ANALYZE ONTOLOGICAL WEB SEARCH ON HEALTHCARE DATA USING DATA MINING TECHNIQUES

J.S.Beulah[1], Mary Metilda[2]

[1]Research scholar, Bharathiar University
[2]Asst Professor, Queen Mary's College

Article History: **Received:** 14.04.2023 **Revised:** 02.05.2023 **Accepted:** 10.06.2023

Abstract: *In the modern era, retrieving healthcare information has become a significant global concern. The identification of biological named entities (NER), which has a variety of applications, is a critical step in text mining of medical information. Deep learning-based ways to tackle this activity have been greatly growing and extending recently because its settings can be entirely academic from start to finish without the requirement for manually created characteristics. The Named Entity Recognition method is used in this study to find disease related synonyms to mine the meaning in medical reports and other applications. The Named Entity Recognition algorithm is one of the automated methods for obtaining medical data from an ontology web search. The desired format is prepossessed and changed to the healthcare text input data after it has been downloaded from the Kaggle repository. After pre-processing, the medical data is recovered using the current clustering techniques Particle Swarm Optimisation (PSO) and Fuzzy C Means (FCM), and the results are evaluated. The research effort also proposes PSFCM, a novel technique for extracting medical data. The results of PSFCM are compared to those of PSO and FCM, which are currently in use. Precision, recall, and f-measure values are a few examples of classification methods that are used to evaluate the PSFCM's performance.*

Keywords: *Healthcare, Information Retrieval, Named Entity Recognition, Ontology, Optimization Algorithm*

DOI: 110.31838/ecb/2023.12.6.214

I. INTRODUCTION

An ontology is a formal specification of the concepts and connections that could be present in an agent or community of agents (much like a formal programme specification). The rising need for trustworthy and high-quality data for information retrieval (IR) investigations has led to an increase in the use of ontologies. Pre-processing of data documents is a step in the IR procedures, which also involve indexing, clustering, and sorting. The same procedures are applied while processing healthcare data [12]. By defining the meaning of healthcare text data using NER algorithms and also by removing blank spaces, extra data, punctuation marks, etc. from the dataset, various pre-processing approaches are employed for ontology web search. Pre-processing is required since the text data's capacity for result prediction will be impacted [40]. Therefore, the specifics of the NER approach's output as well as the pre-processing strategy used to remove this undesired information from the dataset are briefly. Text data are processed by employing simple techniques or processes either to remove unnecessary data or structures such as blank spaces, or to increase the quality of text for ontology search. Text clustering is a fundamental component of information retrieval. Explains the various approaches to clustering used in this research's text data on healthcare.

The main purpose of clustering is to organize the text for processing in a way that information may be accessed using an ontological web search without affecting the text's original content or other qualities or structures. The methodologies or techniques used to solve it depend on the context in which the texts are employed and the area in which they are used. Text data are processed by employing simple techniques or

processes either to remove unnecessary data or structures such as blank spaces, or to increase the quality of text for ontology search [22]. Text clustering is a fundamental component of information retrieval. Explains the various approaches to clustering used in this research's text data on healthcare. The main purpose of clustering is to organize the text for processing in a way that information may be accessed using an ontological web search without affecting the text's original content or other qualities or structures. The methodologies or techniques used to solve it depend on the context in which the texts are employed and the area in which they are used [23].

This proposes a new hybrid PSFCM method for obtaining medical text data for use in healthcare. It also goes into great detail on the PSO algorithm and FCM in order to construct the new hybrid algorithm. The dataset has been given an optimal weight and its weight has been updated, which has enabled this novel approach to achieve its main objective of identifying and retrieving medical information [37]. The resulting data are analysed in this, and the freshly developed hybrid algorithm's efficacy and precision are estimated for upcoming study. When obtaining medical information, the PSFCM algorithm shows that some characteristics present in the final data are more accurate [7] [14].

II. REVIEW OF LITERATURE

Regarding the NER task, several academics propose various strategies. Rule-based techniques were used in the past [19]. These techniques focus on extracting names using various unique rules. Rule-based approaches alone produce better results in constrained environments. Later machine learning approaches, like supervised and unsupervised approaches, alleviate the drawbacks of rule-based systems [33]. These techniques are easy to learn and adaptable to different businesses. However, for testing and training, these approaches require a big annotated corpus. Nowadays, hybrid techniques [26] are routinely applied. Both rule-based and machine learning-based strategies are advantageous for these methods. To reduce computational complexity and rank results, the technique takes use of feature selection, feature clustering, weight computation, and search result clustering.

Karol and others [21] A text document clustering technique based on the PSO algorithm was developed using the clustering algorithms. The Fuzzy C-Means technique and the K-Means algorithm for document clustering were improved by the authors using PSO. PSO has a few flaws that have to be rectified in order to work better. Slow convergence, changing the inertia weight parameter, and less accurate labelling of final clusters were among the drawbacks. A similar technique for grouping documents was developed by Forsati et al. (2013) using the harmony search algorithm and k-means clustering, two powerful stochastic algorithms. A transnational search for the optimal parameters was made possible by harmony search and k-means integration, which improved cluster stability. The population variance of the hybrid algorithm was treated as a Markov chain for the purpose of analysing the behaviour. This method significantly improved the performance of document clustering with higher cluster scalability. Meanwhile, this approach raised the cost and complexity of time [17].

In their analysis, BaiqHaqiqi et al. in [13] went into more depth about the FCM clustering method. Two FCM flaws that cause more irregular clustering results are the partition matrix and random initialization. Although an alternative method called Subtractive Clustering (SC) was considered, it was not known how many clusters there were. The authors presented Subtractive Fuzzy C Means (SFCM), a new hybrid technique that combines FCM and SC, to overcome the limitations of FCM and SC. The experimental results show that the SFCM technique greatly outperforms the FCM algorithm indices in terms of clustering results. In [8], FaqihRofii et al. presented a novel method for determining the bare minimum number of base stations needed for the communication network [35] [39]. In this work, a novel strategy was found to achieve full coverage for the greatest service zone. Finding the number of cells, using GA to determine the election position, and determining the complete coverage area were the three stages of the task. A unique FCM-based approach was developed and applied to the base transceiver station tower problem in order to decrease the number of communication towers.

According to Rajya Gank et al. in [31], the challenging problem in medical image segmentation is the identification of tumour formations based on tumour shape and intensity patterns. The drawbacks of various

methods include their inability to detect tumours and, more especially, their inability to initialise the tumor-affected region's work with the boundary parts that are missing [1]. To remove these limitations, a convergence stable solution is required. This research goes into great detail on the various applications of fuzzy systems. An intelligent system developed by Nanda Gopal et al. in [27] uses IP clustering techniques such FCM, GA, and PSO to identify brain cancers from MRI brain images. The suggested methodology consists of two phases: preprocessing and improvement. The automatic diagnosis of brain cancers provides helpful insights and improves brain tumour accuracy, especially for early-stage brain tumour detection, which is more readily helpful to medical professionals.

A machine learning-based natural language processing technique was developed by Weng et al. (2017) [36] for the classification of medical subdomains in clinical notes. A supervised machine learning-based NLP pipeline was used to develop medical subdomain classifiers of a clinical note in order to quickly direct patients with unresolved medical issues to the best medical specialisations and specialists [28] [38]. This tactic was considered as a general solution because research on the accuracy of medical records is still ongoing. Dess et al.'s (2017) [15] application of cognitive computing and frame semantic properties allowed for the categorising of biomedical documents. . In this approach, features were extracted using two cutting-edge technologies, IBM Watson and Framester, and then reduced via Truncated single-value decomposition to deal with the dimensionality issue. Based on these semantic characteristics, the medical documents were grouped using unsupervised clustering. The problem is that high-level features may also be employed, but our approach solely considered low-level features.

III. MATERIALS AND METHODS

In this section, the work's problem definition is covered. Association of various unstructured story messages in the clinical record is a key problem in clinical content information mining projects. Understanding significant examples and expressions in a subject's clinical history, which can vary greatly, is necessary for pinpointing a subject's illness status from clinical content. Due to the abundance of clinical data, it is necessary to employ effective information evaluation tools in order to extract the most important data. The dataset, which contains extraneous data, missing data, and insignificant characteristics, is preprocessed by methods for name element acknowledgment, and the resulting information is stored in a book record called sentiwordnet. The cleaned coronary illness dataset is then given in a Particle Swarm Fuzzy C-Means (PSFCM), which uses loads taken from the use of the PSFCM to predict which patients are affected intensely and gently.

A. Data Source

High dimensional data elements make up the input documents. 10500 different data sets were obtained in total from the Kaggle library. These unstructured, raw document datasets are pre-processed, and the evaluation data set is then consolidated by removing redundant and less important types of information. After the tokenization procedure during the preprocessing step, the stop-word removal is used to perform the feature extraction, dimensionality reduction, and final clustering. 7000 data were included in the evaluation dataset that was created following pre-processing.

Table 1: Details of Text Data

Date	Patient Number	Patient Name	Doctor	Type	Description
‘Dec 5, 2020’	'P1'	'Kishore.A'	'Dr. Prabhakar'	'investigations'	'BP 120/70 mm of Hg PR 80/min. Weight 64 Kg' 'chest uneasiness'
'Dec 5, 2022'	'P5'	'Vijay.R'	'Dr. Prabhakar'	'complaints'	'Chronic stable angina\nAnterior wall myocardial

					infarction\nInferior wall myocardial infarction'
'Dec 6, 2019'	'P6'	'Saranya.S'	'Dr. Prabhakar'	'observations'	'BP 130/80 mm of Hg PR 93 /min. Weight 57.7 Kg'
'Apr 12, 2022'	'P178'	'Karthick.T'	'Dr. Prabhakar'	'complaints'	'Dental extraction fitness, uneasiness with palpitations.'
'Nov 19, 2019'	'P180'	'Kanimozhi.J'	'Dr. Prabhakar'	'complaints'	'Presurgical evaluation - cataract surgery'

Training samples and testing samples are separated in this pre-processed dataset. In a 9:1 ratio, which indicates that 90% of the samples are used for training and 10% are used for testing. The training datasets were utilised to hone the system's ability to effectively cluster the testing dataset. Using healthcare datasets, the suggested document clustering techniques are evaluated and contrasted with the conventional models. The information contains a range of patient complaints, inquiries, and observations. These textual records include information about past medical treatment, including name, age, gender, date, type, and description. The dataset, which is now in.CSV format, is converted into the format needed for this research project. Each patient's description is different. Pre-processing is carried out on these text data before any algorithm processing. The dataset used for this study, which was converted from its original.CSV file, is in the appropriate format. This includes those who are both severely and barely impacted. Table 1 provides details on the text dataset. The description varies depending on the patient.

B. The Named Entity Recognition (NER) Algorithms

Named Entity Recognition (NER), also known as substance extraction, is an NLP technique that locates and groups the named elements that are present in the document. Named Entity Recognition organises the named entities into pre-characterized classes, such as names of persons, affiliations, regions, quantities, financial attributes, specific phrases, item language, and articulations of periods [32]. By defining the implications of healthcare text data employing NER algorithms in addition to by removing blank spaces, extra data, punctuation marks, etc. from the dataset, various pre-processing approaches are employed for ontology web search. Pre-processing is required since the text data's capacity for result prediction will be impacted.

Therefore, the specifics of the NER approach's output as well as the pre-processing strategy used to remove this undesired information from the dataset are briefly covered in the following sections. Our texts need to be transformed into numerical data so that a quantitative analysis can be carried out. After accepting the conventional wisdom that word order may be discarded with little costs for inference [20] and using a "bag of words" representation, researchers frequently undertake (some subset of) multiple extra binary pre-processing stages in creating the pertinent document term matrix [33].

IV. EXPERIMENTAL RESULTS

This assessment effort aims to use content information (categorical information) to predict the coronary sickness. A few processes were used to collect data on human services and to predict how social insurance will be used in the future, such as anticipating patients' specific needs and health risks. The two key procedures where there have been demonstrable advancements are in clinical choice emotionally supporting networks. Preprocessing of data and prediction. Preprocessing will initially remove duplicate records, missing information, and loud data from the database. Data extraction and recovery will be improved by substance acknowledgment (NER). Here, the optimum deep neural system completes the prediction of

cardiac sickness. In our suggested strategy, methods for simplifying the process change the typical profound neuronal system. The BPNN is used to improve a profound neural system parameter. Figure 1 depicts the PSFCM method's architecture.

A. Preprocessing

Named entity recognition functions as an essential criterion in many natural language applications. It is employed in a pre-processing stage that identifies proper nouns to improve the performance of various natural language applications.

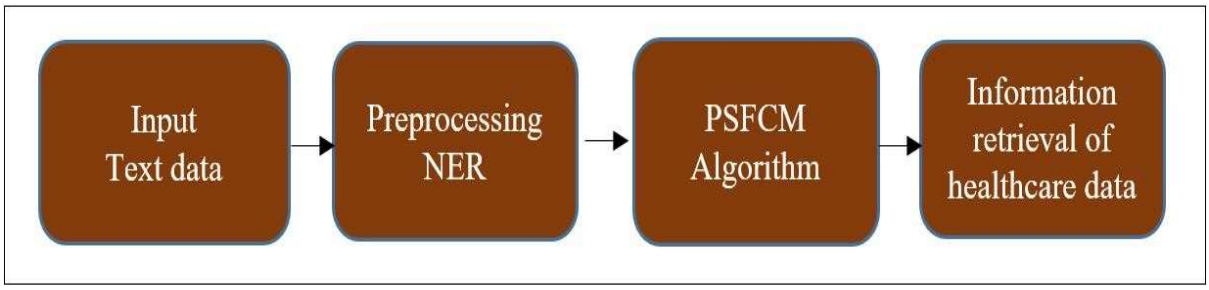

Figure 1: Overall Architecture of PSFCM Algorithm

A.1 Convert Words to Tokens

Tokenization is the process of breaking up a longer piece of text into smaller tokens. Tokens include things like words, characters, numbers, symbols, and n-grams. The most common type of tokenization is whitespace/unigram [25]. The entire text is separated into words in this process by isolating the words from the whitespace. This stage generally separates the sentence into a group of words known as tokens, as opposed to sentence tokenization [9] [10].

A.2 Identify Stopwords in Data

Stopwords are a collection of words that diminish the worth of a text, such as "a," "an," and "the." These are the words that are frequently found in our textual data yet are useless. Many libraries have created lists of stop words for various languages, and we are free to use them directly and for any specific use case if we think we can also add more specialised stop words to the list[6] [11].

A.3 Stemming Words

Finding the word stem of a derived word is done through stemming. For instance, the words "programming," "programmer," and "programmes" can all be translated using the word stem "programme" [4], [5]. In other words, any of the three modification words that came before "programme" can be used in its place. Using this text-processing technique can assist you deal with sparsity and/or standardise terminology.

A.4 Word Lemmatizer

Multiple word inflections with the same meaning are combined into their root forms through lemmatization. Lemmatization is frequently used in web search, information retrieval, SEOs, tagging systems, and indexing. Prior to restoring a word to its dictionary form (the lemma), lemmatization frequently requires removing inflectional endings from words using a lexicon and morphological analysis. For the morphological analysis, it would be necessary to extract the appropriate lemma for each word. [18] [16].

A.5 Vectorizer

Word embeddings, also known as word vectorization, is a method used in natural language processing (NLP) that converts words or phrases from a lexicon into a corresponding vector of real numbers that can be used to determine word predictions and word similarity/semantics. The process of vectorization converts

text data into numerical vectors. This is done to prepare the data for machine learning algorithms, which can only process numerical data. [3] Named entity recognition is used in a variety of different NLP applications, such as ontology population [30], opinion mining [24], semantic search [29], and text clustering [34], in addition to the techniques already mentioned.

B. Particle Swarm Fuzzy C-Means (PSFCM)

Accurate classification is crucial for information retrieval, process monitoring in the healthcare sector, and ontological web search. This research project's main objective is to extract data on healthcare from text datasets. To do this, we develop a hybrid framework that integrates the PSO algorithm with the FCM. Equations 1, 2 and 3, which are mathematical formulas, describe the goal or fitness function of the PSO algorithm.

$$V^{(k+1)} = w.V^{(k)} + c_1.rand_1\left(p_{best}^{(k)} - X^{(k)}\right) + c_2.rand_2\left(g_{best}^{(k)} - X^{(k)}\right) \quad \text{.....(1)}$$

$$X^{(k+1)} = X^{(k)} + V^{(k+1)} \quad \text{.....(2)}$$

$$X_i^{-t+1} = \begin{cases} 1 & if\ r < sig\left(V_{ij}^{t+1}\right) \\ 0 & otherwise \end{cases} \quad \text{.....(3)}$$

Where the attribute should be in the text, ideal; is the attribute's optimal location overall inside the text; and where X signifies the attribute's location within the text, V denotes the attribute's intensity within the text, and W denotes the depth of the attribute. For rand1 and rand2, random numbers between 0 and 1 are set, whereas C1 and C2 are constant positive numbers that are set for every iteration. Equations 4, 5 and 6, which are mathematical formulas, describe the goal function of the FCM algorithm.

$$J^1(u,v) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \left\|X_j - V_i\right\|^2 \quad \text{.....(4)}$$

$$V_i = \frac{\sum_{j=1}^{n} u_{ij}^m X_j}{\sum_{j=1}^{n} u_{ij}^m}, 1 \le i \le c \quad \text{.....(5)}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\left\|X_j - V_i\right\| / \left\|X_j - V_k\right\|\right)^{2/(m-1)}} \quad \text{.....(6)}$$

This work also makes an effort to employ the optimisation concept in conjunction with a deep neural network in order to obtain the best outcome from text data. This approach resolves the issue domain by applying convergent and divergent effects repeatedly. Additional modifications are made using the PSFCM development. The BPNN classifier was used in the current test. When compared to the heart disease text dataset, there are a number of drawbacks to utilising the BPNN method to identify the sick patients. To overcome these shortcomings, we develop an innovative technique based on the conventional BPNN.

$$\vec{U} = \left|\vec{C}.w^{best}(t) - w(t)\right| \quad (7)$$

$$w(t+1) = w^{best}(t) - \vec{A}.\vec{U} \quad (8)$$

Where t denotes the current iteration, $\vec{A}$ and $\vec{C}$ indicates a Coefficient vector, w^{best} indicates a position vector for best solution, w represents a current position Vector, $|\ |$ and represents an absolute value. The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (9)$$

$$\vec{C} = 2.\vec{r} \quad (10)$$

Stage II of the PSFCM algorithm uses the input from stage I. With the help of FCM, the initial optimal weight is determined. The objective of this stage is to convert the weight that was produced from the stage I dataset that was closest to the ideal weight into the current weight after a predetermined number of forecasting iterations. This goal is achieved using the PSO as it is explained in equations 11.

$$E(v,h) = -\sum_{x=1}^{X}\sum_{y=1}^{Y} W_{xy} v_x h_y - \sum_{x=1}^{X} \chi_x v_i - \sum_{y=1}^{y} \xi_y h_y \quad 11)$$

The first stage in the PSFCM technique, which consists of three steps, is to use the BPNN algorithm to identify the ideal weight value. The second stage involves implementing arbitrary weight alterations in the BPNN while iterating the dataset using the backpropagation method, starting with the ideal weight. After a predetermined number of iterations, the PSFCM algorithm's last phases comprise taking the ideal weight and the adjusted weight. The dataset that results is then created empirically. The main objective of the PSFCM method is to use a real-time text dataset to classify patients with heart disease into those who are severely impacted and those who are just slightly affected. To select the best weight from the dataset, update the weight using the backpropagation method to reduce error, and identify patients who are most likely to have heart disease, a novel PSFCM algorithm has been developed.

V. RESULTS AND DISCUSSIONS

Space and speed are essential components in medical TM. Any algorithm's performance is evaluated by looking at a variety of similar types of data to determine how accurate and effective it is. All three techniques use a comparable set of text datasets to investigate and evaluate the best algorithm. The PSFCM algorithm's effectiveness is shown by comparing the output data to those from other clustering techniques. The medical text dataset used in this study was obtained from the Kaggle source. A Windows 10 machine with an Intel Core i3 processor and 8GB of RAM was used for the trials. For this hardware specification, the healthcare dataset produces the outcomes shown below, depending on the required memory and processing time. The algorithm's performance and accuracy are checked, and the top algorithm is chosen based on comparisons of the output datasets. Validation and comparison for the medical dataset are based on time, space, precision, recall, and f-measure.

Given the size of the data and the volume of patient reports, representative results for all methods are shown in figures 2, 3, and 4. Several algorithms are applied on the dataset, and the results are contrasted to determine which algorithm is the most successful.

Figure 2: Results of PSO Algorithm

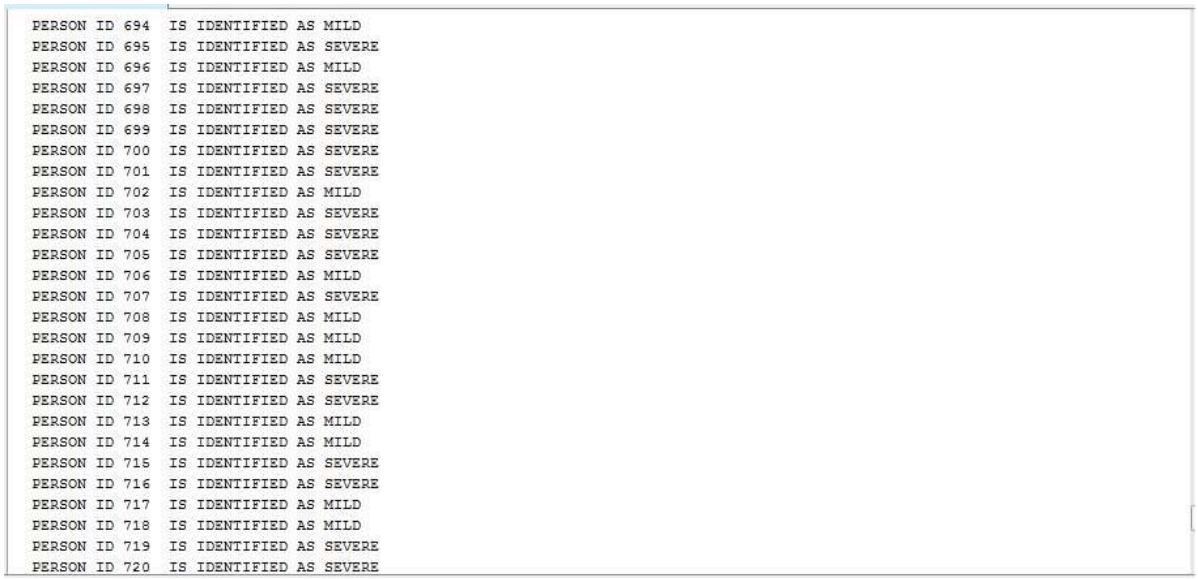

Figure 3: Results of FCM Algorithm

The PSO technique is used after the data has undergone NER preprocessing, and the results are shown in figure 2. Figure 3 displays the results of the FCM clustering technique.
Figure 3 shows the outcomes for those who are both seriously and mildly impacted by the condition after the FCM algorithm is used to pre-process the text dataset using NER.

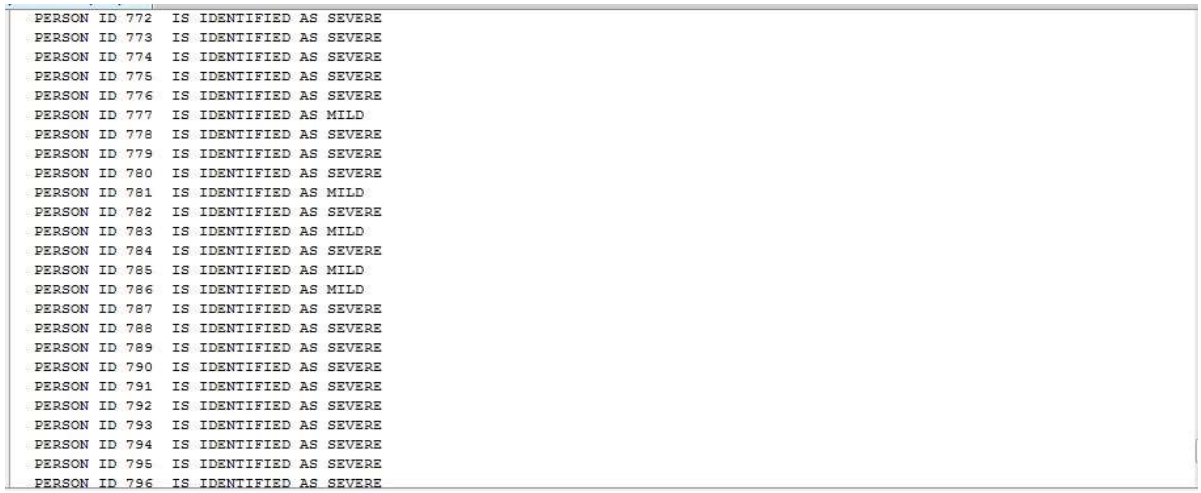

Figure 4: Results of PSFCM Algorithm

Following the pre-processing of the medical text data for Name Entity Recognition, Figure 4 shows the results of the PSFCM implementation. The efficiency and accuracy of the PSFCM algorithm are validated by comparing the results with those of the PSO method and the clustering FCM algorithm. The two factors that determine an algorithm's efficiency are its speed, which is measured by the amount of time it takes to run, and the amount of storage space needed to store the data it generates.

C.1 Efficiency of the PSFCM Algorithm

The average processing time and memory use for the created dataset are shown in Table 1 following the execution of the PSO, FCM, and PSFCM algorithms. The time taken to generate the dataset is expressed in milliseconds (ms), and the memory use is expressed in bits.

Table 1: Average Computational Time and Memory Utilization of Algorithms

Algorithms	Execution time (ms)	Memory utilization (bits)
PSFCM	2754	145469
PSO	3332	188256
FCM	4101	246289

The three algorithms' execution times and memory requirements for the text dataset are shown in Table 1, and it is clear that the hybrid PSFCM technique that has been suggested is preferable. Because the created dataset takes up less time and space, the PSFCM technique outperformed the other two traditional algorithms, PSO and FCM, in terms of resilience.

Figure 5 displays a graphical depiction of the three clustering techniques' execution times for the output dataset. Figure 6 displays a graphical depiction of the memory space utilised by the dataset created by the three clustering algorithms. Figure 5 shows that compared to the PSO and FCM algorithms, the PSFCM approach computes substantially faster. Figure 6 shows that the memory space used by the PSFCM algorithm for the selected text dataset is also much less when compared to the PSO and FCM algorithms.

Figure 5: Comparison Graph Based on Time

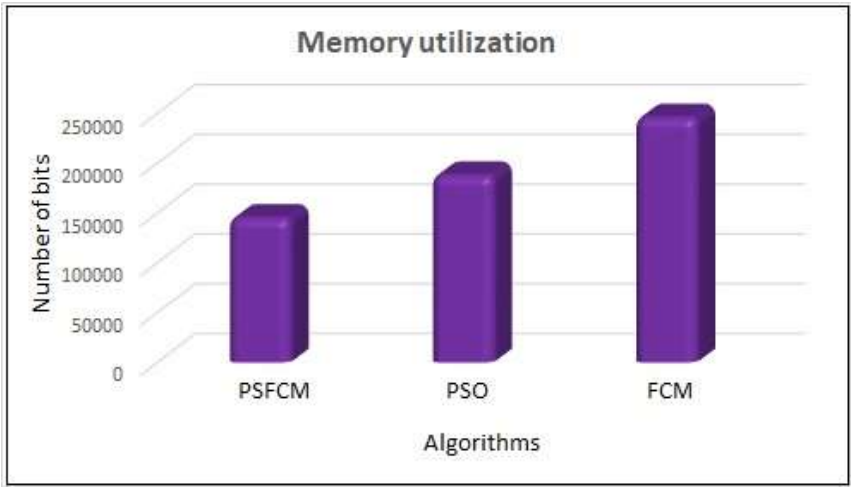

Figure 6: Comparison Graph Based on Memory Space

C.2 Accuracy of PSFCM Algorithm

Accuracy of the generated dataset is one of the most important criteria in determining how effectively an algorithm performs. The final data gathered following the classification of the impacted patients from the medical dataset using the clustering algorithm. The final datasets generated by the PSO, FCM, and PSFCM algorithms are shown in Figures 2, 3, and 4, respectively. Table 2 shows the averages for the three algorithms' precision, recall, and f-measure. When outcomes are empirically seen, the PSFCM algorithm resultant dataset gives improved accuracy and efficiency. The value generated by the PSFCM algorithm offers good precision and recall when compared to values from other algorithms. Thus, the suggested hybrid algorithm provides better accuracy.

Figure 7 shows graphically the average precision, recall, and f-measure of the dataset created by the three techniques. The PSFCM algorithm unquestionably offers improved accuracy, sensitivity, and specificity. The experimental result demonstrates that, when comparing the efficiency and accuracy of the five metrics used, namely time, memory space, precision, recall, and f-measure of the information retrieval of healthcare, our recommended PSFCM method is more effective and generates high-quality information. Results of the PSFCM algorithm and expert delineation are largely comparable.

Table 2: Average performance of all the Algorithms

Algorithms	Precision	Recall	F-measure
PSFCM	81.85	72.81	77.06
PSO	75.16	66.66	70.91
FCM	71.68	61.36	66.52

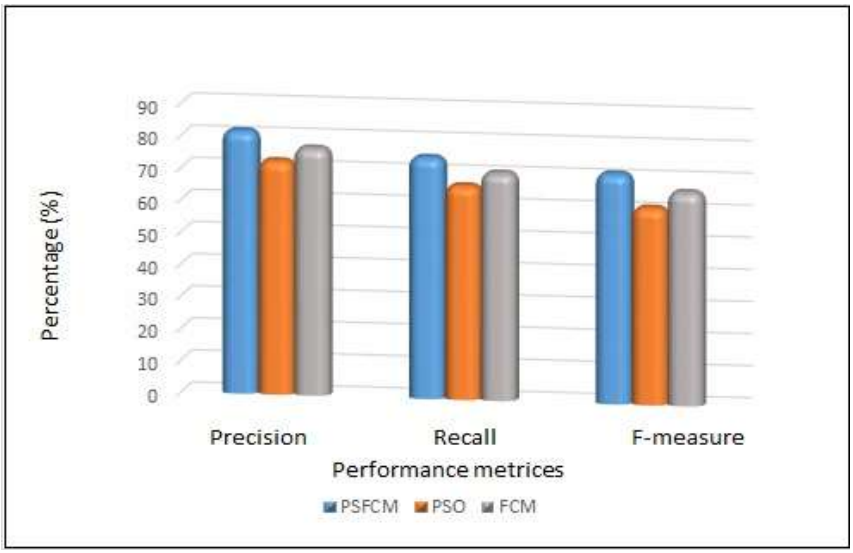

Figure 7: Average performance of all the Algorithms

VI. CONCLUSION

Classifying web searches has recently become a widespread practise. Through a range of automated, computerised methods and processes, the healthcare sector uses textual information to provide "Quality Healthcare" to patients. Today, most healthcare datasets are kept as digital files that are stored electronically. The healthcare text data used for the analysis were downloaded from the Kaggle repository. Pre-processing and clustering are the two steps in the information retrieval process for the healthcare dataset. During the clustering stage, PSO, FCM, and hybrid PSFCM algorithms are evaluated, along with BPNN for classification. Five criteria—computational time, memory space, accuracy, sensitivity, and

specificity—are used to compare the algorithms' outputs. The findings of the proposed PSFCM algorithm are more accurate than those of the current PSO and FCM algorithms when it comes to extracting medical information from text data.

REFERENCES

1. Abbasi, Ahmed, Hsinchun Chen, and Arab Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums", ACM Transactions on Information Systems (TOIS),Vol. 26, Issue 3, pp. 12, 2008.
2. Adak, Chandranath, Bidyut B. Chaudhuri, and Michael Blumenstein, "Named entity recognition from unstructured handwritten document images", IEEE 2016 12th IAPR Workshop on Document Analysis Systems (DAS), pp. 375-380, 2016.
3. Alić, Berina, Lejla Gurbeta, Almir Badnjević, Alma Badnjević-Čengić, Maja Malenica, Tanja Dujić, Adlija Čaušević, and Tamer Bego, "Classification of metabolic syndrome patients using implemented expert system", CMBEBIH 2017, Springer, Singapore, pp. 601-607, 2017.
4. Alić, Berina, Dijana Sejdinović, Lejla Gurbeta, and Almir Badnjevic, "Classification of stress recognition using Artificial Neural Network", IEEE 2016 5th Mediterranean Conference on Embedded Computing (MECO), pp. 297-300, 2016.
5. Aljović, Almir, Almir Badnjević, and Lejla Gurbeta, "Artificial neural networks in the discrimination of Alzheimer's disease using biomarkers data", IEEE 2016 5th Mediterranean Conference on Embedded Computing (MECO), pp. 286-289, 2016.
6. Avdakovic, Samir, Ibrahim Omerhodzic, Almir Badnjevic, and Dusanka Boskovic, "Diagnosis of epilepsy from EEG signals using global wavelet power spectrum", 6th European Conference of the International Federation for Medical and Biological Engineering, pp. 481-484, 2015.
7. Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining", Lrec, Vol. 10, pp. 2200-2204, 2010.
8. Badnjević, A., Lejla Gurbeta, Mario Cifrek, and Damir Marjanovic, "Classification of asthma using artificial neural network", IEEE 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 387-390, 2016.
9. Badnjević, Almir, Dragan Koruga, Mario Cifrek, Hans J. Smith, and Tamer Bego, "Interpretation of pulmonary function test results in relation to asthma classification using integrated software suite", IEEE 2013 36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 140-144, 2013.
10. Badnjevic, Almir, Mario Cifrek, Dragan Koruga, and Dinko Osmankovic, "Neuro-fuzzy classification of asthma and chronic obstructive pulmonary disease", BMC medical informatics and decision making, Vol. 15, Issue 3, pp. 1-9, 2015.
11. Badnjevic, Almir, Mario Cifrek, and Dragan Koruga, "Classification of Chronic Obstructive Pulmonary Disease (COPD) using integrated software suite", XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, pp. 911-914, 2014.
12. Bai Qinghai, "Analysis of Particle Swarm optimization Algorithm", Computer and Information Science, Vol. 3 (1), pp. 180, 2010.
13. Baiq Haqiqi, baiq Nurul, and Robert Kurniawan, "Analisis Perbandingan Metode Fuzzy C-Means Dan Subtractive Fuzzy C-Means", Media Statistika , Vol. 8 (2), pp.59-67, 2015.
14. Cireşan, Dan C., Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks", International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 411-418, 2013.
15. Dessı, D, Recupero, DR, Fenu, G & Consoli, S 2017, „Exploiting Cognitive Computing and Frame Semantic Features for Biomedical Document Clustering", Proceedings of the Workshop on Semantic Web Solutions for Large-scale Biomedical Data Analytics, pp. 20-34.

16. Fojnica, Adnan, Ahmed Osmanović, and Almir Badnjević, "Dynamical model of tuberculosis-multiple strain prediction based on artificial neural network", IEEE 2016 5th Mediterranean Conference on Embedded Computing (MECO), pp. 290-293, 2016.
17. Forsati, R, Mahdavi, M, Shamsfard, M & Meybodi, MR 2013, „Efficient stochastic algorithms for document clustering", Information Sciences, vol. 220, no. 1, pp. 269-291.
18. Faqih Rofii, Dimas Toscanni M., and Diky Siswanto, "Optimization of Coverage and the Number of Base Transceiver Station Towers Using Fuzzy C-Means and Genetic Algorithm", Journal of Theoretical and Applied information Technology, Vol.93 (1), pp. 164-173, 2016.
19. Huang, Jui-Ting, Jinyu Li, and Yifan Gong, "An analysis of convolutional neural networks for speech recognition", 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4989-4993, 2015.
20. Jabbar, M. A., Priti Chandra, and B. L. Deekshatulu, "Cluster based association rule mining for heart attack prediction", Journal of Theoretical and Applied Information Technology, Vol. 32, Issue 2, pp. 196-201, 2011.
21. Karol, S & Mangat, V 2013, „Evaluation of text document clustering approach based on particle swarm optimization", Open Computer Science, vol. 3, no. 2, pp. 69-90.
22. Kim, Yoon, "Convolutional neural networks for sentence classification", arXiv preprint arXiv: 1408.5882, 2014.
23. Lakhani, Paras, and Baskaran Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks". Radiology, Vol. 284, Issue 2, pp. 574-582, 2017.
24. Lodhi, Huma, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins, "Text classification using string kernels", Journal of Machine Learning Research, Vol. 2, pp. 419-444, 2002.
25. Magoulas, George D., and Andriana Prentza, "Machine learning in medical applications", Advanced course on artificial intelligence, pp. 300-307, 1999.
26. Munkhjargal, Zoljargal, Gabor Bella, Altangerel Chagnaa, and Fausto Giunchiglia, "Named entity recognition for Mongolian language", International Conference on Text, Speech, and Dialogue, pp. 243-251, 2015.
27. Nanda Gopal, Nanda N, and Karnan M., "Diagnose Brain Tumor through MRI Using Image Processing Clustering Algorithms such as Fuzzy C Means along with Intelligent Optimization Technique", IEEE International Conference on Computational Intelligence and Computing Research, 2010.
28. Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques", Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pp. 79-86, 2002.
29. Park, Young-Seuk, Régis Céréghino, Arthur Compin, and Sovan Lek, "Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters", Ecological modelling, Vol. 160, Issue 3, pp. 265-280, 2003.
30. Patel, Mr Rahul, and Mr Gaurav Sharma, "A survey on text mining techniques", Int. Journal of Engineering and Computer Science, Vol. 7242, pp. 5621-5625, 2014.
31. Rajya Gank, Mrigank, Sonal Rewri, and Swati Sheoran,"Application of Fuzzy System in Segmentation of MRI Brain Tumor", International Journal of Computer Science and Information Security, Vol. 8 (1), pp. 261-269, 2010.
32. Riedel, Sebastian, Limin Yao, and Andrew McCallum, "Modeling relations and their mentions without labeled text", Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 148-163, 2010.
33. Saha, Sujan Kumar, Shashi Narayan, Sudeshna Sarkar, and Pabitra Mitra, "A composite kernel for named entity recognition", Pattern Recognition Letters, Vol. 31, Issue 12, pp. 1591-1597, 2010.
34. Shehata, Shady, Fakhri Karray, and Mohamed Kamel, "Enhancing text clustering using concept-based mining model", IEEE Sixth International Conference on Data Mining (ICDM'06), pp. 1043-1048, 2006.

35. Turney, Peter D, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417-424, 2002.
36. Weng, WH, Wagholikar, KB, McCray, AT, Szolovits, P & Chueh, HC 2017, „Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach", BMC medical informatics and decision making, vol. 17, no. 1, p. 155, pp. 1-13.
37. Wu, Naiqi, MengChu Zhou, and ZhiWu Li, "Resource-oriented Petri net for deadlock avoidance in flexible assembly systems", IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 38, Issue 1, pp. 56-69, 2007.
38. Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann,"Recognizing contextual polarity in phrase-level sentiment analysis", Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 347-354, 2005.
39. Yang, Kiduk, Ning Yu, and Hui Zhang, "WIDIT in TREC-2007 blog track: Combining lexicon-based methods to detect opinionated blogs", pp. 1-12, 2007.
40. Zaremba, Wojciech, Ilya Sutskever, and Oriol Vinyals, "Recurrent neural network regularization", arXiv preprint arXiv: 1409.2329, pp. 1-8, 2014.