for last 1.5 years	Negative	4
Pneumonia	Neutral	0
High BP and high glucose level	Negative	10
High BP and normal glucose level	Negative	7
Seasonal Cold	Negative	1
BP under control on medication	Positive	8
Body ache	Neutral	7
AIDS Positive	Negative	10

The table above depicts the context clearly. It is

obtained after the implementation of the model. The proper graphs can be generated according to the doctor's needs. The graphs can be any form, i.e. bar graphs, histograms, pie charts, box plots etc. The analysis of the polarity is determined based on RANC [14] technique, while the intensity analysis is done based on the rich pre-

acquired dataset. The dataset was a mixed dataset, i.e. in the form of text and speech. The table concentrated on the medical history of the patients, medical records. The final graph is generated based on the table obtained.

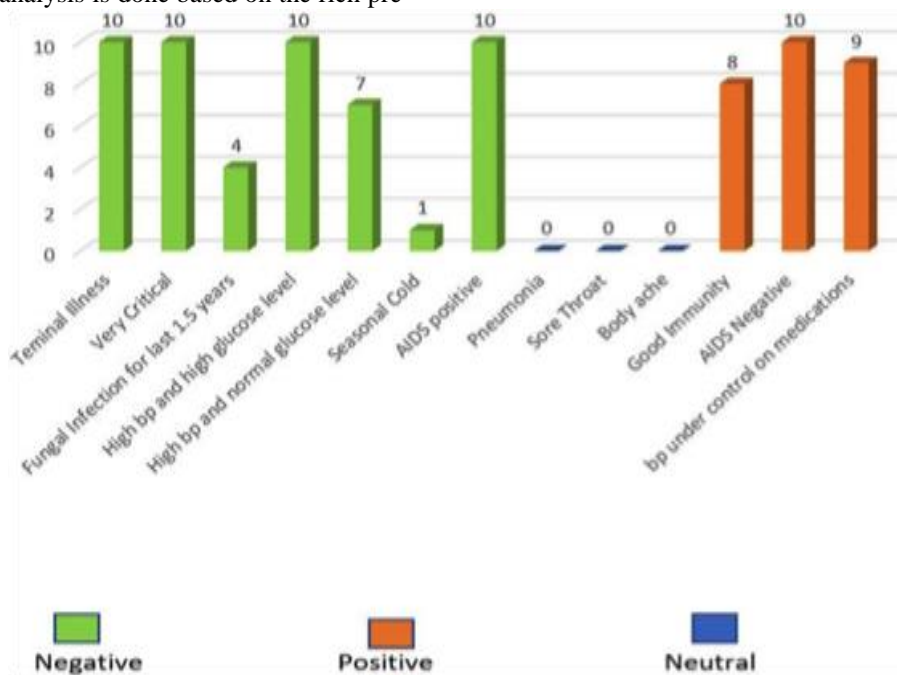


Fig. 2 Graphical representation for Sentiment Analysis

## V. CONCLUSION AND FUTURE SCOPE

The main aim of this article is to design a self-servicing psychometric model or system which is efficient enough of fast computation, thereby creating a fast summary for the health of the patient on the context of emotional ground-based on previous records, medications and accordingly treatments resulting thus saving the time for both patient and doctor. Every time a patient visits a doctor, it is very painful for the patient to say everything to the doctor about their medical history, especially when the patient is suffering from nervous disturbances or depression. Apart from it for doctors, it is quite a time consuming to go through the medical records, medical examinations and different reports. So it can be solved by using this sentiment analysis technique by a psychometric analyzer. The future work of this can be considered as to optimize the time-accuracy-cost so that the result can be found in less time with higher efficiency.

## VI. REFERENCES

- [1] M. Dooley and C. Dickinson, The microbiology of cut-away peat, *Plant and Soil*, 32( 1-3) (1970) 454–467.
- [2] Vij and J. Pruthi, An automated psychometric analyzer based on sentiment analysis and emotion recognition for healthcare, *Procedia Computer Science*, 132 (2018) 1184–1191.
- [3] K. Denecke and Y. Deng, Sentiment analysis in medical settings: New opportunities and challenges, *Artificial intelligence in medicine*, 64(1) (2015) 17–27.
- [4] H. S. Kisan, H. A. Kisan, and A. P. Suresh, Collective intelligence & sentimental analysis of Twitter data by using Stanford nlp libraries with software as a service (saas), in *Computational Intelligence and Computing Research (ICCRIC)*, IEEE International Conference on, (2016) 1–4.
- [5] T. D. Nguyen, L. D-P. Nguyen, and T. Cao., Sentiment analysis on medical text using a combination of machine learning and so-cal scoring, in *Intelligent and Evolutionary Systems (IES)*, 2017 21<sup>st</sup> Asia Pacific Symposium on., (2017) 49–54.
- [6] S. Casale, A. Russo, G. Scebba, and S. Serrano, Speech emotion classification using machine learning algorithms, in *The IEEE International Conference on Semantic Computing*, (2008) 158–165.
- [7] P. Barlas, S. Adam, C. Chatelain, and T. Paquet, A typed and handwritten text block segmentation system for heterogeneous and complex documents, in *Document Analysis Systems (DAS)*, 11th IAPR International Workshop on, (2014) 46–50.
- [8] G. Hales, Visualisation of device datasets to assist the digital forensic investigation, in *Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, International Conference On, (2017) 1–4.
- [9] P. Shahana and B. Omman, Evaluation of features on sentimental analysis, *Procedia Computer Science*, 46 (2015) 1585–1592.
- [10] G. K. Berdibaeva, O. N. Bodin, V. V. Kozlov, D. I. Nefed'ev, K. A. Ozhikenov, and Y. A. Pizhonkov, Pre-processing voice signals for voice recognition systems, in *Micro/Nanotechnologies and Electron Devices (EDM)*, 18th International Conference of Young Specialists on, (2017) 242–245.
- [11] K. Duretec, A. Rauber, and C. Becker, A text extraction software benchmark based on a synthesized dataset, in *Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries*, (2017) 109–118.
- [12] F. Ibraheem and M. Z. Hussain, Visualization of constrained data using trigonometric splines, in *Information Visualisation (IV)*, 21st International Conference., (2017) 400–404.
- [13] B. Gatos, N. Stamatopoulos, and G. Louloudis, Icdar2009 handwriting segmentation contest, *International Journal on Document Analysis and Recognition (IJ DAR)*, 14(1)(2011) 25–33.

- [14] A. Sarker, D. Molla'-Aliod, C. Paris, et al., Outcome polarity identification of medical papers, (2011).
- [15] K. Saeed and M. Albakoor, Region growing based segmentation algorithm for typewritten and handwritten text recognition, *Applied Soft Computing*, 9(2) (2009) 608–617.
- [16] Z. Shi, S. Setlur, and V. Govindaraju, A steerable directional local profile technique for extraction of handwritten Arabic text lines, in *Document Analysis and Recognition. ICDAR'09. 10th International Conference on*, (2009) 176–180.
- [17] M. Soua, A. Benchekroun, R. Kachouri, and M. Akil, Real-time text extraction based on the page layout analysis system, in *Real-Time Image and Video Processing 2017*, International Society for Optics and Photonics, 10223 (2017)1022305.
- [18] N. Stamatopoulos, G. Louloudis, and B. Gatos, Efficient transcript mapping to ease the creation of document image segmentation ground truth with text-image alignment, in *Frontiers in Handwriting Recognition (ICFHR), International Conference on*, (2010) 226–231.
- [19] S. Tokuno, G. Tsumatori, S. Shono, E.Takei, T. Yamamoto, G. Suzuki, S. Mituyoshi, and M. Shimura, , Usage of emotion recognition in military health care , *Defense Science Research Conference and Expo (DSR)*, (2011) 1–5.
- [20] G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis, Text line and word segmentation of handwritten documents, *Pattern Recognition*, 42(12) (2009) 3169–3183.