**Table of Contents**

1. **Abstract**

**2. General Terms**

**3. Key words**

**4. Introduction**

**5. Literature review**

**6. Overview of Hysterectomy**

**6.1. Hysterectomy**

**6.2 Why it is occur?**

**6.3 Type of Hysterectomy**

**6.3.1 Vaginal Hysterectomy**

**6.3.2 Laparoscopic Hysterectomy**

**6.3.3 Minilaparotomy Hysterectomy**

**6.3.4 Abdominal Hysterectomy**

**6.4 Risk factors for Hysterectomy**

**7. Overview of Methodologies**

**7.1 Data Mining**

**7.2 Naïve Bayes:**

**7.3 Support Vector Machine:**

**1. Abstract**

Nowadays, Hysterectomy has become a common major surgical procedure in Gynecology & usually performed to improve quality of life rather than to cure life threatening condition. A Hysterectomy is a surgical remove of uterus. The rate of hysterectomy surgical patient is increasing day by day all over the world. With the rise of information technology and its continued advent into the medical sector, the cases of hysterectomy as well as their symptoms are generally well documented. In health care, the analytical data mining algorithms have great impact on building the machine learning models to carry out disease prediction. As removal of uterus causes loss of an important organ of a female human body, an earlier prediction of the risk of this surgery process may help the treatment as a strong guidance. Considering this issue using the data mining techniques such as classification, clustering, association rules and using various types of software tools like Weka software tool, a proper dataset should be trained and analyzed. In this paper our main objective is to construct an appropriate dataset for this thesis and provide an improved data miming classification based architecture to predict the risk of this diseases by analyzing the clinical data of patients.

**2. General Terms**

Data Mining and techniques.

**3. Key words**

Hysterectomy, Machin Learning Algorithm, Support Vector Machin, Naïve Bayes, Artificial Neural Network, analytical medical data.

**4. Introduction**

Hysterectomy is an operation to remove a woman’s uterus is also known as womb. Thereby, the woman's menstruation end and she can not be pregnant any day later. Doctors try before hysterectomy to solve this problem through any other treatment. If they are not successful and risk free, they advised for hysterectomy. There are many reasons behind the hysterectomy. One of them is the uterine fibroid

uterine fibroid (many times doctors call it leiomyomas or myomas) are muscle tumor or benign lumps which increases in a woman's womb. Very rarely, these tumors have been found to be from cancer. There are many women who did not know if they have such type of things, because sometimes no pain or substratum can be shown for uterine fibroid. Fibroid can be of different size types and in different places. However, in most cases, fibroid can be seen in women's uterus, uterine wall or on its surface. Many times the size of the tumor is small so that the doctor can not see it with the necked eye. In many cases, their size is so large that they have an effect on the uterus for size and shape. Many women have had problems due to the uterine fibroid. There were also several types of symptoms. As many people have been suffering from heavy bleeding and have been suffering a lot. There were several times in a month. There was a lot of pain in the abdomen. Besides, when there was no period of time, it was very painful for sex. Many times the problem of pregnancy in many patients would have become such a miscarriage.

**5. Literature review**

In this section reviews various research works that are related to the proposed work.

**Dr.S.Vijayarani**, **Mr.S.Dhayanand** have predicted about liver disease prediction using SVM and Naïve Bayes algorithms. They have made their data set within five hundred and seventy six instance and ten attributes. They implemented their works in Matlab 2013 tool. After doing their work they showed a result of an experiment that the accuracy of the performance of SVM algorithm is better than Naive Bayes algorithm.

**Pragati Agrawal and Amit Kumar Dewangan** have gave a brief survey on the techniques used for the diagnosis of diabetes –mellitus. They described in their paper about diabetes. They described in their paper about diabetes. They designed their data set on various classification algorithms like SVM, KNN, Naïve Bayes, ID3, C4.5, C5.0 and CART to classify the diabetes data. They have made a comparison of accuracy of these models and they showed that SVM gives best accuracy of classification as 81.77% compare to others.

**Ebrahim Edriuss et.al** clarify the modeling of breast cancer as classification task and describes the implementation of Neural network(NN) and Support Vector Machine approach for classifying breast cancer is either petit or malignant. Dr William H.Wolberg created a dataset to diagnosis breast cancer consist of 400 observations of patients with breast cancer among which 300 are petit band 100 are malignant status. They has 20 features. The work has done in two experiment in which first one is done by using SVM open source tool for multi class SVM, which uses Crammer and singer method. The 2nd one is done by busing Neural Network. The result of the performance and accuracy of both NN and SVM algorithm were compared. In this paper there are shown that the NN technique is more efficient compared to SVM technique in breast cancer detection.

**Dr.D.Asir Antony et.al** conducted a prediction of diabetes using medical data. In his research paper they showed a prediction of diabetes by using WEKA software with the configuration of computer system 4GB RAM ,Intel(R),Core(TM)2,CPU 1.73GHz Processor, Windows 7,64-bit operating system. For preparing this research paper ,the data have been collected from University of California, Irvine and in this paper different machine learning algorithms( Naïve Bayes(NB),Multilayer Perceptron(MLP),decision tree-based random forests(RF) )are used to build the model and those model is tested with different testing methods such as FCV(10-fold cross Validation), PS(Percentage Split) and UTD(Use Training Dataset) to evaluate the accuracy of the model. It is observed that the technique with pre-processing(WPP) increase the accuracy of the machine learning algorithm more than the technique without pre-processing (WOPP)

**Joseph A. Cruz et.al** presented a paper about the applications of Machine learning in cancer prediction and prognosis. In their paper they showed that different types of machine learning algorithm can be used to predict different types of cancer and they showed a histogram to show how many papers are published during years nearly 1994-2005 to predict cancer risk, recurrence and outcome. They have attempted to explain, compare and assess the performance of machine learning algorithms in this field and they also clarify that machine learning methods generally improve the performance or predictive accuracy of most prognosis, especially when compared to conventional statistical or expert-based systems.

**Mike E Janicek, MD et.al** discussed about Cervical Cancer to prevent, diagnosis and therapeutics. They discussed what the Cervical cancer is, what the risk factors will be occur because of Cervical cancer. They described about HPV, types of HPV, role of HPV and its types, Us Epidemiology, Liquid-Based Sampling Techniques and much more. They told in their paper that surgical, radio therapeutic, and chemo radio therapy approaches comprise the successful treatment modalities for invasive cervical carcinoma.

**6. Overview of Hysterectomy**

**6.1 Hysterectomy?**

Hysterectomy is the surgical removal of the uterus. It may also involve removal of the cervix, ovaries, fallopian tubes and other surrounding structures.

**6.2 Why it is occur?**

Hysterectomies are most often done for the following reasons:

* Abnormal uterine bleeding that is not controlled by other treatment methods.
* Uterine fibroids — common, benign (noncancerous) tumors that grow in the muscle of the uterus. More hysterectomies are done because of fibroids than any other problem of the uterus. Sometimes fibroids cause heavy bleeding or pain.[5]
* Uterine prolapse — a benign condition in which the uterus moves from its usual place down into the vagina. Uterine prolapse is due to weak and stretched pelvic ligaments and tissues, and can lead to urinary problems, pelvic pressure or difficulty with bowl movements. Childbirth, obesity and loss of estrogen after menopause may contribute to this problem.[5]
* Cervical or uterine cancer
* Complications during childbirth

**6.3 Type of Hysterectomy**

There are four types of hysterectomies in which can be performed depending on the condition and severity of the patient. These four types are

* Vaginal hysterectomy
* Laparoscopic hysterectomy
* Minilaparotomy hysterectomy
* Abdominal hysterectomy

**6.3.1 Vaginal Hysterectomy**

Vaginal hysterectomy is a surgical procedure to remove the uterus through the vagina.

During a vaginal hysterectomy, the surgeon detaches the uterus from the ovaries, fallopian tubes and upper vagina, as well as from the blood vessels and connective tissue that support it. The uterus is then removed through the vagina.[6]

**6.3.2 Laparoscopic Hysterectomy**

A laparoscopic hysterectomy is a minimally invasive surgical procedure to remove the uterus. A small incision is made in the belly button and a tiny camera is inserted. The surgeon watches the image from this camera on a TV screen and performs the operative procedure. Two or three other tiny incisions are made in the lower abdomen. Specialized instruments are inserted and used for the removal process.[7]

**6.3.3 Minilaparotomy Hysterectomy**

In some cases neither a vaginal hysterectomy nor a microscopic hysterectomy may be possible. In these cases, a minilaparotomy hysterectomy may be required.

Essentially, this type of procedure uses a 5 cm incision point to operate on the uterus. This type of procedure is usually quite quick for the surgeon to perform, as they have a lot more room in which to do the operation. The cost of the procedure may also be less than other types of hysterectomies, so each patient must determine the cost vs. benefit ratio (however, the cost may not be less as it the cost must be reviewed on a per patient basis).

A Minilaparotomy hysterectomy will take more time to recover an a vaginal hysterectomy or a laparoscopic hysterectomy. In addition, there’s an increased chance of scarring in the abdominal area. However, the incision is made below the “bikini line”.[3]

**6.3.4 Abdominal Hysterectomy**

An abdominal hysterectomy is a surgical procedure that removes your uterus through an incision in your lower abdomen. Your uterus or womb is where a baby grows if you're pregnant. A partial hysterectomy removes just the uterus, leaving the cervix intact. A total hysterectomy removes the uterus and the cervix.[6]

**6.4 Risk factors for Hysterectomy**

As with any surgical procedure, complications may occur. Some possible complications include, but are not limited to, the following:

* Hemorrhage
* Injury to the ureters (tubes that carry urine from the kidneys to the bladder) and urinary bladder
* Wound complication (i.e. abscess, disruption)
* Pain with intercourse, decreased libido, or reduced ability to have an orgasm
* Infection
* Injury to the bowel or other intestinal organs
* Difficulty with urination or urinary incontinence
* Blood clot in the veins or lungs

**7. Overview of Methodologies**

**7.1 Data Mining**

The process of digging through data to discover hidden connections and predict future trends has a long history. Sometimes referred to as "knowledge discovery in databases," the term "data mining" wasn’t coined until the 1990s. But its foundation comprises three intertwined scientific disciplines: statistics (the numeric study of data relationships), artificial intelligence (human-like intelligence displayed by software and/or machines) and machine learning (algorithms that can learn from data to make predictions). What was old is new again, as data mining technology keeps evolving to keep pace with the limitless potential of big data and affordable computing power.

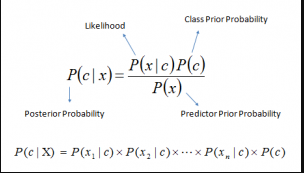
Over the last decade, advances in processing power and speed have enabled us to move beyond manual, tedious and time-consuming practices to quick, easy and automated data analysis. The more complex the data sets collected, the more potential there is to uncover relevant insights. Retailers, banks, manufacturers, telecommunications providers and insurers, among others, are using data mining to discover relationships among everything from pricing, promotions and demographics to how the economy, risk, competition and social media are affecting their business models, revenues, operations and customer relationships.[m3]

**7.2 Naïve Bayes**

The Naïve Bayes algorithm is a probabilistic that is sequential in nature, following steps of execution, classification, estimation and prediction. The Naive Bayesian classifier is based on Bayes’ theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. [m3]

**Algorithm**

Bayes theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence



* P(c|x) is the posterior probability of class (target) given predictor (attribute).
* P(c) is the prior probability of class.
* P(x|c) is the likelihood which is the probability of predictor given class.
* P(x) is the prior probability of predictor

**7.3 Support Vector Machine**

Support vector machine was invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963. [m6] In 1992 Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes. [m6] SVM is a new approach to supervised pattern classification which has been successfully applied to a wide range of pattern recognition problems and it is also a training algorithm for learning classification and regression rule from data. SVM is most suitable for working accurately and efficiently high dimensionality feature spaces in addition to that SVM is based on strong mathematical foundations and result in simple way and very powerful algorithms.[p]

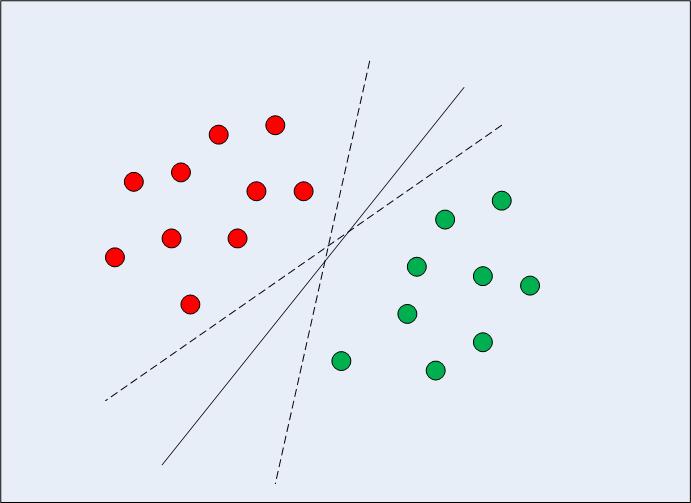


Figure: SVM

The standard SVM algorithm build a binary classifier. a simple way to build a binary classifier is to construct a hyper plane separating class member from non-member in the input space.

**Hyper plane**

The most optimal decision boundary is the one which has maximum margin from the nearest points of all the classes. The nearest points from the decision boundary that maximize the distance between the decision boundary and the points are called support vectors as seen in Figure. The decision boundary in case of support vector machines is called the maximum margin classifier, or the maximum margin hyper plane.

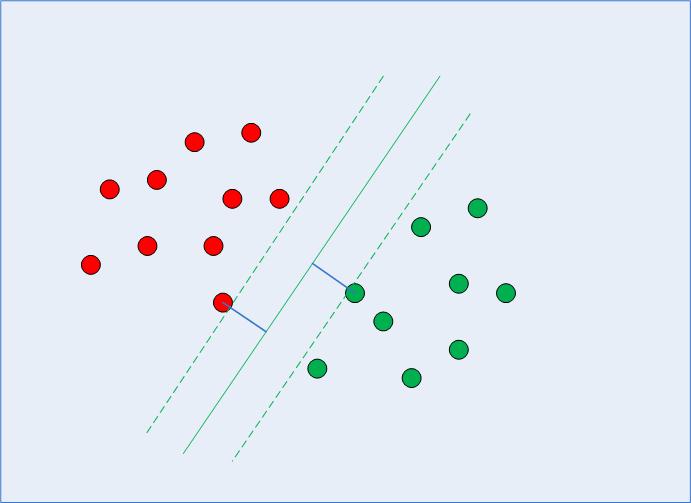


Figure: Hyper Plane