

Proposed Model to Detect Uterine Fibroid By using Data Mining Techniques

Girija D.K¹, Dr. Manish Varshney²

¹Research Scholar, MUIT, Lucknow.

girijadk16@gmail.com

²Professor, MUIT, Lucknow.

itsmanishvarshney@gmail.com

ABSTRACT:

Data mining is a multi-step process that uses a large amount of data. All of the systems components are interrelated. To get started, we need to get some raw data. The data is first pre-processed and then converted into formats that may be used by the various mining tools. The missing values can be supplied either during pre-processing or via a separate tool, as will be discussed later in this WEKA programme section. Diverse data mining techniques uncover patterns in the cleansed material that are common to both. Commonalities in data can be discovered through the identification of patterns and rules in the data. This information can be used to classify the female uterine fibroid with precision.

Keywords: Data Mining, Uterine, Fibroid, Detecting, Techniques and Fibroid.

Introduction:

It is the goal of system architecture efforts to describe a holistic solution based on logically linked principles, concepts, and properties. It is the goal of the solution architecture to meet the problem or opportunity expressed by a set of system requirements (traceable to mission/business and stakeholder requirements) and life cycle concepts (for example, operational, support) as far as possible, while also being implementable through technologies (e.g., mechanics, electronics, hydraulics, software, services, procedures, and human activity).

To achieve the system's goal and life-cycle concepts in an abstract, conceptualization-oriented, global, and concentrated way, it also emphasizes system and system element high-level structure. You'll learn about a variety of different aspects of the system-of-architecture. interest's More than one system can benefit from this approach, as it can serve as a common structure, pattern, and set of requirements for several classes and families of comparable or related systems.

Components of System Architecture:

The underlying and unifying structure of a system characterized in terms of its components, interfaces, processes, restrictions, and behaviors is known as the systems architecture. This includes the system's physical properties, style and structure as well as the systems relationships and purpose.

System Architecture Diagram:

An architecture diagram is a representation of a system that individuals use to abstract the broad structure of a software system and construct restrictions, relationships, and boundaries between components. It provides a comprehensive perspective of the software system's physical deployment and evolution plan.

Function of System Architecture:

A diagram resembles a picture in many ways. Examples of architecture diagrams can be used for many different purposes. The relevant users always benefit from learning about system

architecture and putting it to use in their decision-making processes. It's essential to convey knowledge about architecture effectively. However, before drawing a diagram of architecture, people must follow a set of rules. To name a few:

- Breaking through the barriers of communication.
- Reaching agreement.
- Reducing the amount of room for interpretation.

The Below figure 1 show the “Proposed Model to Detect Uterine Fibroid By using Data Mining Techniques”. The proposed model or architecture includes the following steps that should be used in further works.

- Collected Data.
- Pre- Processing.
- Filling Missing Values.
- Applying the dataset to different data Mining Algorithm.
- Result Obtained and Checking Prediction.

Our Proposed Architecture:

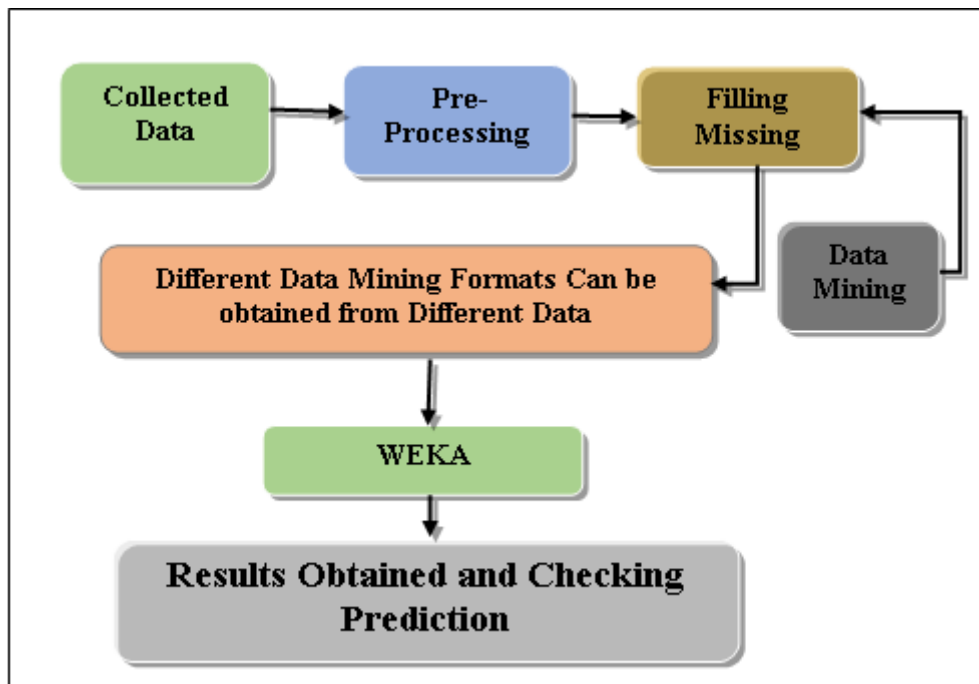


Figure 1: Our Proposed Architecture

1. Collected Data: The dataset of the commonest kind is divided into two sub-datasets: the testing set and the preparation set. The preparation arrangements have been labelled, but the testing set has not been labelled. We have just used the data from the indicated dataset that was prepared. From 450 patients, 450 records from reputable hospitals have been gathered and analysed. Due to the fact that additional data is confidential and cannot be disclosed, the study restricts access to 12 fields, which contain the following attribute names: age, status, HB, PP, FT, LBP, PDI, FU, NEP, SF, CAUSES, and CLASS.

Table 1: Shows the different attributes used in our approach

Data Set Description	
Attribute Name	Description
Age	Age
STATUS	status(Married, Single)
HB	Heavy Bleeding(3 to 4 days - No, More than 7 day-HIGH)
PP	Pelive Pan(High, No)
FT	Fibroid Type(INTRACAVITARY,SUBMUCOSAL,SUBSEROUSAL,PEDUNCULATED,INTRAMURAL)
LBP	Lower Backpain(High, NO)
PDI	Pain During intercourse(High, NO)
FU	Frequet Urination(Yes, No)
NFP	Number of Fibroid Present(Multiple, Single)
SF	size of fibroid(1mm to 20CM -(8 inches) in diameter or even larger)
CAUSES	Causes(INFERTILITY ,ANEMIA ,SWELLING IN THE ABDOME, NO EFFECT OF FERTILITY,PREVEMENT SPERM, NO EFFECT,EFFECT)
CLASS	Class(Eliminate, KEEP)

2. Data Pre- Processing: Data Pre-Processing raw data into a comprehensible format is known as Data pre-processing. In data mining, this is a crucial stage because we can't deal with raw data. Check the data quality before using machine learning or other algorithmic techniques. In this step we translate the collected raw data from the reputed hospital to the format that should be understood by the data mining Weka tools.

3. Filling Missing Values: There are times when some attributes are either missing or imperfect. As the most common technique of indicating missing data, the "-1" number is commonly used. Attributes that are typically vacant might lead to a false impression that the data set represents an undervalued situation. In this example, however, this is not the case because each of the other traits gives essential information about the category of qualities in which it features. It is possible to use a number of methods to fill in any missing qualities, rather than simply leaving them blank. A wide range of data mining processes benefit from accurate ways to filling in missing values. It's possible to enhance data classification with extra training, such as by improving the accuracy of the tool. For instance.

4. Applying the dataset to different data Mining Algorithm: This work necessitates the creation of a model to detect and categories data types and ideas. Classification is the process of determining which category (subpopulation) a new observation belongs to based on a training set

that contains observations and whose category membership is known. A training set of data is a set of data that contains observations and is known to belong to one of the categories. After this observation, I applied my dataset to the weak tool, by using standard different data algorithm like KNN, SVM, Random Forest, J48, and ZeroR.

5. Result Obtained and Checking Prediction: I used conventional data techniques like KNN, SVM, Random Forest, J48, and ZeroR to apply my dataset to the weak tool. We collect results for each data mining algorithm by running the dataset through several data mining tools. Then, based on the results, I can forecast the presence of a uterine fibroid. In addition, by comparing all of the algorithms, I can determine which one is the best.

Conclusion

The architecture of our model system was designed in this section, which was done with the help of the data mining tool WEKA. Collected Data, Pre- Processing, Filling Missing Values, Applying the dataset to different data Mining Algorithm, Result Obtained and Checking Prediction and other preparation assessment activities are coordinated with the intended architecture design in order to get the desired results. Considered properties, together with the definitions and values that relate to them.

Reference

- [1] Elizabeth A. Pritts, M.D.,a William H. Parker, M.D.,b and David L. Olive, M.D.a “Fibroids and infertility: an updated systematic review of the evidence”
- [2] [Neelamadhab Padhy, Dr. Pragnyan Mishra and Rasmita Panigrahi, “The Survey of Data Mining Applications and Feature Scope, *International Journal of Computer Science, Engineering and Information Technology (IJCEIT)*”, vol.2, no.3, June
- [3] Yogeesh N, "Graphical representation of Solutions to Initial and boundary value problems Of Second Order Linear Differential Equation Using FOOS (Free & Open Source Software)-Maxima", *International Research Journal of Management Science and Technology (IRJMST)*, 5(7), 2014, 168-176
- [4] Heikki, Mannila, *Data mining: machine learning, statistics and databases*, IEEE, 1996.
- [5] Yogeesh N, "Mathematical Approach to Representation of Locations Using K-Means Clustering Algorithm", *International Journal of Mathematics And its Applications (IJMAA)*, 9(1), 2021, 2347-1557
- [6] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P, *From Data Mining To Knowledge Discovery in Databases*, The MIT Press, ISBN 0-26256097-6, Fayyad, 1996.
- [7] Piatetsky-Shapiro, Gregory, *The Data-Mining Industry Coming of Age*, "IEEE Intelligent Systems", 2000.
- [8] Jing He, *Advances in Data Mining: History and Future*, Third international Symposium on Information Technology Application, 978-0-7695-3859-4, IEEE, 2009.
http://shodhganga.inflibnet.ac.in/bitstream/10603/22904/7/07_chapter2.pdf
- [9] Flach, P. and Gamberger D. (2001) Subgroup evaluation and decision support for a direct mailing marketing problem. *Aspects of Data Mining, Decision Support and Meta-Learning*,
- [10] Yogeesh N, "Study on Clustering Method Based on K-Means Algorithm", *Journal of Advances and Scholarly Researches in Allied Education (JASRAE)*, 17(1), 2020, 2230-7540
- [11] [Online] available from: <http://www.informatik.unifreiburg>