International Journal of Advanced Trends in Computer Science and Engineering

Research Scenario of Medical Data Mining Using Fuzzy and Graph theory

C. Narmatha[1], Dr. M. Thangamani[2], S. Jafar Ali Ibrahim[3]
[1]Assistant Professor, Kongu Engineering College India, Email: narmathac@gmail.com
[2]Assistant Professor, Kongu Engineering College India, Email: manithangamani2@gmail.com
[3]Research Scholar, Kongu Engineering College, India, Email: jafartheni@gmail.com

ABSTRACT

Fuzzy logic offers a smart periphery for the representation and reasoning of knowledge to address inaccuracy, uncertainty, and vagueness. Because of their capability to infuse human expert expertise and granular computation with the behaviour of complicated systems and their ability to explain them without needing a precise mathematical model, the Fuzzy systems were widely used in healthcare. This paper describes simple fuzzy logic and ways that this logic can be used to perform different decision-making tasks. This article reflects a systematic analysis of fuzzy logic in the area of healthcare applications. The key reasons for this paper are to explain briefly the fuzzy logic implementations of different medical diagnostics systems. We use the benchmark data set of the National Centre of Biotechnology Data for this study in this experimental research

Key words : Computational methods, Fuzzy, Graph, Data Mining, Medical Research and Medical Data.

1. INTRODUCTION

Data mining is useful in industry to uncover trends and connections in data to help make better choices. It is a knowledge-based searching and data mining method that is computer-assisted to search, scan and then retrieve the significance of data from massive data sets. Data mining tools forecast behaviour and future trends which help companies to form proactive decisions based on knowledge.

1.1 Data Mining in Healthcare Industry

Data mining has great potential for promoting health systems to routinely service data and analytics to detect inadequacies and best practices in treatment optimization and cost reduction. Some researchers suggest the opportunity to improve services and also reduce costs will enable access to 30 percent of the total healthcare expenditures. This might be an aggregate win. Data extraction can be different for various people, such as analytics and business intelligence. The most

basic definition of information mining is to analyze large data collections to identify trends and use them to foresee or envision eventual events. However, data mining is not all analyzes of the large amount of data. Research is generally classified as follows:

- ✓ Descriptive analytics- Describing what has happened
- ✓ Predictive analytics - Predicting what will happen
- ✓ Prescriptive analytics - Determining what to do about it

The middle class is alluded to as data mining - predictive analysis. But data mining is now, for the most part, an academic exercise with a few technological successes. Data extraction requires identifying patterns and using this data to create predictive models. Most companies serve data mining effectively. This follows the retail industry's consumer response model. This lets banks forecast the performance of customers. This has related applications in telecommunications, engineering, higher education and life sciences etc. For distribution of their findings, scientists use mining techniques like decision-making trees, groups, neural networks and time series. Nevertheless, health care has been reluctant to include the latest research in daily practice. Leads in the warehouses wonder: How can we affect and impact acceptance periods from bed (research) to bedtime (pragmatic shift of quality)?

For the above problem we identified the following three framework solutions are the most effective strategies for expanding the limits of academic research. In any medical technology project, incorporation of all three technologies is key to improve the real world. Unfortunately, very few healthcare companies are running all these three programs.

1. The method of analysis. The system includes the equipment and the knowledge for data collection, interpretation and standardization. The defining aspect of the framework is the collection of scientific, political, satisfaction of patients and other data in an EDW warehouse.
2. The method in best practice. The best practice system includes standardization of clinical work such as the formal implementation of best practices to the quality of treatment. Scientists draw

significant findings every year about therapeutic best practices, but it takes years to integrate such results in clinical practice as I have previously mentioned. A best practices system allows organisations to easily incorporate the latest medical facts.

3. The program for adoption. This system involves the implementation of transformation through the development of new organisation. This includes in particular the introduction of team systems that enable best practices to be continuously applied across the organization. Nothing can be done easily with this program. In order to promote best practices in the entire organization, it requires real systemic reform.

2. FUZZY GRAPH THEORY

The philosophy of graphics is a very important tool for many challenges in the real world. Today, because of the uncertainty or vagueness of system parameters, the graphs do not represent each system properly. A social network can be chronicled as a graph in which vertices (people, entities, etc.) are described. For example, the accounts ' connection and edges are the ties among them. Fugitive reporting should also be included in cases where the relationship between accounts is classified as good or bad in terms of the level of contact between accounts.   These and many other problems contributed to the concept of fugitive graphs. The interpretation of fuzzy plots was proposed first by Rosenfeld [1]. Following this fuzzy graphic theory, a vast field of research becomes. The use of fluctuating diagrams involves data mining, image segmentation, clustering, image collection, networking, coordination, preparation, scheduling, etc.Crisp and fuzzy graphs are both structurally similar. Nevertheless, when the vertices and edges become unknown, the fuzzy graph is essential separately. The fuzzy graph happens in many real life situations because the world is full of unclearness.

Here we consider a fuzzy graph, Let G = (V, E) be a graph consisting of a non- vacuum finite set of elements known as vertices and a finite set of orders E of separate vertical couples known as boundaries.

A fuzzy set A is defined by a set X, which is recognised as an association function, by a mapping m: X → [0, 1]. The A = (X, m) is a fuzzy package. A fuzzy graph $\xi = (V, \sigma, \mu)$ is an algebraic structure of non-empty set V together with a pair of functions $\sigma : V \rightarrow [0, 1]$ and $\mu : V \times V \rightarrow [0, 1]$ such that for all x, y ∈ V , $\mu(x, y) \leq \sigma(x) \wedge \sigma(y)$ and μ is a symmetric fuzzy relation on σ. Here σ(x) and μ(x, y) represent the membership values of the vertex x and of the edge (x, y) in ξ respectively. The fuzzy graph $\xi 1 = (V, \sigma 1, \mu 1)$ is called a fuzzy subgraph of $\xi = (V, \sigma, \mu)$ if $\sigma 1(x) \leq \sigma(x)$ for all x and $\mu 1(x, y) \leq \mu(x, y)$ for

all edges (x, y), x, y ∈ V. Fuzzy exists in a picture in many respects, two of them in the following:

Type I: crisp vertex set and fuzzy edge set
Type II: crisp vertices and edges with fuzzy connectivity
Type III: crisp graph with fuzzy weights
Type IV: fuzzy set of crisp graphs

We denote $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. For the fuzzy graph $\xi = (V, \sigma, \mu)$, an edge (x, y), x, y ∈ V is called strong [2] if $1/2 \{\sigma(x) \wedge \sigma(y)\} \leq \mu(x, y)$ and it is called weak otherwise. The strength of an edge (u, v) is denoted by $I(u,v) = \mu(u, v) / \sigma(u) \wedge \sigma(v)$. The underlying crisp graph $G = (V 0 , E)$ of a fuzzy graph $\xi = (V, \sigma, \mu)$ is such that $V 0 = \{v \in V \mid \sigma(v) > 0\}$ and $E = \{(u, v)\mid \mu(u, v) > 0\}$.

A path [Eslahchi et al] in ξ is a sequence of vertices $x_0, x_1, \ldots, x_n$, such that $\mu(x_i-1, x_i) > 0$ and i = 1, 2, . . . n. The path is said to have length n. Two nodes that are joined by a path are said to be connected. Another part of a floating graph is a floating subgraph so that a path links any two vertices. Therefore, when one part is linked and otherwise removed, a fluffy graph is called fuzzy diagram.
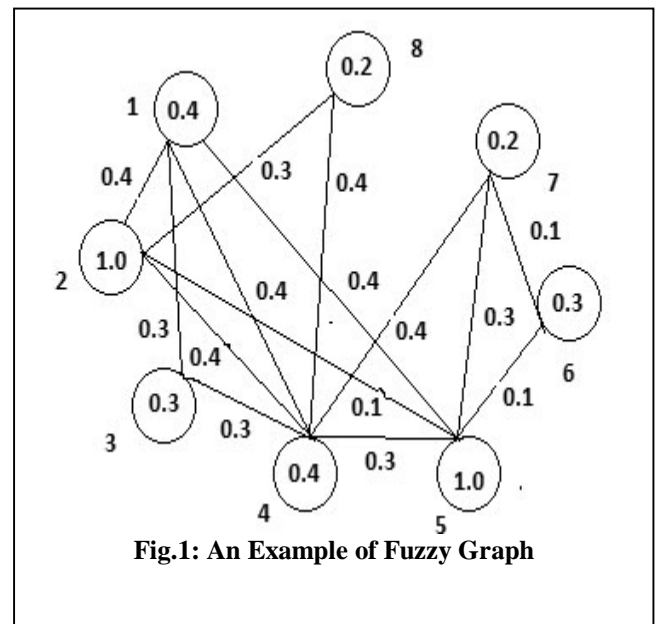

Fig.1: An Example of Fuzzy Graph

Figure.1: An Example of Fuzzy Graph

3. FUZZY GRAPHS IN MEDICAL DATA MINING RESEARCH

The analysis of graphs is a significant multidisciplinary area of physical, statistical, environmental, biological, social and analytical science. If single or directed graphs are used to depict complicated systems, the real-world structures under review are not fully represented. For examples, we can only say whether two individuals interacted or not in the communication network reflecting a specific diagram, but

cannot tell if three or more people connected to one network shared the same message. The communication network as an hypergraph represents the groups of persons in which vertices are persons and hyperedges exchanging the same message. In this document, we will focus specifically on examining the importance of complete fuzzy medical hypergraphs. Here Fig.1&Fig.2 shows the number of research articles published in the past 5 years.

Figure.2: No of Publications in Fuzzy Logic

4. FUZZY HYPERGRAPHS

It is not always "yes" or "no" in real life; nevertheless, it may have an effect in between. In the sense of imprecision, ambiguity, incompleteness of knowledge and contradictory facts partiality of fact and partiality of chance, we must therefore use dynamic reasoning to converse, to think and to make rational decisions. We'll define this segment as the fuzzy hypergraphs, which is different from those provided in ref. in terms of flowing connections[4]. In real life and technology, this definition will be helpful to explain many cases. Let U be a usual nonempty set and A ⊆ U a set. A fuzzy set ˜A on U is the set:

$$\tilde{A} = \{x, \mu \check{A}\ (x)) \mid x \in U\}$$

where the map $\mu\check{A}$: U $\rightarrow$ [0,1] is called membership function and the value $\mu\check{A}\ (x)\ \in [0, 1]$ is called the degree of membership of x in $\tilde{A}$

Let U = {Jafar, Mani, Narmi, Wasi} Define young U $\rightarrow$ [0, 1] such that

$$\text{Young} = \begin{cases} 1 & \text{if } x \leq 20 \\ \dfrac{30 - \text{age}(x)}{10} & \text{if } 20 \leq x \leq 30 \\ 0 & \text{if } x \geq 30 \end{cases}$$

Person	Age	Degree of youth
Jafar	10	1.00
Mani	21	0.90
Narmi	25	0.50
Wasi	83	0.00

Û = { (Jafar, 1), (Mani, 0.9), (Narmi, 0.5), (Wasi, 0) } .

Two sets often have to be compared in several problems. They often need to ask whether two shapes or pictures are identical or similar. Many scientists have studied the problem of measuring the similarity of fuzzy sets. Similarity measures are used in many fields, including image processing, region mining and coding theory. It is not possible to give precise interpretations or explanations of scientific terms and the relations between concepts usually as an essential part of the real-world applications of medicine. There are no clear limits. Remember the argument: "When the patient has a severe back pain and the patient is elderly, then administer the acupuncture for a long time at a certain point." They need more than mathematical skills and false statements to interpret that assertion in a computer system. All words which we need to model are serious, long-lasting, ancient, certain and fuzzy. For instance, the size as variable are defined by "very small," "small," "medium," "larger," "extremely large" and "monstrous," which is why soft computing (SC) methodologies must be used in applications for healthcare information systems. Here figure 3 shows the number of research articles published in Fuzzy Hypergraphs from the past 5 years.
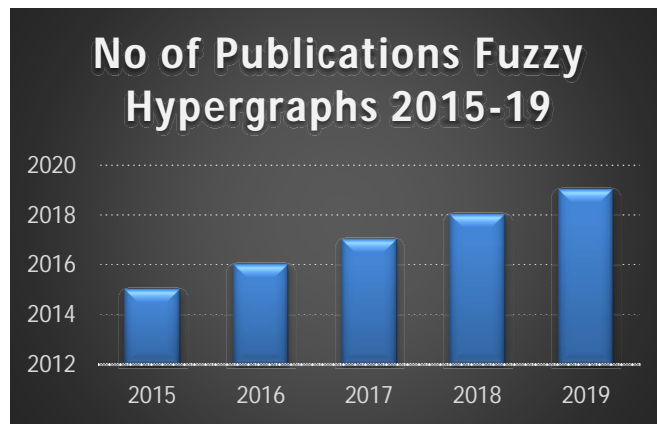

Figure 3: No of Publications Fuzzy Hypergraphs

It can be used alone and hybrid in other It can be used alone and hybrid in other technological applications, such as Neuro-fuzzy applications, Neural network, Fuzzy-Bayesian applications and Bayesian network combinations. FL has a broad range of medical application. Here figure 3 shows the number of research articles published in Fuzzy Hypergraphs from the past 5 years.

4.1 Ranking studies

Classification research is a significant element in medicine. The assessments, risk factors, characteristics, physician providers and results need to be determined for decision-making purposes [5] and [6]. Here figure 4 shows the number of research articles
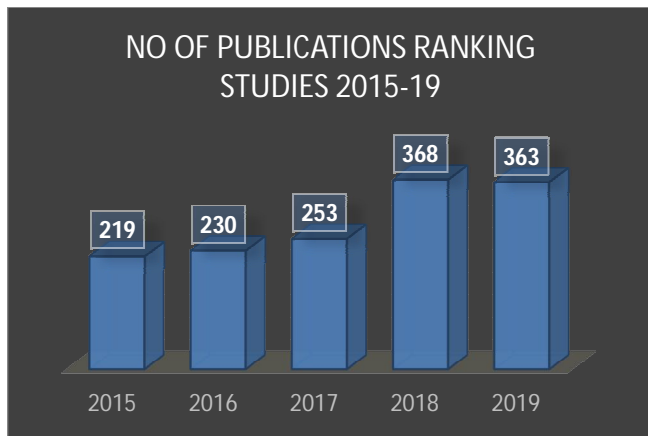

Figure 4: No of Publications Fuzzy Ranking Studies

4.2 Clustering studies

Clustering is a tool for data mining. The data are divided into groups (referred to as clusters) for the purpose of determining valuable correlations in which objects with similar properties within the same cluster and objects from separate clusters have different properties [6]. The main areas of clustering in medicine involve clustering tumours, organisms, genes and photos. Here figure 5 shows the number of research articles published in Fuzzy Clustering Studies from the past 5 years.
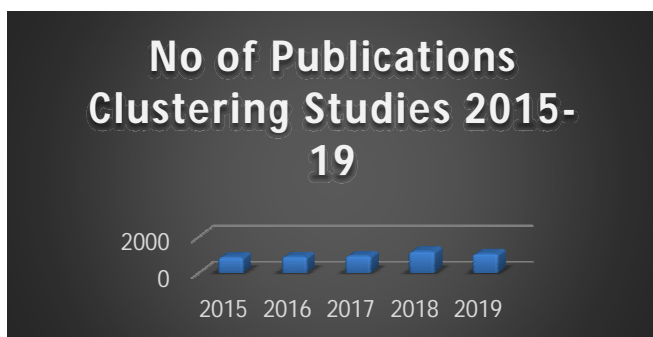

Figure.5: No of Publications Fuzzy Clustering Studies

4.3 Classification studies

Classification is also a method in data mining. Throughout sorting, we attempt to find representatives for the categories (classes) identified and predefined. Here figure 6 shows the number of research articles published in Fuzzy Classification studies from the past 5 years.
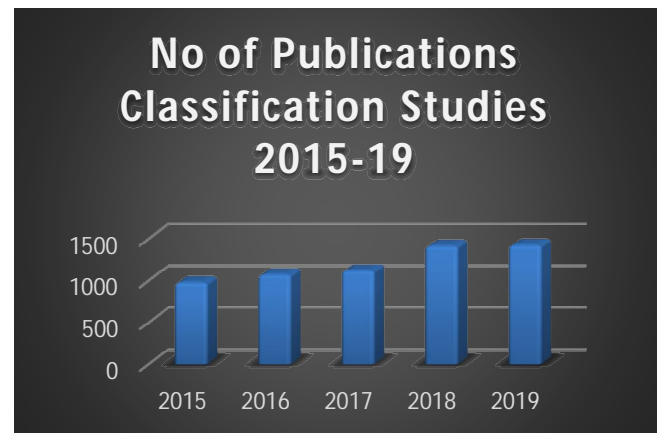

Figure.6: No of Publications Fuzzy Classification Studies

4.4 Pattern recognition studies

The recognition of patterns is a process of machine learning that decrypts the basic patterns in that topic [7]. Analysis of time series attempts to find templates and rules; this type of studies includes recognition of medical images. Here figure 7 shows the number of research articles published in Fuzzy Pattern Recognition from the past 5 years.
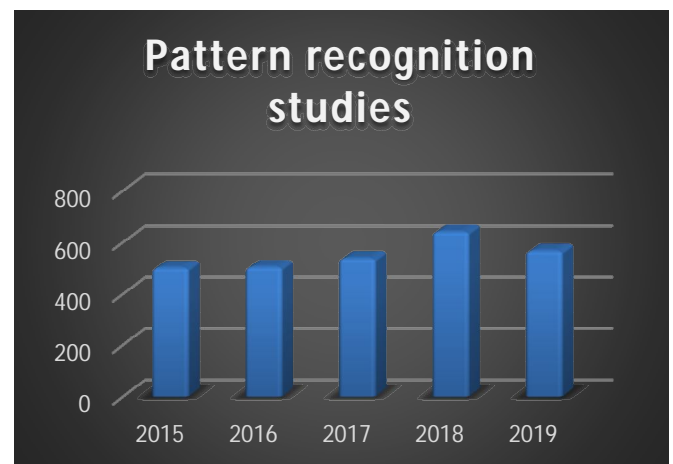

Figure.7: No of Publications Pattern Recognition

4.5 Feature selection studies

The selection of the feature is the approach that determines and removes irrelevant samples from the given sample space to help decision makers decide. This method is used to specifically remove the disease in patients from healthy cells / images / tissues. Here figure 8 shows the number of research articles published in Fuzzy Feature selection studies from the past 5 years.
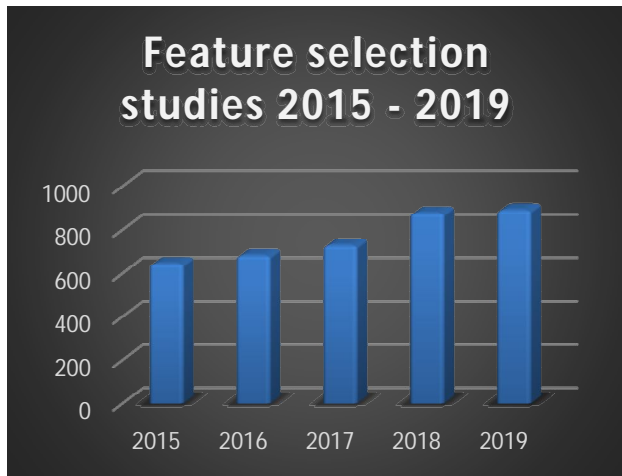
Figure.8: No of Publications Feature Selection Studies

5. MEDICAL IMAGE PROCESSING

Recognition of pattern is the way to identify shapes, cells and images. Classification and clustering are common pattern recognition strategies. The most popular medical identification strategies are fuzzy clustering, fuzzy rules-based approaches, fuzzy object-assimilation methods, and fuzzy interaction methods.One of the most widely employed methods of fuzzy clustering is the fuzzy C-mean algorithm. While many mechanisms have been developed for threshold-based image segmentation, clustering, area formation and so forth, the images taken by magnetic resonance angiography are not suited to the region of the brain tube. The methods of fuzzy granulation have proven to be successful. The description which perceives data as fumigated globules [8] can be narrowly portrayed. Here figure 9 shows the number of research articles published in Fuzzy Medical image processing from the past 5 years.
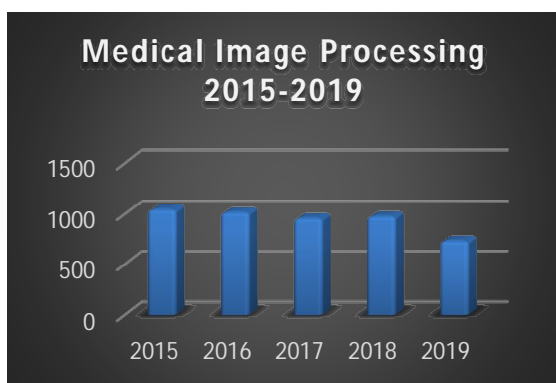

Figure.9: No of Publications Medical Image Processing

6. FUZZY EXPERT SYSTEMS IN MEDICINE

Medical staff can use pc-based health judgment-making software to detect diseases. One of those informatics systems that make medical decision-making simpler is a dynamic expert system (FES). For quantitative research and qualitative evaluations [12] Medical data, FESs proved to be helpful in medical diagnosis, in achieving precision of tests. Here figure 10 shows the number of research articles published in Fuzzy expert systems in medicine from the past 5 years.
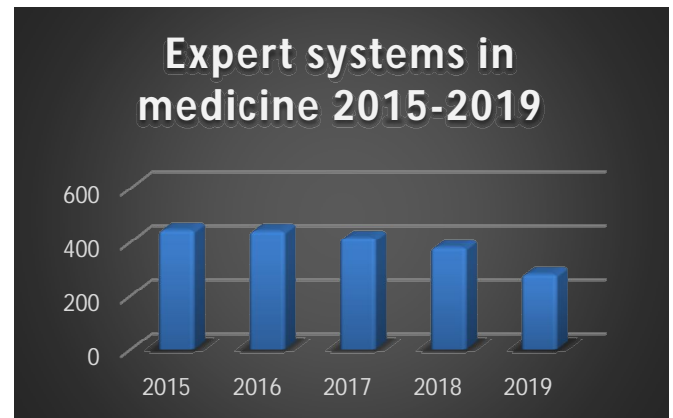

Figure 10: No of Publications expert systems in medicine

6.1 Fuzzy clinical repository information retrieval

Stats are as straightforward as any other computer field in clinical repositories. Additional approaches are needed to quickly obtain vague information from crisp databases. Three different techniques are given in literature to fulfill this need utilizing FL technology:
- ✓ Fuzzy query building
- ✓ Employing fuzzy ontology
- ✓ Constructing fuzzy relational databases.

Here figure 11 shows the number of research articles published in Fuzzy information retrieval in medical databases from the past 5 years.
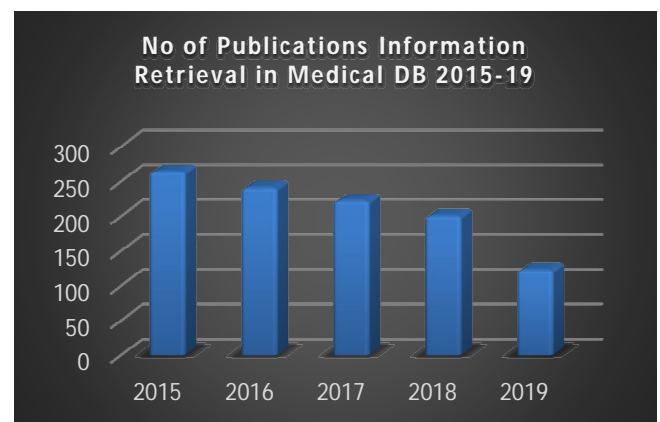

Figure 11: No of Publications Information Retrieval in Medical DB

6.2 Construction of Fuzzy Query In Health Repositories

A significant feature of FL in medicine involves the fluidity of patient records inquiries. Crisp relational query languages ' abstraction level is not adequate in order to provide questions that can meet the requirements of the individual information needs. In order to create a query such as "Mark the low level liver enzyme values in kids suffering from celiac disease," In the query language, we need to be able to designate "small" or "keys" vocabulary concepts. Inaccurate and complicated questions are treated using fuzzy queries. Here figure 12 shows the number of research articles published in Fuzzy query building in medical databases from the past 5 years.
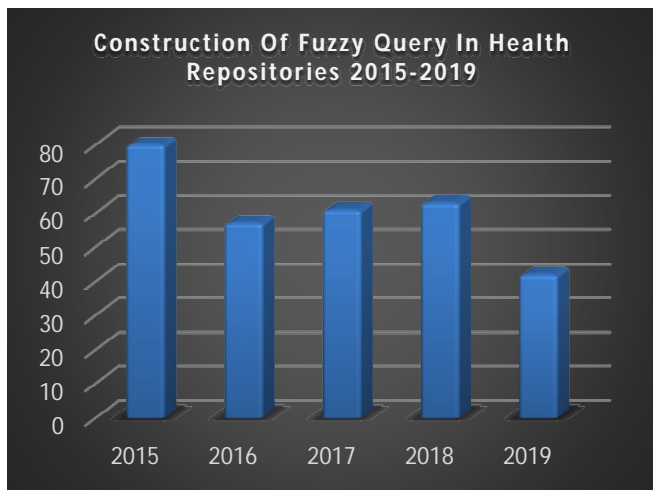

Figure .12: No of Publications Query Building in Medical DB 2015-19

6.2 Fuzzy ontology idea in the clinical domain

Several ontology concepts are present in literature. From the literature we may describe ontology as a structured vocabulary which defines domain knowledge, in a human and computer-readable type, as hierarchical concepts of persons, their attributes and relationships. Based on the above definition, the conceptual representation of medical knowledge in a symbolic form can be defined as a medical ontology. As with all crisp implementations, ontologies are built in a crisp definition and suffer from the same lack in versatility like FL [11]. Once again, the approach is to use and use an FL definition in ontology. The result is an ontology that is fuzzy. In a fuzzy ontology, throughout fact, each object has a certain degree of membership in comparison to another entity. Here figure 13 shows the number of research articles published in Fuzzy ontology concept in the medical domain from the past 5 years.
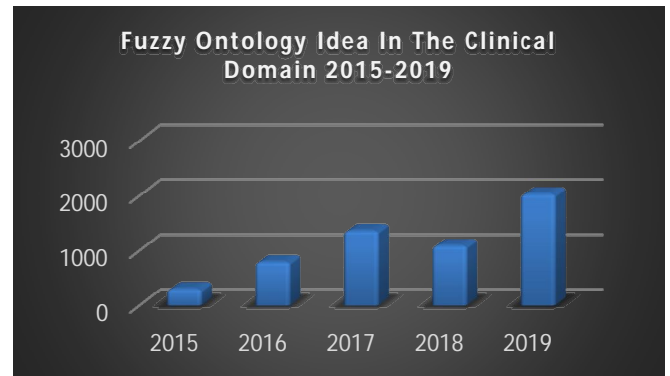

Figure.13: No of Publications Ontology in Medical Domain 2015-19

6.3 Conceptual Fuzzy Data Sources In The Healthcare Field

The Fuzzy Relational Data Base Management System (FRDBMS) is, as in other methods of the FL systems, an expansion of the DBMS. The data interpretation in the database is performed in this method in a soft language fashion. The Fuzzy DB paradigm provides insights on the area of medical information and will allow data management in the areas of medical information simpler and more intelligible. Here figure 14 shows the number of research articles published in Fuzzy relational databases in the clinical sector from the past 5 years.
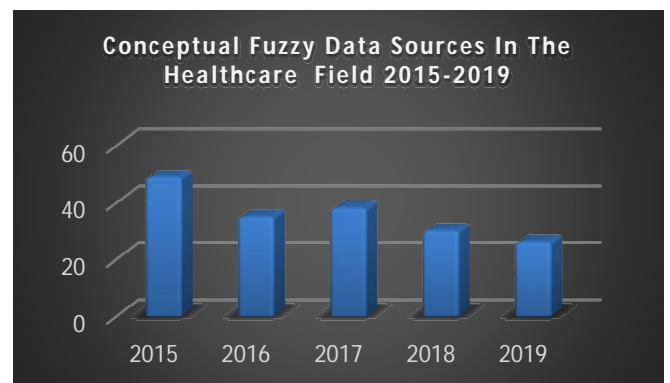

Figure.14: No of Publications Relational DB in Medical Domain 2015-19

6.4 Fuzzy Cognitive Maps In The Surgical Necessity

In the previous section, the complex and complicated complexity of clinical decision-making is established. Such dynamics and tightness should be concerned with in systems that support medical decision making. A medical support network for decision making should be able to draw causal awareness from the patient literature, build up and deduce a causal knowledge base [9]. When designing the FCM, the most tough challenge is to assess the effects and the effect of ideas on other individuals. Historical information is used to

measure these forces and their weights. Throughout our repositories we have huge amounts of medical data. The relationships and weights can be used to create FCM models. In Figure 15, the amount of research papers reported on Fuzzy cognitive maps in surgical necessity in the last 5 years is highlighted.
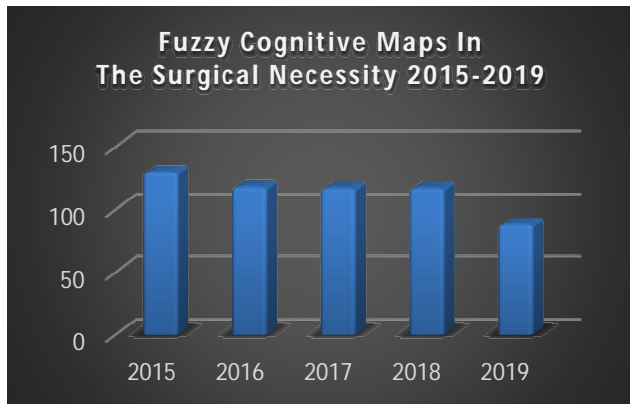

Figure.15: Fuzzy C-Map in Medical Decision Making

7. CONCLUSION

In this analysis we discussed the advanced research and developments in the field of Fuzzy and Graph and the methods used for its medical applications. The persistent conflict between medical personnel and the computer employees, in which medical personnel maintain that they are unable to generalize the illnesses, while computer workers compel them to mold and generalize their definitions, seems to be stopping with SC strategies. With computer staff learning / developing flexible modeling methods and improving / employing techniques that operate smoothly rather than narrowly and one-offly, medical decisions are more supportive. A rapidly developing field of science is the medical field. New medical techniques and devices endive every day. This growth, including medical information technology, is also strongly supported by the participating and supporting divisions. The number of researchers in favor of healthcare is growing every day. Such scholars and their work are also being funded. As a result, the number of studies and published articles in literature increases, as previously stated. With this growing interest, future employment prospects will be given not only in pharmacy, but also in participating and helping areas.

Fuzzy and Graph are used in every critical decision-making process in the health sector from the supply chain to diagnosis, from the mining health information and the obtaining of information. By using FL, we will make the system scalable and stable through taking all potential values into account, including those which are bubbled. Further effort and time are needed to develop a FL system or application. It lengthens the estimation period for the desired

production, however, provides more accurate results in the medical field when working with darkness.

REFERENCES

1. A. Rosenfeld, Fuzzy graphs, in: L.A. Zadeh, K.S. Fu, M. Shimura (Eds.), **Fuzzy Sets and Their Applications**, *Academic Press, New York,* 77-95, 1975.

2. C. Eslahchi and B. N. Onaghe, **Vertex strength of fuzzy graphs**, *International Journal of Mathematics and Mathematical Sciences*, Volume 2006, Article ID 43614, Pages 1-9, DOI 10.1155/IJMMS/2006/43614.

3. M. Akrama and W. A. Dudek, **Information Sciences** 218, 182 (2013).

4. M. W. L. Craine, **Fuzzy hypergraphs and fuzzy intersection graphs**, *Ph.D. Thesis, University of Idaho* (1993).

5. Narasimhan B, Malathi A. **A Fuzzy Logic System with Attribute Ranking Technique for Risk-Level Classification of CAHD in Female Diabetic Patients**. In: *Intelligent Computing Applications (ICICA), International Conference* on March, 2014. IEEE. 2014. p. 179-83.

6. Bhandari P. **Study of various clustering algorithms used by WEKA tool**. *Int J Emerging Res Manag Technol* 2015;4:37-40.

7. Karmakar N. **Generation of Authentic Signature-sets Utilizing Case-Based Reasoning Techniques** (Doctoral dissertation); 2014.

8. Kobashi S, Hata Y, Hall LO. **Fuzzy information granulation of medical images. Blood vessel extraction from 3-D MRA images.** In: *Fuzzy Logic in Medicine, Studies in Fuzziness and Soft Computing*. Vol. 83. Heidelberg: Physica-Verlag; 2002. p. 18-35.

9. Kolçe E, Frasheri N. **A Literature Review of Data Mining Techniques Used in Healthcare Databases**. *Paper Presented at the ICT Innovations* 2012, Ohrid, Macadonia; September, 2012.

10. Ibrahim, S. Jafar Ali, and M. Thangamani. "**Enhanced singular value decomposition for prediction of drugs and diseases with Hepatocellular carcinoma based on Multi-Source Bat Algorithm based random walk."** Journal of *Measurement, Elsevier,* 141 (2019): 176-183.

11. Thangamani M., and S. Jafar Ali Ibrahim. **Ensemble Based Fuzzy with Particle Swarm Optimization Based Weighted Clustering (Efpso-Wc) and Gene Ontology for Microarray Gene Expression.** In *Proceedings of the 2018 International Conference on Digital Medicine and Image Processing*, ACM pp. 48-55. 2018.

12. Ibrahim, S.jafar Ali and M.Thangamani. **Prediction of novel drugs and diseases for hepatocellular carcinoma base of multi-source simulated annealing based random walk**, *Journal of medical systems,*42, no.10 (2018):188.