

How to Cite:

Girija, D. K., & Varshney, M. (2022). Uterine fibroid risk prediction using data analytics and support vector machines in data mining. *International Journal of Health Sciences*, 6(S2), 4125–4133. <https://doi.org/10.53730/ijhs.v6nS2.5929>

Uterine fibroid risk prediction using data analytics and support vector machines in data mining

Girija D. K

Research Scholar, MUIT, Lucknow
girijadk16@gmail.com

Manish Varshney

Professor, Maharishi School of Engineering & Technology, MUIT, Lucknow.
itsmanishvarshney@gmail.com

Abstract--In healthcare, the amount and sensitivity of data is huge. The information must be handled with considerable caution, and no shortcuts should be made. A variety of data mining categorization approaches have been used to estimate the quality of healthcare services. On the basis of 150 patient records, this study describes and evaluates the experience of applying a data mining approach and methods. Using data mining techniques, this unique approach to determining the correctness of a product was devised. Data mining employs a number of techniques, including Decision Tree, Naive Bayes, KNN, Radom Tree Set, Rule Model, ZeroR, and J48 or C 4.5. Using the fibroid data set, I applied the Support Vector Machine data mining classification to estimate the Precision. Our goal is to make a conclusion by analyzing the Precision and other indicator values such as RMSE, Recall, and so on.

Keywords--data mining, healthcare, precision, techniques, uterine fibroid.

Introduction

Increasing amounts of data and the need to extract information and expertise from it have made data mining popular recently. According to the logic of information technology, data mining is the next step in its progression. Database system company's data group and database development, management (such as database storage/recovery/operation processing), as well as fresh data analysis, have all progressed (including data warehousing in addition to data mining).

Data mining is becoming increasingly popular in the healthcare industry due to its enormous success. The fibroid data set was utilised. There are many reasons why we've decided to go with it. It is the most frequent type of tumour in female reproductive organs to be fibroids. Smooth muscle cells and fibrous connective tissue make up fibroids, which are uterine tumours. In this study, numerous categorisation techniques are examined.

Data Mining

Large data sets can be analyzed to help businesses solve problems by recognizing patterns and connections that can be uncovered through data mining. Businesses can use these techniques and technologies to foresee future trends and make more informed decisions.

Types of Data Mining

Knowledge may be extracted from enormous databases using data mining techniques. Descriptive and prescriptive data are both types of data (Hong and Weiss, 1999; Han and Kimber, 2002). The descriptive model is used to summarize facts and draw conclusions from the information. Database summaries and visualizations make advantage of descriptive data mining. Using a variety of abstraction levels, one may examine a data set's overall behavior, which is difficult to deduce from a large database. As opposed to traditional data analysis, predictive analytics aims to create models that can foretell future outcomes. Cataloguing and regression are two of the most common mining jobs. Classification is necessary for data mining. This type of machine learning, known as data mining, is influenced by pattern recognition, which is a study that deals with categorizing things into a number of separate classifications. A pattern is a discrete unit of data that is specific to a particular scenario.

Healthcare Application In Data Mining

When it comes to healthcare transactions, typical systems just can't handle the amount of data that needs to be collected and analysed. To make better decisions, it's important to look for patterns and trends in large amounts of data. As financial constraints have increased, so has the need for healthcare organisations to make decisions based on analysis of medical and financial data. Data mining insights can influence cost, revenue, and operating efficiency while maintaining a high level of care.

Systematic Relevance and Development

A very small number of patients are being studied in this study. In a larger study, there may be a variety of reductions in radiation exposure, both proportionately and cumulatively. Only one out of the 25 patients who participated in the study received additional OAE, even though the collateral flow was only seen in less than one percent of patients. Larger studies, in our opinion, are more likely to capture the true nature of the phenomenon. After embolization, aortograms were investigated to see if they may modify holy arterial flow morphologies and so reduce ovarian artery angiographical empathy. It is our opinion that ovarian

migration to fibroids is best identified after the first supply from the uterine arteries has been terminated.

About Fibroid

Fibroids are benign (non-cancerous) growths found in the uterus of most women (womb). It is most common to find fibroids in the uterine hedge, which is where they are most often found in women. The uterine wall is the only place where they may develop or replicate. One person or a group can spread it. As long as no symptoms like excessive menstrual flow or incontinence are present, obstetric fibroid tumours are considered "benign" (not cancerous).

Fibroids are abnormal growths in or on the uterus that are common in females. uterine fibroids are harmless growths of connective tissue and smooth muscle. Abdominal discomfort and inconsistent sleep patterns can be caused by large-scale tumours of this type. They may not display any symptoms at all in some circumstances.

In the uterus, there are two types of fibroids: those on the inner and outer borders, and those that originate in the middle. Fibroids are one of the most common causes of fibroids, which include tumours, myomas, and leiomyomas. Many problems can be traced back to them, despite the fact that they are not cancerous. The vast majority of women don't discover they have fibroids until they have a problem with them.

Sources of Fibroid

The uterine fibroid exhibits a high degree of reciprocity. Many women have an opinion about uterine fibroids throughout their lives. Small or undetectable uterine fibroids do not represent a danger to the majority of women. It is thought that uterine fibroids may be caused by the presence of oestrogen, a female hormone. After adolescence, fibroids begin to form in the uterus on a sporadic basis until the age of 30. Menopause reduces oestrogen levels, which leads to shrinkage or disappearance of fibroids. Uterine fibroids symptoms are more common in African American women, and they have a higher risk of acquiring fibroids than white women. It's difficult for physicians to say for sure what causes uterine fibroids, however they do know that,

- ❖ Genetics: Genetic changes many fibroids have gene modifications that are different from those found in normal uterine muscle cells.
- ❖ Hormones: Oestrogen and progesterone, two hormones that encourage the formation of uterine covering during each menstrual cycle, appear to increase fibroids. Oestrogen and progesterone receptors are higher in fibroids than in normal uterine muscle cells. Fibroids usually disappear when a woman enters menopause. Uterine fibroids can also grow as a result of, among other things.
- ❖ Pregnancy: Fibroids are less common in women who have children after they have given birth.
- ❖ Early menstruation: If a woman has her first period before the age of ten, her chances of getting uterine fibroids increase.

- ❖ Family history: Fibroids in the uterus are a secondary effect of having a family history of the disease.

Uterine Fibroids Analysis

When a doctor performs a physical pelvic examination, uterine fibroids of all sizes are more likely to be discovered. Imaging exams are performed on a regular basis to determine the presence of uterine fibroids.

- ❖ An ultrasound probe placed in the vagina or above the abdomen's pelvis bounced heavy-frequency sound waves off the uterus and pelvic edifices. The uterus and any uterine fibroids are shown in a cinematic representation.
- ❖ Magnetic resonance imaging (pelvic MRI): Using a strong magnet and a computer, an MRI scanner creates extraordinarily detailed images of the uterus and other pelvic tissues. A pelvic MRI can be utilised to confirm the existence of uterine fibroids if the diagnosis is certain.

Declaration of the Problem

- ❖ Problem The healthcare system in India has long been a source of national pride, but there is still space for improvement in terms of patient outcomes and healthcare expenditures as a whole. Blocking measures are essential in today's healthcare. The efficacy of uterine fibroid treatments is being proven in more and more trials.
- ❖ Excessive bleeding from uterine fibroid tumours has killed many women around the world. There was only enormous agony and death in remote areas where innocent girls could not receive such knowledge. There will be no low-cost diagnosis if there is no easy access to data or symptoms.

What is the learning's objective?

- ❖ In the future, data mining techniques and systems will have a substantial impact on all levels of the 0–1 scale in the Indian healthcare system.
- ❖ Using the Weka data mining tool, we calculated the precision and other parameters by applying the Support Vector Machine classifier to our dataset.

Introduction To Weka Software

Many Data Mining techniques may be found in the Weka workspace, including classification, regression, and clustering. This comprehensive system includes data input, statistical evaluation of learning outlines, and visual depiction of input and learning results. To quickly stab out old ways, it makes use of a variety of methodologies and fresh data. Methods for learning include a wide range of pre-processing approaches. When faced with a given set of issues, users can compare and contrast different ways using these tools.

The structure established at the University of Waikato in New Zealand is known as Weka, which stands for "Waikato Atmosphere for Knowledge Analysis." Outside

the university grounds, the Weka, a non-flying bird native to the New Zealand islands with an inquisitive temperament (its name rhymes with "Mecca"), can be seen. The use and dissemination of the Java software is governed under the GNU General Public License. Under these operating systems, it has been tested on Linux, Windows, Macintosh, and even a personal digital assistant. It provides a consistent boundary for a variety of learning algorithms, as well as methods for pre- and post-processing and computing the results of education structures on any dataset.

Table 1: The different attributes that utilized for my Work

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,SUBSEROSAL,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,PREVENTMENT SPERM, NO EFFECT,EFFECT)
CLASS	Class(Eliminate, KEEP)

The above table: 1 gives the different attributes that utilized for my Work and their descriptions that that should use in this paper for implementation purpose, this data are calm from reputable hospitals have been collected and analysed. Due to the fact that additional data is private 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.

Weka Implementation Using Support Vector Machine Classifier

Support Vector Machine

Both regression and classification employ elastic supervised data mining methods such as Support Vector Machines (SVMs). Classification difficulties are a popular use for this technique. Originally developed in the 1960s, SVMs underwent significant improvements about 1990. SVMs have a unique implementation strategy when compared to other data mining methods. Recently, they've become popular due to their ability to handle values that are both continuous and categorical (variables).

Advantages of SVM classifiers

- ❖ There is a high level of stability since support vectors are used instead of data points.

- ❖ Because they only employ a small subset of training data, SVM classifiers are incredibly memory efficient.
- ❖ High-dimensional settings do not pose a problem for SVM classifiers because of their accuracy.
- ❖ He is unaffected by the outliers.
- ❖ The datasets were devoid of any standards.
- ❖ The problem of numerical predictions may be solved with SVM.

Disadvantages of SVM classifiers

- ❖ The use of the black box method. Predisposition to overfitting extremely complex computations are carried out.
- ❖ Because of their lengthy training period, they are unable to analyse large datasets.
- ❖ Additionally, SVM classifiers do not perform well with crossing classes.

Types of SVM

Linear SVM: A linear SVM may categorise data by dividing it into two categories using a single straight line. Because the data is separable, a linear SVM classifier is applied.

Non-linear SVM: Classifier Non-linear SVM is used for non-linear data that cannot be classified by straight lines, hence non-linear data is non-linear and Non-linear SVM is the classifier utilised.

Working of SVM

Hyperplane representations of separate classes are the SVM ideal in multidimensional space. In order to reduce error, SVM will iteratively build the hyperplane. In order to find the most extreme marginal hyperplane, SVM divides datasets into classes (MMH).

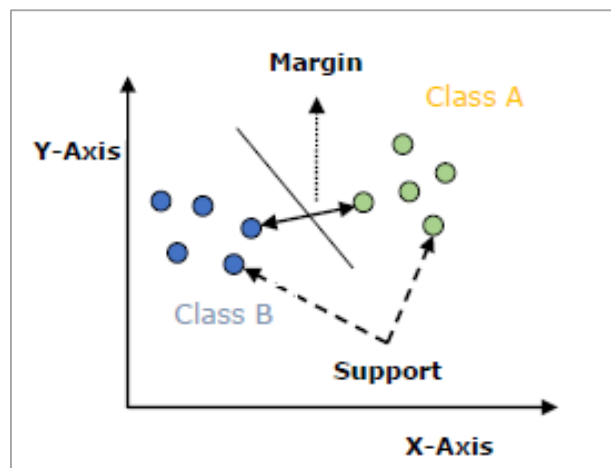


Figure 1: SVM

Some key SVM ideas are outlined in the following paragraphs.

Support Vectors – Because of their proximity to the hyperplane, these locations are evidence. These data points are used to explain the dividing line in the data set.

Hyper plane – You can see this in the figure above, where a set of unique items are separated by a decision plane (space).

Margin – It may be defined as the distance between the nearest data points on two straight lines for different classifications. It is possible to calculate the perpendicular space between a line and the support vectors. A large margin is considered excellent, whereas a small margin is considered poor.

SVM's primary goal is to divide the provided datasets into classes in order to find a maximum marginal hyperplane (MMH), which may be done in two stages,

- ❖ In the beginning, SVM will build hyperplanes that best distinguish the classes.
- ❖ In the second step, it will choose a hyperplane that best divides the groups.
- ❖

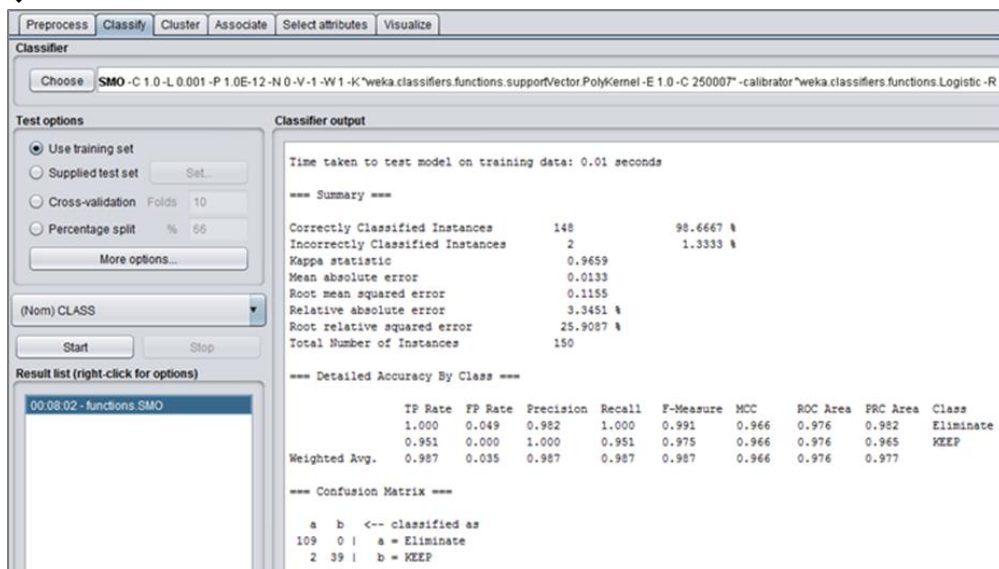


Figure 2: Support Vector Machine Classifier Result

Different Indicators of the Support Vector Machine Classifier

Classifiers/ Indicators	Precision	RMSE	ROC Area	MCC	Recall	F- measure
Support Vector Machine	98.667%	0.1155	0.976	0.966	1.000	0.991

Conclusion

The classifier's precision is calculated using Support Vector Machines. Class-eliminate datasets may be discarded in order to save time in predicting and

treating patients. For medical and instrumentation purposes, the "classes-keep" dataset is used to predict future treatment outcomes.

Precision of the Support Vector Machine classification algorithm is calculated and it is represented in the above figure, from the results Support Vector Machine algorithm classifies the given data 98.6667% correctly, RMSE = 0.1155, Recall = 1.000 and it is useful to decide whether we have to keep the fibroid patient in the hospital for treatment or eliminate if it is starting stage.

Reference

1. Theodore Dalamagas, Panagiotis Bouros, Theodore Galanis, Magdalini Eirinaki, and Timos Sellis. Mining user navigation patterns for personalizing topic directories. In 9th International Workshop on Web Information and Data Management, pages 81-88, 2007.
2. Peter Brusilovsky. "Methods and techniques of adaptive hypermedia". User Modeling and User-Adapted Interaction, 6(2):87-129, 1996. [
3. Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Improving the effectiveness of collaborative filtering on anonymous web usage data. Technical report, 2001.
4. Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. Design science in information systems research. MIS Quarterly, 28(1):75-105, 2004.
5. Yogeesh N. "Study on Clustering Method Based on K-Means Algorithm." Journal of Advances and Scholarly Researches in Allied Education (JASRAE), vol. 17, no. 1, 2020, pp. 485-489(5), www.ignited.in//I/a/305304.
6. Yogeesh N. "Mathematics Application on Open Source Software." Journal of Advances and Scholarly Researches in Allied Education [JASRAE], vol. 15, no. 9, 2018, pp. 1004-1009(6), www.ignited.in/p/304819.
7. Alberto Pan, Juan Raposo, Manuel Alvarez, Paula Montoto, Vincente Orjales, Justo Hidalgo, Lucia Ardao, Anastasio Molano, and Angel Vina. The denodo data integration platform. In 28th International Conference on Very Large Data Bases, pages 986-989, 2002.
8. Aykut Firat. Information Integration Using Contextual Knowledge and Ontology Merging. PhD thesis, Massachusetts Institute of Technology, 2003.
9. Yogeesh N. "Mathematical Approach to Representation of Locations Using K-Means Clustering Algorithm." International Journal of Mathematics And its Applications (IJMAA), vol. 9, no. 1, 2021, pp. 127-136.
10. Ricardo Baeza-Yates and Berthier A. Ribeiro-Neto. Modern Information Retrieval. ACM Press, 1999.
11. Sanjay K. Madria, Sourav S. Bhowmick, Wee Keong Ng, and Ee-Peng Lim. Research issues in web data mining. In 1st International Conference on Data Warehousing and Knowledge Discovery, pages 303-312, 1999.
12. Jose Borges and Mark Levene. Data mining of user navigation patterns. In Workshop on Web Usage Analysis and User Profiling, pages 31-36, 1999.
13. Dominik Flejter, Karol Wieloch, and Witold Abramowicz. Unsupervised methods of topical text segmentation for polish. In Workshop on Balto-Slavonic Natural Language Processing, pages 51-58, 2007.
14. Susan Dumais, John Platt, David Heckerman, and Mehran Sahami. Inductive learning algorithms and Conference on Information and Knowledge Management, pages 148-155, 1998.

15. Michael Steinbach, George Karypis, and Vipin Kumar. A comparison of document
16. Lawrence Kai Shih and David R. Karger. Using URLs and table layout for web classification tasks. In Stuart Feldman, Mike Uretsky, Marc Najork, and Craig Wills, editors, 13th International Conference on World Wide Web, pages 193-202, 2004.
17. Girija D.K & M. S. Shashidhara Data mining techniques used for uterus fibroid diagnosis and prognosis In Kottayam, India, 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), INSPEC Accession Number: 13567035.
18. Girija D.K & M. S. Shashidhara Data mining approach for prediction of fibroid disease using neural networks, Bangalore, India, 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA), INSPEC Accession Number: 14130830.