



CovMIS - APPLICATION OF FUZZY BASED INTEGRATED APPROACH TO SUMMARIZE COVID MEDICAL INFORMATION

¹Ravi Seeta Sireesha, ²D. Lalitha Bhaskari, ³P. S. Avadhani

¹Reserach Scholar, ^{2,3}Professor

^{1, 2, 3}Department of Computer Science and Systems Engineering, AUCE (A), Andhra University, Visakhapatnam, India.

Abstract: COVID-19 outbreak was declared as a Public Health Emergency of International Concern (PHEIC) by WHO on 30 January 2020 [1]. The outbreak of the pandemic COVID-19 has totally shattered the political, social, economic, religious, and financial structures of almost every country of the world. Maximum number of countries have shut their borders, suspended the activities in businesses and educational institutions and mandated their population to self-quarantine [3]. WHO and many other organizations and doctor/expert communities are releasing adequate information about the disease to increase awareness among the people towards the disease which includes the symptoms, tips to be followed to protect while going outside, precautions to be taken, medication, vaccination, and so on. How much information ever was released there is always a chance in getting an updated new symptom/precaution/tip about the disease to educate the people to a completely unknown disease. The pressure and anxiety due to the prevailing uncertain situations, the control on the movement of individuals suggests that it may lead to many nervous and psychiatric disorders like suicidal behavior and may have a negative impact on the individual's mental health [4, 5].

In this pandemic situation reading the entire document related to the disease and understanding it completely may not be possible to all individuals. Medical document summarization techniques can be applied to get a brief but important point of the disease as a summary/overview of the disease which helps the global people to know well about the disease and cope up with the situation. In this paper, Summarization of the Covid Medical Information has been initiated and implemented. A fuzzy based integrated approach which was proposed by the authors and also it was observed to generate good quality summaries for clustered medical documents. The proposed integrated approach by the authors was followed in this paper to generate summaries for Covid Medical Information document. The generated summary was evaluated using ROUGE metric.

IndexTerms - Summarization, Fuzzy based integrated approach, ROUGE metric.

I. Introduction

1.1 Motivation

COVID-19 was found to be a highly infectious disease which can spread fast, endanger the health of huge number of people, and so immediate actions are required to be taken to prevent the disease at the community level [2]. Controlling such an infectious disease is an important, time-critical and also a difficult task. Research is directed towards production of vaccines and governments followed various methods and measures to reduce the spread of the disease. Considering the health of the global population as the crucial aspect, various techniques of Artificial Intelligence can be applied in helping the mankind to combat with such a pandemic circumstances. Natural Language Processing, the popular sub-field of Artificial Intelligence helps in understanding the disease and tackle efficiently with the pandemic situation.

1.2 Background

Summarization of a huge document is one of the useful application areas of Natural Language Processing. The summary is expected to be of short in length when compared to the original document and it should contain only the most predominant sentences.

The Text summarization techniques are broadly classified as: 1) Extractive summarization techniques 2) Abstractive summarization techniques. In Extractive Summarization, the summary includes most important text segments of the actual document. In Abstractive summarization, generated summary is interpreted as the meaning of the original text [7]. The most popular extractive summarization approaches are Statistical Approach and Fuzzy Logic Approach. Identification of most relevant and useful sentences from medical documents which are huge in size is very difficult task [6].

Steps followed in Cov-MIS are:

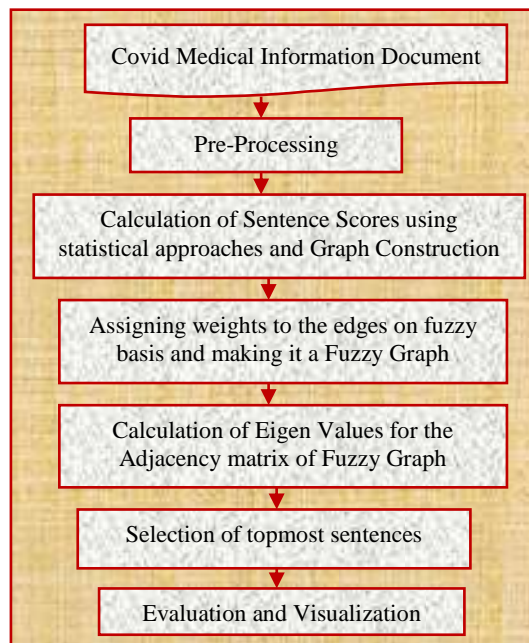


Figure 1: Steps in Cov-MIS

Pre-Processing and Fuzzy graph Construction

Pre-Processing is considered as the initial step in performing any further tasks of the Natural Language Processing. It is the most crucial step as it helps in converting the input document to a more coherent and unambiguous form.

Pre-Processing of the document is performed by the following techniques:

Tokenization, Stemming, Lemmatization, Removing stop words and punctuation.

A document can be represented as a fuzzy graph where the nodes are the sentences and edges are constructed using certain criteria. In this paper, a complete graph is constructed where weights are assigned to the edges based on the importance of the two sentences which are representing the two sentences in the medical document. The edges are assigned weights which further help in generating the summary [6]. In the Integrated approach, the sentence scores of Statistical Approach [9] are utilized to construct the Fuzzy Graph. Graph whose nodes or edges or weights to the edges can be fuzzy is known as a Fuzzy Graph. [11]

Calculation of Eigen Values

For the matrix representation of the Document Graph with assigned fuzzy weights, the Eigen Values can be calculated. The Eigen values and Eigen vectors help in acquiring significant information in a large matrix and in grading items in a dataset. In this paper, Eigen values and Eigen vectors for Fuzzy weighted Medical Document Graph are calculated to generate the summary [6].

Summarization

Summarization helps in condensing the huge data in the document. Two Integrated Approaches are proposed in [6] were found to generate good summaries for medical documents. Between the two approaches, the Fuzzy Graph based Integrated approach which was performing well is considered in this paper. This Integrated Approach utilizes the sentence scores obtained from Statistical Approach [9] and then constructs a Fuzzy Graph to generate summary.

The Statistical Approach generates summary for the medical document by retrieving the sentences with topmost scores [8, 9]. The assignment of sentence scores can be done using five features namely Title feature, Sentence Length feature, Thematic word feature, Numerical Data feature, Proper Noun feature.

Evaluation and visualization Finally, the generated summary of the medical document can be evaluated by using various metrics.

1.3 Objectives and Contributions

The objectives considered and the contributions made during this research are as shown:

Table 1: Objectives and Contributions

Objectives	Contributions
To apply the existing pre-processing techniques to the unstructured/semi-structured medical documents.	An in-depth concentration is laid on the existing pre-processing techniques to represent the medical documents which are unstructured/semi-structured.
To study and understand various existing Statistical Summarization, Fuzzy Logic Summarization Approaches and the proposed Fuzzy Integrated Approaches in [6] to the medical documents.	Understand the existing Statistical Summarization, Fuzzy Logic Summarization Approaches and the proposed Fuzzy based Integrated Approaches in [6] to the medical documents and identify the better approach.
To apply the best approaches to a current scenario to help the mankind.	Application of the novel Fuzzy based Integrated Summarization Approach to the Covid Medical Information document and other documents related to the disease.

II. Literature Review

Kanitha D K , D. Muhammad Noorul Mubarak [14] in their work, investigated about the methods of abstractive and extractive text summarization and their merits and de- merits. In their work they observed that the extractive methods are easy to implement and give better result in summarization whereas the abstractive summarization requires heavy computational models for language generation. Also they have analyzed that the performance of statistical based extractive summarization method gives good result in summarization of the documents.

Ashwitha Dantis, Roshan Fernandes, Anisha P Rodrigues [15] in their work summarized a document by sentence scoring based on term frequency and inverse document frequency matrix. Their technique emphasized on considering the term frequency and inverse document frequency approach and generated a summary without losing the relevancy and meaning of the original document. They concluded that as a future enhancement for their work, machine learning approaches can be applied in order to classify whether a sentence is relevant or not to include it in the generation of summary which is considered as a classification problem.

The authors of this paper Ravi Seeta Sireesha, D. Lalitha Bhaskari, P. S. Avadhani in their work in [6] proposed novel fuzzy based integrated approaches to summarize the clustered medical documents. Their proposed approaches generated better summaries compared to the existing approaches. They concluded that their work can be further enhanced or can be applied to the documents containing information about latest hazardous diseases.

III. Methodology

Covid-19 is totally an unexpected, unknown disease and creating unpredictable challenges to the entire global mankind. Many Computer Science Researchers are doing huge amounts of work related to predictions of the occurrence and many other aspects of the disease. Research work related to Summarization of the Covid medical information has been initiated in this paper which may help the people around the globe to understand the disease in an easier manner and fight against it in an efficient way. In this paper the Integrated Approach [6] is applied for generating the summary from Covid medical information.

Some of the diseases may share commonalities in their symptoms, dosage of drugs, side effects and precautions. All such documents can be grouped using clustering techniques. Various agglomerative and partitional clustering techniques are applied over the medical documents to generate meaningful groups of medical documents [8].

The medical documents contain many important aspects which are supposed not to be neglected. Generation of summary from the clustered medical documents using features identified is performed using Statistical Summarization technique [9].

For any method to be effective selection of features plays a vital role. In this work five features are considered and they are as described in Table 2. All the features are derived using the formulae and the values obtained are also as shown in the Table2 for an example sentence: "Once the virus develops in people, corona viruses can be transmitted from person to person through respiratory droplets in 6meters of distance."

Table 2: Features Considered, their Description, Formulae used and Feature Values obtained

Name of the Feature	Characteristic/Importance of the Feature	Formula used for the Feature	Feature Value obtained
Title feature	Presence of Title word in a sentence secures that sentence much importance.	$SF1(S) = (\text{Number of Title Words in } S) / (\text{Number of Words in Title})$	0.333
Sentence Length	Moderate length sentences are given high priority.	$SF2(S) = (\text{Number of Words in } S) / (\text{Number of Words in longest sentence})$	0.325
Thematic word	Most important words relevant to the context are given much priority.	$SF3(S) = (\text{Number of Thematic Words in } S) / (\text{Max (Number of Thematic Words in all sentences in a document)})$	0.25
Numerical data	Sentence with numerical data is of high importance in medical document.	$SF4(S) = (\text{Number of Numerical Words in } S) / (\text{Number of Words in } S)$	0.077
Proper Noun	Proper nouns are given much importance in a sentence.	$SF5(S) = (\text{Number of Proper Nouns in } S) / (\text{Number of Words in } S)$	0.615

An Integrated Approach named Integrated Approach-1 is proposed by the authors in [6] using the Statistical Approach and the Fuzzy Logic Approach [10]. One more Integrated Approach named Integrated Approach-2 is also proposed which uses the results of Statistical Approach and constructs a Fuzzy Graph where the edges are assigned weights in a fuzzy manner[6]. Between the two proposed integrated approaches the Integrated Approach-2 was found to be more efficient and it was applied for the Covid Medical Information document in generation of summary.

3.1 Integrated Approach

In the Integrated Approach, fuzzy graph is constructed. The sentences are considered as nodes and weights are assigned to the edges on a fuzzy basis.

The feature values and the fuzzy values obtained for the sentences and the weights assigned for the edges for a small chunk of the input Covid Medical Information document and corresponding fuzzy graph are as shown in the subsequent figures


```

f1= [1.0,0.7619047619047619,0.5238095238095238, 1.0476190476190477]
f2= [0.08695652173913043,0.11764705882352941,0.0, 0.038461538461538464]
f3 = [0.0, 0.0, 0.0, 0.0]
f4 = [0.043478260869565216, 0.0, 0.0, 0.0]
f5= [0.2608695652173913, 0.35294117647058826, 0.2, 0.38461538461538464]
f1_fuz = [0.71875, 0.96875, 0.65625, 0.65625]
f2_fuz = [0.94965675, 0.89473684, 0.89473684, 0.96356275]
f3_fuz = [1. ,1. ,1. ,1.]
f4_fuz = [0.95833333 , 0.95833333, 0.95833333, 0.95833333]
f5_fuz = [0.94692264, 0.89763178, 0.84415584 , 0.84415584]
Weights for the edges
[[0, 0.8511252621479608, 0.7911017608895929, 0.8232990874272383],
[0.8511252621479608, 0, 1.0940582846571574, 1.1288099059530983],
[0.7911017608895929, 1.0940582846571574, 0, 1.0175712805101835],
[0.8232990874272383, 1.1288099059530983, 1.0175712805101835, 0]]
    
```

Figure 2: Feature values and weights of the Fuzzy Graph obtained for a small chunk of the Covid Medical Information document

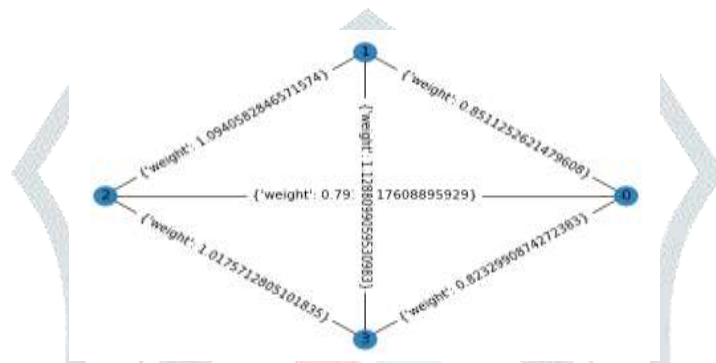


Figure 3: Fuzzy Graph obtained for a small chunk of the Covid Medical Information document



The algorithm to extract summary from Covid Medical Information document is as shown:

Algorithm: Cov-MIS (Integrated Approach for Generation of Covid Medical Information Document Summary)

Input: Covid Medical Information Document

Output: Summary

1. Extract the features from the Covid Medical Information Document.
2. Repeat
 - 2.1 Calculate the sentence score(Sent Score) as:

$$Sent\ Score_{(Statistical)} = \sum_{i=1}^n SF_i(S)$$
 SF_i is the score of the feature and n is the number of features.
 3. Construct complete fuzzy graph
 - 3.1 Calculate (Threshold) = (max (Sent Score_(Statistical)) + second max(Sent Score_(Statistical)))/3
 - 3.2 Assign weights to the edges as

if (Sent Score_i + Sent Score_j)/2 > threshold
then $w(i, j) = (Sent\ Score_i + Sent\ Score_j)/2$
else
 $w(i, j) = threshold/2$
 4. Update the Weights for the edges as

$$w(i, j) * = \frac{(c(f1)*\mu(f1) + c(f2)*\mu(f2) + c(f3)*\mu(f3) + c(f4)*\mu(f4))}{(c(f1) + c(f2) + c(f3) + c(f4))}$$

$c(fi) = (SF_i(S_i) + SF_i(S_j))/2$ and
 $\mu(fi) = (\mu_{fi}(S_i) + \mu_{fi}(S_j))/2$
 f_i is the i^{th} feature, SF_i, SF_j are the sentence scores of nodes S_i and S_j.
The μ is the Triangular Membership Function which generates a Fuzzy Value for the crisp value 'x',
 $Trimf(x, a, b, c) = Max (Min ((x-a) / (b-a)), ((c-x) / (c-b)), 0)$
5. Calculate (Eigen Values) for the Fuzzy Weight Matrix representation of the Covid Medical Information Document Graph.
6. Select the sentences with the highest score as the summary.
Descend (sentences according to the obtained Eigen Values)

Steps in Cov-MIS

All the steps in the algorithms are executed in Python (conda 3). The summary generated for the Covid Medical Information document by implementing the Integrated Approach proposed in [6] was observed to be very encouraging. The input to the summarization algorithm is in text document form. The small portion of the input and the output generated are shown in the table 4.

Table 3: A small portion of the input in the Covid Medical Information document and the summary obtained using the Integrated Approach

<p>The most commonly known symptom of COVID-19 is fever. But COVID affected people in a mild fashion may not experience fever. Most of the COVID-19 affected people come under mild case. The COVID-19 treatment guidelines, people are considered as they come under mild case if they:</p> <ul style="list-style-type: none"> • suffer with any of the prototypical symptoms of COVID-19 (such as cough, fatigue, or loss of taste or smell). • do not suffer with shortness of breath or unusual chest imaging. <p>The long lasting effects of the disease can occur even in the people with mild cases. Long Haulers is the term used for the people who encounter symptoms months after initial diminishing of the virus. According to a research letter in February 2021, almost one-third Trusted Source of people even 9 months after contamination with virus experienced constant symptoms as long as in the succeeding 9 months of viral contamination. Many experts estimated that most of the people affected with COVID-19 are actually asymptomatic. The meaning of asymptomatic is that they have no symptoms at all. In some cases, even patients who require senior care services are also identified to be asymptomatic. The experts and researchers deduced that:</p>	<p>The long lasting effects of the disease can occur even in the people with mild cases. Long Haulers is the term used for the people who encounter symptoms months after initial diminishing of the virus. The experts and researchers deduced that:</p> <ul style="list-style-type: none"> • Asymptomatic cases are at least one-third of all the identified cases. • Even the PCR test result is positive, 75% of the people remain asymptomatic.
---	--

<ul style="list-style-type: none"> Asymptomatic cases are at least one-third of all the identified cases. Even the PCR test result is positive, 75% of the people remain asymptomatic. 	
Number of words in the input Covid Information Medical Document = 4,355	Number of words in the Summary generated through Integrated Approach = 1,560

It was observed that there was a huge decrease in number of words for the Covid Medical Information document that was considered. So, the approach is also applied to the documents containing information about other variants and important consequent aspects of the Covid disease like Delta Plus Variant, Black Fungus, Post Covid Syndromes, Covid in Pregnancy. The small portions of the input and the output generated for the corresponding documents are shown in the table 5.

Table 4: Small portions of the input in the documents of variants of Covid Medical Information and the summary obtained using the Integrated Approach

Delta Plus	
A mutation known as K417N in the spike protein results in formation of Delta Plus. The K417N mutation occurred in combination of AY1 and AY2 is known to be Beta variant. The Delta variant was found to be major responsible for commencement of the second wave of corona virus. The mutating variants are called as AY1 and AY2 and are known as Delta and Delta Plus variants. According to the experts, only a little amount of Delta sequences carrying the K417N are observed on the Global Initiative on Sharing All Influenza Data (GISAID). The initial sequences were found in Europe through normal scanning. There are no proved indications that the Delta variant is severe disease and also more research is needed for this variant. The extremely important points to be followed are double masking, vaccination and following the Covid appropriate behaviour. Experts highly recommend to strictly follow physical distance and practice hand hygiene even in the routine busy life. There is no warning sign that when the third wave may start and when an even more infectious variant emerge and hits the country.	A mutation known as K417N in the spike protein results in formation of Delta Plus. There are no proved indications that the Delta variant is severe disease and also more research is needed for this variant. The extremely important points to be followed are double masking, vaccination and following the Covid appropriate behaviour.
Number of words in the input Delta Plus Variant Covid Medical Document = 1343	Number of words in the Summary generated through Integrated Approach = 636
Black Fungus	
Recently a serious fungal infection is observed among the Covid-19 patients, which was known as Mucormycosis. It was a rear disease and popularly known as Black Fungus. This is a serious disease which may damage skin, lungs and even the brain. Mucormycetes which are found in the environment as a group are the cause of this disease. Experts noticed that this disease mainly affects people who have inability to fight with pathogens due to their medication related to some diseases. Lungs of such patients get affected due to the inhalation of fungal spores. Experts also identified that the effects of the disease are severe in patients who are recovered from Covid-19 and undergone major surgeries. The experts also noticed that the disease has less impact on people with a good immunity system. The disease is considered to be severe if the symptoms are pain and redness around the eyes or nose, fever, headache, cough, shortness of breath, blood vomits. The disease is treated by a group of specialists including microbiologists, neurologists, ENT specialists, ophthalmologists, dentists, surgeons and others. The disease places the people at high risk who are suffering with uncontrolled diabetes, used steroids in excess, stayed at ICU for a long period and undergone through some transplantation.	Recently a serious fungal infection is observed among the Covid-19 patients, which was known as Mucormycosis. This is a serious disease which may damage skin, lungs and even the brain. Experts also identified that the effects of the disease are severe in patients who are recovered from Covid-19 and undergone major surgeries. The disease is considered to be severe if the symptoms are pain and redness around the eyes or nose, fever, headache, cough, shortness of breath, blood vomits.
Number of words in the input Black Fungus Variant Covid Medical Document = 1837	Number of words in the Summary generated through Integrated Approach = 649
Post Covid Syndromes	
Initially the COVID-19 patient's recovery was estimated to be 1-2 weeks. The disease can cause a variety of symptoms with varying levels of severity. Some patients are totally asymptomatic, some may face mild symptoms and some have	The disease can cause a variety of symptoms with varying levels of severity. Some people may suffer with even problems related to kidneys. 'Post-COVID lung disease' is the major harmful side effect of a post-viral infected person.

<p>serious consequences of even using ventilators. The experts are investigating for the period that the symptoms may last in a COVID-19 patient. It was observed that only a few number of people can return to full-time daily routine work within 1 year of time after the viral attack. Breathlessness, fatigue and aches in the muscles are the basic symptoms observed in the patients. Some people may suffer with the loss of the sense of taste and smell partially or totally. Some people may suffer with even problems related to kidneys. Experts advice that there is a danger of multi-organ failure for some of the COVID-19 patients. Some of the symptoms are new and still ongoing which may last for days, weeks and even months. 'Post-COVID lung disease' is the major harmful side effect of a post-viral infected person.</p>	
<p>Number of words in the input Post Covid Syndromes Medical Document = 1617</p>	<p>Number of words in the Summary generated through Integrated Approach =529</p>
Pregnancy in Covid	
<p>All women admitted with COVID-19 and confirmed pregnancy is to be registered. The pregnant women should strictly observe social distancing and practice hand hygiene. The viral attack in pregnant women is not indication of termination of pregnancy. Experts advise that the virus is not teratogen. The viral infection control measures to support pregnant women are to be followed strictly by the health care providers. Continuous supervision should be done to monitor the temperature, respiratory rate & oxygen saturations of the pregnant women at maternal level. During labor the oxygen level should be maintained to be more than 94% and monitoring the woman's sepsis. The foetal monitoring has to be done in a continuous manner. Experts do not have any recommended specifications about the mode of birth. If the woman's respiratory condition is getting worse then caesarean is advised. Epidural or spinal analgesia or anaesthesia is advised in women with suspected/confirmed COVID-19 and whose condition demands urgent delivery. The experts have no proof yet about the spreading of the virus to a child through mother's milk. Even post-delivery if the mother suffers with COVID-19 then they should feed their babies by strictly wearing a double mask. Rest of the time the baby should be kept away from the mother. If the virus is severe in the mother then a breast pump is recommended and a healthy care taker is allowed to supply that milk to the child.</p>	<p>The viral attack in pregnant women is not indication of termination of pregnancy. Continuous supervision should be done to monitor the temperature, respiratory rate & oxygen saturations of the pregnant women at maternal level. If the woman's respiratory condition is getting worse then caesarean is advised. If the virus is severe in the mother then a breast pump is recommended and a healthy care taker is allowed to supply that milk to the child.</p>
<p>Number of words in the input Post Covid Syndromes Medical Document = 1622</p>	<p>Number of words in the Summary generated through Integrated Approach =589</p>

It was observed that there was a huge decrease in number of words for the Covid Medical Information document that was considered and all the documents related to the disease. But as the main aim of the paper is to identify and generate meaningful sentences rather than just simply reducing the words. The rate of compression is taken as 45% of the original medical document size. All the summaries generated by the Integrated Approach are evaluated using the ROUGE scores.

3.2 Evaluation

The summaries generated can be evaluated using Rouge metric which is an intrinsic evaluation method [12]. As the input considered is a medical document, the generated summary is compared with the reference summary which is given by the expert/doctor. ROUGE is a set of metrics called Recall Oriented Understudy of Gisting Evaluation (ROUGE). Based on the similarity in the sequences of words between a model summary which is human-written and the machine generated summary, it generates a score [13]. Automatic evaluation of the summary can be performed with the help of ROUGE. The ROUGE scores are generally mentioned in terms of F-score (f), Precision (p), Recall (r) that help to evaluate the machine generated summary.

Precision = (Number of overlapping words) / (Total Number of Words in machine Summary) (1)

Recall = (Number of overlapping words) / (Total Number of Words in Reference Summary) (2)

F-Score = $2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}))$ (3)

ROUGE-1 mentions the overlap of unigrams, ROUGE-2 mentions the bigrams and ROUGE-l mentions the longest matching sequence between the system summary and reference summary.

IV. Experimental Results

The validation of the summary of the Covid medical information document is analyzed using ROUGE metric with the f-score, precision, recall values for the Integrated Approach. The ROUGE scores for the Covid medical information document is shown in the subsequent table.

Table 5: ROUGE values from the Fuzzy based Integrated Approach for the Covid Medical Information document

	F-Score	Precision	Recall
ROUGE-1 Value	0.842	0.871	0.815
ROUGE-2 Value	0.745	0.779	0.721
ROUGE- <i>l</i> Value	0.854	0.838	0.871

It was observed that the results obtained from the Fuzzy Based Integrated Approach for the summarization of Covid Medical Information are very much encouraging. So, the approach is also applied to the documents containing information about other variants and important consequent aspects of the Covid disease like Delta Plus, Black Fungus, Post Covid Syndromes and Covid in Pregnancy. The ROUGE scores for all the documents related to the Covid disease are as shown in the table 7.

Table 6: ROUGE values from the Fuzzy based Integrated Approach for the documents with Covid related Medical Information

Delta Plus Variant			
	F-Score	Precision	Recall
ROUGE-1 Value	0.881	0.788	0.981
ROUGE-2 Value	0.872	0.785	0.943
ROUGE- <i>l</i> Value	0.903	0.812	0.952
Black Fungus			
	F-Score	Precision	Recall
ROUGE-1 Value	0.818	0.694	0.951
ROUGE-2 Value	0.811	0.688	0.948
ROUGE- <i>l</i> Value	0.842	0.731	0.974
Post Covid Syndromes			
	F-Score	Precision	Recall
ROUGE-1 Value	0.818	0.694	0.964
ROUGE-2 Value	0.805	0.683	0.958
ROUGE- <i>l</i> Value	0.871	0.775	0.978
Covid in Pregnancy			
	F-Score	Precision	Recall
ROUGE-1 Value	0.852	0.744	0.971
ROUGE-2 Value	0.834	0.728	0.966
ROUGE- <i>l</i> Value	0.869	0.771	0.989

V. Discussion

As the F-score is calculated using both the precision and recall, it plays a major role in evaluating the generated summary. The F-scores of Covid Medical Information and all the documents related to the Covid disease are as shown in the subsequent figures.

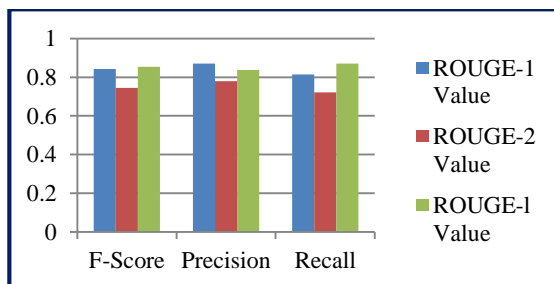


Figure 4: ROUGE values from the Fuzzy based Integrated Approach for the Covid Medical Information document

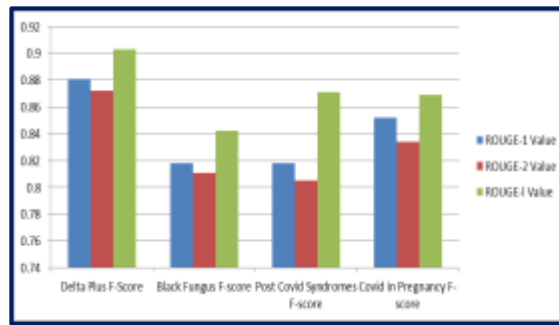


Figure 5: ROUGE values from the Fuzzy based Integrated Approach for the documents with Covid related Medical Information

The ROUGE values lie in the range of [0, 1]. As the ROUGE value approaches to 1, it indicates that the system generated summary is closer to the reference summary and the accuracy is more. It was observed that the ROUGE values are very much closer to 1. Especially the F-score for ROUGE-1 was above 85% for all the documents considered.

5.1 Limitations

The Covid is capable enough to throw continuous challenges to the mankind. So, there is always a scope for obtaining more information about some new variants of the disease. All such information can be collected into a document and can be summarized using the integrated approach [6] for the future reference of a healthy society.

VI. Conclusion and Future Scope

The summary generated from the Covid Medical Information document help the people in a better and prior understanding of the disease, dosage, symptoms and precautions to be taken for the new disease. The generated summary may help many of the people in handling and coping up with the pandemic situation neither directly attending to the doctor/expert nor with physical appointment. Extraction of summary has been initiated in this paper with good results from a document describing the new virus which is totally controlling the entire world.

The results obtained from the Integrated Approach-1 and Integrated Approach-2 proposed in [6] are better than the existing Fuzzy Logic Approach and Statistical Approach in more than 93% of the cases. It was also observed that the Integrated Approach-2 proposed in [6] was efficiently generating summaries not only for clustered medical documents representing various diseases but also for medical document representing a single disease like Covid-19. This Fuzzy based Integrated Approach is also generating good summaries for the other important documents which are closely related to the disease. This work can be further enhanced to generate much better summary which may help the global mankind.

References

- [1] World Health Organization (WHO), "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCov)," WHO, 2020.
- [2] Leesa Lin, Rachel F Mccloud, Cabral A. Bigman et al., "Tuning in and catching on? Examining the relationship between pandemic communication and awareness and knowledge of MERS in the USA," *Journal of Public Health*, Vol. 39, No. 2, pp. 282–289, Advance Access Publication April 15, 2016, doi:10.1093/pubmed/fdw028.
- [3] Pareg Verma, Ankur Dumka, Anuj Bhardwaj et al., "A Statistical Analysis of Impact of COVID19 on the Global Economy and Stock Index Returns", *SN Computer Science* 2021, 2:27, <https://doi.org/10.1007/s42979-020-00410-w>.
- [4] Wenjun Cao, Ziwei Fang, Guoqiang Hou et al., "The psychological impact of the COVID-19 epidemic on college students in China", *Psychiatry Research*, Volume 287, May 2020, <https://doi.org/10.1016/j.psychres.2020.112934> PMID: 32229390.
- [5] Julio Torales, Marcelo O'Higgins, João Mauricio Castaldelli-Maia et al., "The outbreak of COVID-19 coronavirus and its impact on global mental health", *International Journal of Social Psychiatry*, 66, pp. 317–320, <https://doi.org/10.1177/0020764020915212>.
- [6] Ravi Seeta Sireesha, D. Lalitha Bhaskari, P. S. Avadhani, "Novel Fuzzy Based Integrated Approaches For Summarization Of Clustered Medical Documents", *GIS SCIENCE JOURNAL*, ISSN NO : 1869-9391, VOLUME 8, ISSUE 5, 2021, pp 1090-1105.
- [7] Simran kaur, Wg.Cdr AnilChopra, "Document Summarization Techniques", *International Journal of Computer Science Engineering (IJCSE)* ISSN: 2319-7323 Vol. 5 No.02 Mar 2016, pp 71-77.
- [8] Ravi Seeta Sireesha, P. S. Avadhani, "Application of Clustering Algorithms to Group Medical Documents". *International Journal of Computer Applications* (0975 – 8887) Volume 178 – No. 42, August 2019, pp 28-31.
- [9] Ravi Seeta Sireesha, P. S. Avadhani, "Utilization of Summarization Algorithms for a Better Understanding of Clustered Medical Documents", *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-9 Issue-2, December, 2019, DOI: 10.35940/ijeat.B4409.129219 pp 3077-3083.

- [10] Rajesh D. Shinde, Suraj H. Routela, Savita S. Jadhav et al., "Enforcing Text Summarization using Fuzzy Logic", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (6), 2014, pp 8276-8279.
- [11] R.Myna, "Application of Fuzzy Graph in Traffic", International Journal of Scientific & Engineering Research, Volume 6, Issue 2, February-2015 ISSN 2229-5518, pp 1692-1696 .
- [12] Karen Sparck Jones, Julia R. Galliers, "Evaluating Natural Language Processing Systems: An Analysis and Review", In Lecture Notes in Artificial Intelligence 1083, Article in Machine Translation December 1997 DOI: 10.1023/A:1007918307730.
- [13] Saiyed Saziabegum, Priti S. Sajja, "Review On Text Summarization Evaluation Methods", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 8 No. 4 Aug-Sep 2017, ISSN : 0976-5166, pp 497-500.
- [14] Kanitha D K , D. Muhammad Noorul Mubarak International Journal of Computer Engineering & Technology (IJCET) Volume 9, Issue 1, Jan-Feb 2018, pp. 26–36, Article ID: IJCET_09_01_004, "Survey On Text Summarization Methods".
- [15] Ashwitha Dantis, Roshan Fernandes, Anisha P Rodrigues International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019, "An Extractive Summarization Technique for Text Documents".

