Improving Medical Consultation Process through Natural Language Processing Technique

Bolanle F. Oladejo

Abstract—The conventional medical consultation process is prone to the problem of time consumption, on the part of consultants, during the survey of report of a patient's medical history, for diagnosis or treatment purpose. Sequel to the volume of this report, and human error, effective health service delivery is being inhibited. The application of computational techniques, especially Artificial Intelligence algorithms to medical contexts cannot be undermined. Thus, this work develops a system tagged Patient-Centred Medical Consultation System (PCMCS) which applies a Natural Language Processing (NLP)-based algorithm to analyze medical history of a patient in order to reduce the shortcomings of the conventional consultation process. The system specification is modeled and designed with the aid of UML tools. The NLP algorithm is a text processing technique namely, textRank which ranks and sorts all sentences in an input corpus of a patient case report accordingly to get the summarized version of the text. The TextRank algorithm is able to generate a summary of the patient record based on the input corpus. The implemented PCMCS can be deployed to the Electronic Health Records (EHR) used in hospitals to assist in the summarization of patient records. This technique therefore reduces the effort expended by the medical practitioner in comprehending a case report, since s/he is able to view the salient parts promptly. Consequently, resulting to a relatively easier consultation process and reduction of waiting time for both doctors and patients. Cosine similarity is used as the metric to evaluate the accuracy of the output from the algorithm compared to the initial report.

Index Terms— Consultation, Doctor, EHR, NLP, Patient, PCMCS, Ranking, TextRank.

1 INTRODUCTION

THE World Health Organisation (WHO) claims that better health is a prerequisite to the happiness and well-being of humans. It is also pertinent to economic development as healthy populations report a longer life-span, higher productivity and higher savings rate [1].

The medical practice and human health are so intertwined that one cannot make mention of one without the other. Medicine is the application of scientific knowledge to human health and doctors are one important agent through which that scientific understanding is expressed [2].

One has to know where he is coming from in order to know where he is headed. The understanding of a patient's medical history helps to assure that the medical doctor and the individual's health care providers provide the most appropriate and effective treatment and support for the individual's illnesses and health conditions so that the latter maintains the best possible health.

A patient's medical history is made up of many different pieces of information that tells the complete story about that patient's current and past health status. A patient's medical history is very vital as it helps the physician to know what medications he or she is placed on, nature of allergies, and diagnoses to treat the person in an optimal way. The import of the relevance of well comprehended medical history to an effective medical consultation serve as motivation for this work. Hence, it aims at the development of a medical consultation support system based on Natural Language Processing (NLP) technique called Patient-Centred Medical Consultation System (PCMCS).

The paper is grouped into five sections. It is introduced in section 1 and a review of theoretical background with related works follows in section 2. The methodology is presented in section 3 and section 4 discusses the result of implementation and its evaluation. Lastly, it is concluded in section 5.

2. RELATED LITERATURE

This section provides theoretical background and existing works related to this work. Patient-Centred Medical Consultation System (PCMCS) has its background in emerging fields in Computer Science which include Data Science, Data Mining, Artificial Intelligence, Machine learning, Natural Language Processing and Clinical Decision Support System, hence the review in the following section.

2.1 Theoretical Background

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from data in both structured and unstructured forms [3]. It unifies statistics, data analysis, machine learning, and their related methods in order to "understand and analyze actual phenomena" with data.

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. The overall goal of which is to extract information from a data set, and transform it into an understandable structure for further use through data analysis. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD.

Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. Some of the activities computers with artificial intelligence are designed for include Problem-solving, Speech recognition, Learning, Planning among others [4].

Machine learning is an application of AI that provides systems the ability to automatically learn without human intervention, and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves [5].

Natural Language Processing (NLP) is the process of using computer algorithms to identify key elements in everyday language and extract meaning from unstructured spoken or written input. NLP requires skills in artificial intelligence, computational linguistics, and other machine learning disciplines. Some NLP efforts are focused on mimicking human-like responses to queries or conversations. Others try to understand human speech through voice recognition technology, such as automated customer service applications used by many large companies.

Many NLP systems "learn" over time, reabsorbing the results of previous interactions as feedback about which results were accurate and which did not meet expectations. These machine learning programs can operate based on statistical probabilities, which weigh the likelihood that a given piece of data is actually what the user has requested. Based on whether or not that answer meets approval, the probabilities can be adjusted in the future to meet the evolving needs of the end-user [6].

There are diverse applications of NLP systems, ranging from machine translation, named-entity extraction, speech recognition, topic segmentation, sentiment analysis, question-answer session, to automatic summarization among others. Text summarization is the process of producing a concise and coherent versions of a document while preserving the most important information and the overall meaning [7].

Clinical decision support system (CDSS) is an application that analyses data to help healthcare providers make clinical decisions. Physicians, nurses and other healthcare professionals use a CDSS to prepare a diagnosis and to review the diagnosis as a means of improving the final result. Data mining may be conducted to examine the patient's medical history in conjunction with relevant clinical research. Such analysis can help predict potential events, which can range from drug interactions to disease symptoms [8].

2.2 Review of Related Works

In order to analyse the usefulness and technicality behind the development of Text Mining and NLP systems, there is a need for a thorough and careful look at works done by experts and authors who have specialized in Text Mining and Natural Language Processing across different fields, therefore, this paper briefly reviews a number of related works.

In [9], data-to-text summarisation of patient records using computer-generated summaries to access patient histories is developed. It aimed at assessing the efficiency and accuracy of automatically generated textual summaries of patients' medical histories. A computational system (a Report Generator) is developed to produce a range of summarised reports of patient records from patient histories derived from a repository of medical records of cancer patients and composed of narrative documents. Doctors who were tested based on the summaries generated from this system showed that with the summarized text, they are able to quickly understand the nature of presenting illness of patients in question. However, it has a limitation as some key concepts are missed out in the summary.

Another work by [10] centred on 'Mining Physicians' Notes for Medical Insights'. It presents a system for disambiguating the senses of words used in doctors' clinical notes. The system was developed with the general concepts of Topic Modelling in mind. It is 75% accurate in deciphering words with multiple meanings in the freehand portion of a physician's medical note.

Also, [11] presented 'Automated Identification and Predictive Tools to help identify high-risk heart failure patients: pilot evaluation'. The research focused on the development and evaluation of an automated identification and predictive risk report for hospitalized heart failure (HF) patients. NLP was used to analyse free-text reports every 24 hours, in order to improve early identification of patients hospitalized for heart failure. Another application developed by the authors uses an Intermountain Healthcare-developed predictive score to determine each HF patient's risk for 30-day hospital readmission and 30-day mortality period. The clinical decision support helps to improve HF patient identification, significantly reduces 30-day mortality, and increases patient discharges to home care.

In [12], Natural Language Processing (NLP) is used to detect risk patterns related to Hospital Acquired Infections (HAI). The work helps to detect HAI by using risk patterns identification methods in patient records in order to reduce the number of unnoticed cases and time for reaction. NLP is applied to parse Patient Discharge Summaries and identify specific terms and sequences of facts. Thus, the developed NLP algorithm monitors and prevents HAI by scanning patients' records and automatically detects signs of a possible issue.

Another work by [13] centers on a systematic review of extraction of information from the text of Electronic Medical Records (EMR) to improve case detection. It aims at an accurate identification of patient with condition of interest, and examines whether the use of information from EMR text into case-detection algorithms would enhance the quality of research.

A systematic search was carried out, which returned 9659 papers, of which 67 met the eligibility criteria; reporting on the extraction of information from free text of EMRs with the stated intention of detecting cases of a named clinical condition. It is deduced that more harmonization of reporting within EMR studies is needed, as well as the standardized reporting of algorithm accuracy metrics.

Another system named the Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES) by [14] is built on existing open-source technologies—the Unstructured Information Management Architecture (UIMA) framework and OpenNLP natural language processing toolkit to allow extraction of information from the EHR. It is meant to build and evaluate an open-source NLP system for information extraction from EHR clinical free-text.

The cTAKES has an accuracy of over 0.8 which is measured over a range of individual component. The cTAKES annota-

ISSN 2229-5518

tions are the foundation for methods and modules for higher-level semantic processing of clinical free-text.
In [15], a review of data processing and text mining technologies on Electronic Medical Records is presented. An in-depth study on the applications developed based on text mining together with the open challenges and research issues for future work are brought to fore. The source data is first pre-processed in order to improve the quality and data mining results. The task of information extraction for medical texts mainly includes NER (named-entity recognition) and RE (relation extraction).

The reviewed works revolve around clinical records with application of NLP usually for information extraction. This work though related with the revived works differs with the emphasis on the use of NLP to facilitate faster comprehension of patients' case report during medical consultation process.

3 METHODOLOGY

This work develops a system named, Patient-Centred Medical Consultation System (PCMCS) in order to reduce the shortcomings of the manual consultation process. The system architecture, system models based on Unified Modelling Language (UML) tools and the logic design of the text summarization technique are considered in the subsequent sections.

3.1 Architecture of PCMCS

Figure 1 presents the architecture of PCMCS. The flow begins with the consultation between a doctor and a patient. The notes taken by the doctor are saved in a consultation repository. The notes are fed into the NLP system, in which the bulk of the work takes place. The NLP system converts the unstructured text sent by the doctor, into tokens and then filters the tokens for the keywords, which are analysed and classified to give an accurate summary. The summary is then viewed by the doctor. Based on the summary, further insight is provided for the doctor by the use of charts and graphs. This is also stored in the repository which stores the consultation notes (input) and result of the NLP process, and feeds the requested summary (output) to the doctor.

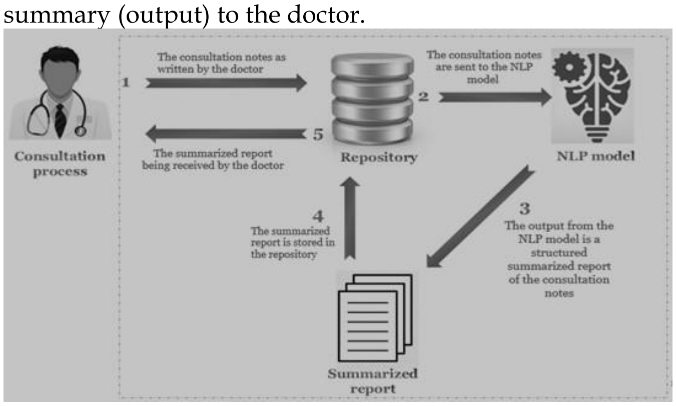

Fig. 1 General Architecture of PCMCS

3.2 Context Diagram of PCMCS

Context models are used to show system boundaries. Context models normally show that the environment includes several other automated systems. However, they do not show the types of relationships between the systems in the environment and the system that is being specified.
Figure 2 depicts the context diagram of PCMCS. PCMCS interacts with the patient's records, which contains the bio-data of the patient. The summary generated from PCMCS further becomes an element of the patient records, which is stored in the patient record system. The consultation system handles the exchange between the doctor and the patient, wherein the doctor sends the details of the consultation and also receives the summary of the patient's record.

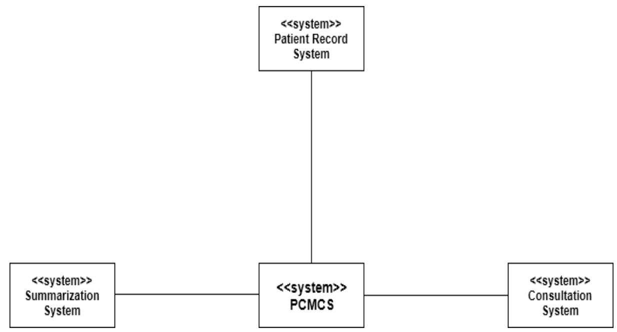

Fig. 2 Context Diagram of PCMCS

3.3 Use Case Diagram of PCMCS

Use-case modelling is applied to analyse the functional requirements of a system without worrying about how those requirements will be implemented. Use case diagram models the interactions between the system and external actors (users or other systems).
Figure 3 shows the principal actor in the system (the doctor) and the different functionalities that can be carried out by the actor.

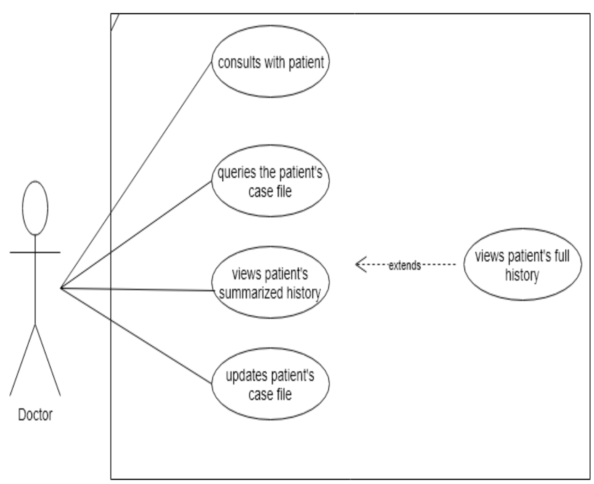

Fig. 3 Use Case Diagram of PCMCS

Table 1 elucidates each of the system functionalities depicted by the use cases in figure 3.

TABLE 1
DESCRIPTION OF THE USE CASES

USE CASE	DESCRIPTION
Consults with patient	The doctor inputs the details from the consultation process into the system. Data: Patient's complaints, treatment summary Response: Confirmation that the data has been saved
Queries patient's case file	The doctor can query the patient's case file if he wants to view the history of the patient.
Views case history	This shows the full history of the patient for a more in-depth analysis
View's patient's summary	This shows the summarized history of the patient for a quick overview.
Updates patient's case file	The doctor is able to update the patient's case file by saving the details of the most recent consultation

3.4 Activity Diagram of PCMCS

Activity diagrams are intended to show the activities that make up a system process and the flow of control from one activity to another. It gives a high level view of a system's functionalities. It is therefore required to model the requirements of the system.

Figure 4 shows the flow of the different activities within PCMCS. From the initiation of the consultation process, a doctor requests to view a patient's history. S/he may decide to view the summary if he wants the high-level history or if he wants to see the full history, he has the option of seeing that too. Then, s/he is able to input a new consultation note base on his conversation with the patient. Once s/he is done with the patient, he saves his new note and thus, that particular patient's history is updated. For subsequent consultations, the saved case note becomes a part of the summarized report to be presented to the doctor.

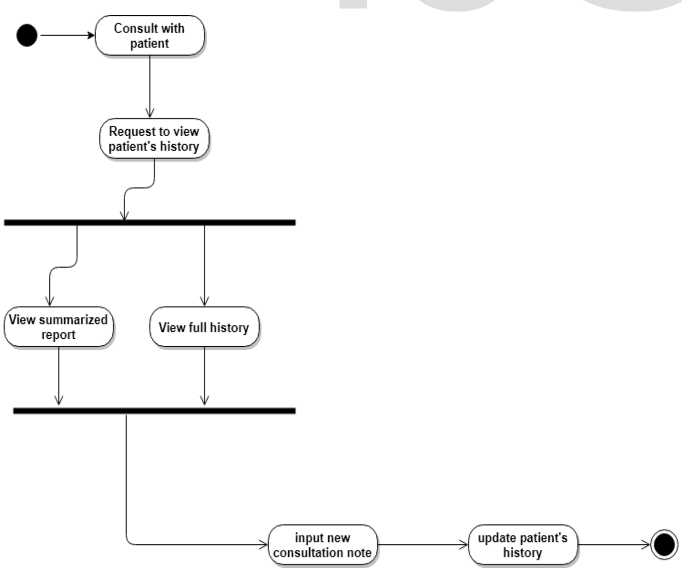

Fig. 4 Activity Diagram of PCMCS

3.5 System Design of PCMCS

This section describes the processes involved in the design of PCMCS, to convert the input to the required output, which is the summary of a patient's record. Figure 5 shows the system design using TextRank Algorithm. TextRank is an unsupervised text summarization technique. The algorithm calculates sentence vectors, calculates the similarities between sentence vectors and stores it in a matrix. The similarity matrix is converted to a graph, with sentences as vertices and similarities as edges, for sentence rank calculation. A certain number of top ranked sentences form the final summary.

Figure 6 shows an alternative approach using a neural network. The word embeddings are calculated using a pre-trained word embedding and the vectors are fed into the neural network. The neural network then summarizes the data based on the labelled input fed into it (a form of supervised learning).

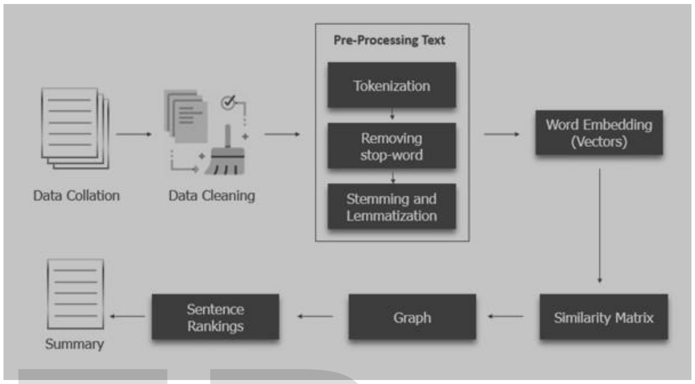

Fig. 5 TextRank algorithm process

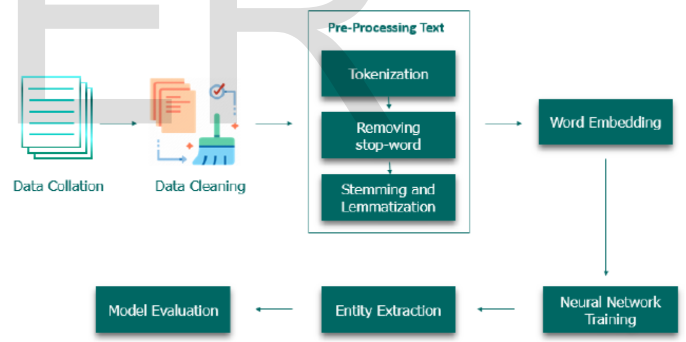

Fig. 6 System Design of PCMCS using a Neural Network

3.6 Pseudocode for Text Summarizer

Figure 7 shows the pseudocode for the text summarization algorithm. The input text corpus is first loaded into the system. Next, each paragraph is tokenized into its respective sentences. Stopwords are then taken out of the sentences. The TextRank Algorithm works based on the ratio of frequency of occurrence of words in a particular sentence, to the frequency of occurrence within the entire text. Each sentence is then attributed a weight based on this ratio. Sentences which contain the most frequent words within the text were assigned the highest weight. For the remaining words, the frequency of occurrence of each is calculated, so is the weighted frequency. The weighted frequency is obtained by dividing the frequency of each word with the highest frequency. Each word in the sentence is replaced by its weighted frequency in the original sentence and each sentence is added up. The sentences are then sorted in decreasing order of sum to get the summarized version of the input text of patients' report.

```
#Sorting the sentences in decreasing order of sum
summary_sentences = largest(7, sentence_scores, key=sentence_scores.get)
automated_summary = ' '.join(summary_sentences)
print(automated_summary)

# Using cosine similarity to evaluate the summary
Human_summary = read(‘human_summary.csv’)
Similarity_ratio = []
For i in range(0, len(automated_summary))
      Similarity_ratio.append(Get_cosine_similarity(human_summary[i],
automated_summary[i]))

# Compute the average similarity ratio
Sum(similarity_ratio) / len(similarity_ratio)
```

```
#Load data into program
File = read(“project.csv”)

#Convert Paragraphs to sentences
Sentence_list = sententence_tokenize(file)

#Text Preprocessing
No_punctuation = re.sub(/[^\w\s]/gi , ‘’, sentence_list)
Stopwords = nltk.stopwords(‘english’)

#Finding the frequency of occurrence of each word
For word in word_tokenize(no_punctuation)
      If word not in stopwords
            Word_frequency[word] = 1
      Else
            Word_frequency[word] +=1
Finding the weighted frequency
for word in word_frequencies.keys():
word_frequencies[word] = (word_frequencies[word]/maximum_frequency)

#Replacing words by their weighted frequency in the original sentence and
summing
For sent in sentence_list
      sentence_scores[sent] += word_frequencies[word]
```

Fig. 7 Pseudocode for the Text Summarization algorithm

4. RESULTS AND DISCUSSION

This section documents the results of the implementation of PCMCS based on the NLP technique. Anonymized patients case reports from online source were collated into the corpus as input dataset. An instance of a case report is depicted by Figure 8. The data is collated into a CSV file. Subsequently, preprocessing of the corpus using the preprocessing tools available in the NLTK library follows. Thereafter, TextRank algorithm is used to summarize the corpus. The auto-generated summary extracts the relevant sentences from the bulk of text.

> A 65-year-old male presented to an emergency department in Canada with complaints of nausea and vomiting for 72 hours. The patient had been overseas on business where he travelled throughout South Africa, Zimbabwe, and the Mauritius region. He stated that his vaccines were up to date but that he was not taking anti-malarial medication as he was only briefly travelling through these African countries. He also noted that he had many "bug bites". The patient further stated that he also had a questionable meal before he started feeling nauseous. There was no evidence of blood in the vomit and no diarrhea. Patient appears diaphoretic and flush in the face. No signs of respiratory distress. GI exam in normal with no abdominal tenderness or evidence of organomegaly; however, the patient complained of extreme nausea during examination. Laboratory Results showed increased PT and INR, elevated ALT and positive Rapid Diagnostic Test for falciparum. The patient was diagnosed with uncomplicated falciparum malaria. The patient was admitted to internal medicine and Infectious Disease was notified. The patient was started on Atovaquone-proguanil 4 tabs qd for three days. Patient was discharged 5 days later.

Fig. 8 Sample of Input data

4.1 Computer Generated Summary

The textRank algorithm enables the ordering of the text summary in terms of the most frequently used words in the text. The sentence which contains most words with the highest frequency is assigned the highest weight, and as such, is ordered first. Figures 9 and 10 present the summary in 5 and 7 sentences respectively.

> 'The patient had been overseas on business where he travelled throughout South Africa, Zimbabwe, and the Mauritius region.The patient further stated that he also had a questionable meal before he started feeling nauseous.GI exam in normal with no abdominal tenderness or evidence of organomegaly examination; however, the patient complained of extreme nausea during.Respiratory exam is normal with no evidence of accessory muscle use to adventitious breath sounds.The patient was started on\xa0Atovaquone-proguanil\xa04 tabs qd for three days.'

Fig, 9 Summary in 5 sentences

> 'The patient had been overseas on business where he travelled throughout South Africa, Zimbabwe, and the Mauritius region.The patient further stated that he also had a questionable meal before he started feeling nauseous.GI exam in normal with no abdominal tenderness or evidence of organomegaly; however, the patient complained of extreme nausea during examination.Respiratory exam is normal with no evidence of accessory muscle use to adventitious breath sounds.Laboratory Results showed increased PT and INR, elevated ALT and positive\xa0Rapid Diagnostic Test for\xa0falciparum.The patient was admitted to internal medicine and Infectious Disease was notified.The patient was started on\xa0Atovaquone-proguanil\xa04 tabs qd for three days.'

Fig. 10 Summary in 7 sentences

In solving the problem of time consumption on the part of consultants, this system allows consultants view a condensed version of a patient's medical history. The consultant still has the option of viewing the detailed history for further understanding, if need be. The system also allows consultants search for particular keywords in the course of the consultation process. This leads to a higher service quality received by patients as they do not have to spend additional time on the waiting queue. As only authorized personnel are able to log into the system, PCMCS ensures security of patients' record.
Finally, the auto generated summary is compared with a human generated summary.

4.2 Human Summary of Corpus Text

The manually generated summary for the sample patient case report is as shown in figure 11. This serves as a means of evaluating the accuracy of the computer generated version using the text processing technique.

> A 65-year-old male presented complaints of nausea and vomiting for 72 hours. He also noted that he had many bug bites. The patient further stated that he also had a questionable meal before he started feeling nauseous. Patient appears diaphoretic and with flushed-face. The patient complained of extreme nausea during examination. Laboratory results showed increased-PT, increased-INR and elevated-ALT and **positive** rapid-diagnostic-test for falciparum. The patient was diagnosed with uncomplicated falciparum-malaria. The patient was treated with Atovaquone-proguanil 4 tabs qd for three days.

Fig. 11 Human generated summary of Sample Case Report

4.3 System Evaluation of PCMCS

Both qualitative and quantitative evaluation methods are applied to the output of PCMCS. The former method entails the comparison between the human generated summary and the the NLP - based algorithm generated summary.

The algorithm generated summary and the human generated summary both reduce the volume of the initial input text of patients' case report. For instance, evaluating the sample case of figure 8, the human summary highlighted the major aspects of the case note such as the symptoms, tests taken, results, and the prescribed treatment. The algorithm generated summary highlighted the complaints, the physical examination done and the treatment.

With more input to the algorithm, it would be able to further understand which parts of the entire input are more relevant than the others.

Applying quantitative evaluation method, the textRank algorithm is evaluated using cosine similarity between the summary generated by the algorithm and the initial corpus text of case report. The similarity metric is as depicted in (1).

$$\text{Similarity}(A, B) = \frac{A \,.\, B}{\|A\| * \|B\|} = \frac{\sum_{i=1}^{n} A_i * B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \; * \; \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

Where A and B are representations of a system summary and its reference document (initial case report) based on the vector space mode respectively [16].

There is a 59% similarity rate when five sentences are used for the summary and a 61% similarity when seven sentences are used. Further increment of the number of sentences does not yield a significant increase in the similarity rate.

5 Conclusion

This study shows that although medical data/record is highly unstructured, inconsistent and somewhat ambiguous, there is a lot of insight that could be uncovered from it. The developed system, PCMCS is able to read through a doctor's consultation note, a patient's medical record, or a case report, and then extract the relevant keywords from the input data/text report. This work uses a Natural Language Processing-based technique to analyze medical history of a patient in order to extract the relevant keywords and then present to a doctor a concise summary of the previous consultation notes. Thus, this technique improves medical consultation process, thereby reducing the time spent in the consultation room and leading to better quality of treatment received by patients. Further work would consider more input data on a deep learning driven system for an improved accuracy of summary.

Acknowledgment

The author wishes to thank Ebhomielen Ofure and Ayetuoma Isaac for their contributions towards the success of this work.

References

[1] World Health Organization, Electronic Health Records: "Manual for Developing Countries. Manila: WHO Regional Office for the Western Pacific", 2006

[2] Working Party of the Royal College of Physicians, Doctors in Society. Medical Professional in a Changing World, Pubmed., Retrieved on November 21 2018 from *https://www.ncbi.nlm.nih.gov/pubmed/1640840, 2005*

[3] C. Ley and S.P.A. Bordas, "What makes Data Science different? A discussion involving Statistics2.0 and Computational Sciences". *Int J Data Sci Anal 6*, 167–175 (2018). https://doi.org/10.1007/s41060-017-0090-x

[4] J. Copeland, "Artificial Intelligence". *Encyclopaedia Britannica, Inc.* Retrieved on October 4 2018 from *https://www.britannica .com/technology/artificial-intelligence*

[5] E. Burns, Macine learning (ML). *TechTarget.* Retrieved on November 5 2018 from *https://searchenterprisea.techtarget.com/definition/machine-learning-ML*

[6] J. Bresnick, "What is the Role of Natuarl Language Processing in Healthcare?" *HealthITAnalytics.* Retrieved on November 5 2018 from https://healthitanalytics.com/features/what-is-the-role-of-natural-language-processing-in-healthcare, 2016

[7] M. Allahyari, S. Pouriyeh, M. Safaei, E. D. Trippe, J. B. Gutierrez and K. Kochut. "Text summarization techniques: a brief survey". *arXiv preprint arXiv: 1707.02268.* Retrieved on April 27 2019 from https://arxiv.org/pdf/1707.02268.pdf

[8] M. Rouse, "Clinical Decision Support System (CDSS)" *Search Health IT*, Retrieved on September 10 2018 from https://searchhealthit.techtarget.com/definition/clinical-decision-support-system-CDSS, 2014

[9] D. Scott, C. Hallett, and R. Fettiplace, "Data-to-Text Summarisatino of Patient Records: Using Computer-generated summaries to access Patient Histories." *Patient Education and Counselling, 92(2), 153-159. http://doi.org/10.1016/j.pec.2013.04.019*, 2013

[10] L. Hardesty, "Mining Physicians' Notes for Medical Insights." *MIT News Office.* Retrieved on April 19 2019 from http://news.mit.edu/2012/digital-medical-records-offer-insights-1031, 2012

[11] R.S. Evans, J. Benuzillo, B.D. Horne, J.F. Lloyd, A. Bradshaw, D. Budge, K.D. Rasmusson, C. Roberts, J. Buckway, N. Geer, T. Garrett and D.L. Lappe, "Automated Identification and Predictive Tools to Help Identify High-Risk Heart Failure Patients: Pilot Evaluation", *Pubmed, https://www.ncbi.nlm.nih.gov/pubmed/26911827*, 2016

[12] D. Proux, M. Pierre, F. Segond, I. Kergourlay, S. Darmoni, S. Pereira, Q. Gicquel and M. Metzer, "Natural Language Processing to Detect Risk Patterns Related to Hospital Acquired Infections," Retrieved on September 10 2018 from *http://www.aclweb.org/anthology/W09-4506*. 2010

[13] E. Ford, J. Carroll, H. Smith, D. Scott, and J. Cassell, "Extracting Information from the Text of Electronic Medical Records to Improve Case Detection: A Systematic Review" *Journal of the American Medical Informatics Association, 23(5),pp. 1007-1015, https://doi.org/10.1093/jamia/ocv180*. 2016

[14] G. Savova, J. Masanz, P. Ogren, J. Zheng, S. Sohn, K. Kipper-Schuler, and C. Chute, "Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications". *Journal of the American Medical Informatics Association, 17(5), pp. 507-513, https://doi.org/10.1136/jamia.2009.001560*. 2010

[15] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, and G. Wang, "Data Processing and Text Mining Technologies on Electronic Medical Records: A Re-

view", *Journal of Healthcare Engineering, vol 2018, Article ID 4302425, 9 pages.* [https://doi.org/10.1155/2018/4302425. 2018](https://doi.org/10.1155/2018/4302425)

[16] J. Steingerger, and K. Jezek, "Evaluation Measures for Text Summarization" *Computing and Informatics, 28(2)*, pp. 251-275, 2012