**American Journal of Data Mining and Knowledge Discovery**
2023; 8(1): 1-10
http://www.sciencepublishinggroup.com/j/ajdmkd
doi: 10.11648/j.ajdmkd.20230801.11
ISSN: 2578-7810 (Print); ISSN: 2578-7837 (Online)

# Medical Factors-Based Mobile Application for Choosing Contraceptives: A Data Mining Approach

**Bala Pwa'anda Bulus[1], Yusuf Musa Malgwi[2]**

[1]Department of Computer Science, Federal College of Education, Yola, Nigeria
[2]Department of Computer Science, Moddibbo Adama University, Yola, Nigeria

## Email address:
Bpbala@fceyola.edu.ng (Bala Pwa'anda Bulus), Yumalgwi@mautech.edu.ng (Yusuf Musa Malgwi)

## To cite this article:
Bala Pwa'anda Bulus, Yusuf Musa Malgwi. Medical Factors-Based Mobile Application for Choosing Contraceptives: A Data Mining Approach. *American Journal of Data Mining and Knowledge Discovery*. Vol. 8, No. 1, 2023, pp. 1-10. doi: 10.11648/j.ajdmkd.20230801.11

**Received**: April 1, 2023; **Accepted**: April 28, 2023; **Published**: July 24, 2023

**Abstract:** Women's negative experience with contraception and understanding of the experience differentials, coupled with Limited information accessibility on contraceptives to healthcare centers, is responsible for the unwillingness and discontinuation in the use of contraceptives in Nigeria. The study aims at developing a Medical Factor based Mobile application Model for contraceptive implants using K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) techniques for the prediction of discomfort, and blood type. KNN and SVM techniques in R were the two methods used to develop a model to classify blood group types based on discomfort and vice versa. 10-fold cross-validation was carried out and was repeated 3 times and the optimal values were selected. The model was tested by the use of Predict function in which test data was used as the new data (input data) of the model. Experimental results showed the prediction accuracy of the KNN model was 85.72% and the SVM model was 92.2%. SVM outperformed KNN. However, the performances of models imply that the application can be used by women as the means for accessing information on discomforts associated with contraceptive implants as well as blood type. Most women with similar blood types have similar experiences (discomforts). Therefore this model can be used to choose the right contraceptive that is friendly to one blood type. The prediction mobile application of the tested model frontend was implemented in Android built-in with Java as the programming language. The backend was designed using structural query language in the WAMP server.

**Keywords:** Contraceptives, Data Mining, k-NN Algorithm, Support Vector Machine

## 1. Introduction

Contraceptive usage is the simple way of birth control in this contemporary time. It is a way of preventing unwanted pregnancy [2], and a medium for child spacing for those who are not ready to stop childbearing completely. Contraception enables women to exercise their human right to choose the number and spacing of their children [13] and achieve birth control. Birth control can be achieved using either natural methods or contraception methods. In an attempt to do family planning, most couples before now, and because of cultural and religious beliefs, opt for natural ways of contraception, which involves withdrawal or woman observing and documenting fertility and infertility days.

Methods of Natural Family Planning (NFP) seem to be acceptable by some couples, and interestingly because of their high effectiveness, in that it does not lead to any complication with women's menstrual cycle compared to hormone contraceptive methods. However, as effective as they are, men find it difficult to withdraw sometimes. This act sometimes leads to an unwanted pregnancy. Women, on the other hand, don't observe and document their fertility and infertility days properly. Lack of idea on fertility and infertility days contributes to complaints of inconveniences of the NFP methods. The necessity of daily observations and documentation of fertility symptoms, and adequate education to determine fertility and infertility days are some complaints on inconveniences of NFP methods by some women [30]. Understanding fertility and infertility days is key in NFP. Lack of this knowledge and good education on this has over the years been a problem to people that preferred NFP methods. However, governments in both developed and developing countries are not relenting in their quest for the best family planning methods that will be accepted and used by all with little or no side effects at all. The quest led to the introduction of what is termed modern or contemporary methods of family planning.

The contraception methods are hormonal-based. There are various kinds; ranging from the use of Oral Pill, injection, diaphragm, implant, intrauterine, etc. Recently, the United States (US) Food and Drug Administration approved a new oral contraceptive known as Nextstellis [7, 17]. The Use of modern-day contraception requires the expertise of medical professionals; doctors and midwives in counseling and administering most of the contraceptives. The methods seem to be promising, and a surest way of preventing unwanted pregnancy.

Some of the advantages of these methods are that men do not need to worry about withdrawal during sexual intercourse, and women do not as well need to worry about observing and documenting fertility and infertility days once the right thing is done. Consequently, women are beginning to accept these methods for child spacing and birth control, but as it were with natural family planning, the problem of inconveniences associated with these methods is one of the major factors mitigating its full adoption.

The inconveniences include dizziness, headache, loss of appetite, too much sleep, weight loss, weight gain, increase appetite, heavy flow during the cycle, menstruation lasting more than usual, or fear of not being able to give birth again, etc. [18]. These inconveniences often lead to the removal or stoppage of contraception by some women earlier than planned period. The experience differs from woman to woman.

Hormonal and blood group differences may contribute to different reactions to different contraceptive methods. Giving attention to knowing certain conditions like blood group, BP, heart health, etc of women when administering contraceptive methods of any kind, may help greatly in reducing the inconveniences associated with contraceptives. Understanding the conditions will help in knowing how each person reacts to each method of contraceptives, and appropriate contraceptives will be given to clients without guesses. Therefore, the role of information accessibility and dissemination cannot be overemphasized. According to IGI Global, information dissemination is to distribute or broadcast information. Active distribution and spreading of information of all kinds to users that deserve it. Information dissemination is key to the success of family planning but the channel of dissemination is not enough. As it stands today, the main channels of disseminating contraceptive information are healthcare service providers; gynecologists, midwives, and nurses. The services are provided usually during antenatal care. This is to say that information is limited to married and single parents. Another channel is electronic media.

The study's objective is to develop a model for a mobile-based informatics system for contraceptives for women. Collect relevant data on contraceptives, and perform normalization on the data to form Patient Discomfort Symptoms–Blood Group matching and Patient–hospital/clinic location matching models for data mining. Apply K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) techniques in R. in the prediction of contraceptives, and blood type. Evaluate the performance of the models using the K fold cross-validation, to determine the best technique and design and develop a knowledge base for patients for ease of access to contraceptive information on mobile apps.

## 2. Overview of Data Mining

Data mining is a logical process that is used to search through large amounts of data to find useful data [22]. Reddy, et al. define data mining as "the process of evaluating the databases to extract new insights from them" [24]. It is a tool by which knowledgeable and hidden data can be obtained to access the hidden patterns of the data set [27]. Data mining is becoming more popular in healthcare nowadays. It offers great potential to the healthcare industry for enabling health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. It turns a large collection of data into knowledge [11]. According to Baker, "Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data" [6]. Data mining usage is numerous and is applied in many different ways. For example, Hessami, et al. used data mining techniques to provide enhanced cost estimating and project development procedures for Metropolitan Organizations [14]. They came up with a decent model at the end of their project. Data mining has been used in numerous fields, including retail sales, bioinformatics, and counter-terrorism.

Data mining, a subfield of artificial intelligence that makes use of vast amounts of data to allow significant information to be extracted through previously unknown patterns, has been progressively applied in healthcare to assist clinical diagnoses and disease predictions [16]. This information has been known to be rather complex and difficult to analyze. Furthermore, data mining concepts can also perform the analysis and classification of colossal bulks of information, grouping variables with similar behaviors, and foreseeing future events, amid other advantages for monitoring and managing health systems ceaselessly seeking to look after the patients' privacy [21]. The knowledge resulting from the application of the aforesaid methods may potentially improve resource management and patient care systems, and assist in infection control and risk stratification [25]. Several studies in healthcare have explored data mining techniques to predict incidence [5] and characteristics of patients in pandemic scenarios, identification of depressive symptoms [31], prediction of diabetes, and cancer [8]. Scenarios in emergency departments [23], among others. Thus, the utilization of data mining in health organizations ameliorates the efficiency of service provision [1], and quality of decision-making, and reduces human subjectivity and errors [29].

The application of data mining in the investigation of scientific questions in medical fields has been of great interest in recent times. Medical data mining is defined as the area of scientific inquiry centered on the development of methods for making discoveries within the unique kinds of data that come from medical fields and using those methods to better understand medical issues. The goal of this technique is to find patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for new development.

Data mining is about extracting information, hidden patterns and discovering fundamental relationships from a wide range of databases [27]. Data mining is very useful to mine healthcare data. Electronic records explore the possibilities related to data mining which leads to data analysis. Ordinary databases cannot find the trends and relationships in the data. One of the data techniques that extracts and captures the relationship is clustering, also known as cluster analysis. Cluster analysis is not a specific task. Clustering is a process of discovering iterative knowledge, by way of multilayer interactive optimization that consists of failures and trials instead of automated processes. Clustering can be categorized into soft and hard clustering. The K-mean algorithm is a process of clustering used to group the data according to the criteria. The data is categorized into k groups where there is a predefined positive integer. It has three steps which include center initializing, making clusters from closer objects, and the relocation of new clusters [10].

# 3. Review of Related Work

## *3.1. Related Work on Contraception*

Unmarried adolescents are not commonly included in global monitoring of contraceptive use despite the more severe consequences of unintended childbearing for them [4]. They carried out Demographic Research on Adolescent contraceptive use and its effects on fertility. The objectives of their study were to document levels and trends of contraceptive prevalence and demand for married and sexually active unmarried adolescent women aged 15–19 in Latin America and sub-Saharan Africa. They proposed a fertility model informed by the proximate determinants framework separating adolescents by marital status. The result of the study shows that increasing contraceptive prevalence has already reduced adolescent fertility by 6.8% in Latin America and 4.1% in sub-Saharan Africa. They concluded that Contraceptive demand and prevalence are generally higher for sexually active unmarried adolescent women than for those married.

United Nations 2019 produced a data booklet that presents estimates of the prevalence of contraceptive use by a method based on the World Contraceptive Use 2019 based on data from 1,247 surveys for 195 countries or areas of the world. The estimates are presented for female and male sterilization, intrauterine devices (IUD), implants, injectables, pills, male condoms, withdrawal, rhythm, and other methods combined. The estimates of contraceptive prevalence (any, modern or traditional) for 1994 and 2019 are from Estimates and Projections of Family Planning Indicators 2019. The Booklet indicated Women who are only sometimes sexually active and who want to delay pregnancy for a few months or a couple of years, may prefer a short-acting method, one that they can start and stop on their own, over an IUD or an implant, both of which usually require a visit to a service provider to obtain and remove the device, or a permanent method such as sterilization. The experience, or awareness, of side effects and inconveniences of using specific contraceptive methods as well as their effectiveness at preventing pregnancy, play a role in the choice of the method used.

Durowade, et al. study awareness, practices, preferred methods of contraception, emergency contraceptive and Medical Termination of Pregnancy (MTP), Awareness of family planning services in the vicinity& the Decision making regarding contraceptive use [9]. The study is a community-based cross-sectional observational conducted among married women in the reproductive age group. 342 married women were interviewed in the local language using a pre-tested questionnaire. Data were analyzed using SPSS version 17.87.7% of women were aware of at least one method of contraception. 68.4% of women were using a contraceptive at the time of the study. 14% of women were unaware of any healthcare facility providing contraceptives in the vicinity. Knowledge and practice of Emergency Contraceptives were very low. Although there is a high level of awareness, contraceptive use is not very high. They said new methods of motivating people to adopt and sustain Family Planning methods should be considered.

Rahman et al. did a qualitative study using focus group discussions and in-depth interviews of women having two or more children was conducted in an urban area of Central Delhi to explore the perception and attitude of women towards family planning and barriers to using currently available contraceptives [26]. The findings reveal that the majority of the women in the current study did not favor early-age marriage and preferred smaller family sizes. However, the attitude of the husband and family was mostly considered to be unfavorable for the use of contraception and to limit the family size. Religious beliefs were the most commonly cited barrier to using contraceptives, especially surgical sterilization. Other barriers include fear of side effects of IUDs and prejudiced behavior of health care providers. These women need a contraceptive that they can use confidentially and is devoid of adverse effects. Education of women can help a lot in the long-term for improving women's reproductive health.

## *3.2. Related Work on the Application of Data Mining on Contraceptives*

Hassan et al. study investigated the performance of six different machine learning (ML) algorithms applied to predict unintended pregnancies among married women [12]. They considered six popular ML algorithms, such as logistic regression (LR), random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), naıve Bayes (NB), and elastic net regression (ENR) to predict the unintended pregnancy. They found various performance parameters for the classification of unintended pregnancy: LR accuracy 79.29%, LR AUC 72.12%; RF accuracy 77.81%, RF AUC 72.17%;

SVM accuracy 76.92%, SVM AUC 70.90%; KNN accuracy 77.22%, KNN AUC 70.27%; NB accuracy 78%, NB AUC 73.06%; and ENR accuracy 77.51%, ENR AUC 74.67%. Based on the AUC value, they concluded that the ENR algorithm provides the most accurate classification for predicting unwanted pregnancy among Bangladeshi women. On the other hand, Hag et al. identified the best model selection procedure and predicted contraceptive practice among women. To identify the best model, they applied a hierarchical logistic regression classifier in the machine learning process [13]. Seven well-known ML algorithms, such as logistic regression (LR), random forest (RF), naïve Bayes (NB), least absolute shrinkage and selection operation (LASSO), classification trees (CT), AdaBoost, and neural network (NN) were applied to predict contraceptive practice. The validity computation findings showed that the highest accuracy of 79.34% was achieved by the NN method. According to the values obtained from the ROC, NN (AUC = 86.90%) is considered the best method for this study. Moreover, NN (Cohen's kappa statistic = 0.5626) shows the most extreme discriminative ability. Similarly, Jaitley proposed a Based data mining classification approach to produce a model that can predict the duration of contraceptive use by productive couples. Through Cross-Industry Standard Process for Data Mining (CRISP-DM), it tested four experiments with seven data mining techniques [16]. The result shows that the Adaboost data mining technique produced the best performance of the contraceptive-used prediction model, with the accuracy score of the classification model as 85.1%, and a precision score of 85.1%.

Therefore, this study adopted K-Nearest Neighbour (KNN) and SVM. The motivation for choosing these algorithms is that there are mixed reports on their performance. KNN works best on small datasets and supports multiclass classification. KNN performance then compares with the SVM classifier which by default does not support multiclass classification.

In addition, none of these studies use the models in determining factors responsible for women's experiences in using contraceptives. Also, Hag et al. did not use either SVM or KNN. Hossain et al. used a support vector machine (SVM), and k-nearest neighbor (KNN) but the models' performance was poor [12] whereas Musa, et al. used the KNN algorithm and the model performs very well [19]. KNN is one of the successful data mining techniques used in classification problems [28].

## 4. Methodology

This section presents the proposed system life cycle and its architecture, analysis of the system, selected areas used for data collection, the method used in collecting data, and the design and implementation of the proposed system. The overview of the entire proposed model life cycle and its architecture is shown in Figure 1.

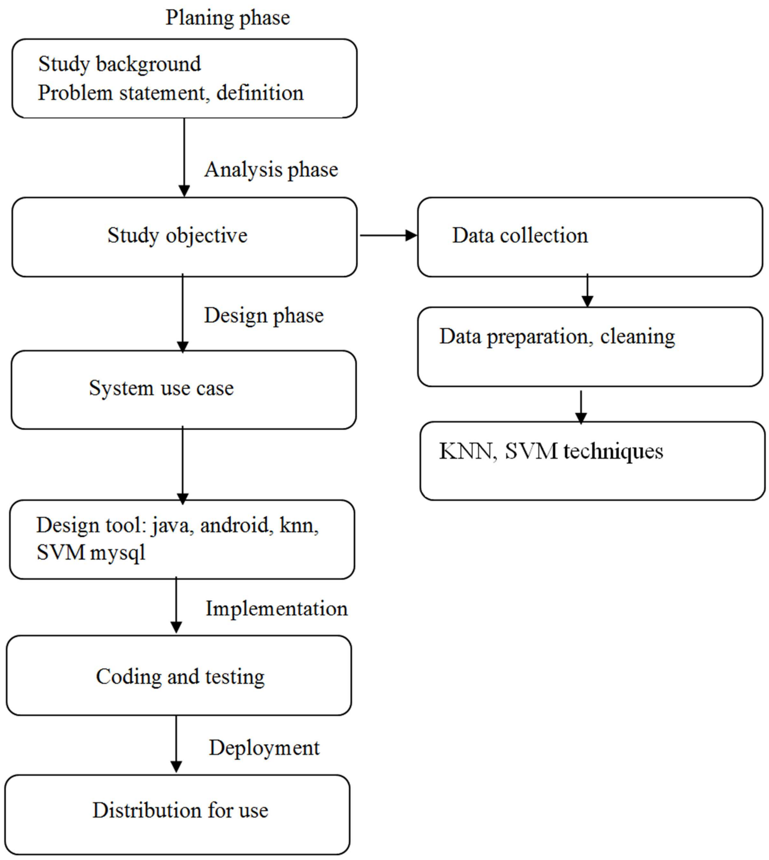


***Figure 1.*** *Proposed system model life cycle.*

### *4.1. Analysis of the System*

The existing system as analyzed, is a conventional system of administering contraceptives in a manual type, in which medical experts request for a pregnancy test to be carried on a patient to ascertain the woman is not pregnant. After testing a woman's pregnancy status, the expert will then mention the various contraceptives available and the choice of particular contraceptive decision-making that lies with the woman. However, if a patient chooses a particular type, she will be cautioned of the possible inconveniences. Finally, the woman's name, age, contraceptive type, and duration are recorded except for oral contraceptives.

The operation shows that most of the inconveniences are associated with implant contraceptives. Implantation is one of the commonly used methods among married women that visit primary health care centers in Yola North. Although the inconveniences could be the result of a hormonal reaction to contraceptives, the existing system has no records to show how a particular blood group reacts to a contraceptive hence the caution for the possible inconveniences which are not defined based on certain conditions, thus, the negative impact on people's willingness to use them.

### *4.2. New System*

The new design is a model based on a data mining technique implemented on mobile application devices to provide information on blood type and contraceptive services will help women to know likely discomfort based on blood type and also know contraceptives that fit particular through a user-friendly interface. The interface contains blood group information, contraceptive information, and clinic information to provide contraceptive services.

### *4.3. Sampling Technique*

Andrade views purposive sampling as the one whose characteristics are defined for a purpose that is relevant to the study [3]. The sampling method relies on the researcher's judgment when identifying and selecting the individuals, and cases that can provide the best information to achieve the study's objectives [20]. A purposive sampling technique was used to select six primary healthcare centers in Yola North. These primary health care centers include Malamre, Demsawo, Doubeli, Ladi Atiku, Gwadabawa, and Jambutu primary health care centers respectively. The centers were selected based on proximity and contribution to providing implant contraceptive service as advised by an expert. A total of 550 data were collected.

### *4.4. Procedure for Data Collection*

Questionnaires were used to collect relevant data for this study. A twenty-six (27) item questionnaire instrument on women's experiences in the use of contraceptives was designed by researchers and was validated by experts. The question items were designed for each contraceptive method. The contraceptive methods considered include implants, excluton, IUD, diaphragm, oral pill, nostril, depo, and combine pill. This gave an in-depth understanding of the inconveniences and their peculiarity to contraceptives and women's health conditions. This was necessary because there were no records of patients' discomfort/ negative experiences with the use of contraceptives in all the selected centers.

A qualitative method was used in collecting data for transformation into useful information for the study. The researchers synergized with the selected health centers to collect data from respondents. Their involvement allowed easy access to respondents who may not be willing to provide information to the researchers. 600 questionnaires were administered from 15$^{th}$ September to 22$^{nd}$ November, 2022; 100 questionnaire per centre. 550 of 600 administered questionnaires returned that is to say response level was 92% success.

## 5. Building the Models

### *5.1. Pre-processing*

The dataset consists of 550 observations and seventy (70) attributes. The dataset contained some records where several attributes do not have any value and is referred to as missing values. Records with many attributes missing values were first removed from the dataset as these records did not provide valuable knowledge. Attributes were reduced to eleven (11) after cleaning. Attributes are headache, nausea, dizziness, weight loss, Weight gain, Loss of appetite, Body weakness, Fever, Irregular Monthly cycle, Menstruation lasting more than usual, and body weight. The reason for choosing the eleven attributes as parameters is because they are the main variables for prediction. Five predictive variables (class) which are the predictive results were used for prediction. patient experience in using contraceptives (PEC) was coded as Headache (HD), Dizziness (DN), stomach-ache (SA), Weight loss (WL), Weight gain (WG), Loss of appetite (LA), Increase of apatite (IA), M. Last Unusual (IM), blood pressure (PB), weight (W) and was scored 1 to 10; HD (1), DZ (2), SA (3), WL (4), WG (5), LA (6), IA (7), IM (9), PB (10), BW (11), and 0 means no PEC, and class variable (BGT) are $O^+$, $B^+$, $AB^+$, $A^+$, $O^-$.

### *5.2. Data Normalization*

Normalization is a technique used as part of preparation in machine learning and the goal is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values [15].

The dataset was arranged into a suitable format and normalized to ensure that the output remain unbiased. It was imperative to do so because of the variance in the variables especially in BP and BW. BP and BW values are in 3 and 2 digits, whereas the rest of the variables are in single digits.

### *5.3. Split Dataset*

Data splitting is the process commonly used to divide data into a train, and test set. The dataset was divided into train data and test data. 70% of the data was used for training and 30% for testing. The model is trained on the training set and then examined using the testing set. This approach helps to evaluate model performance.

### *5.4. Training and Testing Data*

During the training and testing process, KNN and SVM Models were analyzed 10 times and repeated three (3) times each where a part of the data set is randomly selected for the training process and the other part for the testing. This mechanism is known as 10-fold cross-validation. 385 (70%) of the data was used for training and 165 (30%) testing set was used as an input to model training set to examine performance using the testing set.

### *5.5. Building the Model*

The classification models were built in R version 4.2.3 using the Caret library in the Caret package to perform KNN classification. Firstly, the trainControl() function was used to define the method of cross-validation. Train function is one of the primary tools used to evaluate, and resample the effect of model tuning parameters on performance, choose the "optimal" model across these parameters, and estimate model performance from a training set.

After building the model, its accuracy in KNN was obtained at K=5. K is a parameter of the algorithm with which an input is closely related within the dataset determined. Similarly, the "e1071" library in "e1071" package was used to build SVM. The kernel radial function tackled a multi-class classification problem and the model is tuned with arguments; gamma and cost, where gamma is the argument for use by the kernel function, and cost allow for the specification of the cost of a violation to the margin. Usually, when the cost is small, the margins will be wide resulting in many support vectors. SVM-Type: C-classification, SVM-Kernel: radial, and cost: 10 were used to build the SVM model.

### *5.6. Models Performance Evaluation*

The confusion matrix indicates the actual values vs. predicted values and summarizes the true negative, false positive, false negative, and true positive values in a matrix format [12]. Therefore, to evaluate the performances of the different classification models, the following performance matrix is adopted in this study.

Let, TP = the number of true positives of A+, AB+, O+, O-, and B+, i.e., the classifier predicted a positive outcome when the actual outcome was also positive.

FN = the number of false negatives; the classifier predicted a negative outcome when the actual outcome is positive;

Classification Accuracy (ACC): classification accuracy is one of the widely used performance metrics to evaluate a classifier. ACC is defined as the ratio of all samples that are classified correctly to the total number of test samples.

$$\mathrm{ACC} = \frac{\sum TT}{\sum N} \quad (1)$$

Where TT= sum of all classifier true predictions N = is the sum of all true and false predictions of the classifier.

Sensitivity (true positive rate) is defined as the ability of the classifier to accurately predict a successful blood type outcome from the attributes of the patient.

$$\text{Sensitivity} = \frac{\mathrm{TC}}{\mathrm{TC+FC}} \quad (2)$$

Where TC= true prediction of a class and FC= false prediction of a class.

Precision is the proportion of the predicated successful blood type cases that are successful.

$$\text{Precision} = \frac{\mathrm{TT}}{\mathrm{TT+TF}} \quad (3)$$

Where TT is the target variable true prediction and TF is target variable false prediction

## 6. Result

### *6.1. KNN Model Prediction*

The models were tested by the use of Predict function in which test data were used as the new data (input data) of the models. Table 1 shows KNN Row and Column totals of the Confusion Matrix of the classifier used to assess its performance.

***Table 1.*** *Row and Column totals of the Confusion Matrix.*

| N=169 | prediction | | | | | |
|---|---|---|---|---|---|---|
| | A+ | AB+ | B+ | O+ | ON | Total |
| | 16 | 3 | 1 | 2 | - | 22 |
| | - | 26 | 1 | 5 | - | 32 |
| | - | - | 29 | 3 | - | 32 |
| | 15 | - | 1 | 13 | - | 29 |
| | 5 | 1 | - | 4 | 44 | 54 |
| Total | 36 | 30 | 32 | 27 | 44 | 169 |

From table 1, a total of 169 test data was used for prediction, and the classifier predicted 169. The 5 classes are "A+", "AB+", "B+", "O-", "and O+". Classifier predicted 16 as "A+" meaning that they possessed the attribute of "A+" PEC but number 22. 26 was predicted as "AB+", meaning they possessed the attribute of "AB+" PEC but the actual prediction is 32. 29 predicted as "B+" and it is 32, 13 predicted as "O+", but it is 29, and 44 as "o+", but it's actual prediction is 54. The prediction accuracy of the KNN model on the test was found to be 75.74%. The accuracy was obtained by the use of the Confusion Matrix. Table 2 shows the confusion matrix score percentage of KNN.

***Table 2.*** *Confusion matrix percentage for all classes.*

| terms | A+ % | AB+ % | B+% | O+% | ON% |
|---|---|---|---|---|---|
| Sensitivity | 44.4 | 86.67 | 90.62 | 48.15 | 100 |
| Specificity | 95.49 | 95.68 | 97.81 | 88.73 | 92 |
| Pos Pred Value | 72.72 | 81.25 | 90.63 | 44.83 | 81.48 |
| Neg Pred Value | 86.40 | 97.08 | 97.81 | 90.00 | 100.00 |
| Balanced Accuracy | 69.97 | 91.18 | 94.22 | 68.44 | 96.00 |

KNN classifier had a total of 41 wrong predictions; 6 were predicted as "O+". This means the classifier could have classified them as "A+", "AB+", "B+", or "O-" because they possessed PEC attributes of "A+", "AB+", "B+", and "O-". Similarly, 2 were predicted as "B+". This means the classifier could have predicted them in "A+", and "O-" because they possessed PEC attributes of "A+", and "O-".

### 6.2. KNN Model Accuracy

The experiment result of KNN model accuracy was found to be 83.6% on train data. The accuracy Optimal of K was selected at the value K= 5. Figure 2 shows the graph of the K optimal value.

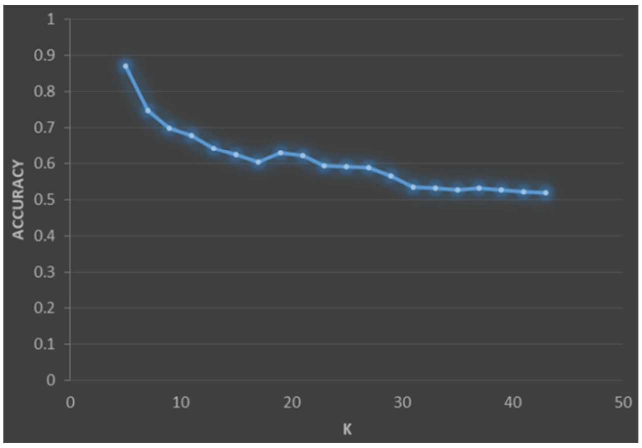


***Figure 2.*** *Graph of the optimal value of K.*

### 6.3. SVM Model Prediction

The SVM model confusion matrix of the classifier is shown in Table 3.

***Table 3.*** *Row and Column totals of the Confusion Matrix.*

| N=165 | prediction | | | | | |
|---|---|---|---|---|---|---|
| | A+ | AB+ | B+ | O- | O+ | Total |
| | 23 | - | - | - | - | 23 |
| | 5 | 29 | - | 4 | - | 38 |
| | - | - | 27 | - | - | 27 |
| | - | 2 | 2 | 32 | - | 36 |
| | - | - | - | - | 41 | 41 |
| Total | 28 | 31 | 29 | 36 | 41 | 165 |

From table 4 the classifier had a total of 13 wrong predictions; 9 data were classified as "B+", instead of being classified as "A+", and "O-". This is a false prediction because they possessed PEC attributes of "A+", and "O-". Similarly, 4 data were predicted in "O-" instead of "AB" and "A". The miscalculation rate of SVM is 7.8%.

The performance accuracy of the SVM model on the test is 92.2%. Cost value influenced the performance as it was held at c= 5 because of the data set size and the fact that the smaller the C, the more support vectors. The accuracy was obtained by the used of the confusion Matrix and presented in table 4.

***Table 4.*** *Confusion matrix score for all classes.*

| terms | A+ % | AB+ % | B+% | O+% | ON% |
|---|---|---|---|---|---|
| Sensitivity | 91.67 | 96.55 | 96.08 | 71.43 | 100 |
| Specificity | 100 | 99.28 | 98.31 | 100 | 88.98 |
| Pos Pred Value | 100 | 96.30 | 96.08 | 100 | 79.69 |
| Neg Pred Value | 95.65 | 96.48 | 98.31 | 98.78 | 100 |
| Balanced Accuracy | 76.67 | 91.57 | 97.19 | 85.71 | 94.49 |

### 6.4. SVM Model Accuracy

SVM model accuracy of the train data set was 90% on C-classification, SVM-Kernel radial maintained at a cost: 5. figure 3 shows the graph of cost for best support vectors.

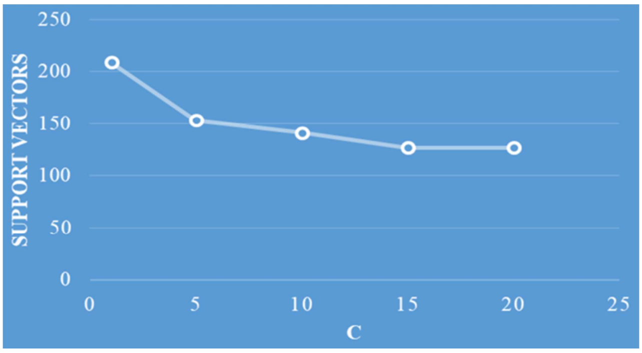


***Figure 3.*** *Graph of the optimal value of C.*

## 7. Implementation

The proposed model (MOBACO) was developed in an Android Studio version 4.3. Android Studio is a platform for Android app development incorporated with Java. It has an inbuilt emulator for testing applications virtually on a local computer. It has a user-friendly interface: the Graphical User Interface (GUI). The Blood group information interface and the Contraceptives information interface are the two implemented results.

### 7.1. Dashboard

The dashboard offers a range of options that users can utilize when engaging with the system. These activities are visually represented in Figure 4. For example, if a user seeks information on contraceptives, they can select "contraceptives," or if they need information on blood group types, they can click on "blood." Additionally, the option to search for the nearest clinic is available under the "clinic" section.

Moreover, the popup menu on the page enables users to perform various other operations, such as updating their user profile, adjusting settings, and logging out. To update their profile, users can simply click on the user profile option within the menu. Similarly, modifying settings can be achieved by selecting the respective option, and to conclude their session, users can click on "logout."


*Figure 4. Dashboard.*

### *7.2 Blood Group Information Interface*

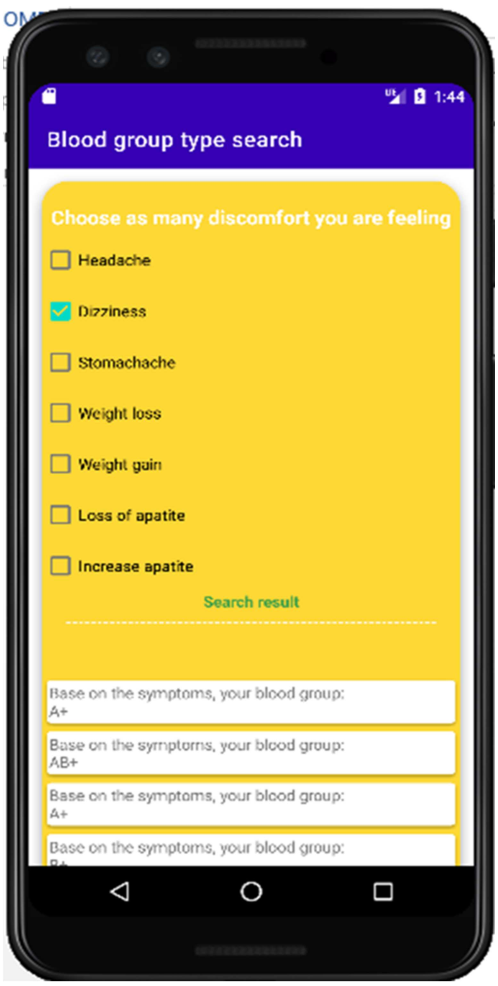


*Figure 5. Blood group information interface.*

This page is the graphical user interface (GUI) of the KNN prediction model. The user follows on-screen instructions to choose discomforts. The system predicts the user's blood group type that matches the selections by checking the neighbors that are most closely to the input selections. Figure 5 Contraceptives information interface.

### *7.3 Contraceptives Information Interface*

Unlike the blood group information, contraceptives provide information on likely possible discomforts, especially for a user who has the intention of having an implanted contraceptive. Users select the contraceptive type and blood group type from the dropdown menu. Based on the selections, the system displays possible discomfort. Figure 6 shows the contraceptive information page.

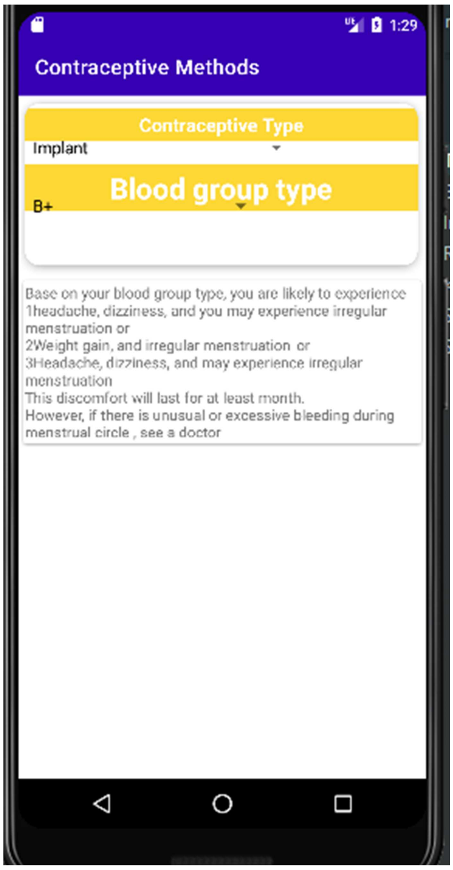


*Figure 6. Contraceptive information page.*

## 8. Conclusion

Experimental result shows that women's experience in the use of contraceptives differs and the differentials depend on blood type. Women with similar blood types are most likely to have many similar experiences (discomforts).

In this study, KNN and SVM were used to develop a model with 5 classes for the prediction of contraceptives and blood groups based on certain medical conditions. The two models work well. SVM performed better on the test data set. The result of the work elucidates an accuracy score of 90.8% on train data and 92.2% on test data. The accuracy was determined using a confusion matrix. KNN performs better on "train" data with an accuracy of 77.68% but the performance is reduced on "test". The accuracy of "test" data is 85.72%.

SVM has better Sensitivity, Specificity, Positive prediction, negative prediction, and balance accuracy compared to KNN as shown in table 2 and table 4. However, the performances of models imply that the application can be used by women as a means of accessing contraceptive information.

Therefore women's willingness to use contraceptives can greatly be improved and sustained by providing adequate and reliable information, peculiar to individual service needs. A woman that opts for modern contraceptives should know either before or after, the possible inconveniences she will experience based on her condition, and know how long it will last.

# References


[1] Ahmad, P., Qamar, S. & Rizvi, S. (2015). Techniques of Data Mining In Healthcare: A Review. *International Journal of Computer Applications*. 120. 38-50. 10.5120/21307-4126.

[2] Alano, A., & Hanson, L. (2018). "Women's perception about contraceptive use benefits towards empowerment: A phenomenological study in Southern Ethiopia". 13 (9), PLoS ONE, e0203432.

[3] Andrade C. (2021). The Inconvenient Truth About Convenience and Purposive Samples. *Indian Journal of psychological medicine*, *43* (1), 86–88. https://doi.org/10.1177/0253717620977000

[4] Antonio, D., Sanchez, P. & Jose, A. O. (2018). Adolescent Contraceptive Use and its Effects on Fertility. V38, Article 45, Pages 1359–1388 http://www.demographiresearch.org/Volumes/Vol38/45/ DOI: 10.4054/DemRes.2018.38.45.

[5] Ayyoubzadeh, S. M., Ghazisaeedi, M., Kalhori S. R. N., Hassaniazad M., Baniasadi T., Maghooli K., & Kahnouji K. (2020). A study of factors related to patient's length of stay using data mining techniques in a general hospital in southern Iran. Health Inf. Sci. Syst. doi: 10.1007/s13755-020- 0099-8.

[6] Baker, R. S. J. d. (2010 in press). Data Mining for Education. McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.

[7] Carr L. (2021). FDA Approves New Oral Contraceptive Nextstellis. Retrieved on 2/2/2021 from https://www.contemporaryobgyn.net/view/fda-approves-new-oral-contraceptive-nextstellis

[8] Chen, S. I., Tseng, H. T., & Hsieh C. C. (2020). Evaluating the impact of soy compounds on breast cancer using the data mining approach. Food Funct. 11: 4561–4570. doi: 10.1039/C9FO00976K.

[9] Durowade, K. A., Omokanye, L. O., Elegbede, O. E., Adetokunbo, S., Olomofe, C. O., Ajiboye, A. D., Adeniyi, M. A., & Sanni, T. A. (2017). Barriers to Contraceptive Uptake among Women of Reproductive Age in a Semi-Urban Community of Ekiti State, Southwest Nigeria. Ethiopian Journal of health sciences, 27 (2), 121–128. https://doi.org/10.4314/ejhs.v27i2.4

[10] Ghiya, N., Godbole, S., Hol, P., Deortare, G. & Chavan, M. (2015) "Forecasting Diseases by Classification and Clustering Techniques", *International Journal of Innovative Research in Science, Engineering, and Technology*, 4 (1), pp. 18958-18961.

[11] Han, J., Pei, J., & Kamber, M. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann publisher: USA, 3rd ed.

[12] Hassan, R., Al-Insaif, S., Hossain, M. I., & Kamruzzaman, J. (2020). A Machine Learning Approach for Prediction of Pregnancy Outcome following IVF Treatment. Neural Computing and Applications DOI: 10.1007/s00521-018-3693-9.

[13] Haq, I., Hossain, I., Rahman, M., Methun, I. H, Talukder, A., Jakaria Habib, J & Hossain, S., (2022). Contraceptive Practice among Ever-Married Women in Bangladesh: A Hierarchical Machine Learning Classification Approach. DOI: http://dx.doi.org/10.5772/intechopen.103187

[14] Hessami, A. R., Sun, D., Odreman, G. J., Zhou, X., Nejat, A., & Saeedi, M. (2017). Project Scoping Guidebook for Metropolitan Planning Organization Transportation Projects. Kingsville, Texas: Texas A&M University Kingsville (TAMUK).

[15] Jaitley, U. (2018). Why Data Normalization is necessary for Machine Learning models. Retrieved on 2/2/2022 from https://medium.com/@urvashilluniya/why-data-normalization-is-necessary-for-machine-learning-models-681b65a05029

[16] Jothi, N. & Husain, W. (2015). Data mining in healthcare: A review. Procedia Comput. Sci. 2015; 72: 306–313. doi: 10.1016/j.12.145. [CrossRef] [Google Scholar].

[17] Mayne pharma. (2021). U.S. FDA Approves NEXTSTELLIS®, New Oral Contraceptive. Retrieved on 2/2/2022 from https://www.prnewswire.com/news-releases/us-fda-approves-nextstellis-new-oral-contraceptive-301270669.html

[18] Metcalf, E. (2021). Birth Control Implants. Retrieved on 5/2/2022 from https://www.webmd.com/sex/birth-control/birth-control-implants-types- safety-side- effects.

[19] Musa, M., Y., Wajiga, M. G. & Garba, E. J. (2019). Multi-Agent-Based Diagnostic Model for Breast Tumour Classification. American Journal of Data Mining and Knowledge Discovery. 4. 1-7. 10.11648/j.ajdmkd.20190401.11.

[20] Nikolopoulou, K. (2022). What Is Purposive Sampling? | Definition & Examples. Available at https://www.scribbr.com/methodology/purposive-sampling/

[21] Pika A., Wynn, M. T., Budiono, S., Ter Hofstede, A. H., van der Aalst, W. M.,& Reijers H. A. (2020). Privacy- Preserving Process Mining in Healthcare. Int. J. Environ. Res. Public Health.; 17: 1612. doi: 10.3390/ijerph17051612.

[22] Ramageri, B. M. (2010). Data Mining Techniques And Applications. *Indian Journal of Computer Science and Engineering 1 (4)* 301-305. ISSN: 0976- 5166.

[23] Rahman, A., Honan B., Glanville T., Hough P., & Walker K. (2019). Using data mining to predict emergency department length of stay greater than 4 hours: Derivation and single-site validation of a decision tree algorithm. Emerg. Med. Australas. 32: 416–421. doi: 10.1111/1742- 6723.13421.

[24] Reddy, R. P., Mandakini, C. h. & Radhika, C. H. (2020). A Review on Data Mining Techniques and Challenges in Medical Field. *International Journal of Engineering Research & Technology (IJERT), 9 (8).*

[25] Ricciardi, C., Amboni, M., de Santis, C., Improta, G., Volpe G., Iuppariello L., Ricciardelli G., D'Addio G., Vitale, C., Barone P., et al. (2019). Using gait analysis parameters to classify Parkinsonism: A data mining approach. Comput. Methods Programs Biomed. 2019; 180: 105033. doi: 10.1016/j.cmpb.2019.105033.

[26] Rustagi, N. & Taneja, D. K., Kaur, R. & Ingle, G. K.. (2010). Factors affecting contraception among women in a minority community in Delhi: A qualitative study. 33. 10-15.

[27] Saeed, S., Shaik, A.& Memon, M. A. (2018). Impact of Data Mining Techniques To Analyze Health Care Data. *Journal of Medical Imaging and Health Informatics. 8 (4)*: 682-690. DOI: 10.1166/jmihi.2018.2385.

[28] Shouman, M., Turner, T., & Stocker, R. (2012). Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients. International Journal of Information and Education Technology, V 2 (3).

[29] Ţăranu I. (2016). Data mining in healthcare: Decision making and precision. Database Syst. 6: 33–40.

[30] Urbaniak, A. (1996). Expert system for diagnosis of women's menstrual cycle using natural family planning method. Presented at the 17th IFIP Conference on System ModeUing and Optimization.

[31] Yim, S. J., Lui L. M., Lee Y., Rosenblat, J. D., Ragguett R.-M., Park C., Subramaniapillai M., Cao B., Zhou A., & Rong C., et al (2020). The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J. Affect. Disord. 2020; 274: 602–609. doi: 10.1016/j.jad.2020.05.116.