

## Using Fuzzy Logic and Ranking Algorithms Techniques for Automatic Summarization and Extraction of Clinical Text

Moges Tsegaw Melesse<sup>1\*</sup>, Gizatie Desalegn Taye<sup>1</sup>, Gezahegn Mulusew<sup>2</sup>

<sup>1</sup>Department of Computer Science, Debre Tabor University, Debre Tabor, Ethiopia; <sup>2</sup>Department of Information Technology, Debre Tabor University, Debre Tabor, Ethiopia

### ABSTRACT

Information and knowledge management has become a serious issue in the endeavor to serve the medical society due to the growing volume of data, the absence of structured information, and the diversity of information. Clinical doctors may need to know the information included in any piece of clinical free text, but do not have the time to read the entire item. This problem can be mitigated by using an automatic text summarizing technique that reduces the amount of time required while maintaining the integrity of the information. Recognizing the redundancy is a problem that has yet to be solved, and fragmentation makes creating an effective clinical summary even more difficult. We propose an automatic clinical free text summarizer in this work. The researcher utilizes five extraction rates for both rank and fuzzy logic algorithms to summarize the clinical free texts. As a result, the summarizing rates are ten percent, twenty percent, thirty percent, forty percent, and fifty percent. The ranking algorithm had the highest accuracy of 43.52 percent among the five extractive summaries, while the fuzzy logic method had the best accuracy of 43.88 percent. The outcome shown that fuzzy logic extractive summarization outperforms rank algorithm extractive summarization. Fuzzy logic is founded on the idea of computing with words rather than numbers, because words are less accurate than numbers. Using linguistic variables, fuzzy logic seeks to imitate human reasoning. The result is too little; thus we advocate using supervised algorithms to produce a satisfying performance that medical practitioners will approve. The system's performance can be improved further by looking into a variety of domain-specific aspects and enhancing the methods for detecting medical entities.

**Keywords:** Fuzzy logic; Rank Algorithm; Text extraction; Text summarization

### INTRODUCTION

In everyday existence, humans' nature develops a yearning for abstraction. People have less time to read large amounts of written documents as a result of technological advancements and lifestyle changes, and they have a much lower information need. Users concentrate on compressing information by employing abstract perception and avoiding specifics if they are uninterested. As a result, unless automated summary techniques are utilized, finding the most significant and currently needed information from many text documents is challenging. Automated text summarizing methods produce short summaries of the most important information from a document [1].

Extractive summarization approaches depend entirely on the extraction of sentences from the original text, and they work by detecting the key areas of the text. Selecting essential phrases, paragraphs, and other elements from the original document and concatenating them into a shorter form is an extractive summarizing method. To create the summary, extractive algorithms select a subset of existing words, phrases, or sentences from the original text. Because the way information is structured and presented to clinicians can have a significant impact on their decision-making, an accurate, well-designed, and context-specific summary can possibly save time, increase clinical accuracy, and reduce the risk of errors. Medical information, on the other hand, is frequently fragmented, residing in a variety of locations

**Correspondence to:** Moges Tsegaw Melesse, Department of Computer Science, Debre Tabor University, Debre Tabor, Ethiopia, Tel: 251913220071; E-mail: moges.dtu@gmail.com

**Received:** 11-Apr-2022, Manuscript No. JITSE-23-22185; **Editor assigned:** 14-Apr-2022, PreQC No. JITSE-23-22185 (PQ); **Reviewed:** 28-Apr-2022, QC No. JITSE-23-22185; **Revised:** 12-May-2023, Manuscript No. JITSE-22-22185 (R); **Published:** 19-May-2023, DOI:10.35248/2165-7866.23.13.333

**Citation:** Melesse MT, Taye GD, Mulusew G (2023) Using Fuzzy Logic and Ranking Algorithms Techniques for Automatic Summarization and Extraction of Clinical Text. J Inform Tech Softw Eng. 13:333.

**Copyright:** © 2023 Melesse MT, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and formats, putting patients at risk of errors, adverse events, and inefficient care. The goal of this research is to create an automatic clinical text summarizer for free texts from the university of Gondar pediatric clinic [2].

## MATERIALS AND METHODS

To create a summary, we must first examine the document to determine what information is important to include in the summary. The extract summary is created by repurposing words and sentences from the main content. The main text's most important material is copied to the final summary. The summarized text in an abstract summary is an interpretation of the original text. Text summarization, as shown in Figure 1, is an attempt to summarize a document using the formula Summarization=identification+interpretation+generation.

Corpus preparation, preprocessing, text selection, summary creation, and model evaluation are the primary components of the architecture for both subjective and objective evaluation techniques [3,4].

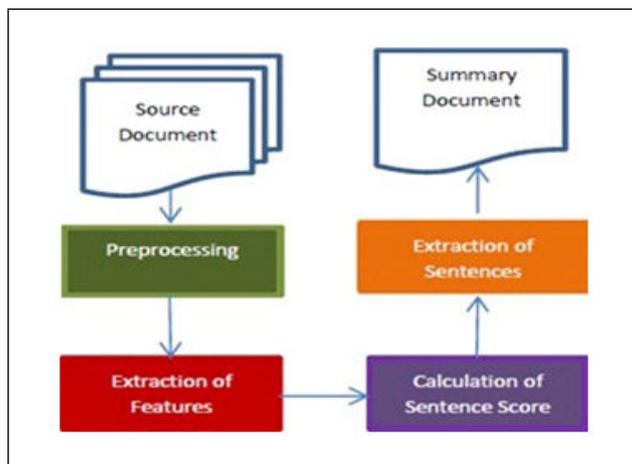


Figure 1: Text summarizing system architecture.

Table 1: Summary of experimental data generated by fuzzy logic algorithms.

Percentage	150 data%	Number (r) size
10%	78.35	P%
	11.21	R%
	19.62	F%
20%	70.88	P%
	11.96	R%
	20.48	F%
30%	75.63	P%
	12.92	R%
	22.08	F%
40%	42.95	P%

## Experimental data source and preparation

The data was gathered from the medical records of children. Following the collection of the patient's medical history, the data was formatted in txt format and made available for experimentation. The patient's text history was preserved in.txt format. It can be thought of as an extension of a single document summarization of a collection of documents taken from several sources.

After transforming the text into sentences, all special characters, stop words, and numerals must be removed from all sentences except the period [5].

## RESULTS AND DISCUSSION

To get all of the words in the sentences, all of the phrases are tokenized. These codes tokenize the sentence and use a single quotation to identify words that are stop words. The sentence is tokenized into words and quote marks are used to divide them.

The above chart history tokenizes the provided statement into words, separated by quotation and word frequency (Table 1) [6].

	44.87	R%
	43.88	F%
50%	43.67	P%
	42.52	R%
	43.07	F%

All of the experiments were carried out using fuzzy logic using compression ratios (extraction rates) of 10%, 20%, 30%, 40%, and 50% for each of the clinical free documents (Table 2).

**Table 2:** Summary of experimental data by generated rank algorithm.

Percentage	150 data%	Number (r) size
10%	42.07	P%
	45.08	R%
	43.52	F%
20%	41.27	P%
	43.16	R%
	42.19	F%
30%	41.21	P%
	42.31	R%
	41.75	F%
40%	40.28	P%
	43.16	R%
	41.67	F%
50%	40.60	P%
	42.94	R%
	41.75	F%

All of the experiments were carried out utilizing ranking algorithms summary generations with compression ratios (extraction rates) of 10%, 20%, 30%, 40%, and 50% for each of the clinical free documents.

The summarizer is evaluated using both objective and subjective techniques in this study. To assess the system's performance, two types of summaries (system summary and reference summary) are employed [7]. This table has 150 chronological charts with varying numbers of sentences, words, and total characters. As a result, the total number of sentences used in the experiment for both the fuzzy and rank algorithms is 1212, and the total number of words used in the experiment is 34089. The rationale for choosing this sentence is because it is based on a

history of 150 charts, with a total sentence of 1212 derived from the charts [8-10].

Sentence extraction is based on key term frequency and sentence position methods to determine the score of each sentence, and then the sentence is extracted based on their score ranking algorithms. The researcher gathered 150 unique clinical texts, which were structured and formulated into extraction rates of 10%, 20%, 30%, 40%, and 50%, which were employed in both ranking and fuzzy logic algorithms. Summaries for the 150 clinical free texts were extracted using bespoke methods at various rates [11].

The result is displayed using a graphical user interface with extraction rates, a work area, and summary generations. The systems were evaluated and found to be promising. A performance of 43.88% for fuzzy logic algorithms and 43.52% for rank algorithms. As a result, fuzzy logic outperforms rank algorithms in terms of efficiency. The researchers conducted five alternative summaries for each of the experimental corpuses, with a difference of five extraction rates of different data sets, in order to determine the best extraction rate. 10%, 20%, 30%, 40%, and 50% were the five extraction rates. The outcome shown that fuzzy logic extractive summarization outperforms rank algorithm extractive summarization. Fuzzy logic is founded on the idea of computing with words rather than numbers because words are less accurate than numbers. Using linguistic variables, fuzzy logic seeks to imitate human reasoning [12,13].

Professional professors and authors in the language of medical science and health science are required to construct this system. The most difficult aspect of this thesis is gathering data and extracting information from the source (chart), which is extremely neat and intricate. Because it is so difficult to understand handwritten notes from several physicians. Some of the information is incomplete, and some of the descriptions are abbreviated [14]. As a result, professionals and researchers collected the essential information from clinical free text, which is a tedious and time-consuming task. One of the most difficult difficulties is using domain knowledge in order to obtain more relevant content from a collection based on a user's query. Because each expert system generates a different summary, the accuracy of each summary varies, which has an impact on the system's outcome [15,16].

In analyzing automatic text summarizing, several severe obstacles have been recognized, making summarization evaluation a particularly fascinating subject. These difficulties are listed below. Summarization is the process of a machine creating output that is communicated in plain language [17]. There may be a proper answer in circumstances where the result is an answer to a question, but in other cases, it is difficult to determine what the correct output is. There's always the chance that a system will produce an excellent summary that differs significantly from any human summary used as a rough approximation to the correct result. Because people may be necessary to judge the system's output, the cost of a review could escalate. Because summarization contains compression, it's critical to be able to assess summaries at various compression rates. The evaluation's scope and complexity grow as a result. As a result, the evaluation design becomes more difficult.

## CONCLUSION

In everyday existence, humans' nature develops a yearning for abstraction. People have less time to read large amounts of written documents as a result of technological advancements and lifestyle changes, and they have a much lower information need. Automated text summarizing systems create short summaries of the most important information from a document. The researchers discussed the state of the art in the area of text summarization, different approaches used in text summarization, types of text summarization, evaluation

techniques in text summarization, and related works in order to do clinical free text summarization for the university of Gondar hospital pediatric clinic.

So, the researcher takes 150 data sets from pediatric clinic charts and uses graphical user interfaces to preprocess the data, eliminate stop words, and provide a summary using ranking and fuzzy logic algorithms. As a result, in the case of the university of Gondar hospitals in Ethiopia, the topic involves both fuzzy logic and ranking algorithms based models for text summarization in medical, clinical free text. Researchers employed fuzzy logic and a ranking algorithm to find the best technique for summarizing a given text in this study. The proposed fuzzy logic approach outperforms the ranking algorithms systems it is compared against. In general, this clinical summary is based on patient charts from the university of Gondar institutions, primarily the pediatric clinic, from 2006 to 2010, with a total of 150 charts covering a variety of disorders.

Information overload exists among the many medical history documents. This difficulty can be handled if strong text summarizers are available, which generate a document summary for consumers to use. The types of summarization methods that could be utilized in a system to provide a summary of extractive approaches are discussed in this thesis. An extracted summary is a condensed version of the original text that highlights key points. Furthermore, extractive base summarization approaches such as hidden markov model, ranking algorithm, and fuzzy logic have to some extent succeeded in producing an effective document summary.

## RECOMMENDATION

This work presents an in-depth examination of various extractive text summarization strategies, with a particular emphasis on recent efforts and advancements in unsupervised approaches. Furthermore, the researcher discusses various text summarizing types as well as the evaluation task. While many academics are working to improve extractive summarization using supervised algorithms based on dataset attributes, others are working to improve it using unsupervised approaches. Unsupervised techniques try to find hidden characteristics in documents or learn document semantic representation without having to train the model on datasets. Furthermore, because summaries and labels are in short supply, unsupervised techniques can be utilized to generate these labels automatically. As a result, there is room to develop unsupervised extractive summarizing algorithms in order to uncover novel document aspects.

## REFERENCES

1. Moen H, Peltonen LM, Heimonen J, Airola A, Pahikkala T, Salakoski T, et al. Comparison of automatic summarisation methods for clinical free text notes. *Artif Intell Med.* 2016;67:25-37.
2. Allahyari M, Pouriyyeh S, Assefi M, Safaei S, Trippe ED, Gutierrez JB, et al. Text summarization techniques: A brief survey. *Int J Adv Comput Sci Appl.* 2017;8(10).
3. Feblowitz JC, Wright A, Singh H, Samal L, Sittig DF. Summarization of clinical information: A conceptual model. *J Biomed Inform.* 2011;44(4):688-699.

4. Indu M, Kavitha KV. Review on text summarization evaluation methods. In 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS). 2016; pp. 1-4.
5. Zelingher J, Rind DM, Caraballo E, Tuttle MS, Olson NE, Safran C. Categorization of free-text problem lists: An effective method of capturing clinical data. *Proc Annu Symp Comput Appl Med Care*. 1995; pp. 416-420.
6. Devarakonda M, Zhang D, Tsou CH, Bornea M. Problem-oriented patient record summary: An early report on a Watson application. 2014 IEEE 16<sup>th</sup> International Conference on e-Health Networking, Applications and Services (Healthcom). 2014; pp. 281-286.
7. Bisandu DB. Design science research methodology in computer science and information systems. *Int J Inform Technol*. 2016;5(4): 55-60.
8. Yeasmin S, Tumpa PB, Nitu AM, Uddin MP, Ali E, Afjal MI. Study of abstractive text summarization techniques. *Am J Eng Res*. 2017;6(8):253-260.
9. Moradi M, Ghadiri N. Text summarization in the biomedical domain. *arXiv preprint arXiv:1908.02285*. 2019.
10. Moradi M, Ghadiri N. Quantifying the informativeness for biomedical literature summarization: An itemset mining method. *Comput Methods Programs Biomed*. 2017;146:77-89.
11. Kiyomarsi F. Evaluation of automatic text summarizations based on human summaries. *Procedia Soc Behav Sci*. 2015;192:83-91.
12. Narayan S, Cohen SB, Lapata M. Ranking sentences for extractive summarization with reinforcement learning. *Association for Computational Linguistics, Louisiana*, 2018, pp. 1747-1759.
13. PadmaLahari E, Kumar DS, Prasad S. Automatic text summarization with statistical and linguistic features using successive thresholds. In 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. 2014; pp. 1519-1524.
14. Erdirencelebi D, Yalpir S. Adaptive network fuzzy inference system modeling for the input selection and prediction of anaerobic digestion effluent quality. *Appl Math Model*. 2011;35(8):3821-3832.
15. Inderjeet MA. Summarization evaluation: An overview. *Proceedings of the NTCIR Workshop*, 2009.
16. Fang C, Mu D, Deng Z, Wu Z. Word-sentence co-ranking for automatic extractive text summarization. *Expert Syst Appl*. 2017;72:189-195.
17. Tesema FB. Afaan oromo automatic news text summarizer based on sentence selection function. Addis Ababa University, Ethiopia, pp. 1-174.