ACHSM Asia-Pacific Journal of Health Management

RESEARCH ARTICLE

PREDICTING FERTILITY DETERMINANTS IN BANGLADESH USING MACHINE LEARNING MODELS

Shayla Naznin[1], *Md Jamal Uddin*[2,3], *Ahmad Kabir*[2]

1. Department of Statistics, Mawlana Bhashani Science and Technology University, Bangladesh
2. Department of Statistics, Shahjalal University of Science and Technology, Bangladesh
3. Faculty of Graduate Studies, Daffodil International University, Savar, Dhaka, Bangladesh

*Correspondence: shayla.naznin@gmail.com

ABSTRACT

BACKGROUND:

Fertility is a key demographic indicator that influences population dynamics, economic growth, social welfare, and public policy. Understanding fertility determinants is crucial for addressing population challenges in developing countries like Bangladesh. This study aims to fill a gap in existing research by applying advanced machine learning (ML) methods to predict fertility outcomes and identify key fertility determinants in Bangladesh using data from the Bangladesh Demographic Health Survey (BDHS) 2017–18.

METHODS:

The dataset comprised 20,127 women, reflecting the fertility landscape in Bangladesh. We employed both conventional statistical methods and various machine learning models including Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, LightGBM and Neural Networks (NN). The performance of the models was evaluated based on metrics such as accuracy, precision, recall, F1 score, Area Under the Receiver Operating Characteristic Curve (AUROC) and alongside K-fold cross-validation.

RESULTS:

The SVM model emerged as the most effective, with macro average recall (89%), precision (89%) and F1 score (89%) as well as weighted average recall (96%), precision (95%), F1 score (95%) and accuracy was 95.5%. The results revealed that access to mass media, father's education level, mother's highest educational attainment, number of household members, maternal Body Mass Index (BMI), number of living children and child mortality (i.e., incidents of sons or daughters dying) were the most influential factors of fertility in Bangladesh.

CONCLUSIONS:

Our findings demonstrate that machine learning models, particularly SVM, provide superior predictive power compared to traditional statistical methods in identifying key fertility determinants. These insights contribute to a deeper understanding of fertility dynamics in Bangladesh and suggest that targeted interventions—focusing on media access, parental education, and healthcare improvements could significantly influence fertility rates. Future research should explore the use of ensemble models and longitudinal data to enhance predictive accuracy and capture temporal changes in fertility determinants.

KEYWORDS

fertility, boruta algorithm, machine learning, SVM, best model

INTRODUCTION

Fertility is a critical demographic indicator shaping the socio-economic and public health landscape of nations. As a core measure of population growth, it is intricately linked to various facets of human development, including health, education, and economic prosperity. Globally, significant reductions in fertility rates have been observed over recent decades largely driven by improvements in healthcare expanded access to education and the increased availability of family planning services [1, 2].

In Bangladesh, fertility has undergone substantial changes in recent decades. In 1975 the Total Fertility Rate (TFR) was 6.3 births per woman [3] reflecting the socio-economic barriers and limited access to family planning services at the time. By the 2017-18 period the TFR had decreased dramatically to 2.3 births per woman [4] underscoring significant progress in reproductive health, women's empowerment and public health initiatives. Yet, despite these advances Bangladesh's TFR remains relatively high compared to other developing countries, signaling persistent socio-economic and cultural challenges that continue to shape reproductive behaviors [5].

Research has consistently identified key factors that influence fertility rates. Age at first marriage and first birth are prominent determinants: early marriages often lead to early childbearing and higher fertility while delayed marriage tends to reduce family size. Education, especially for women plays a pivotal role in fertility reduction [1], [6], [7] . Additionally, women's employment [8]–[11] and economic empowerment have contributed to fertility declines as working women often prioritize career development and tend to have fewer children. Access to contraception [1], [12]–[15] enables couples to plan family size more effectively reducing overall fertility rates. Urbanization, with improved healthcare and education, leads to smaller families, while access to healthcare and family planning in rural areas is crucial for reducing fertility rates by empowering informed reproductive choices [8], [10], [16]–[20].

While these factors have been well-established in the literature much of the existing research relies on traditional statistical methods such as regression analysis to explore the relationships between fertility determinants [7], [13], [21–24]. These methods, while useful for identifying linear relationships, often fall short in capturing complex interactions and non-linear dynamics among fertility-related factors. Fertility behavior is shaped by a range of socio-economic, demographic and cultural influences that interact in intricate ways patterns that traditional models may not detect [25–27].

Machine Learning (ML) offers a powerful alternative for analyzing fertility data. ML techniques excel at processing large datasets and identifying non-linear trends and complex interactions that traditional methods might miss. They can model relationships between variables with greater accuracy and handle high-dimensional data, allowing researchers to evaluate multiple factors simultaneously without overfitting [28]–[31]. This is particularly useful for studying fertility, where factors like education, employment, and healthcare interact in complex ways.

The rationale for this study stems from the growing recognition that ML techniques due to their flexibility and capacity to manage large and complex datasets frequently outperform conventional statistical methods. While traditional approaches have been successful in identifying simple relationships, they often struggle to account for the non-linear and multidimensional nature of fertility determinants. This research fills a critical gap by applying ML methodologies to fertility analysis offering more robust and precise insights compared to previous studies that relied predominantly on conventional techniques.

This study aims to leverage ML techniques to: (i) identify the key factors influencing fertility in Bangladesh and (ii) predict fertility using data from the Bangladesh Demographic and Health Survey (BDHS) 2017-18 [4]. By integrating advanced computational methods with demographic data this research aims to offer a more comprehensive understanding of fertility determinants in Bangladesh with important implications for public health planning and policy development

METHODS AND MATERIALS

ETHICAL STATEMENT

Bangladesh Demographic and Health Survey (BDHS) secondary data were used in this study; this is a nationally representative survey conducted in Bangladesh at regular intervals. The BDHS collects crucial information on health, family planning, and socio-demographic factors. It is carried out by the National Institute of Population Research and Training (NIPORT) under the Ministry of Health and Family Welfare, with technical support from ICF International. Permission for the survey protocol was secured from the ORC Macro (Macro International Inc.) Institutional Review Board and the National Research Ethics Committee in Bangladesh. The study relied on existing public domain survey datasets freely available online. Before analysis, all identifying information of the respondents was removed. The dataset for this research was acquired after obtaining permission by the authors from the Demographic and Health Surveys (DHS) program.

DATA

This study utilizes Bangladesh Demographic and Health Survey (BDHS) conducted in 2017-18 data which is part of an ongoing series initiated in 1993. The nationally representative surveys employed a two-stage stratified sampling design covering all administrative divisions in Bangladesh. The BDHS ensured consistency in comparing different demographic data by employing nearly identical questionnaires over time. Comprehensive details of sampling and survey methodology regarding the BDHS have been previously documented [4] and all survey data considered secondary data are publicly accessible. For this research we specifically extracted information from the BDHS (2017-2018) datasets with a focus on female respondents excluding temporary (de jure) residents and cases with missing values. The dataset comprising 20,127 weighted observations serves as a robust foundation for our analysis. This data cleaning process ensured consistency across all variables used in the machine learning models and prevented bias that could arise from incomplete records.

TARGET VARIABLE

The primary target variable in this study was Children Ever Born (CEB), representing the total number of children a woman had given birth to at the time of the survey. For descriptive analysis (Table 2), CEB was grouped into three categories: no children (0), 1–3 children, and four or more children (≥4) to facilitate interpretation of fertility patterns. However, for the machine learning models, CEB was retained as seven categories (0, 1, 2, 3, 4, 5, and 6 or more children) to capture finer variation in fertility outcomes.

INDEPENDENT VARIABLES

The list of independent variables appears in Table 1

TABLE-1: INDEPENDENT VARIABLES

Variables	Variables Coding
Division	0= Barishal, 1= Chattogram, 2= Dhaka, 3= Khulna, 4= Rajshahi, 5= Mymensingh, 6= Rangpur, 7= Sylhet
Type of place of residence	1= Rural, 0= Urban
Religion	1=Muslim, 2=Non-Muslim
Wealth index	0=Poorest, 1=Poorer, 2=Middle, 3=Richer, 4=Richest
Mother's educational level	0=No education, 1=Primary, 2=Secondary, 3=Higher
Father's education level	0=No education, 1=Primary, 2=Secondary, 3=Higher
Father's occupation	1=Agriculture, 2=Worker or Labor, 3=Professional Worker, 4=Business & Others
Mother's occupation	0= Housewife, 1=Worker or Labor, 2=Professional Worker, 3=Business & Others
Type of Toilet Facilities	1=Toilet With Flush, 2= VIP latrine, 3=Pit Latrine, 4=Hanging toilet and Other
Sources of Drinking Water	1=Other than Tubewell or Borehole, 2=Tubewell or Borehole
Number of household members	1=1-3 Member, 2=4-6 Member, 3=Over 6 Member

Type of Cooking Fuel	1= Kerosene or Natural Gas, 2=Wood, 3=Agricultural Crop, 4=Animal Dung and Others
Number of living children	0=No Child, 1=One Child, 2=Two Child, 3=Three Child, 4=Four Child, 5=Five or more Children
Body Mass Index (BMI)	1=Thin(<18.5), 2=Normal (18.5-24.99), 3=Overweight (25-29.99), 4=Obese(>=30)
Age at first Marriage	1=Below 18years old, 2=18 years and above
Contraceptive use and intention	1=Using modern method, 2=Using traditional method, 3=Non-user - intends to use later, 4=Does not intend to use
Son or daughter died	0=No, 1=Yes
Access to Media	0=No, 1=Yes
Sex of household head	1=Male, 2=Female
Ideal number of children	1=0-2 Children, 2=3-4 Children, 3=Over 4 Children
Currently breastfeeding	0=No, 1=Yes
Desire for more children	1=Wants within 2 years, 2=Wants after 2+ years, 3=Wants, unsure timing/Undecided, 4=Wants no more, 5=Sterilized (respondent or partner)/Declared infecund
Father's desire for children	1=Both want same, 2=Husband wants more, 3=Husband wants fewer, 4=Don't know

STATISTICAL ANALYSIS

This study integrates both traditional statistical methods and modern machine learning techniques to explore fertility determinants combining SPSS for statistical analysis, R for feature selection and Python for model building and evaluation. This study aims to explore risk factors associated with fertility by employing various machine learning classification models namely Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, LightGBM and Neural Network (NN). The dataset was divided into an 80% training set and a 20% test set. The training set was used to train the machine learning models, while the test set evaluated their performance. The models were trained using the training dataset and evaluated using the independent test dataset. The evaluation of predictive models was conducted and compared with metrics such as the macro average and weighted average of the Confusion Matrix, Accuracy, F1 Score, Precision, Recall, Area Under the Receiver Operating Characteristics Curve (AUROC) and K-fold cross-validation.

In this study, several software tools were employed to achieve a comprehensive analysis of fertility determinants. SPSS Version 25 was utilized for bivariate analysis specifically chi-square tests to identify significant variables related to fertility. These tests provided a foundation for comparing traditional statistical approaches with machine learning techniques.

For feature selection the Boruta algorithm was employed in R. The Boruta algorithm is a powerful all-relevant feature selection method based on the Random Forest classifier. Its unique approach involves creating shadow features (randomly permuted versions of the original variables) and comparing their importance with the actual features. By retaining only those variables that outperform their shadow counterparts Boruta ensures that all relevant features are included in the model, providing a robust selection of fertility-related factors.

Machine learning models were built using Python (version 3.0) specifically employing various libraries including pandas, scikit-learn, XGBoost, LightGBM and matplotlib.

The dataset was processed using the scikit-learn Pipeline framework to ensure a consistent and reproducible preprocessing workflow for all machine learning models.

The preprocessing procedure included the following steps:

1. Removal of observations containing missing values to ensure a complete dataset for model training and evaluation.
2. Encoding categorical variables using OneHotEncoder to convert categorical features into numerical format suitable for machine learning algorithms.
3. Standardization of numerical variables using StandardScaler so that all numerical features had a mean of 0 and a standard deviation of 1.
4. Feature selection using the Boruta algorithm to retain only the most relevant predictors for fertility.
5. Splitting the dataset into training (80%) and testing (20%) subsets to evaluate model performance on unseen data.

Standardization was particularly important for distance-based and kernel-based models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), as these models are sensitive to differences in feature scale.
This combination of SPSS, R and Python allowed for a multi-faceted analysis facilitating a rigorous comparison between traditional statistical methods and modern machine learning approaches in predicting fertility determinants.

Decision Tree (DT)

The Decision Tree (DT) algorithm uses a tree-like structure to classify data based on recursive splitting [32]. Its intuitive design makes it easy to interpret and useful for both categorical and continuous variables. The Decision Tree (DT) algorithm using the CART method classifies data by recursively splitting it based on feature values. Each internal node represents a decision and leaf nodes represent outcomes making the model intuitive and easy to interpret. DTs are effective for both categorical and continuous data but can overfit if the tree becomes too complex. In this study it helps identify key decision points related to fertility predictions.

Random Forest (RF)

Random Forest (RF) is an ensemble method that builds multiple decision trees and averages their outputs to improve prediction accuracy. It reduces overfitting and handles high-dimensional data efficiently [33]. RF's robustness makes it ideal for analyzing complex fertility determinants.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful classification method that constructs hyperplanes to separate data into distinct classes [34]. It excels in high-dimensional spaces and performs well with small, complex datasets. In fertility prediction, SVM's flexibility ensures high accuracy.

Logistic Regression (LR)

Logistic Regression (LR) is a probabilistic model for binary classification, often used to predict outcomes like fertility determinants. It estimates the probability of an event based on input features. LR's simplicity and interpretability make it an excellent choice for categorical data [35].

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a non-parametric, instance-based algorithm that classifies data points based on the majority class among the k-nearest neighbors in feature space [36]. Its simplicity makes it easy to implement, but it can struggle with high-dimensional data and is sensitive to feature scaling and outliers.

Extreme Gradient Boosting (XGBoost)

XGBoost (Extreme Gradient Boosting) is a powerful and efficient gradient boosting framework that enhances decision tree performance through regularization techniques to prevent overfitting [37]. It excels in handling large datasets and offers high predictive accuracy, making it a popular choice for data science competitions and complex structured data problems.

Light Gradient Boosting Machine (LightGBM)

LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework designed for speed and efficiency, using a novel tree-building algorithm. It supports large datasets with faster training times and lower memory usage compared to traditional methods. LightGBM can also directly handle categorical features, improving model performance [38].

Neural Network (NN)

Neural Networks are advanced computational models inspired by the human brain, consisting of interconnected layers that learn to extract patterns from data. The Neural Network (NN) [39] used here is a Multi-Layer Perceptron (MLP) consisting of input hidden and output layers. It uses the ReLU activation function in the hidden layers and is trained with backpropagation. NN models are powerful for capturing complex non-linear relationships though they require more data and computational resources. In this study the MLP is applied to predict fertility outcomes based on socio-economic and demographic factors.

Hyperparameter Tuning

To optimize model performance, hyperparameter tuning was performed using GridSearchCV with 10-fold cross-validation on the training dataset. Key parameters were tuned for each algorithm, including n_estimators and max_depth for Random Forest, C and gamma for Support Vector Machine, and the number of neighbors (k) for K-Nearest Neighbors. For gradient boosting models (XGBoost and LightGBM), parameters such as learning rate, max_depth, and number of estimators were optimized. For the Neural Network, the hidden layer size and regularization parameter (alpha) were tuned. The final model parameters were selected based on the highest cross-validated accuracy, ensuring improved predictive performance and reduced risk of overfitting.

SENSITIVITY ANALYSIS

A sensitivity analysis was conducted to evaluate the robustness of the machine learning models. The models were re-estimated under alternative preprocessing settings and cross-validation configurations. The results showed that model performance remained consistent across these variations, indicating that the findings were stable and not sensitive to minor changes in model specification.

Confusion Matrix Performance Parameters

A confusion matrix is used to evaluate classification performance by comparing predicted versus actual outcomes. It provides metrics like accuracy, precision, sensitivity, and specificity. These metrics offer insight into model effectiveness, especially in fertility predictions.

Receiver Operating Characteristic (ROC) Curve

The ROC curve visually represents a model's ability to differentiate between classes by plotting true positive rates against false positives. The Area Under the Curve (AUC) measures overall performance [40]. Higher AUC values indicate a better-performing fertility prediction model.

Boruta Algorithm

Boruta, a Random Forest-based feature selection method, was chosen for its ability to retain all relevant features by comparing them with shadow features. Unlike other methods like Lasso or Recursive Feature Elimination (RFE), Boruta [41] ensures that no potentially important fertility-related variable is overlooked making it ideal for high-dimensional data in this study.

RESULTS

DESCRIPTIVE STATISTICS

The results presented in Table 2, Division, Type of residence, Religion, Wealth index, Maternal and Father's education, Occupation, Toilet facilities, Water sources, Household size, Cooking fuel, and several others are significant fertility determinants at the 5% level ($p<0.001$). Factors like Mother's Age, Body Mass Index, Age at first Marriage, Contraceptive

use, Son or daughter death, Access to media, Marital status, Ideal children number, and Desire for more children also show significant relationships (p<0.001), with Father's desire for children (p<0.001) and Sex of household head (p=0.020).

TABLE-2 DESCRIPTIVE ANALYSIS OF BACKGROUND CHARACTERISTICS

Characteristics	Total	Children Ever Born			χ2	p-value
	n(%)	No children n(%)	1 to 3 children n(%)	4 or More Children n(%)		
Division						
Barishal	1125(5.59)	102(9.07)	757(67.29)	266(23.64)		
Chattogram	3622(18)	355(9.8)	2325(64.19)	942(26.01)		
Dhaka	5123(25.45)	601(11.73)	3577(69.82)	945(18.45)		
Khulna	2336(11.61)	237(10.15)	1802(77.14)	297(12.71)	323.96	<0.001
Mymensingh	1546(7.68)	160(10.35)	1006(65.07)	380(24.58)		
Rajshahi	2801(13.92)	261(9.32)	2064(73.69)	476(16.99)		
Rangpur	2380(11.83)	196(8.24)	1681(70.63)	503(21.13)		
Sylhet	1193(5.93)	114(9.56)	705(59.09)	374(31.35)		
Type of place of residence						
Urban	5729(28.46)	628(10.96)	4196(73.24)	905(15.8)		
Rural	14398(71.54)	1397(9.7)	9722(67.52)	3279(22.77)	121.81	<0.001
Religion						
Muslim	18251(90.67)	1889(10.35)	12435(68.13)	3927(21.52)	94.98	<0.001
Non-Muslim	1877(9.33)	137(7.3)	1483(79.01)	257(13.69)		
Wealth index						
Poorest	3743(18.6)	277(7.4)	2407(64.31)	1059(28.29)		
Poorer	3956(19.66)	349(8.82)	2575(65.09)	1032(26.09)		
Middle	4059(20.17)	387(9.53)	2828(69.67)	844(20.79)	464.30	<0.001
Richer	4183(20.79)	497(11.88)	2917(69.73)	769(18.38)		
Richest	4183(20.79)	515(12.31)	3190(76.26)	478(11.43)		
Maternal educational level						
No education	3333(16.56)	112(3.36)	1663(49.89)	1558(46.74)		
Primary	6290(31.25)	342(5.44)	4148(65.95)	1800(28.62)	3218.38	<0.001
Secondary	7973(39.61)	979(12.28)	6225(78.08)	769(9.65)		
Higher	2531(12.58)	592(23.39)	1882(74.36)	57(2.25)		
Father's education level						
No education	4077(21.54)	151(3.7)	2381(58.4)	1545(37.9)		
Primary	6081(32.12)	508(8.35)	4127(67.87)	1446(23.78)	1599.57	<0.001
Secondary	5675(29.98)	694(12.23)	4249(74.87)	732(12.9)		
Higher	3098(16.36)	520(16.79)	2402(77.53)	176(5.68)		
Father's occupation						
Agriculture	4902(25.87)	295(6.02)	3103(63.3)	1504(30.68)		
Worker/Labor	9075(47.9)	1078(11.88)	6441(70.98)	1556(17.15)	621.71	<0.001
Professional Worker	1046(5.52)	187(17.88)	787(75.24)	72(6.88)		

Business & Others	3924(20.71)	319(8.13)	2838(72.32)	767(19.55)		
Mothers's occupation						
Housewife	10026(49.83)	1361(13.57)	7008(69.9)	1657(16.53)		
Worker/Labor	6909(34.34)	283(4.1)	4640(67.16)	1986(28.75)	730.96	<0.001
Professional Worker	2480(12.33)	265(10.69)	1764(71.13)	451(18.19)		
Business & Others	705(3.5)	115(16.31)	501(71.06)	89(12.62)		
Type of Toilet Facilities						
Toilet With Flush	5724(30.95)	650(11.36)	4194(73.27)	880(15.37)		
VIP latrine	2748(14.86)	252(9.17)	1954(71.11)	542(19.72)	302.31	<0.001
Pit Latrine	9673(52.3)	745(7.7)	6377(65.93)	2551(26.37)		
Hanging toilet and Other	351(1.9)	16(4.56)	236(67.24)	99(28.21)		
Sources of Drinking Water						
Other than Tubewell or Borehole	195(10.36)	1368(72.69)	319(16.95)	1882(100)	32.94	<0.001
Tubewell or Borehole	1468(8.84)	11393(68.57)	3753(22.59)	16614(100)		
Number of household members						
1-3 Member	4258(21.16)	601(14.11)	3169(74.42)	488(11.46)		
4-6 Member	11287(56.08)	831(7.36)	7976(70.67)	2480(21.97)	517.49	<0.001
Over 6 Member	4581(22.76)	593(12.94)	2773(60.53)	1215(26.52)		
Type of Cooking fuel						
Kerosene or Natural Gas	3769(20.4)	415(11.01)	2838(75.3)	516(13.69)		
Wood	8436(45.65)	790(9.36)	5690(67.45)	1956(23.19)	223.78	<0.001
Agricultural Crop	4918(26.61)	360(7.32)	3306(67.22)	1252(25.46)		
Animal Dung and Others	1356(7.34)	96(7.08)	916(67.55)	344(25.37)		
Number of living children						
No Child	2138(10.62)	2025(94.71)	109(5.1)	4(0.19)		
One Child	4574(22.73)	0(0)	4556(99.61)	18(0.39)		
Two Child	6205(30.83)	0(0)	6038(97.31)	167(2.69)		**<0.001***
Three Child	4053(20.14)	0(0)	3215(79.32)	838(20.68)		
Four Child	1863(9.26)	0(0)	0(0)	1863(100)		
Five or more Children	1294(6.43)	0(0)	0(0)	1294(100)		
Body Mass Index						
Thin	2285(11.55)	315(13.79)	1453(63.59)	517(22.63)		
Normal	11173(56.47)	1237(11.07)	7578(67.82)	2358(21.1)	158.29	<0.001
Overweight	5036(25.45)	342(6.79)	3683(73.13)	1011(20.08)		
Obese	1290(6.52)	76(5.89)	969(75.12)	245(18.99)		
Age at first Marriage						
<18	15166(75.35)	1192(7.86)	10315(68.01)	3659(24.13)	638.86	<0.001
18+	4962(24.65)	834(16.81)	3603(72.61)	525(10.58)		

Contraceptive use and intention						
Using modern method	9854(48.96)	401(4.07)	7518(76.29)	1935(19.64)		
Using traditional method	1889(9.38)	77(4.08)	1229(65.06)	583(30.86)	2685.50	<0.001
Non-user - intends to use later	5081(25.24)	1275(25.09)	3382(66.56)	424(8.34)		
Does not intend to use	3304(16.41)	273(8.26)	1789(54.15)	1242(37.59)		
Son or daughter died						
No	17036(84.64)	2025(11.89)	12779(75.01)	2232(13.1)	4071.07	<0.001
Yes	3091(15.36)	0(0)	1139(36.85)	1952(63.15)		
Access to Media						
No	6847(34.02)	529(7.73)	4202(61.37)	2116(30.9)	658.04	<0.001
Yes	13280(65.98)	1497(11.27)	9715(73.16)	2068(15.57)		
Sex of household head						
Male	17166(85.29)	1767(10.29)	11822(68.87)	3577(20.84)	7.78	0.020
Female	2960(14.71)	258(8.72)	2096(70.81)	606(20.47)		
Ideal number of children						
0-2 Children	15520(77.61)	1803(11.62)	11518(74.21)	2199(14.17)		
3-4 Children	4361(21.81)	205(4.7)	2322(53.24)	1834(42.05)	1823.94	<0.001
Over 4 Children	117(0.59)	5(4.27)	33(28.21)	79(67.52)		
Currently breastfeeding						
No	15989(79.44)	2025(12.66)	10287(64.34)	3677(23)		
Yes	4138(20.56)	0(0)	3631(87.75)	507(12.25)		**<0.001***
Desire more children						
Wants within 2 years	2253(11.87)	934(41.46)	1285(57.04)	34(1.51)		
Wants after 2+ years	3928(20.69)	791(20.14)	3095(78.79)	42(1.07)		
Wants, unsure timing/Undecided	577(3.04)	102(17.68)	457(79.2)	18(3.12)	6110.81	<0.001
Wants no more	10256(54.02)	21(0.2)	7313(71.3)	2922(28.49)		
Sterilized (respondent or partner)/Declared infecund	1970(10.38)	34(1.73)	1045(53.05)	891(45.23)		
Father's desire for children						
Both want same	14118(79.05)	1588(11.25)	10150(71.89)	2380(16.86)		
Husband wants more	1939(10.86)	65(3.35)	1232(63.54)	642(33.11)	565.93	<0.001
Husband wants fewer	1089(6.1)	65(5.97)	805(73.92)	219(20.11)		
Don't know	714(4)	162(22.69)	343(48.04)	209(29.27)		

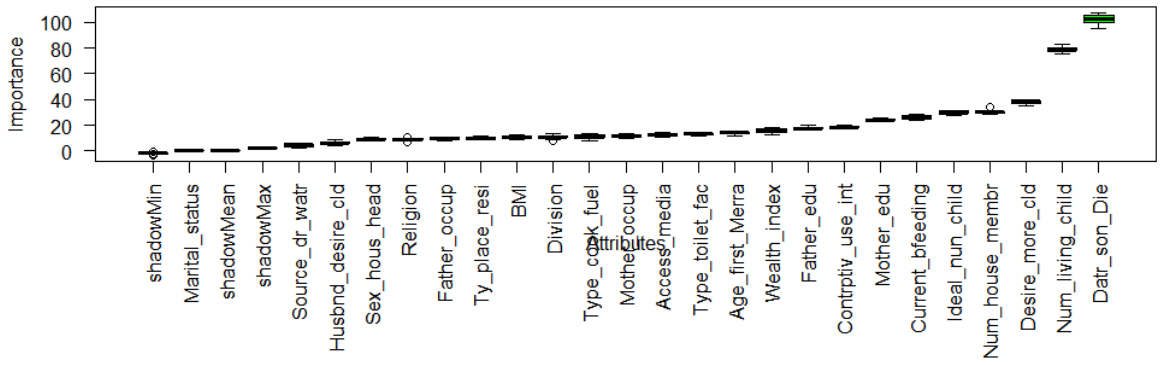
* Fisher's Exact test was utilized in place of chi-square.

SELECTION OF THE BEST FEATURES

Figure-1 the Boruta algorithm illustrates that we opted to keep 23 variables as a determinant of fertility i.e., Division, Type of place of residence, Wealth index, Maternal educational level, Father's education level, Father's occupation, Mothers' occupation, Number of household members, Type of Toilet Facilities, Sources of Drinking Water , Type of Cooking, Number of living children, Maternal age at first birth, Contraceptive use and intention, Body Mass Index, Son or daughter died, Access to Media, Sex of household head, Ideal number of children, Currently breastfeeding, Desire for more children,

Husband's desire for children out of 23 variables are determined most influential factors to predict the fertility in Bangladesh.

FIGURE 1: FEATURE SELECTION USING BORUTA ALGORITHM



PREDICTING FERTILITY USING MACHINE LEARNING APPROACHES

The machine learning models namely Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), XGBoost, LightGBM and Neural Network (NN) were utilized to build a model for prediction of fertility in Bangladesh. After training each predictive model with 80% of the data then testing was conducted with the remaining 20% of the dataset.

The Support Vector Machine (SVM) achieved the highest fertility prediction accuracy (95.55%), followed by LightGBM (94.91%) and Random Forest (94.85%), with strong weighted average metrics in recall, precision, and F1 score (all above 95%). Logistic Regression and Decision Tree also performed well, with accuracies of 89.04% and 88.16% respectively, while K-Nearest Neighbors (76.06%) showed the lowest accuracy. Across models, SVM, LightGBM, and Random Forest had the least misclassifications in predicting the number of children ever born (table-3).

TABLE-3: PREDICTIVE MODELS OF PERFORMANCE OF FERTILITY

Predictive Models Performances								
	RF		**DT**		**KNN**		**LR**	
Accuracy	94.85%		88.16%		76.06%		89.04%	
	Macro Average	Weighted Average	Macro Average	Weighted Average	Macro Average	Weighted Average	Macro Average	Weighted Average
Recall	88%	95%	79%	89%	66%	78%	81%	90%
Precision	88%	95%	82%	88%	69%	76%	82%	89%
F1 score	88%	95%	80%	89%	67%	76%	80%	89%

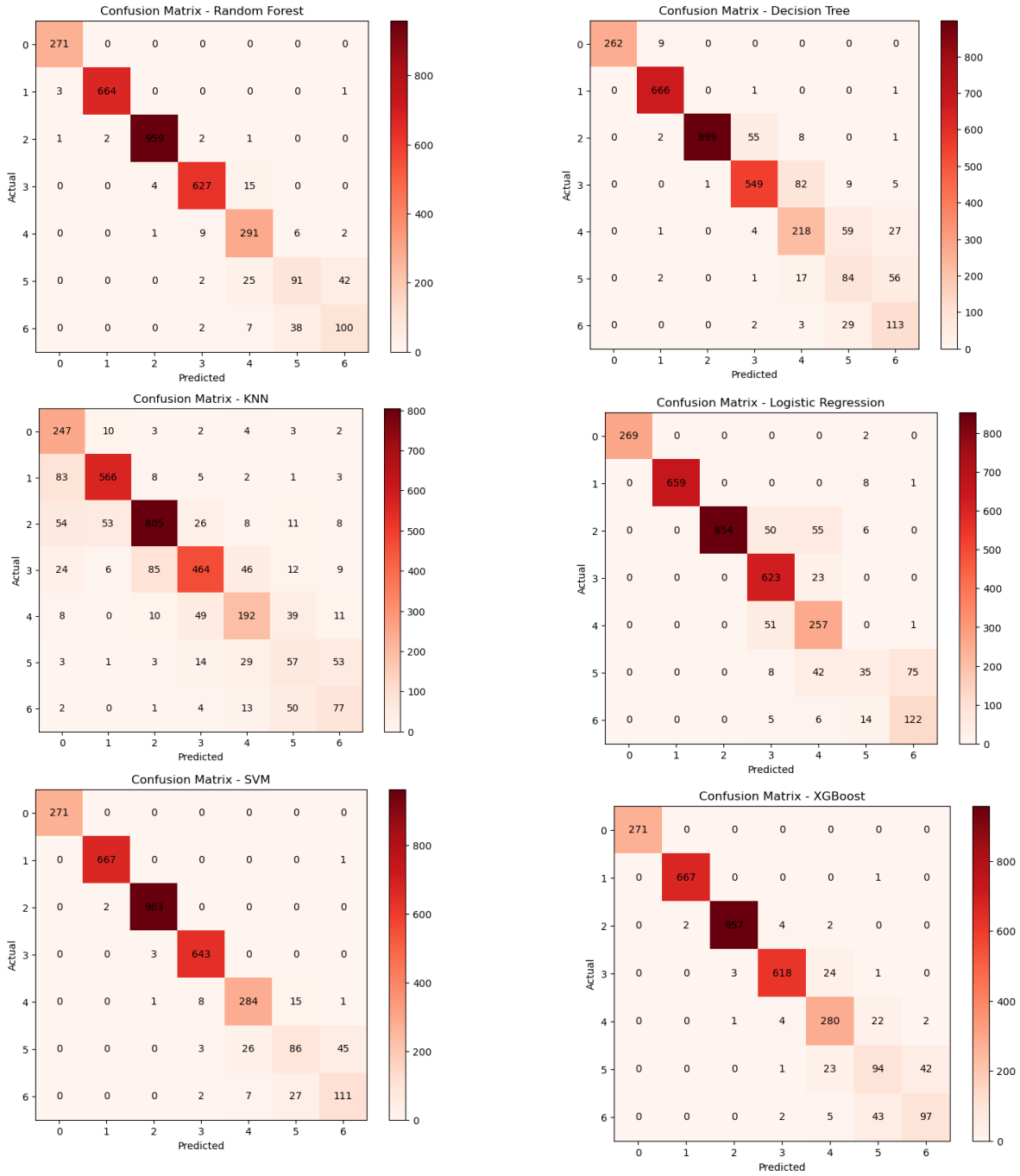
	SVM		**XGBoost**		**LightGBM**		**NN**	
Accuracy	95.55%		94.25%		94.91%		92.74%	
	Macro Average	Weighted Average	Macro Average	Weighted Average	Macro Average	Weighted Average	Macro Average	Weighted Average
Recall	89%	95%	87%	94%	88%	95%	85%	93%
Precision	89%	96%	87%	94%	88%	95%	86%	93%
F1 score	89%	95%	87%	94%	88%	95%	86%	93%

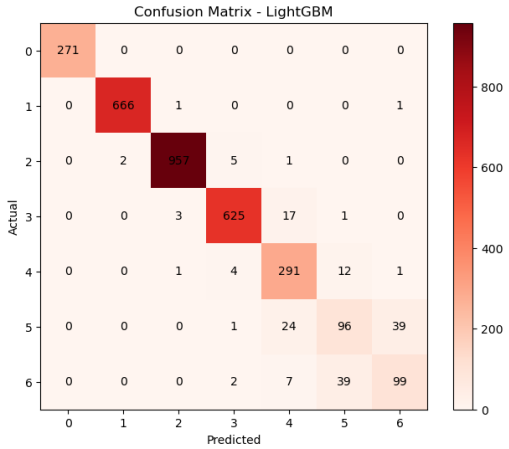
Models with higher recall might be preferred if correctly identifying positive cases (fertility cases in this context) is of utmost importance. Models with higher precision are crucial when reducing false positive cases is a priority. The weighted averages provide insights considering class imbalances, while macro averages treat all classes equally. This analysis

suggests that while all models perform reasonably well the choice among them might be based on the specific needs computational complexity and the importance of correctly identifying fertility cases versus minimizing false positives. Figure-2 presents the confusion matrix and Figure-3 presents AUROC of all models comparative performances. According to the findings presented in Table-3 the predictive model performance results indicate that the Support Vector Machine (SVM) outperformed other models utilized in this study as the most effective predictor of fertility.

Figure 2 presents confusion matrices for eight machine learning models, showing their performance in predicting fertility determinants. The Random Forest, Neural Network, and SVM models perform best, with more correct predictions (darker diagonal lines), while models like KNN and LightGBM show higher misclassification rates.

FIGURE-2: CONFUSION MATRIX OF CLASSIFICATION RESULTS.





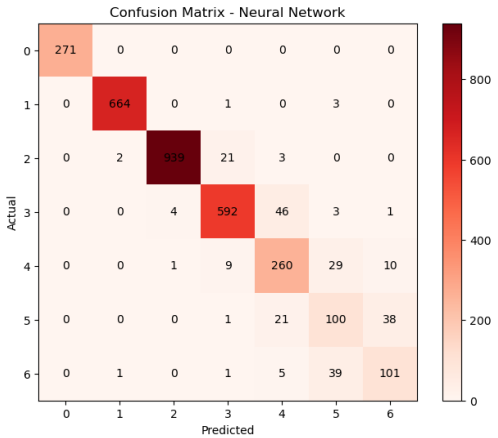
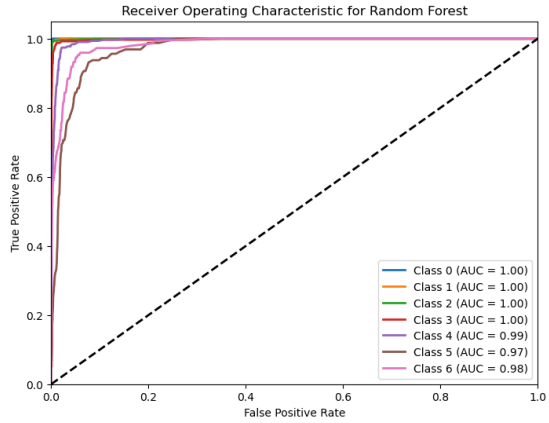
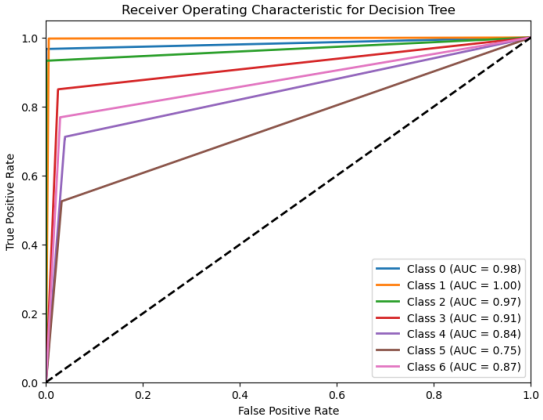
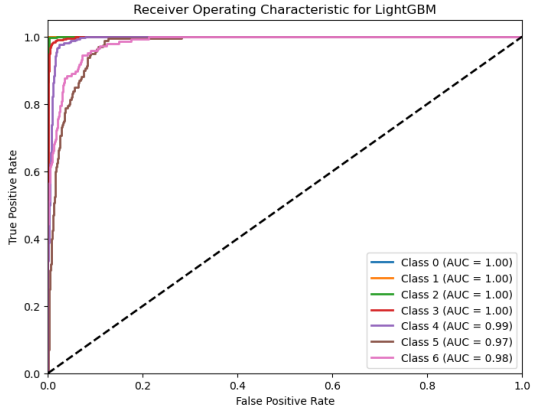
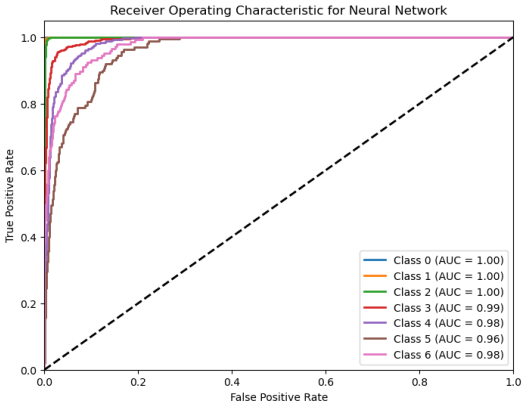


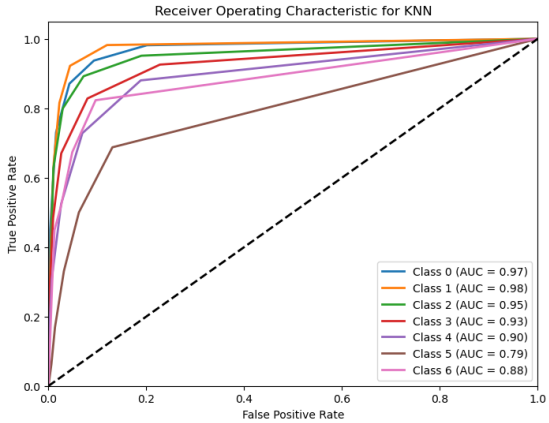
FIGURE-3: THE ROC CURVES TO PREDICT FERTILITY IN BANGLADESH USING RF, DT, KNN, LR, SVM, XGBOOST, LIGHTGBM AND NN MODELS

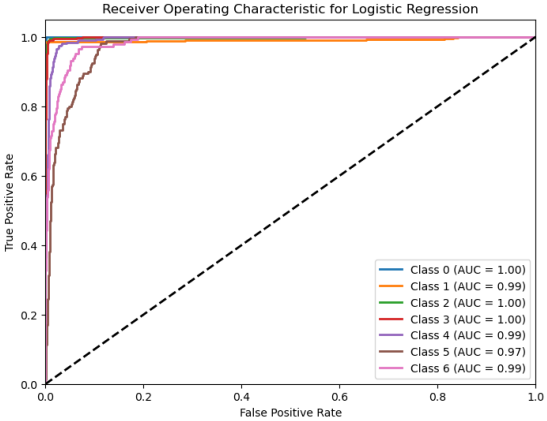


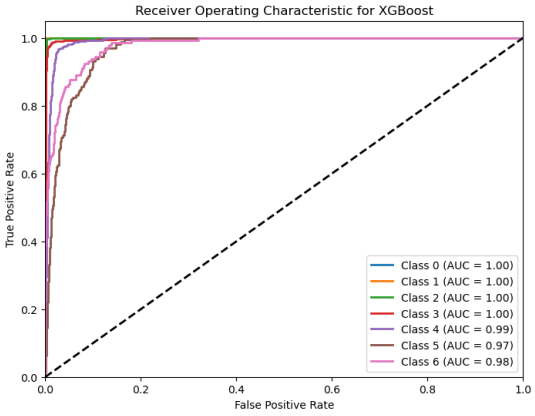


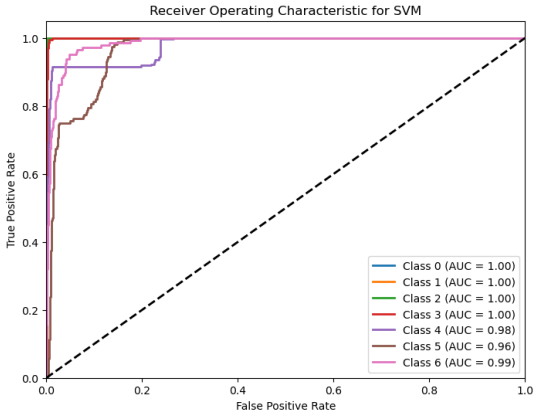












SENSITIVITY ANALYSIS

To assess the robustness of the predictive models, a sensitivity analysis was conducted by comparing model performance using the original dataset and a dataset balanced using the Synthetic Minority Oversampling Technique (SMOTE). The results presented in Table 4 indicate that model performance remained largely consistent across both scenarios. In the original dataset, the Support Vector Machine (SVM) achieved the highest accuracy (96.08%), followed by Random Forest (96.02%) and LightGBM (95.67%). After applying SMOTE, SVM continued to demonstrate the strongest performance with an accuracy of 95.55%, followed by LightGBM (94.91%) and Random Forest (94.85%). Although SMOTE slightly improved recall for some minority fertility categories, overall accuracy and weighted performance metrics showed only minor changes across models, suggesting that the predictive results are robust to class imbalance.

TABLE-4 SENSITIVITY ANALYSIS RESULTS OF MACHINE LEARNING MODELS

Models	Without SMOTE	With SMOTE
RF	96.02%	94.85%
DT	94.15%	88.16%
KNN	80.20%	76.06%
LR	89.80%	89.04%
SVM	**96.08%**	**95.55%**
XGBoost	95.48%	94.25%
LightGBM	95.67%	94.91%
NN	95.17%	92.45%

K-FOLD CROSS-VALIDATION RESULTS

We performed k-fold cross-validation with 5-fold, 10-fold, 15-fold, 20-fold, and 30-fold splits to evaluate the robustness of the machine learning models. Table 5 presents the mean accuracy scores for each model across these fold values. After testing various fold values, k=10 emerged as the most consistent and reliable across most models, and was therefore chosen for the final evaluation in this study.

Among the models, the Support Vector Machine (SVM) consistently delivered the highest accuracy of 95.90% across all fold values, demonstrating strong robustness in predicting fertility determinants. Logistic Regression (LR), a widely recognized model in fertility prediction, also performed well, achieving 95.72% accuracy under 10-fold cross-validation. Given its established reputation, Logistic Regression was used as the benchmark ("gold standard") for comparing the performance of more advanced machine learning models.

Random Forest (RF), LightGBM, and XGBoost all achieved strong results, with mean accuracies exceeding 95%. Models like Decision Tree (DT) and Neural Networks (NN) performed slightly lower, with accuracies around 94%. However, K-

Nearest Neighbors (KNN) lagged significantly, with an accuracy of around 57%, indicating it may not be suitable for this dataset.

TABLE-5: K-FOLD CROSS VALIDATION OF MACHINE LEARNING MODELS

Models	Mean Accuracy of K-Fold				
	5-Fold	10-Fold	15-Fold	20-Fold	30-Fold
RF	0.9496	0.9512	0.9503	0.9496	0.9511
DT	0.9348	0.9375	0.9363	0.9354	0.9349
KNN	0.5647	0.5709	0.5707	0.5727	0.5737
LR	0.9577	0.9572	0.9579	0.9574	0.9574
SVM	0.9590	0.9590	0.9590	0.9590	0.9590
XG Boost	0.9519	0.9534	0.9526	0.9533	0.9537
Light GBM	0.9529	0.9535	0.9534	0.9536	0.9545
NN	0.9351	0.9380	0.9375	0.9352	0.9373

In summary, SVM provided the best predictive performance for fertility determinants in this dataset, with Logistic Regression serving as a solid baseline. The selection of k=10 for cross-validation was supported by the stability and consistent performance of the models.

FIGURE-4 IMPORTANT FEATURES VISUALIZATION USING ML SVM CLASSIFIER

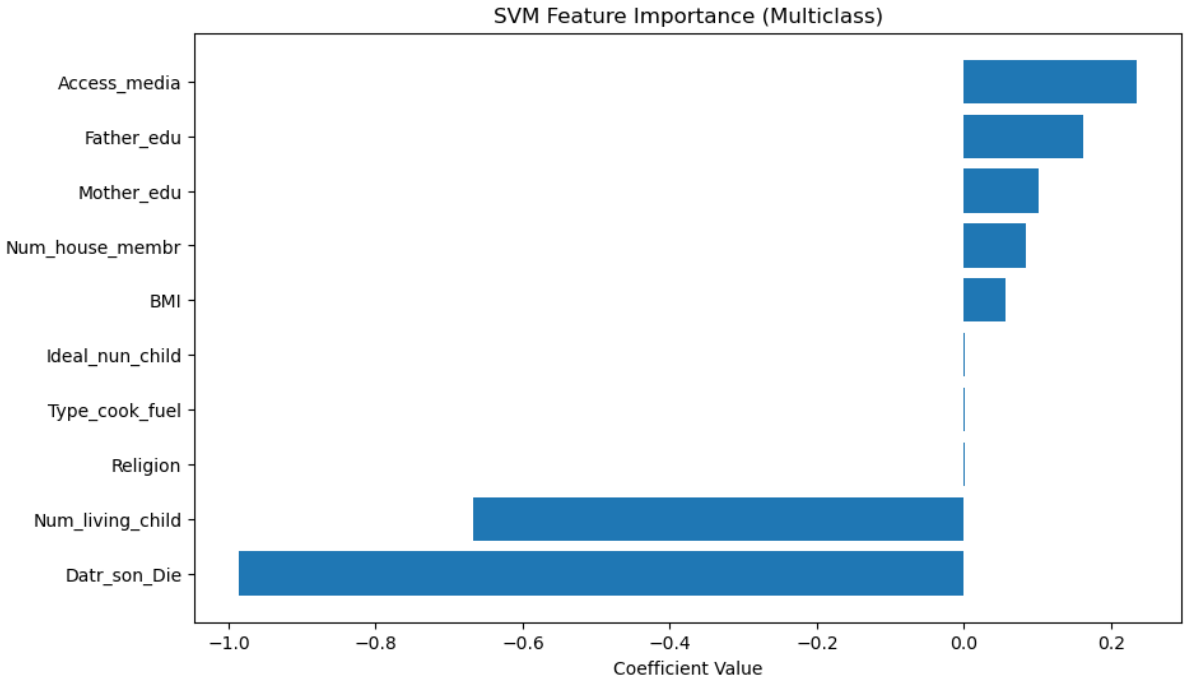


Figure-4 presents a visualization of the most influential features identified through the SVM (Support Vector Machine) classifier in the context of predicting fertility-related outcomes in Bangladesh. These features are crucial in determining the predictions made by the machine learning model. There are several factors such as access to mass media, parents' education levels, household size, Body Mass Index (BMI), number of living children, and experiences of child mortality have been found to influence the model's predictions, this does not imply a direct causal relationship between these factors and fertility. These features are significant within the confines of the SVM algorithm, but further investigation is required to understand their role in the broader socio-economic and cultural determinants of fertility.

DISCUSSION

This study makes a significant contribution to fertility research in Bangladesh by combining traditional statistical methods with advanced machine learning techniques to identify key determinants of fertility. While traditional chi-square tests revealed several important factors such as Division, Type of Residence, Religion, Wealth Index, and Parental Education, machine learning allowed for a deeper analysis of their relative importance and provided a more focused perspective on the most critical predictors.

At the 5% level of confidence, traditional chi-square tests confirmed that determinants like

Division, Type of residence, Religion, Wealth index, Maternal and Father's education, Occupation, Toilet facilities, Water sources, Household size, Cooking fuel, Mother's Age, Body Mass Index, Age at first Marriage, Contraceptive use, Son or daughter death, Access to media, Marital status, Ideal children number, and Desire for more children, Father's desire for children and Sex of household head play a statistically significant role in fertility. These factors align with existing literature [5], [9], [12]–[14], [42], but the application of machine learning allowed us to assess their relative importance compared to other variables. The identification of media access as a significant predictor emphasizes the role of mass communication in shaping reproductive behaviors [9]. This insight opens new opportunities for policy interventions that leverage media-based education especially in promoting family planning.

Moreover, the importance of parental education, household size and maternal BMI underscores the multifaceted nature of fertility decisions. Larger households as our analysis suggests, tend to have higher fertility rates and educational attainment remains a crucial factor [15], [30], [43]–[45]. The positive correlation between maternal BMI and fertility also aligns with broader health-related findings, indicating that nutritional and health status plays a key role in reproductive outcomes [46], [47] .

While conventional approaches such as the chi-square test revealed a wide range of significant factors including Division, Type of Place of Residence, Religion, Wealth Index and Parental Education and many more, on the other hand the Support Vector Machine (SVM) offered a refined and focused most influential predictors. Figure 2 illustrates the confusion matrices for the various machine learning models used in this study, providing a comparative view of their classification performance. Notably, the SVM model outperformed others, showing higher accuracy with fewer misclassifications, as seen by the concentration of predictions along the diagonal.

The Support Vector Machine (SVM) model has identified seven essential determinants: access to mass media, father's education, mother's education, number of household members, maternal Body Mass Index (BMI), number of living children and child mortality events (i.e., incidents of sons or daughters dying). These findings not only confirm prior research but also offer new insights into the specific factors influencing fertility dynamics in Bangladesh.

Moving forward, there are several avenues for future research incorporating real-time data from mobile health apps and wearable devices could enhance fertility predictions. Social media sentiment analysis may reveal shifts in attitudes toward family planning. Incorporating economic indicators like employment and income levels could broaden the understanding of fertility trends. Data from healthcare systems and electronic records can improve model accuracy and personalization. Advancements in AI will further refine models, offering actionable insights for policy interventions.

In conclusion, our study demonstrates the powerful potential of machine learning in fertility research. This study highlights key factors like media access, parental education, and maternal BMI, number of living children, and child mortality events offering valuable insights that can help policymakers and health professionals make informed decisions. These findings suggest that targeted interventions focusing on media-based education and improving healthcare access could significantly impact fertility rates in Bangladesh. Future research should continue exploring advanced machine learning

models and longitudinal datasets to further refine our understanding of fertility determinants and inform data-driven policymaking.

STRENGTHS AND LIMITATIONS

The country representative BDHS-2017-18 dataset were used in this study. Causal inference is not feasible as this analysis is cross-sectional. The results we examined rely on self-reporting and are thus vulnerable to variation in memory and social desirability. We have used 08 (eight) different types of Machine Learning models: Random Forests (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), XGBoost, LightGBM and Neural Network (NN) for examining and seeking the best output.

Another limitation of this study is the absence of external validation using an independent dataset. Although internal validation was performed through k-fold cross-validation to assess the robustness of the machine learning models, the models were not tested on an external dataset from a different survey wave or population. Future studies should consider validating the proposed models using independent datasets, such as more recent Bangladesh Demographic and Health Survey (BDHS) rounds, to further assess the generalizability and stability of the predictive models across different populations and time periods.

CONCLUSION

This study demonstrates that Machine Learning (ML) models, particularly Support Vector Machines (SVM), provide more reliable and precise predictions of fertility determinants in Bangladesh compared to traditional statistical models. By leveraging training data for model construction and test data for validation, ML techniques enabled the identification of key fertility predictors with greater accuracy. Our findings reveal that access to mass media, father's education level, mother's education, household size, maternal Body Mass Index (BMI), number of living children, and child mortality events (i.e., incidents of sons or daughters dying) are the most influential factors in predicting fertility.

The integration of ML into fertility research represents a significant methodological advancement, allowing for the handling of complex, non-linear relationships between variables. This approach enhances the understanding of fertility dynamics and provides insights that may be overlooked by conventional statistical methods. These findings offer valuable evidence for policymakers to design targeted interventions to address fertility issues in Bangladesh.

Future research should explore more advanced ensemble ML models or utilize longitudinal data to capture temporal changes in fertility determinants. Expanding the scope of study to other regions or countries could also offer comparative insights into global fertility patterns.

DECLARATIONS

Acknowledgements

The authors express their gratitude to the MEASURE DHS project for their support and for providing free access to the original data.

Ethics approval and consent to participate

No ethical approval for this study because this analysis is secondary and research is based on BDHS (2017-18) data. The data were completely anonymous and there is no identifiable information was there and available DHS website.

Funding

The study does not receive any funding.

Consent for publication

Consent for publication is not required for this study, as no personal information, images, or videos of individual participants were used. Furthermore, all the data utilized in the study are publicly available.

Availability of data and materials

Sourced from the publicly available dataset at https://dhsprogram.com/data/available-datasets.cfm the manuscript incorporates secondary data.

Competing interests

The authors declared no competing interests exist.

References:

1. C. J. A. Bradshaw, C. Perry, M. A. Judge, C. M. Saraswati, J. Heyworth, and P. N. Le Souëf, "Lower infant mortality, higher household size, and more access to contraception reduce fertility in low- and middle-income nations," PLoS One, vol. 18, no. 2 February, pp. 1–16, 2023, doi: 10.1371/journal.pone.0280260.
2. N. V. Bhattacharjee et al., "Global fertility in 204 countries and territories, 1950–2021, with forecasts to 2100: a comprehensive demographic analysis for the Global Burden of Disease Study 2021," Lancet, vol. 403, no. 10440, pp. 2057–2099, 2024, doi: 10.1016/S0140-6736(24)00550-6.
3. S. Islam and M. K. Nesa, "Fertility transition in Bangladesh: The role of education," Proc. Pakistan Acad. Sci., vol. 46, no. 4, pp. 195–201, 2009.
4. National Institute of Population Research and Training (NIPORT); ICF International, "Bangladesh Demographic and Health Survey 2017-18," Dhaka, Bangladesh, Rockville, Maryland, USA NIPORT ICF, pp. 1–511, 2019.
5. M. B. Asresie and G. W. Dagnew, "Association of maternal high-risk fertility behavior and under-five mortality in Ethiopia: Community-based survey," PLoS One, vol. 17, no. 5 May, pp. 1–11, 2022, doi: 10.1371/journal.pone.0267802.
6. [6] R. Singh and V. Tripathi, "Maternal factors contributing to under-five mortality at birth order 1 to 5 in India: A comprehensive multivariate study," Springerplus, vol. 2, no. 1, pp. 1–12, 2013, doi: 10.1186/2193-1801-2-284.
7. D. G. Ayele, "Determinants of fertility in Ethiopia," Afr. Health Sci., vol. 15, no. 2, pp. 546–551, 2015, doi: 10.4314/ahs.v15i2.29.
8. I. Haq, "Relationship Between Age at Marriage, Education and Fertility Among Residence of Bangladesh," Am. J. Soc. Sci. Res., vol. 4, no. 2, pp. 33–39, 2018, [Online]. Available: http://www.aiscience.org/journal/ajssrhttp://creativecommons.org/licenses/by/4.0/.
9. A. M. F. Rabbi, "Factors influencing fertility preference of a developing country during demographic transition: Evidence from Bangladesh," South East Asia J. Public Heal., vol. 4, no. 2, pp. 23–30, 2015, doi: 10.3329/seajph.v4i2.23691.
10. R. Adhikari, "Demographic, socio-economic, and cultural factors affecting fertility differentials in Nepal," BMC Pregnancy Childbirth, vol. 10, pp. 1–11, 2010, doi: 10.1186/1471-2393-10-19.
11. J. K. Bora, N. Saikia, E. B. Kebede, and W. Lutz, "Revisiting the causes of fertility decline in Bangladesh: the relative importance of female education and family planning programs," Asian Popul. Stud., vol. 19, no. 1, pp. 81–104, 2023, doi: 10.1080/17441730.2022.2028253.
12. S. Abedin and D. Arunachalam, "Maternal autonomy and high-risk pregnancy in Bangladesh: The mediating influences of childbearing practices and antenatal care," BMC Pregnancy Childbirth, vol. 20, no. 1, pp. 1–8, 2020, doi: 10.1186/s12884-020-03260-9.
13. A. C. Church, M. Ibitoye, S. Chettri, and J. B. Casterline, "Traditional supports and contemporary disrupters of high fertility desires in sub-Saharan Africa: a scoping review," Reprod. Health, vol. 20, no. 1, pp. 1–17, 2023, doi: 10.1186/s12978-023-01627-7.
14. Y. Fitaw, Y. Berhane, and A. Worku, "Impact of child mortality and fertility preferences on fertility status in rural Ethiopia," East Afr. Med. J., vol. 81, no. 6, pp. 300–306, 2004, doi: 10.4314/eamj.v81i6.9179.
15. C. Baynes et al., "The Impact of Childhood Mortality on Fertility in Rural Tanzania: Evidence From the Ifakara and Rufiji Health and Demographic Surveillance Systems," Demography, vol. 60, no. 6, pp. 1721–1746, 2023, doi: 10.1215/00703370-11048233.
16. S. Roy and S. M. I. Hossain, "Fertility differential of women in Bangladesh demographic and health survey 2014," Fertil. Res. Pract., vol. 3, no. 1, pp. 3–7, 2017, doi: 10.1186/s40738-017-0043-z.
17. M. R. K. Chowdhury, M. S. Rahman, B. Billah, M. Rashid, M. Almroth, and M. Kader, "Prevalence and factors associated with severe undernutrition among under-5 children in Bangladesh, Pakistan, and Nepal: a comparative study using multilevel analysis," Sci. Rep., vol. 13, no. 1, pp. 1–12, 2023, doi: 10.1038/s41598-023-36048-w.

18. X. Zhu et al., "Prediction models and associated factors on the fertility behaviors of the floating population in China," Front. Public Heal., vol. 10, 2022, doi: 10.3389/fpubh.2022.977103.
19. D. Ganguly, S. Goli, and A. Rammohan, "Explaining the diminishing of Muslim advantage in child survival in India," Genus, vol. 78, no. 1, 2022, doi: 10.1186/s41118-022-00177-3.
20. O. Gruebner et al., "Spatial variations and determinants of infant and under-five mortality in Bangladesh," Heal. Place, vol. 47, no. December 2016, pp. 156–164, 2017, doi: 10.1016/j.healthplace.2017.08.012.
21. M. A. Khan et al., "Trends and projections of under-5 mortality in Bangladesh including the effects of maternal high-risk fertility behaviours and use of healthcare services," PLoS One, vol. 16, no. 2 February 2021, pp. 1–17, 2021, doi: 10.1371/journal.pone.0246210.
22. M. R. Sajid, F. Maqsood, and M. Rani, "Determinants of Fertility : A Neural Network Approach," vol. 8, no. 2, pp. 440–449, 2011.
23. M. M. Islam, M. A. Islam, and N. Chakroborty, "Fertility transition in Bangladesh: Understanding the role of the proximate determinants," J. Biosoc. Sci., vol. 36, no. 3, pp. 351–369, 2004, doi: 10.1017/S0021932003006333.
24. M. H. Howlader, H. O. Roshid, S. Kundu, H. R. Halder, S. K. Chanda, and M. A. Rahman, "Determinants associated with high-risk fertility behaviours among reproductive aged women in Bangladesh: a cross-sectional study," Reprod. Health, vol. 19, no. 1, pp. 1–10, 2022, doi: 10.1186/s12978-022-01333-w.
25. S. S. Panesar, R. N. D'Souza, F. C. Yeh, and J. C. Fernandez-Miranda, "Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database," World Neurosurg. X, vol. 2, p. 100012, 2019, doi: 10.1016/j.wnsx.2019.100012.
26. A. E. Adegbosin, B. Stantic, and J. Sun, "Efficacy of deep learning methods for predicting under-five mortality in 34 low-income and middle-income countries," BMJ Open, vol. 10, no. 8, pp. 1–8, 2020, doi: 10.1136/bmjopen-2019-034524.
27. F. H. Bitew, S. H. Nyarko, L. Potter, and C. S. Sparks, "Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian Demographic and Health Survey," Genus, vol. 76, no. 1, 2020, doi: 10.1186/s41118-020-00106-2.
28. C. Mangold, S. Zoretic, K. Thallapureddy, A. Moreira, K. Chorath, and A. Moreira, "Machine learning models for predicting neonatal mortality: A systematic review," Neonatology, vol. 118, no. 4, pp. 394–405, 2021, doi: 10.1159/000516891.
29. M. Mansur, A. Afiaz, and M. S. Hossain, "Sociodemographic risk factors of under-five stunting in Bangladesh: Assessing the role of interactions using a machine learning method," PLoS One, vol. 16, no. 8 August, pp. 1–17, 2021, doi: 10.1371/journal.pone.0256729.
30. T. Dendup, Y. Zhao, and D. Dema, "Factors associated with under-five mortality in Bhutan: An analysis of the Bhutan National Health Survey 2012," BMC Public Health, vol. 18, no. 1, pp. 1–15, 2018, doi: 10.1186/s12889-018-6308-6.
31. R. K. Saroj, P. K. Yadav, R. Singh, and O. N. Chilyabanyama, "Machine Learning Algorithms for understanding the determinants of under-five Mortality," BioData Min., vol. 15, no. 1, pp. 1–22, 2022, doi: 10.1186/s13040-022-00308-8.
32. P. Gulati, A. Sharma, and M. Gupta, "Theoretical Study of Decision Tree Algorithms to Identify Pivotal Factors for Performance Improvement: A Review," Int. J. Comput. Appl., vol. 141, no. 14, pp. 19–25, 2016, doi: 10.5120/ijca2016909926.
33. M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," Stata J., vol. 20, no. 1, pp. 3–29, 2020, doi: 10.1177/1536867X20909688.
34. S. Huang, C. A. I. Nianguang, P. Penzuti Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of support vector machine (SVM) learning in cancer genomics," Cancer Genomics and Proteomics, vol. 15, no. 1, pp. 41–51, 2018, doi: 10.21873/cgp.20063.
35. M. Maalouf, "Logistic regression in data analysis: An overview," Int. J. Data Anal. Tech. Strateg., vol. 3, no. 3, pp. 281–299, 2011, doi: 10.1504/IJDATS.2011.041335.
36. K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," 2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019, no. January 2021, pp. 1255–1260, 2019, doi: 10.1109/ICCS45141.2019.9065747.

37. M. Noorunnahar, A. H. Chowdhury, and F. A. Mila, "A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh," PLoS One, vol. 18, no. 3 March, pp. 1–15, 2023, doi: 10.1371/journal.pone.0283452.
38. M. R. Machado, S. Karray, and I. T. De Sousa, "LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry," 14th Int. Conf. Comput. Sci. Educ. ICCSE 2019, no. Nips, pp. 1111–1116, 2019, doi: 10.1109/ICCSE.2019.8845529.
39. H. Chiroma et al., "Progress on Artificial Neural Networks for Big Data Analytics: A Survey," IEEE Access, vol. 7, pp. 70535–70551, 2019, doi: 10.1109/ACCESS.2018.2880694.
40. S. Yang and G. Berdine, "The receiver operating characteristic (ROC) curve," Southwest Respir. Crit. Care Chronicles, vol. 5, no. 19, p. 34, 2017, doi: 10.12746/swrccc.v5i19.391.
41. M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," J. Stat. Softw., vol. 36, no. 11, pp. 1–13, 2010, doi: 10.18637/jss.v036.i11.
42. M. Chen1, S. M. A. Haq, K. J. Ahmed, A. H. M. B. Hussain, and M. N. Q. Ahmed, "The link between climate change, food security and fertility: The case of Bangladesh," PLoS One, vol. 16, no. 10 October, pp. 1–18, 2021, doi: 10.1371/journal.pone.0258196.
43. M. S. Rahman, M. M. Rahman, S. Gilmour, K. T. Swe, S. Krull Abe, and K. Shibuya, "Trends in, and projections of, indicators of universal health coverage in Bangladesh, 1995–2030: a Bayesian analysis of population-based household data," Lancet Glob. Heal., vol. 6, no. 1, pp. e84–e94, 2018, doi: 10.1016/S2214-109X(17)30413-8.
44. J. M. K. Aheto, "Predictive model and determinants of under-five child mortality: Evidence from the 2014 Ghana demographic and health survey," BMC Public Health, vol. 19, no. 1, pp. 1–10, 2019, doi: 10.1186/s12889-019-6390-4.
45. A. Razzaque et al., "Levels, trends and socio-demographic determinants of infant and under-five mortalities in and around slum areas of Dhaka city, Bangladesh," SSM - Popul. Heal., vol. 17, no. September 2021, p. 101033, 2022, doi: 10.1016/j.ssmph.2022.101033.
46. J. Lai et al., "Association between Waist-Hip Ratio and Female Infertility in the United States: Data from National Health and Nutrition Examination Survey 2017-2020," Obes. Facts, no. February, pp. 1821–1831, 2024, doi: 10.1159/000538974.
47. J. R. Khan and N. Awan, "A comprehensive analysis on child mortality and its determinants in Bangladesh using frailty models," Arch. Public Heal., vol. 75, no. 1, pp. 1–10, 2017, doi: 10.1186/s13690-017-0224-6.