Machine Learning based Comparative Analysis for Menstrual Abnormalities

Yamuna Devi S[1], Rathika A[2],Vivek M[3]

[1]PG scholar, Department of Computer science & Engineering, Jansons Institute of Technology
[2] Assistant Professor, Department of Computer Science& Engineering, Jansons Institute of Technology

---***---

Abstract - *In today's hasty world, the technology improves where the disease increases. Improvements in technology brought changes in the daily life style behaviors which ends in disturbing the activities of the normal biological clock of human being. Particularly, adolescent females are mostly affected by various health issues regarding their reproductive system. Various factors such as excess stress, sedentary behaviors and improper food habits affects the normal menstrual cycle of the females and imbalances the hormone levels. Due to this they suffer with menstrual disorders like amenorrhea and oligo menorrhea. This paper focuses on factors causing menstrual abnormalities and comparative analysis is made using various machine learning methods to predict the abnormality for the collected data set.*

Key Words: Menstrual abnormalities, Oligomenorrhea, Amenorrhea, Machine learning,Supervised Learning.

1. INTRODUCTION

Menstrual irregularity or menstrual abnormality is defined as having abnormal pattern in the menstrual cycles.Where, Infertility is the condition defined as inability to give childbirth. How ever, these two are well connected implicitly.

Due to today's life style, less attention is paid to this menstrual problems in adolescent period which causes trouble in later. Adolescence is the transitional phase where maturation starts physically as well as mentally. It is also referred to as development from childhood to adulthood. Puberty starts at this age where physical, sexual, cognitive and emotional changes occur gradually. Especially a female child goes through various levels of changes.

Brain releases the hormone called gonadotropin-releasing hormone(GnRH). This GnRH also releases other two puberty hormones named luteinizing hormone (LH) and follicle-stimulating hormone(FSH) which inturn initiates the process called Menstruation. Menstruation refers to a regular cyclical discharge of blood and mucosal tissue from the inner lining of the uterus through the vagina which is considered to be a readjustment of uterus.

Every woman faces any one type of menstrual disorder at her life time. But the rate of prevalence of menstrual disorders and abnormalities is increasing day by day among girls. Particularly, Adolescent age girls and early adulthoods are at a high risk of menstrual abnormalities which could lead to infertility during their reproductive age.Menstrual abnormalities shold not be taken as a common one, which may be the reason of other diseases or lead to some chronic diseases like CAD. Amenorrhea(Absence of menstrual cycles), Oligomenorrhoea(Irregular menstrual cycles), Dysmenorrhea(Menstrual pain) , Premenstrual symptoms such as Breast tenderness, nausea, Stress and Irritation are some of the menstrual disorders. This natural phenomenon is affected by various factors such as life style behaviour, Eating disorders, Obesity, Chronic diseases such as Diabetes Mellitus, Tuberclosis, CAD, Sedentry behaviours, Lack of physical activities, Body parameters, Excess stress and so on.

Eating disorder is all about improper food habit which affects the normal functioning of the body system as well as hormonal functions of the body which ends in Hormonal imbalance. Having low nutrient foods such as junk foods and other packed items leads to obesity which causes to increase of BMI. Sedentry behaviour also ends with obsetity. Menstrual disorders are mainly associated with food habits, Physical activities and life style behaviours.

Menstrual abnormalities can be prevented and treated when they are detected earlier. Various studies are conducted all over the world to find the menstrual irregularities caused by various health conditions. This study concentrates on the various factors causing menstrual irregularities and tries to find the effect of various Chronic illness to the menstruation.

In this paper various supervised machine learning algorithms are used to predict the menstrual abnormality.The algorithms used are Logistic Rregression, K-Nearest Neighbours(KNN), Support Vector Machines(SVM) and Naïve Bayes. Accuracy, precison, recall for the algorithms are found and compared.

2. LITERATURE SURVEY

Amouda Venketesan et al[1] proposed a data mining approach to extract all the major factors that causes menstrual disorders in students. They made a prediction on association with academic stress and lifestyle factors of Pondicherry students. Through the questionnaire, they have collected menstrual and lifestyle data of students with a sample size of 483 whose age falls between 18 to 25. Random sampling technique is used by the authors for selecting the sample population.They had collected the data and analysed it using SPSS software. They performed Chi-

Square test to find association between categorical variables. They used t-test to compare the mean of quantitative variables. Multivariate Analysis of Variance(MANOVA) reveals that poor diet has a negative impact on the menstrual cycle. Logistic Regression is also to find the other associated risk factors that affects menstrual regularity. Finally, they conclude that students who are having a balanced diet and good food habits are with the normal menstruation cycles.

N Vijayalakshmi et al[2] made a statistical analysis on the data of infertile females through the clinical database. They created a structured questionnaire and collected samples from 575 patients who were visiting endometrial centre in Trichy for monthly or weekly checkups. Out of 575 samples, 154 are of fertile and 421 are of infertile women. Data regarding environmental, physical and hormonal factors are taken to the analysis they used Weka tool for mining and work with data and analysing the most significant factors. They used K-means clustering for grouping objects that are of the same class. J48 pruned tree technique and random tree techniques were used for classification. Association rule mining is done for the data set collected. Various hormone levels such as TSH, FS, LH values are considered for evaluation of infertility. They found that BMI greater than 29 plays vital role in infertility. Their study revealed that patients with the history of diabetes mellitus and tuberculosis are really in danger. Accuracy prediction of their study is 86%.

Dr Mayur et al[3], presented a cross-sectional design study of structured questionnaire format, with the sample size of 384. This study is conducted among the females of west zone of Rajkot city, Gujarat. In this study, they have correlated factors like BMI, physical activity, junk food consumption to the menstrual activities. They found an association between many variables such as physical activities and menstrual irregularities. Chi-Square test is used for analysis by them. According to their study, premenstrual symptoms like dysmenorrhoea and mood swings are the causes of absence. 11.5% of the sample population were reported with the irregular menstrual cycles. They finally conclude that physical excercise has a significant association with the menstrual abnormalities.

Janula Raju et al[4] presented a descriptive research design to calculate the effect of menstrual symptoms on the academic performance among nursing students of 200 samples. Their research study revealed that 67% of the sample population don't have interest to go to college and 71% of are having lack of concentration during menstrual periods. 58% of students answered that they were not able to remember the contents and 76% of are not interested in attending the practical classes during menses. They have concluded that menstrual symptoms has a negative impact on the academic performance of the students.

Hanan Elzeblawy Hassan et al[5] presented a says an approach for menstrual disorders necessitating counselling among students. They created a questionnaire and distributed to the students of various faculties of Beni-Suef University. They have collected responses from 1,519 students. They found that dysmenorrhoea has high rate (94.1%) when compared to other menstrual symptoms. 48.8% faced irritation, 40.9% faced fatigue and acne flare is faced by 49.2% of sample population during menses.

Malay Kumar Patsa et al[6] explains a relationship between menstrual disorders and nutritional status. They take survey to the college girls are of age 18 to 22 in the Bankura district, West Bengal. They formed a semi-structured questionnaire which consists of socio-economic data and menstrual pattern related data. They have collected anthropometric details such as BMI(Body Mass Index), BAI(Body Adiposity Index), BPF(Body Fat Percentage) and MTC(Mid-Thigh Circumference) are measured and calculated for the selected students. They used one way- ANOVA to analyse the factors. From this study, they found that age at menarche and socioeconomic status are inversely related. And also, they found that menarche age is inversely related to some parameters such as BMI, BAI, BPF, MTC. In their study, they found that PMS, Dysmenorrhoea and leucorrhoea has significant impact such as 67.01%, 73.20% and 85.57% respectively. They got a U-shaped relationship between BMI and Dysmenorrhoea, which proves that body fat influence the dysmenorrhoea.

Geetha et al[7] made a survey on 752 rural women in Chittoor district, Andhra Pradesh aged 20 to 40 with a random sampling technique. Through their questionnaire, they have analysed data of various factors such as socioeconomic factors, lifestyle and menstrual characteristics. Their study revealed that 32% of the sample population are having menstrual problems. Also they found that oligomenorrhea and hypermenorrhea are having more impact such as 12% and 9% respectively.

Ahmed M Nooh[8] presents a method on prevalence of menstrual abnormalities and disorders among female students of the zagazig University, Egypt. They made purpose-based questionnaire which focuses on demographic data and menstrual data. They had analyzed the various factors from the questionnaire of 283 selected samples. They analysed the data using SPSS software version 20. F-test and chi-square test is used for analysis of mean, standard deviation and variance(ANOVA). From their study, they revealed that 11.3% of the total population comes under one kind of menstrual disorders. Oligomenorrhoea was reported by 6% of the population. Relationship between various biological factors and dysmenorrhoea was calculated and analysed. This study also revealed that Increase of BMI has a significant association with Pre Menstrual Syndrome.

Rose Mary Nakame et al[9] presented a study that access the major factors associated with dysmenorrhoea among female students. They conducted a study among females aged with 18-45 as an inclusion criteria in selected universities of Uganda. Purposive sampling technique was used to select the participants. They have collected data using a self-

administered questionnaire with a sample population of 351. Epi-info version 7 is used to enter the data. The main factors considered to analyse are sugary food consumption, having children and family history of having dysmenorrhoea. SPSS software was used for analysis with a confidence interval of 95%. Chi-square test and Analyse of Variance(ANOVA) is used to analyse the factors. Bivariate analysis and Multivariate analysis also used for analysis. Their study revealed that among the sample population 75.8% experienced dysmenorrhoea at least once with a standard deviation of 0.793.

Awoke Giletew et al[10] proposed a study that concentrates on primary dysmenorrhoea and factors that has an effect on academic performance. They conducted an Institution based cross-sectional study with a sample population of 183. Descriptive analytics and Logistic Regression were used for analysis. They found that various factors such as family history of dysmenorrhoea, menarche at an early age and menstrual irregularities has a significant impact on the primary dysmenorrhoea. Their study revealed that the prevalence of dysmenorrhoea is of 62.3%.

Soodabeh Darvish et al[11] proposed a study that concentrates on the female reproductive factors that causes Coronary Artery Disease (CAD). 80 premenopausal and 100 postmenopausal females were attended the survey. Various factors such as menstrual characteristics, demographic data, CAD risk factors were included in the questionnaire. Their study revealed that Oligomenorrhoea has a significant on the Cardio Vascular Diseases(CAD) which has association with insulin resistance and diabetes.

Husnah [12]presented a study that explains the importance of intake of breakfast. He surveyed the students of SMAN 2, Banda Aceh, Indonesia with a sample population of 132. He selected the students based on the nonprobability sampling with stratified random sampling method. His inclusion criteria was about students with the history of dysmenorrhoea from the menarche age and exclusion criteria was students who are not suffering from any kind of chronic illness such as diabetes, Hypertension. Confidence interval was set to be 95%. Visual analogue questionnaire scale (VAS) with pain ruler 0–10 cm was used to measure the dysmenorrhoea. Data related to breakfast was obtained through Food Frequency Questionnaire (FFQ). His study revealed that degree of dysmenorrhoea has a relationship with breakfast patterns.

Siddhesh R. Rajiwade et al[13] compares the various endocrine factors susch as PCOS, Thyroid dysfunction, hyperprolactinemia and hyperandrogenism with the menstrual abnormalities. Adolescent females aged 10-19 with menstrual disorders such as primary amenorrhea, secondary amenorrhea, polymenorrhea, hypomenorrhea, oligomenorrhea, menorrhagia, metropathia and irregular bleeding are participated in this study. Venous blood of 5ml was taken from the sample population. Thyroid function test, total amount of testosterone, serum prolactin were analyzed

using the blood samples by the chemiluminescence system. Their study revealed that Oligomenorrhoea(61.0%) is most affecting factor in adolescents. 13.6% were reported with thyroid dysfunction. Biochemical hyperandrogenism was reported by 9.04% of the sample population. There is no endocrine abnormalities detected in the case of polymenorrhea, hypomenorrhea and intermenstrual bleeding. They finally conclude that immaturity of hypothalamic pituitary ovarian axis is the cause for menstrual irregularities.

Naz et al[14] analyses the hormonal disorders of the female reproductive system that leads to the dysfunction of the hypothalamic-pituitary-ovarian axis. They designed a pattern to analyze the endocrinological disorders that leads to infertility. They conducted a retrospective study at National Health Research Complex (NHRC) and department of Obstetrics and Gynecology, Shaikh Zayed Hospital, Lahore. Their inclusion criteria is all about the females with infertility, oligomenorrhoea, Amenorrhea disorders with the age in range of 14-40 . They have extracted the information from the patient's medical history files and laboratory records. Through their study,they found that Luteinizing Hormone(LH) and Follicle Stimulating Hormone(FSH) levels were elevated in the hormonal imbalance subgroup of infertility, oligomenorrhea and amenorrhea.

Louise A. Brinton, Ph.D., et al[15] made a retrospective cohort study to define the relationship between androgen excesses and cancer risk. They used five large infertility practices. They analyzed women evaluated for infertility during 1965-1988 , from that 2,560 had androgen excess or menstrual disorders. 412 met the established criteria for the polycystic ovarian syndrome. Through their method, they found that androgen excess or menstrual disorder patients showed significant SIRs for breast cancer(1.31; 95% CI, 1.05– 1.62) , uterine cancer (2.02; 95% CI, 1.13–3.34) and melanoma(1.96; 95% CI, 1.12–3.18). They finally concluded that, Previous findings linking androgen excess disorders to elevated uterine cancer risks might largely reflect underlying risk profiles.

3. DATA DESCRIPTION AND PRE PROCESSING:

3.1 Data collection

A cross-sectional study has been conducted through Google forms from the adolescent girls and early adulthoods whose age between 10 to 25. Before that, a pilot study has been conducted with a sample size of 10 to know about how the collected data has an impact on the other and to get optimized results. Sample questions are distributed to 10 students initially to check for removing the inefficiencies and complexities of the questionnaire. Female students who are studying in school and colleges and those who are working are participated based on their willingness is all about the inclusion criteria. Exclusion criteria is students with chronic illness such as diabetes mellitus, tuberculosis, are avoided. Ethical approval was obtained from my college. The

systematic questionnaire was framed with the relevant questions regarding to the demographic details, menstrual patterns, dietary habits, physical activities, mental abnormalities.

Information regarding anthropometric parameters such as Height,Weight, Body Mass Index, Waist circumference are taken. Information regarding food habits such as Proper intake of breakfast, intake of iron rich food, citrus fruits, interval of junk foods in taken, intake of fruits, vegetables and greens , Intake of tea and coffee are also collected from the participants. Mental health and abnormalities related questions were taken from the STAI and MDQ questionnaire.

3.2 Data pre-processing

Collected data from the Google form responses are entered to the Microsoft Excel(Ms-Excel) which will be saved as CSV file. Rows containing duplicate values and missing values are removed.From sklearn library preprocessing package is imported for this operations. Preprocessed data is saved as separate file for further operations.

4. ALGORITHMS:

4.1 Logistic Regression

Logistic regression is all about a classification algorithm which is also a predictive analysis algorithm that is based on the concept called 'probability'. Predicting whether the email is spam or not is the example for the classification problem. Logistics regression is used for assigning observations to a discrete set of classes.

Logistic Regression uses a function called 'complex cost function' which also known as 'Sigmoid function'. The hypothesis of the logistic regression tends it to limit the cost function lies in between 0 and 1. It transforms its output with the help of logistic sigmoid function for returning a probability value.

4.2 K-Nearest Neighbours

It is also a supervised machine learning algorithm, where it is used for both classification and regression problems. Value of K will be initialized at first for number of nearest neighbours. For each data, distance will be calculated from the group where it will be added with similar data. After that,based on the distance value data will be sorted. The algorithm then chooses the top k rows from the sorted array.

4.3 Support Vector Machines

SVM is a supervised machine learning algorithm in which hyperplane is created that splits the data set in to different classes. The data points which lie closer to the hyperplane are known as 'support vectors'. Distance between the nearest data point and the hyper plane is called 'Margin'. The function used for training the model is fit(). Support Vector Classifier is used to implement the algorithm. Term named

'SVM Kernal' represents a function that converts not separable problem in to separable one. It doesn't work well for larger data set.

4.4 Naïve Bayes

Naïve Bayes is one of the fast classification algorithm used for large amount of data.It is based on Bayes theorem for predicting unknown class. Prior probability for class levels is calculated.Likelihood probability for each attribute will be found with other each attribute. These values are applied to the following formula to calculate the posterior probability. Find the class with high probability.

$$P(h|D)=P(D|h)P(h) \, / \, P(D)$$

5. RESULTS AND DISCUSSION

The collected data set is is processed and predictive model is made and analysed using supervised learning algorithms. Prediction results reveals that Excess BMI levels, less physical activities are connected with menstrual irregularity. Using four various algorithms the data set is processed. Misclassified samples are removed and the model is rebuilt again.

This predictive model generated using logistic regression gives the accuracy of 88%. And another predictive model that is generated using KNN gives the accuracy of 80%. Accuracy score of Support Vector Machine is 79%. Another algorithm named 'Naïve Bayes'gives accuracy of 84%.

Table 1 : Results

Method	Accuracy in %
Logistic Regression	88.46
KNN	80.76
SVM	79.48
Naïve Bayes	84.61

6. CONCLUSION

This study is made to bring awareness among young adolescence females about the menstrual abnormalities and the impact of lifestyle behaviors, sedentary behaviors and eating habits to the reproductive health which causes infertility. Predictive model is compared using four different supervised machine learning algorithms. For further levels, clinical dataset of the adolescent females can be used to get accurate prediction.

REFERENCES

1) Amouda Venketesan, Mathavan Muthaiyan and Lavanya Ravichandran(2018),'Data Mining to Elicit the Critical Factors causing Menstrual Disorders in Students of Pondicherry University', International Journal of Statistics and Reliability Engineering (IJSRE) Vol. 5(2), pp. 96-100, 2018 (ISSN(P): 2350-0174 ; ISSN(O):2456-2378)

2) Ms. N. Vijayalakshmi, Ms. M. UmaMaheswari(2016) ,'Data Mining To Elicit Predominant Factors Causing Infertility In Women' International Journal of Computer Science and Mobile Computing(IJCSMC), Vol:5 Issue:8.

3) Dr. Mayur V. Savsani1, Riddhi R. Solanki2 A Study of Menstrual Disorders in Reproductive Age Group and Its Correlation with Biological Variables 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X

4) Janula Raju and Suguna M A study to assess the effect of menstrual symptoms on academic performance among nursing students at selected colleges in Tamil Nadu, India., ISSN Print: 2394-7500 ISSN Online: 2394-5869 Impact Factor: 5.2 IJAR 2017; 3(3): 78-80

5) Hanan Elzeblawy Hassan∗1, Wafaa Mostafa Ahmed Gamel2, Eman Ali Abd El Moaty Sheha3, Mervat Amin Sayed3, Ahmed Emad Eldin Arafa4 Menstrual disorders necessitating counselling among students in Beni-Suef University, Clinical Nursing Studies 2019, Vol. 7, No. 2.

6) Malay Kumar Patsa1,2, Dibya Pal1, Srimanta Sen1, Mousumi Ganguly1, Nirmalya Kumar Sinha2,3,4, Dulal Chandra Das5, Sumendev Chakrabortty6, Monalisa Das7,8 Relationship of menstrual disorders with nutritional status of college girls from Bankura District, West Bengal, India. International Journal of Bioassays 5.4 (2016): 4515-4523., ISSN: 2278-778X

7) Geetha P1, Chenchuprasad C2, Sathyavathi RB2, Bharathi T3, Surendranadha Reddy K1 and Kodanda Reddy K, "Effect of Socioeconomic Conditions and Lifestyles on Menstrual Characteristics among Rural Women., Journal of Women's Health Care 2016, 5:1

8) Ahmed M Nooh, Menstrual disorders among Zagazig University Students, Zagazig, Egypt.

9) Rose Mary Nakame, Frank Kiwanuka, Afayo Robert," Dysmenorrhoea among students aged 18–45 years attending University in Uganda: A cross-sectional multicenter study of three Universities in Uganda"(2018)

10) Awoke Giletew, Workaferahu Bekele, "Prevalence and Associated Factors of Primary Dysmenorrhea Among Debre Tabor University Students, North Central Ethiopia", IJBECS, 2018; 4(4): 70-74, ISSN: 2472-1298 (Print); ISSN: 2472-1301 (Online).

11) Soodabeh Darvish 1*, Saeed Sadeghian 2, Azizeh Ghaseminejad 3, Reza Mohebi, "The Effect of Reproductive Factors on Coronary Artery Disease in Women", Summer 2019, Volume 28, Issue 2 (9-13).

12) Husnah, "Association between Breakfast and Dysmenorrhea Degrees of SMAN 2 Students in Banda Aceh, Indonesia"(2018), World Nutrition Journal, eISSN 2580-7013.

13) Siddhesh R. Rajiwade1 • Haritha Sagili2 • R. Soundravally3 • L. Subitha4, "Endocrine Abnormalities in Adolescents with Menstrual Disorders"(2017), The Journal of Obstetrics and Gynecology of India.

14) Naz S, Ghafoor F And Mukhtar S.3, "Reproductive Hormone Profiles Of Women With Infertility And Menstrual Disorders: A Retrospective Study" , D:\Biomedica Vol. 34, Issue 3, Jul. – Sep., 2018\Bio-2.Doc.