A Comparative Study of Feature Selection Techniques in Predicting School Attrition Among Female Students in Bangladesh

Mohammad Nayeem Hasan   
*Department of Statistics*  
*Shahjalal University of Science and Technology*  
Sylhet, Bangladesh  
[nayeem5847@gmail.com](mailto:nayeem5847@gmail.com)

Shomi Khan  
*Department of Electrical and Electronic Engineering*  
*Shahjalal University of Science and Technology*Sylhet, Bangladesh  
[nkskl6@gmail.com](mailto:nkskl6@gmail.com)

Faizul Islam  
*Department of Statistics*  
*Shahjalal University of Science and Technology*  
Sylhet, Bangladesh  
[faizulislam095@gmail.com](mailto:faizulislam095@gmail.com)

Sabrin Sultana  
*Department of Banking and Insurance*  
*University of Chittagong*Chittagong, Bangladesh  
[sabrinsultana1060@gmail.com](mailto:sabrinsultana1060@gmail.com)

*Abstract*— In real world scenarios, such as school attrition, it is important to consider relevant features when applying machine learning algorithms. However, not all features are essential, and including unnecessary or irrelevant ones can diminish the effectiveness of the algorithm. Feature selection (FS) is a crucial task in machine learning to reduce data dimensionality and improve algorithm performance. Evolutionary algorithms (EA) have gained significant attention as a means to solve FS problems. Surprisingly, no previous research has applied these methods specifically to predict factors related to school attrition. Therefore, the objective of this study is to employ various feature selection techniques (Genetic Algorithm, Chi-squared, Mutual information, Exhaustive feature selection) and compare their performance. Data for this study were collected from the Multiple Indicator Cluster Surveys, 2019. Among the methods tested, artificial neural network (ANN) demonstrated the highest accuracy in predicting school attrition. With the use of the GA, Chi-squared, MI, and EFS techniques, the ANN model's accuracy was 76.50%, 76.67%, 76.15%, and 76.98%, respectively. Using the same feature selection methods, SVM produced results of 76.00%, 74.06%, 75.42%, and 75.73% whereas LR produced accuracy numbers of 75.83%, 75.94%, 76.25%, and 76.25%. The findings indicate that all models performed well in classifying students as in-school or out-of-school. Thus, tailoring filter measures based on specific FS technique characteristics can significantly enhance efficiency and effectiveness.

Keywords—feature selection, evolutionary algorithms, genetic algorithm, machine learning, classification

# Introduction

Girls play a vital role in fostering a nation's progress as they lay the foundation for education. Educated women tend to have higher incomes, are more confident in their decision-making, and raise healthier and better-educated children, surpassing their male counterparts in these aspects [1]. In numerous low-income countries at present, there remains a substantial proportion of females who face school dropouts. Over 50% of girls commence their primary education but are unable to successfully complete it [2]. A student's family poverty level exerts a significant influence on their likelihood of experiencing attrition in education. Among the factors that contribute to high attrition rates in secondary schools, school-to-school transitions play a prominent role. The academic and social environment of the school directly impacts institutional attrition rates [3], [4].

The process of selecting a subset of features in feature selection (FS), which aims to reduce data volume and enhance the efficiency of classification algorithms, poses a challenging task [5]. Reducing the amount of data, particularly by selecting a limited number of relevant features, contributes to the development of improved models characterized by higher accuracy. In the domains of machine learning (ML), data mining, and statistics, evolutionary computations (ECs) provide effective solutions for optimizing problems by emulating certain aspects of biological evolution. Evolutionary algorithms (EA) adopt the strategies derived from ECs, benefiting from the inspiration drawn from biological evolution. Notable optimization techniques employed by EA include principal component (PC) analysis, particle swarm (PS) optimization, ant colony (AC) optimization, and genetic algorithm (GA) optimization [6]. Filter methods have the ability to assess individual features or evaluating entire feature subsets where the constructed measures for feature filtering can be broadly categorized into statistical measures, information, consistency, distance, and similarity [7]. As filtering techniques, we implemented mutual information (MI) and Chi-squared test. Wrapper method assess a subset of features based on their predictive power using a learning algorithm and an approach to search to find the best values (based on specified criteria) for certain parameters [8]. We have used Exhaustive Feature Selection (EFS) technique of wrapper method.

# Literature Review

A considerable body of research has utilized genetic algorithms (GA) to effectively classify models across various domains. For instance, Barrios et al. [9] utilized GA to train a network and developed a classification technique for breast cancer classification. Another study focused on utilizing GA for feature selection in support vector machines (SVM) to classify defective and normal instances [10]. GA has also been applied by Beiko and Charlebois [11] to identify optimal outputs for artificial neural networks (ANN) in the classification of DNA sequences. Similarly, Karzynski et al. [12] employed GA to reduce data for weight classification in ANN for microarray classifications. Additionally, Cho et al. [13] applied GA to enhance the prediction and classification of cDNA microarrays using ANN outputs. Kaushalya and Gapar has applied different feature selection methods (filter, wrapper, embedded) for heart disease prediction on classification algorithm [14]. For the prediction of risk in hepatitis disease Pinar has also used filter-based feature selection method. [15]

# Methods

## **Study Design**

This study utilized data from a cross-sectional survey, specifically the Multiple Indicator Cluster Surveys conducted in 2019 (source: https://mics.unicef.org/surveys). The response variable of interest is school attrition, categorized as "Yes" or "No." The explanatory variables used in this study are consistent with those included in previously published research [16], [17].

## **Data Preparation**

The entire dataset was evenly divided into two subsets: the training data and the validation (testing) data. The validation data played a crucial role in identifying any issues that might arise in the classification models, such as overfitting or premature termination of the training process. To address potential outliers in the data matrix, this study employed the standardization of Z-scores.

# Feature Selection Methods

Feature selection (FS) holds significant importance not only in large datasets but also in dealing with various challenges. One such challenge arises from frequent feature interactions observed in many cases. Features can exhibit diverse patterns, ranging from general characteristics to complex multi-way interactions [18]. Furthermore, an appropriate feature can effectively serve multiple purposes concurrently when utilized in parallel. However, the primary objective of FS is to identify the key terms or features [19], facilitate the creation of graphical representations in image analysis [20], and determine crucial business indicators [21].

## **Genetic Algorithm**

The concept of genetic algorithms (GA) was initially proposed by [22] as a technique for addressing a wide range of optimization problems [23]. This method's design draws inspiration from the observation of natural Darwinian evolutionary processes and the principle of survival of the fittest [24]. GA involves the implementation of several parameters, including mutation probability, crossover probability, and their interrelationships [25]. In order to comprehend the underlying theory of genetic algorithms, it is important to grasp the process of recombination. Recombination involves two parents and can be illustrated by considering a binary string, such as 1101001100101101, and another binary string represented as,

yxyyxyxxyyyxyxxy, where x and y represent 0 and 1.

GA can be effectively utilized for training purposes in logistic regression (LR), artificial neural networks (ANN), and support vector machines (SVM), as well as for determining optimal model structures. Notably, GA does not rely on gradient-based error functions, making it less susceptible to the problem of local minima [26]. The step-by-step process of GA is depicted in Figure 1. Initially, a set of randomly generated codes is created and divided into an initial population. Subsequently, fitness values are assigned to each code by evaluating the fitness function. This is followed by performing a crossover, which involves exchanging features among the selected subset, and introducing mutations that randomly modify selected features. The process then loops back to the second step and continues until the exit criteria are met.

|  |
| --- |
|  |
| Figure. 1. Steps of genetic algorithm  B**. Filter methods**  Without using classification algorithms, filter methods assess features based on the properties of the data. Information theory, correlation, consistency, distance, fuzzy-set, and rough-set can all be used as bases for filter measures. First, features are chosen and ordered in the multivariate case in a batch (naturally handling redundancy) and in the univariate case independently of feature space. In the second step, the features with the highest ranking are selected based on an evaluation criterion. IG (Information gain), CHI, multiple scores aggregation, and Fisher score are a few prominent univariate ranking techniques; multivariate techniques include minimum redundancy maximum relevance and Relief [27]. Mutual Information (MI) is a measure of feature relevance and redundancy-based feature selection techniques [28]. When determining the relationship between two categorical outcome features in feature selection, a Chi-square test is often used. It allows you to verify whether the features are independent of each other. When two features are independent and the outcome count is close to the expected value, we have a minimum chi-square value [29]. Mutual information (MI) in information theory can be used to assess any arbitrary dependence between random variables. The MI between two random variables, X and Y, represents the knowledge that X has provided about Y (or, conversely, the knowledge that Y has provided about X). If X and Y are independent, meaning that neither one knows anything about the other, then their mutual information is zero [30]. Mutual information (MI) is defined as,  **I (X; Y) = H(X) - H(X|Y)**  Where, H(X) is the marginal entropy, H(X|Y) is the conditional entropy and I(X; Y) is the joint entropy of X and Y [31] C. **Wrapper method:** Inductive algorithms are used for selecting feature subsets. The accuracy of the training model is estimated by this selected feature subset. The approach determines whether to add or remove a feature from the selected subset based on the accuracy measured in the previous step [32]. In order to choose the best set, wrapper methods search the dataset. The feature subsets are assessed using a prediction model, which then calculates a score depending on the classifier accuracy. For large dimensional datasets, the approach is computationally costly. A Exhaustive search of the 2n feature subset is carried out by the wrapper method [33]. The wrapper method searches the space of potential parameters. A state space, a starting state, a termination condition, and a search engine are necessary for a search [34]. |

## D. **Prediction Models**

1. **Artificial Neural Network (ANN):** ANN presents an alternative approach for comprehending and identifying intricate patterns within datasets. It can be viewed as an iterative process that converges towards solutions for various classification problems.
2. **Support Vector Machine (SVM):** SVM is a method specifically designed to handle noisy patterns within large datasets. This technique involves finding a function in a multidimensional space that effectively separates the training data based on known class labels.
3. **Logistic Regression (LR):** LR is utilized to analyze binary response data, where the outcome variable Y is represented by either 1 ("success") or 0 ("failure"). The logistic model for a variable Y is defined as

(1)

## E. **Fitness Evaluation**

The fitness values are evaluated based on the accuracy metric, which is calculated using the formula:

Here, TP refers to true positives, FP represents false positives, TN indicates true negatives, and FN corresponds to false negatives. True positives (TP) are instances that are correctly classified as positive. False positives (FP) are instances that are incorrectly classified as positive. True negatives (TN) are instances that are correctly classified as negative, while false negatives (FN) are instances that are incorrectly classified as negative.

# Results and Discussion

A subsample of 4800 girls was taken for the analysis, comprising different age groups. Among them, 15-year-olds accounted for 36.40%, 16-year-olds for 34.10%, and 17-year-olds for 29.50%. Within the group of 15-year-olds, 7.20% were out of school, while 29.20% were attending school. Similarly, the percentages of out-of-school girls for the 16 and 17-year-old groups were 10.60% and 11.30%, respectively.

In this study, the GA utilized single-point crossover, a mutation threshold of 0.55, a maximum iteration of 10, a population size of 40, and a number of offspring of 40 in each iteration. Previous research by T. Liu et al. employed a population size of 20 and stopped all algorithms after 200 iterations [35]. The parameters used by the GA for all models can be found in Table I.

1. Parameters Used in GA

| GA Parameter | Value |
| --- | --- |
| Crossover | Single point crossover |
| Mutation threshold | 0.55 |
| Maximum iteration | 10 |
| Number of population (Initial) | 40 |
| Number of offspring in each iteration | 40 |

Considering dropout as the target variable, the features that the Genetic Algorithm (GA) selected are Girl Age, Marital Status, Area, Divisions, Household Wealth Index, Mother Alive, Religion, and Household Education Level. Chi-squared selected the following variables: Household Wealth Index, Divisions, Girl Age, Marital Status, and Household Education Level. The following were highlighted by Mutual Information: Mother Alive, Religion, Marital Status, and Household Wealth Index. Marital Status, Mother Alive, Household Education Level, Divisions, and Household Wealth Index were determined by Exhaustive Feature Selection (EFS). All three models—ANN, SVM, and LR—were taken into account when choosing these features.

1. Classsification Accuracy in different Techniques

| Model | GA | Chi-squared | MI | EFS |
| --- | --- | --- | --- | --- |
| **ANN** | 76.50% | 76.67% | 76.15% | 76.98% |
| **SVM** | 76.00% | 74.06% | 75.42% | 75.73% |
| **LR** | 75.83% | 75.94% | 76.25% | 76.25% |

The ANN model achieved accuracy of 76.50%, 76.67%, 76.15%, and 76.98% with the GA, Chi-squared, MI and EFS techniques, respectively. In terms of accuracy, SVM delivered results of 76.00%, 74.06%, 75.42%, and 75.73%, while LR obtained values of 75.83%, 75.94%, 76.25%, and 76.25% using the identical feature selection techniques. The Artificial Neural Network (ANN) performed better than the other models. ANN continuously provided better accuracy and overall performance in predicting dropouts.

# Conclusion

Using different feature selection techniques such as Genetic Algorithm (GA), Chi-squared, Mutual Information (MI), and Exhaustive Feature Selection (EFS), we built an independent classification system in this study to categorize data on school attrition. The research's feature selection techniques provided insightful information that improved the classification models' overall resilience and accuracy. However, the feature selection techniques consistently produced improved classification accuracy without introducing complexity to the method. An important advantage of these approaches is its controllability, as the models can be fine-tuned to achieve better results by adjusting the hyperparameters.

Ultimately, the selection of the best model methodology depends on the application of the model and the likelihood of the required assumptions being true. In conclusion, when choosing which approach to utilize, careful consideration must be given to the assumptions underlying each method.

# Ethics Approval

This research utilized a secondary dataset obtained from the UNICEF website, which is publicly accessible and does not contain any identifying information. As a result, institutional ethics approval was not required for this study (<https://mics.unicef.org/surveys>).

[[1]](#footnote-1)

##### References

[1] M. A. Amadi, E. Role, and L. N. Makewa, “Girl Child Dropout: Experiential Teacher and Student Perceptions,” *International Journal of Humanities and Social Science*, vol. 3, no. 5, p. 8, 2013, [Online]. Available: http://www.ijhssnet.com/journals/Vol\_3\_No\_5\_March\_2013/13.pdf.

[2] R. Sabates, A. Hossain, and K. M. Lewin, *Consortium for Research on Educational Access, Transitions and Equity School Drop Out in Bangladesh: New Insights from Longitudinal Evidence*, no. 49. 2010.

[3] J. W. Alspaugh, “The Effect of Transition Grade to High School, Gender, and Grade Level Upon Dropout Rates,” *American Secondary Education*, vol. 29, no. 1, pp. 2–9, 2000, doi: 10.2307/41064411.

[4] P. Goldschmidt and J. Wang, “When can schools affect dropout behavior? A longitudinal multilevel analysis,” *American Educational Research Journal*, vol. 36, no. 4, pp. 715–738, 1999, doi: 10.3102/00028312036004715.

[5] B. Xue, M. Zhang, W. N. Browne, and X. Yao, “A Survey on Evolutionary Computation Approaches to Feature Selection,” *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, Aug. 2016, doi: 10.1109/TEVC.2015.2504420.

[6] B. de la Iglesia, “Evolutionary computation for feature selection in classification problems,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 6, pp. 381–407, Nov. 2013, doi: 10.1002/widm.1106.

[7] Jović, A., Brkić, K., & Bogunović, N. (2015, May). A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1200-1205). Ieee.

[8] Sánchez-Maroño, N., Alonso-Betanzos, A., & Calvo-Estévez, R. M. (2009, June). A wrapper method for feature selection in multiple classes datasets. In *international work-conference on artificial neural networks* (pp. 456-463). Berlin, Heidelberg: Springer Berlin Heidelberg.

[9] D. Barrios, A. Carrascal, D. Manrique, and J. Ríos, “Cooperative binary-real coded genetic algorithms for generating and adapting artificial neural networks,” *Neural Computing and Applications*, vol. 12, no. 2, pp. 49–60, Nov. 2003, doi: 10.1007/s00521-003-0364-1.

[10] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, “Bearing Fault Detection Using Artificial Neural Networks and Genetic Algorithm,” *Eurasip Journal on Applied Signal Processing*, vol. 2004, no. 3, pp. 366–377, Mar. 2004, doi: 10.1155/S1110865704310085.

[11] R. G. Beiko and R. L. Charlebois, “GANN: Genetic algorithm neural networks for the detection of conserved combinations of features in DNA,” *BMC Bioinformatics*, vol. 6, no. 1, p. 36, Feb. 2005, doi: 10.1186/1471-2105-6-36.

[12] M. Karzynski, Á. Mateos, J. Herrero, and J. Dopazo, “Using a genetic algorithm and a perceptron for feature selection and supervised class learning in DNA microarray data,” *Artificial Intelligence Review*, vol. 20, no. 1–2, pp. 39–51, Oct. 2003, doi: 10.1023/A:1026032530166

[13] H. S. Cho, T. S. Kim, J. W. Wee, S. M. Jeon, and C. H. Lee, “cDNA microarray data based classification of cancers using neural networks and genetic algorithms,” in *2003 Nanotechnology Conference and Trade Show - Nanotech 2003*, 2003, vol. 1, pp. 28–31, Accessed: Aug. 07, 2020. [Online]. Available: https://briefs.techconnect.org/papers/cdna-microarray-data-based-classification-of-cancers-using-neural-networks-and-genetic-algorithms/.

[14] Dissanayake, K., & Md Johar, M. G. (2021). Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Applied Computational Intelligence and Soft Computing*, *2021*(1), 5581806.

[15] Yildirim, P. (2015). Filter based feature selection methods for prediction of risks in hepatitis disease. *International Journal of Machine Learning and Computing*, *5*(4), 258.

[16] M. N. Hasan, “A Comparison of Logistic Regression and Linear Discriminant Analysis in Predicting of Female Students Attrition from School in Bangladesh,” Dec. 2019, doi: 10.1109/EICT48899.2019.9068776.

[17] M. N. Hasan, “Factors Associated with Attrition of Girls Students from School in Bangladesh,” *Journal of Scientific Research*, vol. 12, no. 1, pp. 29–38, Jan. 2020, doi: 10.3329/jsr.v12i1.41579.

[18] I. Guyon and A. M. De, “An Introduction to Variable and Feature Selection André Elisseeff,” 2003.

[19] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, “Text feature selection using ant colony optimization,” *Expert Systems with Applications*, vol. 36, no. 3 PART 2, pp. 6843–6853, Apr. 2009, doi: 10.1016/j.eswa.2008.08.022.

[20] A. Ghosh, A. Datta, and S. Ghosh, “Self-adaptive differential evolution for feature selection in hyperspectral image data,” *Applied Soft Computing Journal*, vol. 13, no. 4, pp. 1969–1977, Apr. 2013, doi: 10.1016/j.asoc.2012.11.042.

[21] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, Apr. 2005, doi: 10.1109/TKDE.2005.66.

[22] J. H. Holland, “Adaptation in Natural and Artificial Systems | The MIT Press,” *The University of Michigan Press*, 1975. https://mitpress.mit.edu/books/adaptation-natural-and-artificial-systems (accessed Jul. 31, 2020).

[23] H. Chiroma *et al.*, “Neural networks optimization through genetic algorithm searches: A review,” *Applied Mathematics and Information Sciences*, vol. 11, no. 6. pp. 1543–1564, 2017, doi: 10.18576/amis/110602.

[24] M. Hamdan, “A heterogeneous framework for the global parallelisation of genetic algorithms,” *International Arab Journal of Information Technology*, vol. 5, no. 2, pp. 192–199, 2008, Accessed: Jul. 31, 2020. [Online]. Available: https://www.researchgate.net/publication/220413569\_A\_Heterogeneous\_Framework\_for\_the\_Global\_Parallelisation\_of\_Genetic\_Algorithms.

[25] C. T. Capraro, I. Bradaric, G. T. Capraro, and T. K. Lue, “Using genetic algorithms for radar waveform selection,” 2008, doi: 10.1109/RADAR.2008.4720947.

[26] “BP Neural Network Algorithm Optimized by Genetic Algorithm and Its Simulation,” *International Journal of Computer Science Issues*, vol. 10, no. 1, pp. 516–519, 2013, Accessed: Jul. 31, 2020. [Online]. Available: https://www.researchgate.net/publication/303102486\_BP\_Neural\_Network\_Algorithm\_Optimized\_by\_Genetic\_Algorithm\_and\_Its\_Simulation.

[27] Cherrington, M., Thabtah, F., Lu, J., & Xu, Q. (2019, April). Feature selection: filter methods performance challenges. In *2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-4). IEEE.

[28] Estévez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. (2009). Normalized mutual information feature selection. *IEEE Transactions on neural networks*, *20*(2), 189-201.

[29] Vikas, P. K. (2021). Lung cancer detection using chi-square feature selection and support vector machine algorithm. *International Journal of Advanced Trends in Computer Science and Engineering*.

[30] Amiri, F., Yousefi, M. R., Lucas, C., Shakery, A., & Yazdani, N. (2011). Mutual information-based feature selection for intrusion detection systems. *Journal of network and computer applications*, *34*(4), 1184-1199.

[31] Hoque, N., Bhattacharyya, D. K., & Kalita, J. K. (2014). MIFS- ND: A mutual information-based feature selection method. *Expert systems with applications*, *41*(14), 6371-6385.

[32] Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybernetics and information technologies*, *19*(1), 3-26

[33] Sumi, M. S., & Narayanan, A. (2019, February). Improving classification accuracy using combined filter+ wrapper feature selection technique. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-6). IEEE.

[34] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, *97*(1-2), 273-324.

[35] T. Liu, H. Zhang, H. Zhang, and A. Zhou, “Information Fusion in Offspring Generation: A Case Study in Gene Expression Programming,” *IEEE Access*, vol. 8, pp. 74782–74792, 2020, doi: 10.1109/ACCESS.2020.2988587.

[

1. [↑](#footnote-ref-1)