Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Efficient Model for Child Mortality Prediction Using Ensemble Feature Selection and Classification Algorithms

Sayed Zabihullah Musawi

sayedzabimusawi@stu.yazd.ac.ir

Kunduz University

Mohammad Farshi

Yazd University

Research Article

Keywords: Machine Learning, Feature Selection, Optimization Algorithm, Under-five Child, ADHS2015, Afghanistan

Posted Date: May 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9743242/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Efficient Model for Child Mortality Prediction Using Ensemble Feature Selection and Classification Algorithms

Sayed Zabihullah Musawi[1,2*] and Mohammad Farshi[1]

[1]Department of Information Technology, Kunduz University, Daneshgah Blvd, Kunduz, Kunduz 3501, Afghanistan.
[2]Department of Computer Science, Yazd University, Daneshgah Blvd, Yazd, Yazd 89158-741, Iran.

*Corresponding author(s). E-mail(s): sayedzabimusawi@stu.yazd.ac.ir;
Contributing authors: mfarshi@yazd.ac.ir;

Abstract

The Under-5 Mortality Rate (U5MR) is a critical indicator of global health, yet millions of preventable child deaths occur annually. Machine learning offers a powerful approach for predicting U5MR and identifying risk factors, though high-dimensional data often requires effective feature selection to enhance model performance and interpretability. This study proposes a novel two-stage feature selection framework. First, an Ensemble Feature Selection (EFS) method aggregates results from five filter-based techniques. Second, the EFS output is optimized using six binary Evolutionary Algorithms (EAs), including the Binary Whale Optimization Algorithm (BWOA) and Binary Sine Cosine Algorithm (BSCA). The resulting EFS-EAs frameworks were evaluated with three machine learning classifiers—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gaussian Naïve Bayes (GNB)—and three deep learning models—CNN, RNN, and BiLSTM. Experimental results demonstrate that the KNN classifier, when paired with the EFS-BWOA or EFS-BSCA feature subset, achieved perfect performance 100% accuracy, significantly outperforming baseline machine learning models and state-of-the-art deep learning architectures. This study conclusively shows that sophisticated feature selection can elevate simple classifiers to surpass complex models, providing an highly effective methodology for U5MR prediction.

Keywords: Machine Learning, Feature Selection, Optimization Algorithm, Under-five Child, ADHS2015, Afghanistan

1 Introduction

The U5MR is a critical indicator of population health, representing the probability of a child dying between birth and their fifth birthday. According to UNICEF, 4.9 to 5.6 million children died before reaching age five in 2022, with approximately half (2.5 to 3.0 million) occurring between 1–59 months [1]. This extensive loss of life, largely from preventable causes, reflects global disparities in access to healthcare, nutrition, sanitation, and other socioeconomic determinants of health. Despite a declining global trend due to healthcare advancements and targeted interventions, U5MR remains disproportionately high in low-income countries.

As a low-income country, Afghanistan continues to experience one of the highest U5MR in the world, making it among the most perilous places for a child to be born. According to UNICEF's State of the World's Children report, the U5MR in Afghanistan was 55.7 deaths per 1,000 live births in 2021 [1, 2]. Although this represents a substantial decline from 178.5 in 1990, child mortality remains a critical public health concern amid the country's persistent socio-economic and political challenges. Consequently, data-driven approaches have become essential, with machine learning offering a powerful set of tools to analyze complex factors, inform decision-making, and optimize intervention strategies.

Machine learning is a transformative technology widely applied across scientific and analytical domains. Its strength lies in uncovering patterns, generating predictions, and extracting insights from complex datasets to support evidence-based decision-making. Within this context, feature selection is a critical step that identifies the most informative subset of features to enhance model performance. Given an initial feature set X of n features, the objective is to select m features ($m < n$) to reduce computational cost, prevent overfitting, and improve predictive accuracy. However, feature selection remains a challenging task due to three factors: (i) the exponentially large search space (2^n possible subsets); (ii) the presence of redundant or irrelevant features that can degrade performance; and (iii) its formulation as a multi-objective, NP-hard optimization problem [3]. Therefore, effective feature selection relies on efficient algorithms, commonly categorized into Filter, Wrapper, and Embedded methods [4, 5].

Existing research employs diverse feature selection strategies, each characterized by specific advantages and limitations. Some studies rely on a single feature selection method; however, such an approach may inadvertently eliminate relevant features due to the inherent bias of any single technique [6–8]. To address this issue, other researchers have adopted a two-step framework, in which a filter method is first applied for preliminary dimensionality reduction, followed by an optimization algorithm for final feature selection [9–11]. Nevertheless, this sequential approach can still result in the exclusion of potentially informative features during the initial filtering stage. Alternatively, hybrid models that integrate two optimization algorithms have been proposed to improve selection performance, though such combinations often lead to increased computational complexity [12–14].

More recently, EFS methods have been employed to enhance diversity and reduce the likelihood of omitting important features [15, 16]. A promising approach integrates EFS with optimization techniques. For instance, Rouhi et al. [17] proposed a two-stage framework in which an EFS method is applied in the first stage, followed by an

optimization algorithm in the second, to identify the optimal feature subset. A key research challenge in such hybrid frameworks lies in designing an effective architecture, particularly ensuring that the initial stage selects and forwards a robust set of features to the subsequent optimization process.

In this study, we propose an effective approach for predicting child mortality by leveraging the increased diversity achieved through both feature selection and learning algorithms. The analysis is based on the Afghanistan Demographic and Health Survey 2015 (ADHS2015) dataset, conducted under the DHS program. As the ADHS2015 dataset is high-dimensional, Principal Component Analysis (PCA) is first applied to reduce data dimensionality and eliminate noise. Feature selection is then performed in two stages. In the first stage, an EFS method is employed to identify relevant features and minimize the likelihood of discarding informative attributes. In the second stage, unlike previous studies, the selected features are provided to EAs to extract the most significant subset for model training.

The main contributions of this study are summarized as follows:

(i) A multi-stage feature selection framework, referred to as EFS-EAs, is proposed to enhance the efficiency and accuracy of U5MR prediction.
(ii) The study directly addresses the critical public health challenge of U5M in Afghanistan. By developing and applying the proposed model to the ADHS2015 dataset, it provides valuable insights into the key determinants influencing child mortality in this region.
(iii) The performance of the proposed EFS framework is comprehensively evaluated using both deep learning and traditional machine learning models.
(iv) The selected EFS subset is further optimized through six EAs, which significantly reduce feature dimensionality and improve the predictive accuracy of U5MR outcomes.

The remainder of this paper is organized as follows: Section 2 reviews the related work. Section 3 presents a concise overview of the research methodology. Section 4 discusses the experimental results and analysis. Finally, Section 5 concludes the study and outlines directions for future research.

2 Related Works

Understanding trends in U5M is critical for guiding public health policies and interventions. Numerous studies have examined U5M patterns across various countries [18–23]. At the global level, two major initiatives systematically document and estimate child mortality. The Institute for Health Metrics and Evaluation (IHME), through its Global Burden of Disease (GBD) study, provides comprehensive and regularly updated estimates of U5MR. Similarly, the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME) compiles, harmonizes, and disseminates authoritative datasets on child mortality, thereby facilitating international comparisons and monitoring progress toward global health targets [2, 23–27].

Nasrat et al. [21] developed a predictive model for child mortality using five widely applied machine learning classifiers: KNN, GNB, Artificial Neural Network (ANN),

Random Forest (RF), and SVM. The dataset, obtained from ADHS2015, was preprocessed using the Correlation-Based Feature Selection (CFS) algorithm to select 13 key attributes. The classification involved two categories, Alive and Dead, with balanced samples. Model performance was evaluated using accuracy, precision, recall, and AUROC. The RF achieved the highest accuracy of 85.47%, making it the most effective classifier. In another study, Pal et al. [28] used the ADHS2015 dataset, focusing on 1,640 women aged 15–49 who had been married at least once and had at least one child under the age of five at the time of the survey. The study employed bivariate and multivariate regression analyses to identify factors contributing to U5M. The findings revealed that children born small or very small, those born within less than two years of a previous birth, children living in rural areas, and those whose mothers did not receive antenatal care were at a higher risk of U5M. These significant risk factors underscore the need for targeted interventions to reduce under-five mortality in Afghanistan.

In 2020, Bitew et al. [18] addressed the limited use of machine learning models for predicting under-five mortality (U5M) in Ethiopia. Using data from the 2016 Ethiopian Demographic and Health Survey (EDHS2016), they evaluated RF, LGR, and KNN, with the RF model achieving the highest accuracy of 67.2%. Tesfaye et al. [29] developed a web-based U5M prediction system in an Ethiopian local language using Decision Tree (J48) and rule induction (PART) algorithms on 11,654 records, achieving 94.3% accuracy. The study identified key determinants of child mortality, including breastfeeding status, maternal education, family planning, birth interval, and the presence of diarrhea, and suggested further validation with advanced algorithms. Similarly, Mohapatra et al. [30] analyzed factors influencing infant and child mortality using machine learning methods on 10,641 records from EDHS2016 to identify key predictors and develop supervised models.

Saroj et al. [31] aimed to assess the accuracy of machine learning models in predicting U5M and identify key factors associated with it using data from the NFHS-IV of Uttar Pradesh. Various models, including DT, RF, GNB, KNN, Logistic Regression (LGR), SVM, ANN, and ridge classifier, were evaluated for accuracy, precision, recall, and other performance metrics. Results showed that the ANN model was the best predictor of U5M, with high accuracy (95.29% to 95.96%) and strong AUROC and precision-recall scores, while LGR also performed well. Key factors influencing U5M included the number of living children, wealth index, child size at birth, mother's education, and birth order.

Adegbosin et al. [32] analyzed data from the DHS of 21 Low-and-Middle-Income Countries (LMICs), with a focus on U5M and the efficacy of different deep learning models in predicting child survival. Key predictors of U5M included breastfeeding duration, household wealth, and maternal education level. Deep learning models like Deep Neural Network (DNN), Convolution Neural Network (CNN), and a hybrid CNN-DNN outperformed traditional LGR in predicting child survival, with higher sensitivity and specificity. The findings highlight the need for targeted interventions to reduce U5M in LMICs and demonstrate the effectiveness of deep learning techniques over traditional methods. Ashwini et al. [33] analyzed a global dataset from 1967 to 2019, consisting of 30,940 entries, to predict U5MR using regression algorithms

such as XGBoost, CatBoost, RF, AdaBoost, and DT Regressor. They assess model performance through evaluation metrics including R^2, adjusted R^2, Mean Absolute Error (MAE) , Mean Square Error (MSE), and Rood MSE (RMSE). Visualization tools like Matplotlib and Seaborn are used to present findings via pie charts and graphs. The study aims to determine the most accurate algorithm for predicting child mortality, contributing to improved healthcare planning and intervention strategies to reduce U5MR worldwide.

Recently, Iqbal et al. [20] have used predictive analytic and the Pakistan DHS (2017-2018) dataset to predict child mortality. The scholars found that the most risk factors, such as the number of children under-five in a household, birth intervals, family size, mother's age, antenatal care, breastfeeding, child size at birth, and place of delivery are the most important ones. The RF classifier demonstrated a high accuracy of 93.8% and could be assisting child health programs. In another study, the Mbunge et al. [34] study, used in Zimbabwe, employs machine learning algorithms, such as DT, RF, LGR, and XGBoost, to make predictive models for U5M using Pakistan DHS data. The XGBoost model outperforms others with an accuracy of 81%, high precision, and recall, making it a valuable tool for policymakers and healthcare professionals to enhance child health, access to quality care, preventive measures, and immunization programs, ultimately reducing child mortality.

Despite extensive global research on U5M, limited studies have focused specifically on Afghanistan using advanced machine learning and multi-step feature selection techniques. Existing works often rely on single feature selection methods or traditional models, which may overlook important predictors. Given the high U5MR in Afghanistan and the complexity of the ADHS2015 dataset, there is a clear need for a more robust, data-driven approach to accurately identify key risk factors and improve prediction performance supporting more effective health interventions and policy decisions.

3 Methodology

The research methodology is structured around five main phases. First, the raw data undergoes preprocessing to address missing values, outliers, data balancing, and to perform normalization. Second, a dimensionality reduction technique is applied to mitigate initial data complexity. The core of the analytical framework is a two-step feature selection process: an initial EFS method consolidates multiple feature rankings to create a robust candidate subset, which is then refined by an EAs to identify the most predictive features. Finally, the optimal feature subset is used to train machine learning models, whose performance is rigorously evaluated and compared against established benchmarks to determine the most effective approach. The schematic presentation of the research method is presented in Figure 1.

In this research we employed ADHS2015 [23], a recent series of demographic and health surveys conducted every five years by DHS program [35]. The ADHS2015, designed as a nationally representative household survey, aims to collect data spanning a wide range of population, health, and nutrition indicators, specially focusing on enhancing maternal and child health in all countries. ADHS employed a multi-stage

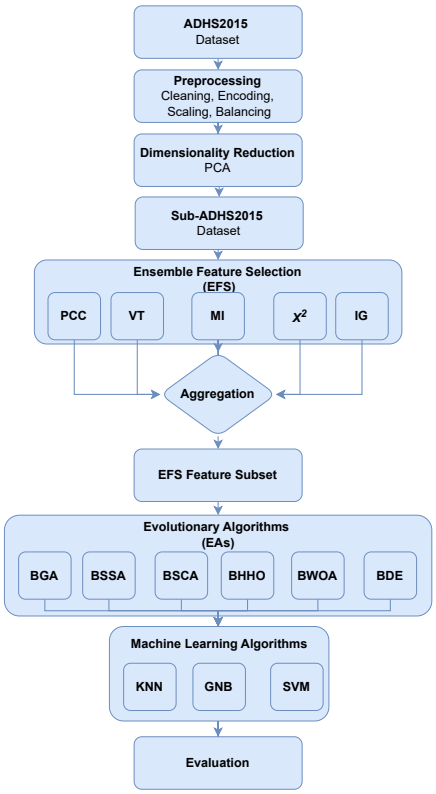


Fig. 1: Workflow of the Proposed Method for U5MR Prediction.

stratified sampling technique, selecting respondents from all thirty four provinces of Afghanistan.The primary focus of this study's analysis is on children under the age of five, with a total sample size of 32,712 drawn from rural and urban areas across Afghanistan. More details about the original dataset can be found readable in [23]. The data were derived from retrospective information provided by mothers regarding their children who either reached the age of five or died in the five years preceding the survey. Consequently, we adopted a binary classification approach, extracting a set of independent variables X_i to predict the dependent variable y_i. The target values are represented by Equation 1.

$$y_i = \begin{cases} i = 0 & \text{if child is Died,} \\ i = 1 & \text{if child is Alive.} \end{cases} \tag{1}$$

3.1 Data Preprocessing

The data preprocessing phase is a critical step in machine learning, as it ensures data quality and prepares datasets for effective model training. In the ADHS2015

dataset, features included ordinal, nominal, and scale variables. The process began with label encoding to transform nominal and ordinal features into numerical form. Subsequently, the Interquartile Range (IQR) method was applied to detect and handle outliers. Missing values, which posed a significant challenge, were imputed using a linear regression-based approach. Addressing class imbalance in the target variable was also a key focus; the Synthetic Minority Over-sampling Technique (SMOTE) was employed to achieve a balanced dataset. Finally, the data were split into training and testing subsets using a hold-out strategy, allocating 70% of the data for training and 30% for evaluating model performance.

3.2 Dimensionality Reduction

Given the high-dimensional nature of the ADHS2015 dataset, which can lead to computational inefficiency and the curse of dimensionality, PCA was employed for dimensionality reduction. PCA transforms the original features into a new set of uncorrelated variables, known as principal components, which are ordered by the amount of variance they capture from the data. This technique allows for the retention of the most significant information while reducing noise and redundancy. Applying PCA resulted in a condensed set of 35 principal components that collectively preserve the majority of the dataset's variance. The characteristics of these components, including the explained variance for each, are detailed in Table 1.

3.3 Feature Selection

Feature selection refers to the process of identifying, from a given dataset, the subset of features that have the greatest influence on model performance [4]. Suppose the feature set X contains n features; the objective is to select m features ($m < n$) to reduce computational complexity while improving model accuracy. This process enhances the accuracy, efficiency, and execution time of machine learning algorithms by reducing data dimensionality and eliminating redundant or irrelevant features, thereby retaining only those that most strongly affect the target variable. In this study, feature selection was performed in two stages using an EFS approach combined with EAs, referred to as the EFS-EAs method.

3.3.1 Ensemble Feature Selection

The EFS techniques are inspired by ensemble learning, where the results of multiple selectors or feature subsets are combined to produce a single, robust decision. Based on the combination strategy, EFS can be categorized into homogeneous and heterogeneous approaches [36]. In the homogeneous approach, multiple subsets of data are processed by the same selector to identify relevant features, which are then aggregated to form the final selection [37, 38]. In contrast, the heterogeneous approach applies different feature selection techniques to the same dataset, and the results from the various selectors are combined in a manner analogous to the homogeneous approach [36].

In this study, a heterogeneous EFS approach was employed due to its ability to introduce diversity. However, a key challenge in heterogeneous EFS is combining and selecting features from the aggregated results. To address this, we formulated a scoring

Table 1: Description of the Sub-ADHS2015 Dataset A Obtained through PCA.

No	**Risk factors**	**Type**	**Values**
1	Breastfeeding	Categorical	2
2	Age of mother at first birth	Numeric	30
3	Mothers age	Numeric	35
4	Father education level	Categorical	2
5	Birth order number	Numeric	15
6	Source of drinking water	Categorical	3
7	Sex of child	Categorical	2
8	Antenatal care visits	Numeric	20
9	Place of delivery	Categorical	3
10	Child twin	Categorical	2
11	Gender of the child	Categorical	2
12	Preceding birth interval	Numeric	146
13	Occupation of mother	Categorical	2
14	Wealth index	Categorical	2
15	Husband occupation	Categorical	2
16	Mother age group	Categorical	4
17	Mothers education level	Categorical	3
18	Contraceptive use	Categorical	2
19	Number of antenatal visit	Categorical	3
20	Received family planning	Categorical	2
21	Residence	Categorical	2
22	Type of toilet facility	Categorical	3
23	Presence of diarrhea	Categorical	2
24	Child size at birth	Categorical	3
25	Current breastfeeding	Categorical	2
26	Region	Categorical	34
27	Current marital status	Categorical	3
28	Baby postnatal check-up	Categorical	2
29	Family members	Numeric	38
30	No. of under-five children in a household	Categorical	10
31	Vaccination of child	Categorical	2
32	Total children ever born	Numeric	15
33	Iron folic acid supply during pregnancy	Categorical	2
34	Births in the last 5 years	Numeric	5
35	Religion	Categorical	3

system in which a selected feature receives a score of 1, while an unselected feature receives a score of 0. The results of each feature selection technique are presented in separate columns, and the aggregated scores for all features are shown in Table 2.

In the proposed EFS-EAs method, the Sub-ADHS2015 dataset obtained through PCA is processed by five filter-based feature selection techniques PCC, MI, χ^2, IG, and VT which are combined to identify the most relevant features. This combination is necessary because most filter-based techniques rely on a single criterion; for example, PCC selects features solely based on their correlation with the target variable, without accounting for redundancy. In our approach, the techniques complement each other: if one method misses an important feature, the others may capture it. Conversely, a feature not selected by any of the five methods is considered uninformative and is excluded from further analysis. Consequently, the features Received family planning

Table 2: Features Selected by Individual Filter Methods and Their Scores.

No	**Risk factors**	**VT**	χ^2	**MI**	**PCC**	**IG**	**Score**
1	Breastfeeding	1	0	1	1	1	4
2	Age of mother at first birth	1	1	1	1	0	4
3	Mothers age	1	1	1	1	1	5
4	Father education level	1	0	0	1	1	3
5	Birth order number	1	1	1	0	1	4
6	Source of drinking water	1	1	1	0	0	3
7	Sex of child	0	0	0	0	0	0
8	Antenatal care visits	1	1	1	1	0	4
9	Place of delivery	0	1	1	1	1	4
10	Child twin	0	1	1	1	0	3
11	Gender of the child	1	1	1	1	1	5
12	Preceding birth interval	0	0	1	1	0	2
13	Occupation of mother	1	1	1	0	0	3
14	Wealth index	0	1	1	0	1	3
15	Husband occupation	0	1	1	1	1	4
16	Mother age group	1	0	1	1	1	4
17	Mothers education level	0	1	1	1	1	4
18	Contraceptive use	1	1	1	1	1	5
19	Number of antenatal visit	1	1	1	0	1	4
20	Received family planning	0	0	0	0	0	0
21	Residence	1	0	1	1	0	3
22	Type of toilet facility	1	1	0	0	1	3
23	Presence of diarrhea	0	1	1	1	0	3
24	Child size at birth	1	1	1	1	1	5
25	Current breastfeeding	0	0	1	0	1	2
26	Region	1	1	1	1	1	5
27	Current marital status	1	1	1	1	1	5
28	Baby postnatal check-up	1	0	1	1	0	3
29	Family members	0	1	1	0	0	2
30	No. of under-five children in a household	1	1	1	0	0	3
31	Vaccination of child	1	1	0	1	1	4
32	Total children ever born	0	0	1	1	1	3
33	Iron folic acid supply during pregnancy	0	1	1	1	1	4
34	Births in the last 5 years	0	1	1	0	0	2
35	Religion	1	0	1	1	0	3

and Sex of child were eliminated, and 33 features were retained, forming the EFS Feature Subset, as presented in Table 2.

3.3.2 Evolutionary Algorithms

Feature selection is a multi-objective optimization problem (MOP) that aims to simultaneously maximize classification performance and minimize the number of selected features two objectives that are often conflicting. Although EAs are effective for solving MOPs, their computational complexity can be prohibitive for high-dimensional datasets. To mitigate this, an EFS preprocessing step is first applied to select a relevant subset of features, forming the EFS Feature Subset. This subset is then used as input for binary EA variants, including the Binary Whale Optimization Algorithm

(WOA) [39], Salp Swarm Algorithm (SSA) [40], Sine Cosine Algorithm (SCA), Genetic Algorithm (GA) [41], Harris Hawks Optimization (HHO) [42], and Differential Evolution (DE) [43], to identify the most informative features. Parameter settings for these algorithms are adopted from their respective references. Moreover, feature selection is a binary optimization problem. Each solution is represented by a vector of length d, where a value of 1 indicates a selected feature and 0 indicates an excluded feature, as illustrated in Figure 2. To enable this representation, a sigmoid transfer function Equation 2 is used to convert continuous search space values into binary form, allowing the effective application of the algorithms to the feature selection task.

$$S = \frac{1}{1 + e^{-x}},$$

$$X_{binary} = \begin{cases} 1, & \text{if } rand \geq S, \\ 0, & \text{if } rand < S, \end{cases} \tag{2}$$

where x represents a continuous feature value, S is the continuous output obtained from the sigmoid function, and X_{binary} denotes the resulting binary value, and $rand$ is a randomly generated number within the range [0,1].

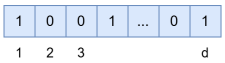


Fig. 2: Solution Presentation for Feature Selection Problem.

Feature selection is a MOP that seeks to minimize the number of selected features while maximizing model accuracy. Accordingly, the fitness function in this study is formulated to balance these competing goals, as defined in Equation 3:

$$\text{Fitness function} = \text{argmin}\left(\alpha, Err + \beta\frac{|SFs|}{|N|}\right), \tag{3}$$

where Err denotes the classification error, SFs is the number of selected features, and N is the total number of features. The parameters α(0.99) and β (0.01) control the trade-off between accuracy and feature reduction. In wrapper-based feature selection, the choice of classifier strongly affects the quality of the selected features. The KNN classifier [44] is widely adopted due to its effectiveness in high-dimensional datasets [45], and it was therefore employed to evaluate the performance of feature selection in this study.

3.4 Machine Learning Algorithms

- **Support Vector Machine (SVM):** SVM is a supervised machine learning algorithm for classification and regression tasks, making it versatile across various applications. SVMs exhibit adaptability and efficiency, particularly in handling high-dimensional data and capturing nonlinear relationships. The primary goal of the SVM algorithm is to identify the optimal hyperplane within an N-dimensional space

that effectively separates data points belonging to different classes in the feature space. This hyperplane aims to maximize the margin between the closest points of distinct classes. For more detailed information, refer to [46, 47].

- **K-nearest neighbors (KNN):** KNN is employed for both classification and regression tasks. As a non-parametric algorithm, it operates by identifying the KNN of a given data point and classifying it based on the majority class among those neighbors. The hyperparameter K dictates the number of neighbors and must be predetermined [48].
- **Gradient Näive Bayes (GNB):** GNB algorithm is a probabilistic machine learning algorithm utilized for classification tasks and regression [49, 50]. This is based on the Bayes theorem, which says that the probability of a hypothesis (class) given some evidence (features) is proportional to the probability of the evidence given the hypothesis multiplied by the prior probability of the hypothesis.

3.5 Deep Learning Algorithms

- **Convolutional Neural Networks (CNNs):** CNNs are a specialized class of deep neural networks designed to process data with a grid-like topology, such as images [51]. Their core functionality stems from convolutional layers that employ learnable filters instead of weight matrix which slide across the input to compute feature maps, effectively capturing local spatial correlations and translational-invariant patterns [32].
- **Recurrent Neural Networks (RNNs):** Recurrent Neural Networks (RNNs) are a fundamental architecture designed for sequential data processing. Their recurrent nature allows them to handle variable-length sequences and incorporate contextual information from previous time steps[52].
- **Bidirectional Long Short-Term Memory (BiLSTM):** BiLSTM networks are an advanced RNN architecture designed to overcome the limitations of standard recurrent networks [52]. A BiLSTM incorporates two separate LSTM layers that process the input sequence in both forward and backward directions. This allows the model to capture contextual information from both past and future states simultaneously [53]. The bidirectional approach provides a more comprehensive contextual understanding, making BiLSTMs particularly powerful for tasks that require full-sequence comprehension.

3.6 Evaluation Metrics

The model's performance was evaluated using accuracy, precision, recall, F1-score, specificity, and the AUC, computed using Equations 4–8,

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100, \tag{4}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100, \tag{5}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100, \tag{6}$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \times 100, \tag{7}$$

$$\text{F1-score} = \frac{2 * Recall * Precision}{Recall + Precision} \times 100, \tag{8}$$

where the TP, FN, FP, and TN stand from True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN), respectively.

4 Results and Discussions

The experiment was organized into five main steps: (i) preprocessing, (ii) dimensionality reduction, (iii) multi-step feature selection, (iv) classification, and (v) evaluation. In the data preprocessing step, nominal features were encoded using label encoding, and missing values were imputed using a logistic regression algorithm. Additionally, the data were balanced using SMOTE and normalized with standard scaling techniques. Given the high dimensionality of the ADHS2015 dataset, PCA was applied for dimensionality reduction. Feature selection was implemented as a two-step process. The first step involved EFS using five filter-based methods: PCC, χ^2, MI, VT, and IG. In the second step, binary EAs including Binary Whale Optimization Algorithm (BWOA), Binary Salp Swarm Algorithm (BSSA), Binary Sine Cosine Algorithm (BSCA), Binary Genetic Algorithm (BGA), Binary Harris Hawks Optimization (BHHO), and Binary Differential Evolution (BDE) were applied to identify the most informative features. In the classification step, three machine learning algorithms SVM, KNN, and GNB and three deep learning models RNN, CNN, and BiLSTM were used to evaluate the selected feature subsets. Finally, model performance was assessed using multiple metrics, including accuracy, precision, recall, F1-score, and AUC.

The experiments were conducted on a desktop computer running Windows 11, equipped with an Intel® Core i7 CPU @ 2.6 GHz, 8 GB RAM, and a 1 TB hard drive. All simulations were implemented in Python 3.10, Jupyter Notebook, and MATLAB R2021b. The feature selection was configured with 100 iterations, 20 agents, and 20 independent runs to ensure a robust evaluation.

4.1 Result of Machine Learning Algorithms on EFS Feature Subset

The performance of three machine learning algorithms using the EFS feature subset is compared in Table 3, with their corresponding confusion matrices shown in Figure 3. The mean accuracy of 75.92% is misleading, as it obscures a critical precision-recall trade-off. The high average precision (93.17%) coupled with low recall (57.85%) indicates a high false negative rate. This imbalance is reflected in the more representative average F1-score of 68.51%, a trend heavily influenced by the GNB model. In contrast, the KNN and SVM algorithms exhibit superior and more balanced performance.

4.2 Result of Deep Learning Algorithms EFS Feature Subset

Table 4 presents the performance of deep learning models—CNN, RNN, and BiLSTM—on the EFS Feature Subset. The deep learning models demonstrate exceptionally strong and balanced performance. The key strength is the harmony between all

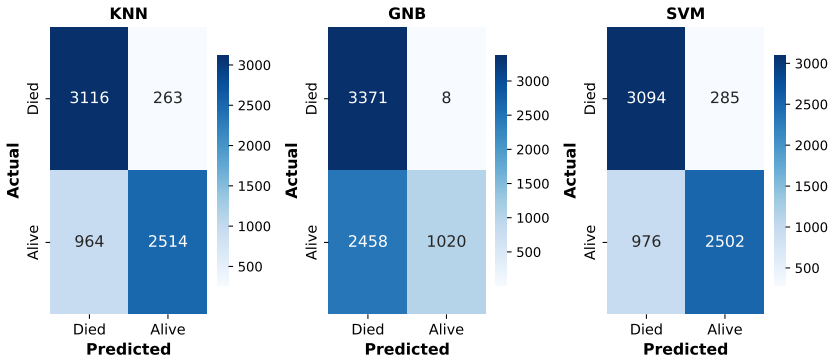


Fig. 3: Confusion Matrices for Machine Learning Models Trained on the EFS Feature Subset.

Table 3: Classification Performance of Machine Learning Models using the EFS Feature Subset.

Machine Learning Models	Accuracy	Precision	Recall	F1-Score	AUC
KNN	82.11	90.53	72.28	80.38	82.25
GNB	64.04	99.22	29.33	45.27	64.55
SVM	81.61	89.77	71.94	79.87	81.75
Average	75.92	93.17	57.85	68.51	76.18

metrics. The average accuracy is 90.86%, and this is a reliable indicator of performance because the average precision (90.99%), recall (90.87%), and F1-score (90.86%) are virtually identical. This indicates the models make very few false positive and false negative errors. Furthermore, the exceptional average AUC of 96.59% confirms an outstanding ability to distinguish between the classes. Performance improves with model complexity, with BiLSTM emerging as the top performer, achieving near-perfect scores across all metrics.

Table 4: Classification Performance of Deep Learning Models using the EFS Feature Subset.

Deep Learning Models	Accuracy	Precision	Recall	F1-Score	AUC
CNN	82.75	83.55	83.26	83.23	91.57
RNN	92.37	93.35	93.34	93.34	98.59
BiLSTM	97.46	96.07	96.01	96.01	99.62
Average	90.86	90.99	90.87	90.86	96.59

The Figure 5 shows model performance over 50 epochs for CNN, RNN, and BiLSTM. The BiLSTM consistently achieves the highest accuracy and lowest loss, indicating superior learning and generalization. The RNN shows moderate improvement,

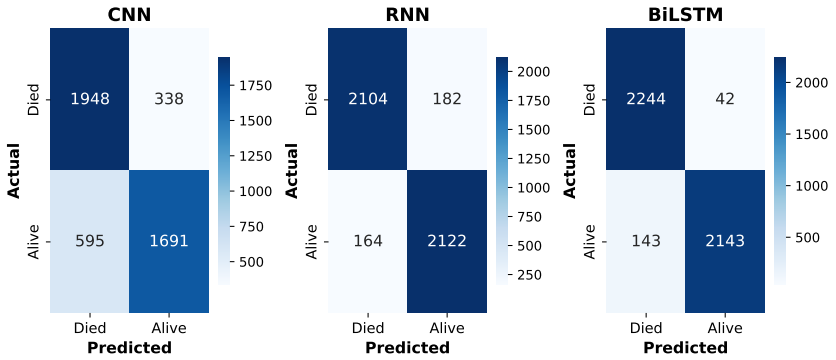


Fig. 4: Confusion Matrices for Deep Learning Models Trained on the EFS Feature Subset.

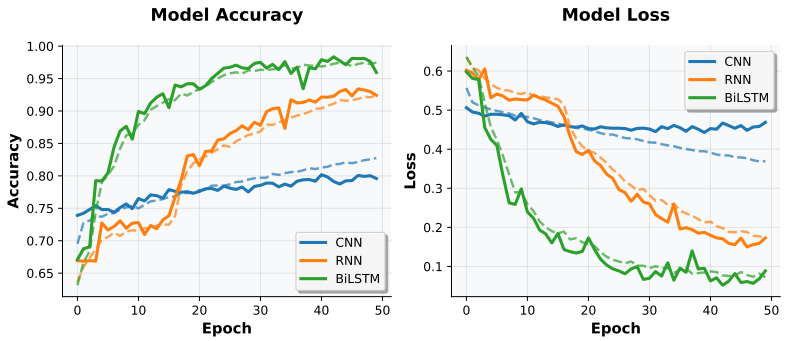


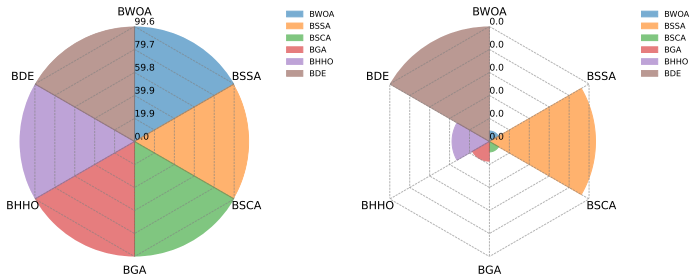
Fig. 5: Model Accuracy and Loss over 50 Epochs for the EFS Feature Subset.

while the CNN lags behind both in accuracy and loss reduction. Overall, BiLSTM demonstrates the most efficient convergence and stability among the models.

In short, the deep learning models represent a substantial improvement over the traditional machine learning models. They completely resolve the critical precision-recall trade-off that plagued the machine learning models, resulting in a 15-point increase in average accuracy (90.86% vs. 75.92%).This makes the deep learning models, particularly BiLSTM, vastly more reliable and effective for a balanced classification task.

4.3 Result of Proposed EFS-EAs Feature Selection

The results in Table 5 demonstrate that the EFS-based feature selection method significantly enhances classification accuracy, with most bio-inspired algorithms achieving near-perfect performance. In terms of accuracy, algorithms like BWOA and BSCA reached a consistent 99.98%, showing flawless stability, while others, though highly accurate, exhibited minor fluctuations. This exceptional accuracy is corroborated by



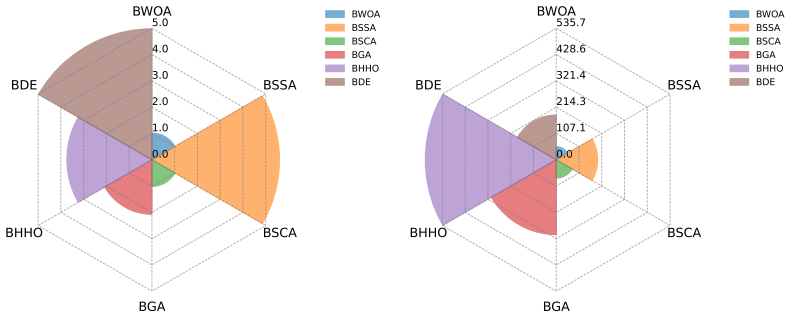
(a) Classification Accuracy. (b) Fitness Values.

Fig. 6: Comparison of Feature Selection Algorithms on EFS Feature Subset Based on (a) Classification Accuracy and (b) Fitness Values.

fitness values, where most algorithms converged optimally to zero, confirming effective minimization of the objective function. A critical differentiator is the feature selection ratio; the most efficient algorithms, notably BWOA and BSCA, maintained elite accuracy using a minimal feature subset, whereas others required a larger feature set for similar results. Computational efficiency varied substantially, with BWOA achieving the fastest convergence, making it highly suitable for practical applications. The comparative performance of the algorithms across four key metrics—classification accuracy, fitness values, feature selection ratio, and computation time—is shown in Figures 6 and 7.

Table 5: Comparison of Feature Selection Algorithm based on Classification Accuracy, Fitness Values, Number of Selected Features, and Computation Time (s) on EFS Feature Subset.

Measurement	Statistics	EFS-BWOA	EFS-BSSA	EFS-BSCA	EFS-BGA	EFS-BHHO	EFS-BDE
Accuracy	AVG	99.98	99.49	99.98	99.97	99.90	99.44
	STD	0.00	0.46	0.00	0.02	0.06	0.91
	Worst	99.98	98.23	99.98	99.91	99.78	96.17
	Best	99.98	99.96	99.98	99.98	99.98	99.98
Fitness Values	AVG	0.00	0.01	0.00	0.00	0.00	0.01
	STD	0.00	0.00	0.00	0.00	0.00	0.01
	Worst	0.00	0.02	0.00	0.00	0.01	0.04
	Best	0.00	0.00	0.00	0.00	0.00	0.00
NSFs	AVG	1.00	4.80	1.00	2.05	3.20	4.95
	STD	0.00	1.54	0.00	0.60	1.44	1.93
	Worst	1.00	9.00	1.00	3.00	7.00	8.00
	Best	1.00	3.00	1.00	1.00	1.00	2.00
Time	AVG	52.75	168.53	74.57	304.65	535.73	182.04
	STD	8.56	44.10	17.65	9.24	17.23	39.23
	Worst	70.38	271.87	108.67	327.90	572.49	257.38
	Best	34.37	104.80	47.60	287.46	502.44	126.15



(a) Number of Selected Features. (b) Computation Time(s).

Fig. 7: Comparison of Feature Selection Algorithms on EFS Feature Subset Based on (a) Number of Selected Features and (b) Computation Time(s).

4.4 Results of Proposed EFS-EAs Methods with Machine Learning Algorithms

Table 6 demonstrates that the performance of the classifiers is predominantly determined by the feature subset selected by the bio-inspired algorithms, underscoring that feature selection is a more critical factor than the classifier choice for this specific dataset. Specifically, algorithms such as BWOA, BSSA, and BSCA enable the KNN classifier to achieve perfect scores of 100% across accuracy, precision, recall, and F1-score, indicating the identification of an optimally discriminative feature subset. This result constitutes a substantial improvement over the baseline KNN model, which exhibited significant performance trade-offs with an accuracy of 82.11% and a recall of 72.28%. Moreover, the optimized KNN model surpasses the performance of the most advanced deep learning model previously evaluated, the BiLSTM, which attained 97.46% accuracy. However, as illustrated in Figure 8, this peak performance is not universal; both GNB and SVM exhibit highly variable and suboptimal results with the same feature subsets. The consistently low AUC values for KNN further suggest a potential anomaly in probability calibration. Ultimately, these findings demonstrate that sophisticated feature selection can elevate a simple model such as KNN to outperform complex deep learning architectures, thereby highlighting its paramount importance within the machine learning pipeline.

4.5 Statistic test

The Friedman test, a non-parametric statistical method, was employed to determine if the performance differences among the feature selection algorithms were statistically significant. The results in Table 7, processed via the STAC platform [54], yield a P-value below the standard significance level of 0.05, leading to the rejection of the H_0. The Friedman test confirms that five of the six EAs significantly outperform the EFS method. The BDE algorithm, however, performs no better than EFS, with both showing statistically inferior results.

Table 6: Performance of Machine Learning Algorithms using EFS-EA Optimized Feature Subsets.

Classifiers	Measurement	EFS-BWOA	EFS-BSSA	EFS-BSCA	EFS-BGA	EFS-BHHO	EFS-BDE
KNN	Accuracy	100	100	100	99.98	99.93	99.28
	Precision	100	100	100	99.98	99.93	99.29
	Recall	100	100	100	99.98	99.93	99.28
	F1-Score	100	100	100	99.98	99.93	99.28
	AUC	25.00	25.00	25.00	25.01	25.03	25.36
GNB	Accuracy	79.70	79.54	79.48	79.22	70.82	64.95
	Precision	79.72	79.55	79.48	79.22	80.77	79.26
	Recall	79.70	79.54	79.48	79.22	70.81	64.95
	F1-Score	79.71	79.55	79.48	79.22	75.46	71.39
	AUC	29.18	29.28	29.46	29.57	30.00	29.06
SVM	Accuracy	52.66	53.29	53.71	53.23	69.20	57.25
	Precision	52.68	53.35	53.75	53.26	76.11	57.29
	Recall	52.66	53.29	53.71	53.23	69.20	57.25
	F1-Score	52.67	53.32	53.73	53.25	72.49	57.27
	AUC	50.32	49.05	49.19	49.29	39.35	44.91

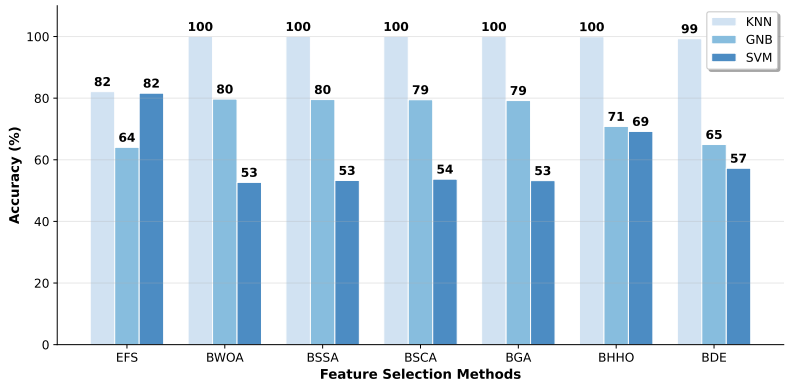


Fig. 8: Comparison of Machine Learning Algorithms on the EFS Feature Subset versus EFS-EA Optimized Subsets by Average Accuracy.

4.6 Compared to stat-of-the-art

A review of the literature reveals various methodologies for predicting U5MR. Bitew et al.[18] employed classifiers such as LGR, KNN, and RF without feature selection, achieving a baseline accuracy of 67.2%. Subsequent studies integrated feature selection to improve performance: Nasrat and Lavangnananda [21] used CSF with multiple classifiers to reach 85.47% accuracy, while Iqbal et al. [20] utilized IG to achieve 93.8%. Other works, such as Saroj et al. [31] and Tesfaye et al. [29], reported high accuracies of 95% and 94.3%, respectively, without employing feature selection. In comparison, the proposed method, which integrates EFS-EAs for feature selection with a KNN classifier, achieves a superior accuracy of 100%, outperforming all existing state-of-the-art approaches.

Table 7: Friedman Test Results for the Feature Selection Algorithms.

Algorithms	**Rank**	
BSCA	2.42857	
BWOA	2.57143	
BGA	4.07143	
BSSA	4.07143	
BHHO	4.21429	
EFS	5.28571	
BDE	5.35714	
Statistic: 2.42579	P-Value: 0.04513	H_0 is rejected

5 Conclusion and Future Works

This study presents a comprehensive evaluation of a hybrid feature selection framework, termed EFS-EA, which integrates EFS with EAs for predicting U5MR using the ADHS 2025 dataset. The experimental findings lead to two principal conclusions. First, the proposed EFS-EA framework significantly outperforms EFS, demonstrating that EA is highly effective for refining feature subsets to achieve superior predictive performance. The statistical significance of this superiority was confirmed by a Friedman test, which ranked algorithms like BSCA and BWOA consistently highest. Second, the results underscore that the feature selection strategy is a more critical determinant of performance than the choice of classifier for this domain. This is compellingly evidenced by the KNN classifier achieving perfect scores in accuracy, precision, recall, and F1-score when paired with an optimal feature subset from EFS-BWOA or EFS-BSCA. This outcome not only resolves the performance trade-offs observed in baseline machine learning models but also surpasses the results achieved by more complex deep learning architectures, such as CNNs and BiLSTMs.

While the proposed method shows exceptional promise, this study is limited by its evaluation on a single dataset, which affects the generalizability of the findings. Consequently, several avenues for future work are opened. The primary next step is to assess the generalizability of the EFS-EA framework by applying it to other public health datasets from different regions or time periods, as well as to different predictive tasks, such as disease outbreak forecasting. Furthermore, integrating the optimized feature subsets from EFS-EAs as input for deep learning architectures presents a compelling research direction, potentially enhancing the performance, efficiency, and interpretability of complex neural networks.

Authors' contribution

Sayed Zabihullah Musawi analyzed the data and did the experiments on them, and was a major contributor in writing the manuscript. Mohammad Farshi contribute in checking the results, writing part of the paper and proofreading the paper. All authors read and approved the final manuscript.

Funding

There were no funding for this research.

Conflicts of Interest

The authors declare no conflicts of interest.

Availability of data and materials

The dataset is already available on https://dhsprogram.com/methodology/survey/survey-display-468.cfm. The authors clarify the dataset used after legal regimentation.

References

[1] unicef: UNICEF Data: Monitoring the situation of children and women. https://data.unicef.org/country/afg/ (2023)

[2] UNDP: What are the Sustainable Development Goals? (2023). https://www.undp.org/sustainable-development-goals

[3] Zabihullah Musawi, S., Farshi, M., Ebrahimi Mood, S., Souri, A.: Enhancing the harris hawks optimization algorithm with ambush-based operators for feature selection in uav-based intrusion detection systems. Concurrency and Computation: Practice and Experience **37**(21-22), 70207 (2025)

[4] Choubey, D.K., Kumar, P., Tripathi, S., Kumar, S.: Performance evaluation of classification methods with PCA and PSO for diabetes. Network Modeling Analysis in Health Informatics and Bioinformatics **9**, 1–30 (2020) https://doi.org/10.1007/s13721-019-0210-8

[5] Karimzadeh, M., Moradi Dowlatabadi, P., Ghaderzadeh, A.: Unsupervised feature selection by integration of regularized self-representation and sparse coding. AUT Journal of Mathematics and Computing **5**(2), 91–109 (2024)

[6] Asghari Varzaneh, Z., Hosseini, S., Javidi, M.M.: A novel binary horse herd optimization algorithm for feature selection problem. Multimedia Tools and Applications **82**(26), 40309–40343 (2023) https://doi.org/10.1007/s11042-023-15023-7

[7] Kiran, T.R., Srisaila, A., Shankar, G.S., Sowjanya, B., Lakshmanarao, A.: Machine learning approach for diabetes prediction using genetic algorithm based feature selection. In: 2024 3rd International Conference for Innovation in Technology (INOCON), pp. 1–5 (2024). https://doi.org/10.1109/INOCON60754.2024.10511558 . IEEE

[8] Gao, Y., Zhou, Y., Luo, Q.: An efficient binary equilibrium optimizer algorithm for feature selection. IEEE Access **8**, 140936–140963 (2020) https://doi.org/10.1109/ACCESS.2020.3013617

[9] Ansari Shiri, M., Mansouri, N.: Novel correlation-based feature selection approach using manta ray foraging optimization. Computer and Knowledge Engineering **6**(1), 81–100 (2023) https://doi.org/10.22067/CKE.2023.79741.1071

[10] Saidi, R., Bouaguel, W., Essoussi, N.: Hybrid feature selection method based on the genetic algorithm and pearson correlation coefficient. Machine learning paradigms: theory and application, 3–24 (2019) https://doi.org/10.1007/978-3-030-02357-7_1

[11] Varzaneh, Z.A., Hosseini, S., Ebrahimi-Mood, S., Javidi, M.M.: A new hybrid feature selection based on improved equilibrium optimization. Chemometrics and Intelligent Laboratory Systems **228**, 104618 (2022) https://doi.org/10.1016/j.chemolab.2022.104618

[12] Hussain, K., Neggaz, N., Zhu, W., Houssein, E.H.: An efficient hybrid sine-cosine harris hawks optimization for low and high-dimensional feature selection. Expert Systems with Applications **176**, 114778 (2021) https://doi.org/10.1016/j.eswa.2021.114778

[13] El-Kenawy, E.-S., Eid, M.: Hybrid gray wolf and particle swarm optimization for feature selection. International Journal of Innovative Computing, Information and Control **16**(3), 831–844 (2020) https://doi.org/10.24507/ijicic.16.03.831

[14] Vijayanand, R., Devaraj, D.: A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network. IEEE Access **8**, 56847–56854 (2020) https://doi.org/10.1109/ACCESS.2020.2978035

[15] Akhiat, Y., Touchanti, K., Zinedine, A., Chahhou, M.: IDS-EFS: ensemble feature selection-based method for intrusion detection system. Multimedia Tools and Applications **83**(5), 12917–12937 (2024) https://doi.org/10.1007/s11042-023-15977-8

[16] Kolukisa, B., Bakir-Gungor, B.: Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis. Computer Standards & Interfaces **84**, 103706 (2023) https://doi.org/10.1016/j.csi.2022.103706

[17] Rouhi, A., Nezamabadi-pour, H.: A hybrid feature selection approach based on ensemble method for high-dimensional data. In: 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), pp. 16–20 (2017). https://doi.org/10.1109/CSIEC.2017.7940163 . IEEE

[18] Bitew, F.H., Nyarko, S.H., Potter, L., Sparks, C.S.: Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian demographic and health survey. Genus **76**, 1–16 (2020) https://doi.org/10.1186/s41118-020-00106-2

[19] Chelangat, M.: Feature selection methods and resampling techniques in survival data: determination of risk factors of under-five child mortality. PhD thesis, University of Nairobi (2021)

[20] Iqbal, F., Satti, M.I., Irshad, A., Shah, M.A.: Predictive analytics in smart healthcare for child mortality prediction using a machine learning approach. Open Life Sciences **18**(1), 20220609 (2023) https://doi.org/10.1515/biol-2022-0609

[21] Nasrat, N., Lavangnananda, K.: Implementing predictive model for child mortality in Afghanistan. In: Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 4, pp. 331–342 (2022). https://doi.org/10.1007/978-981-16-2102-4_31 . Springer

[22] Ntukidem, S., Chukwu, A., Oyamakin, O., James, C., Habimana-Kabano, I.: Trend analysis and determinants of under-5 mortality in nigeria: A machine learning approach. Asian Journal of Probability and Statistics **24**(2), 10–23 (2023) https://doi.org/10.9734/ajpas/2023/v24i2520

[23] UNICEF: Afghanistan Demographic Health Survey 2015. https://www.unicef.org/reports/every-child-alive (2018)

[24] IGME, unicef, Organization, W.H., GROUP, W.B., Nations, U.: Level and Trends in Child Mortality Report 2022 (2023). https://www.unicef.org/

[25] Organization, W.H.: World health statistics 2023: monitoring health for the SDGs, sustainable development goals (2023). https://www.who.int/publications/i/item/9789240074323

[26] UNICEF: Every Child Alive. https://www.unicef.org/reports/every-child-alive (2018)

[27] UNICEF: The State of the Worldâ€™s Children 2023 (2023). https://www.unicef.org/

[28] Pal, S.K., Vijay, J., Patel, K.K.: Prevalence of under-5 mortality and its associated risk factors in afghanistan. Children and Youth Services Review **120**, 105801 (2021) https://doi.org/10.1016/j.childyouth.2020.105801

[29] Tesfaye, B., Atique, S., Elias, N., Dibaba, L., Shabbir, S.-A., Kebede, M.: Determinants and development of a web-based child mortality prediction model in resource-limited settings: a data mining approach. Computer methods and programs in biomedicine **140**, 45–51 (2017) https://doi.org/10.1016/j.cmpb.2016.11.013

[30] Mohapatra, S.K., Prasad, S., Habtemariam, G.M., Siddique, M.: Identify determinants of infant and child mortality based using machine learning: Case study on ethiopia. Big Data Analytics and Machine Intelligence in Biomedical and

Health Informatics: Concepts, Methodologies, Tools and Applications, 21–45 (2022) https://doi.org/10.1002/9781119792376.ch2

[31] Saroj, R.K., Yadav, P.K., Singh, R., Chilyabanyama, O.N.: Machine learning algorithms for understanding the determinants of under-five mortality. BioData mining **15**(1), 20 (2022) https://doi.org/10.1186/s13040-022-00308-8

[32] Adegbosin, A.E., Stantic, B., Sun, J.: Efficacy of deep learning methods for predicting under-five mortality in 34 low-income and middle-income countries. BMJ open **10**(8), 034524 (2020) https://doi.org/10.1136/bmjopen-2019-034524

[33] Ashwini, C., Bose, S.R., Padmavathy, M.D., Raj, C., Ajay, J.C.: Predicting child mortality with diverse regression algorithms using a machine learning approach. Advancing Intelligent Networks Through Distributed Optimization, 329–352 (2024) https://doi.org/10.4018/979-8-3693-3739-4.ch017

[34] Mbunge, E., Fashoto, S.G., Muchemwa, B., Millham, R.C., Chemhaka, G., Sibiya, M.N., Dzinamarira, T., Buwerimwe, J.: Application of machine learning techniques for predicting child mortality and identifying associated risk factors. In: 2023 Conference on Information Communications Technology and Society (ICTAS), pp. 1–5 (2023). https://doi.org/10.1109/ICTAS56421.2023.10082734 . IEEE

[35] Mumtaz, S., Bahk, J., Khang, Y.-H.: Current status and determinants of maternal healthcare utilization in Afghanistan: analysis from Afghanistan demographic and health survey 2015. PloS one **14**(6), 0217827 (2019) https://doi.org/10.1371/journal.pone.0217827

[36] Tsai, C.-F., Sung, Y.-T.: Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches. Knowledge-Based Systems **203**, 106097 (2020) https://doi.org/10.1016/j.knosys.2020.106097

[37] Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., Alonso-Betanzos, A.: Ensemble feature selection: homogeneous and heterogeneous approaches. Knowledge-Based Systems **118**, 124–139 (2017) https://doi.org/10.1016/j.knosys.2016.11.017

[38] Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A.: Data classification using an ensemble of filters. Neurocomputing **135**, 13–20 (2014) https://doi.org/10.1016/j.neucom.2013.03.067

[39] Mirjalili, S., Lewis, A.: The whale optimization algorithm. Advances in engineering software **95**, 51–67 (2016) https://doi.org/10.1016/j.advengsoft.2016.01.008

[40] Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., Mirjalili, S.M.: Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. Advances in engineering software **114**, 163–191 (2017) https://doi.org/10.1016/

j.advengsoft.2017.07.002

[41] Mirjalili, S., Mirjalili, S.: Genetic algorithm. Evolutionary algorithms and neural networks: theory and applications, 43–55 (2019)

[42] Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., Chen, H.: Harris hawks optimization: Algorithm and applications. Future generation computer systems **97**, 849–872 (2019) https://doi.org/10.1016/j.future.2019.02.028

[43] Deng, W., Shang, S., Cai, X., Zhao, H., Song, Y., Xu, J.: An improved differential evolution algorithm and its application in optimization problem. Soft Computing **25**, 5277–5298 (2021) https://doi.org/10.1007/s00500-020-05527-x

[44] Cunningham, P., Delany, S.J.: K-nearest neighbour classifiers-a tutorial. ACM computing surveys (CSUR) **54**(6), 1–25 (2021)

[45] Agrawal, P., Abutarboush, H.F., Ganesh, T., Mohamed, A.W.: Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). Ieee Access **9**, 26766–26791 (2021)

[46] Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**, 273–297 (1995)

[47] Vapnik, V.N.: The support vector method. In: International Conference on Artificial Neural Networks, pp. 261–271 (1997). https://doi.org/10.1007/BFb0020166 . Springer

[48] Sun, S., Huang, R.: An adaptive k-nearest neighbor algorithm. In: 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, vol. 1, pp. 91–94 (2010). https://doi.org/10.1109/FSKD.2010.5569740 . IEEE

[49] Frank, E., Trigg, L., Holmes, G., Witten, I.H.: Naive bayes for regression. Machine Learning **41**, 5–25 (2000) https://doi.org/10.1023/A:1007670802811

[50] Murphy, K.P.: Naive bayes classifiers. University of British Columbia **18**(60), 1–8 (2006)

[51] Liu, Y., Zhao, Q.: Ensemble learning. In: Handbook on Computer Learning and Intelligence: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation, pp. 635–660. World Scientific, ??? (2022)

[52] Al-Selwi, S.M., Hassan, M.F., Abdulkadir, S.J., Muneer, A., Sumiea, E.H., Alqushaibi, A., Ragab, M.G.: Rnn-lstm: From applications to modeling techniques and beyondâ€"systematic review. Journal of King Saud University-Computer and Information Sciences **36**(5), 102068 (2024)

[53] Siami-Namini, S., Tavakoli, N., Namin, A.S.: The performance of lstm and bilstm in forecasting time series. In: 2019 IEEE International Conference on Big Data

(Big Data), pp. 3285–3292 (2019). IEEE

[54] Rodríguez-Fdez, I., Canosa, A., Mucientes, M., Bugarín, A.: STAC: a web platform for the comparison of algorithms using statistical tests. In: Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2015)