Mining and Interpretation of Critical Aspects of Infant Health Status Using Multi-Objective Evolutionary Feature Selection Approaches

JAYASHREE PIRI[1], PUSPANJALI MOHAPATRA[1], (Member, IEEE), DEBABRATA SINGH[2],
DEBABRATA SAMANTA[3], (Member, IEEE), DILBAG SINGH[4], (Senior Member, IEEE),
MANJIT KAUR[4], (Member, IEEE), AND HEUNG-NO LEE[4], (Senior Member, IEEE)
[1]Department of CSE, International Institute of Information Technology, Bhubaneswar, Bhubaneswar, Odisha 751029, India
[2]Department of Computer Application, Institute of Technical Education and Research (ITER), Siksha 'O' Anusandhan (SOA) Deemed to be University, Bhubaneswar, Odisha 751030, India
[3]Department of Computer Science, CHRIST University, Bangalore, Karnataka 560029, India
[4]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Heung-No Lee (heungno@gist.ac.kr)

ABSTRACT The rate of infant mortality (IMR) in a population under one year of age is a marker for infant mortality. It is a major sensitive marker of a community's overall physical health. Protecting the lives of newborns has become a challenging issue in public health, development programs, and humanitarian initiatives. Almost 10.1% infants died in the United States of America (USA) in 2021. Therefore, this paper aims to extract and understand the various influential factors causing infant deaths in the USA. A crowding distance-based multi-objective ant lion optimization (MOALO-CD) is proposed here with statistical evidence for feature selection. The proposed technique is compared with competitive metaheuristic models such as multi-objective genetic algorithm based on crowding distance (MOGA-CD), multi-objective filter approaches, and recursive feature elimination. Various machine learning classifiers are applied to the selected feature subset obtained from MOALO-CD on the USA's infant dataset. Extensive experimental results indicate that the proposed model outperforms the existing metaheuristic approaches in terms of Generational Distance, Inverted Generational Distance, Spread, and Hyper volume. Also, the comparative analysis of various machine learning models reveals that random forest achieves significantly better performance on the feature subset obtained from MOALO-CD.

INDEX TERMS Infant mortality, feature selection, multi objective optimization, genetic algorithm, ant lion optimization.

I. INTRODUCTION

Infants are human offspring who range in age from birth to one year old. Child deaths under the age of one year are called infant deaths. This mortality rate is estimated according to the number of children under one year old, which is the number of deaths/1,000 live births. Infant mortality can be categorised into three classes: perinatal mortality, neonatal mortality, and post-neonatal mortality. Fetal death is the mortality of the fetus during the gestation of 22 weeks to a newborn's birth or death up to one week after birth (perinatal). Neonatal deaths occur within 28 days after the baby is part of the family (Neonatal), and child deaths occur within 29 days to one year after the baby is part of the family (Postneonatal). Malnutrition, respiratory disease, maternity complications, unexpected infant death disorders, and home complications are the main causes of post-neonatal mortality. India recorded an IMR of 44/1000 live births[1] in 2016. In 2017, 802,000 infant deaths occurred due to a lack of access to water, sanitation,

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

[1]https://www.indiaspend.com/indias-infant-mortality-down-42-in-11-years-yet-higher-than-global-average/

appropriate nutrition, and necessary health services. The US mortality rate is 5.8 [2] which, in the face of extreme poverty, extreme risky health outcomes, and a fragmented healthcare system, is substantially more advanced than other comparable countries. Further, there are significant variations in infant death rates between various racial groups, with the most striking disparities between babies born to black women and white women. In the past 50 years, substantial improvements have been made to reduce infant deaths in the USA, but more needs to be done.

Infant mortality relies heavily on factors such as the mother's health, safety, and access to healthcare facilities, social and economic circumstances, and policies on public healthcare. It is therefore a significant measure of a country's well-being. Mani *et al.* [1] have used data mining techniques for routine collection of data from 299 admitted infants in the Monroe Carell Jr. Children's Hospital at Vanderbilt. This data was tested by nine ML-based models for the early diagnosis of late-onset sepsis. Wilson *et al.* [2] have focused on identifying a complex Bayesian network of causes that contribute to neonatal mortality, including preterm birth, maternal death, movement, and breathing at birth. Here the authors also aim to estimate the causal effects of neonatal mortality from the Global Network Maternal and Child Health Registry using logistic regression models. Rittenhouse *et al.* [3] have focused on the development of algorithms that integrate maternal factors linked to SGA (Small Gestational Age). The data used in this research was assembled from an ongoing obstetric cohort of early pregnancy ultrasounds for the evaluation of GA (Genetic Algorithm) in Lusaka, Zambia. The results of this study include a mixture of six parameters that have been correctly classified as being above 94% of newborns, such as baby weight, twin birth, maternal height, labor hypertension, and HIV serostatus, and that reached 0.9796 AUC. Saravanou *et al.* [4] in their research are interested to find how various socioeconomic factors like race or mother's education can improve predicting infant mortality. They used a highly imbalanced dataset with complete birth certificate information from 2000–2002 in the USA, for solving binary infant mortality classification problem. Kefi *et al.* [5] have applied ML methodologies for prediction of short-term mortality in the NICU so that better clinical decisions can be made by doctors. The data supplied in this analysis were obtained from the Medical Information Mart III (MIMIC-III) for Intensive Care, which was developed by the MIT Laboratory for Computational Physiology. Unfortunately, the prognosis of infant health utilising the characteristics selected by evolutionary techniques is rarely addressed.

The number of infants in the USA born prematurely and with low birth weight continues to rise, and as infants born too early or too young have higher mortality rates, this ongoing increase has a major impact on infant mortality. However,

substantial disparities between racial and ethnic groups in child mortality rates continued to rise and not all groups have received social and health gains equally. Therefore, in this proposal, we are analyzing U.S. Territories Birth Data (2018) (https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm) recorded by the Centers for Disease Control (CDC) and the National Center for Health Statistics (NCHS) for mining sensitive causes of infant death, as we found from literature that there was no or very limited feature selection (FS) work on this dataset. Again, FS tasks based on evolutionary multi-objective techniques have not been explored in this dataset. This is the motivation behind the selection of the U.S. Territories Birth Data to extract the critical factors related to infant health.

Mothers, infants, and children are critical public health priorities in the United States. Predicting public health concerns for families, communities, and the healthcare system depends on how well children are doing now. The main objective is to mine the most influential factors to predict whether an infant will survive or not. Removing non-important factors from the dataset is a significant goal to achieve. The field of metaheuristic algorithms has grown significantly in the last two centuries as an alternative to real-world optimization problems. They are capable of performing well in circumstances where exact optimization methods fail to produce a good result. Metaheuristic techniques can provide high-standard solutions in comparatively much less time than conventional optimizers for challenging optimization issues. Applications for metaheuristics are available in a variety of areas, including finance, planning, scheduling, and engineering design. These Metaheuristics/Population-based strategies have been used as a wrapper in recent years in solving FS tasks [6], and have proven their strengths. Various population-based approaches have been used to solve FS problems, including several recent algorithms: Firefly and grey wolf optimization [7], and many variants of genetic algorithms [8], [9], etc. This motivated us to propose a crowding distance-based MOALO (MOALO-CD) for selecting important features from the infant dataset.

The major contributions of this paper are:
1) A crowding distance-based multi-objective ant lion optimization (MOALO-CD) is proposed for the feature selection (FS) task.
2) The concept of crowding distance is used to maintain diversity and to select the most unique feature subset for mining the factors affecting infant health status.
3) Various multi-objective performance metrics are used to evaluate the quality of Pareto fronts obtained from the proposed model.
4) To evaluate the role of selected features for the classification of infants as survived or not, various state-of-the-art classifiers are trained on the infant's dataset by considering the same selected features only.

The remaining portion of this paper is organized as follows: Section II discusses the materials and methods. Section III demonstrates the performance analysis. Section IV presents
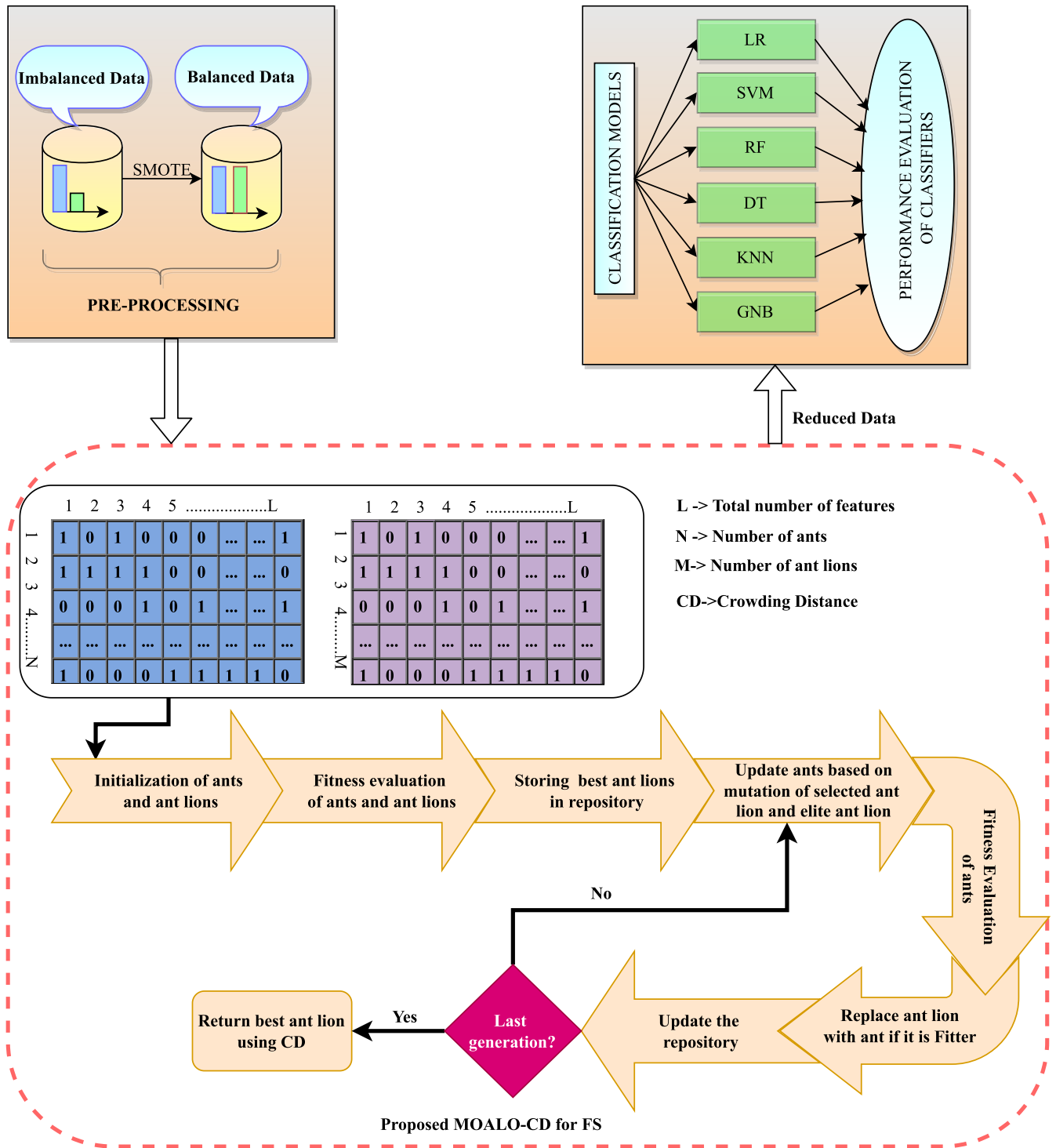
---

[2] https://worldpopulationreview.com/countries/infant-mortality-rate-by-country/

FIGURE 1. Architecture of the proposed model.

the discussion of the obtained results. Section V concludes the proposed work and also presents the future directions.

II. MATERIALS AND METHODS

The suggested system is intended to predict infant mortality by using the most influential characteristics of the 2018

birth data file for the US Territories. The architecture of the proposed model is given in FIGURE. 1. This section is split into six subsections. First, the details of the 2018 birth data file for the US Territories are explained. The second subsection introduces the notion of multi-objective optimization. The third subsection explains the proposed feature selection

methodology. A short description of the FS methods used for comparison purposes is presented in the fourth section. Explanations of the basic classifiers used to decide whether or not the infant survives depending on the features selected are given in the fifth section. Finally, the setups for conducting all the experiments are specified.

A. DATASET

The 2018 birth data file of the US Territories recorded by the CDC and the NCHS is the dataset used in this proposed research. There are a total of 25,919 records, each having a length of 1330 including filler. The different characteristics present in this dataset are: i) Demographic characteristics, ii) Medical and Public Services Utilization, iii) Maternal Behavior and Health Characteristics, and iv) Infant Health Characteristics. After removing fillers, recode variables, flag variables, and some irrelevant columns such as Paternity_Acknowledged, Mother's_Marital_Status_Imputed, etc., 99 relevant columns, including a class column, are derived and set for further research. The class column ''Infant_livin'' has three values: Yes, No, and Unknown. We removed the records having an unknown Infant_living value and finally, we began our work with a preprocessed dataset of 25868 rows and 99 columns. The number of records having infant_living = Yes (25799) is greater as compared to the number of records of infant_living = No (69), so the dataset is a highly imbalanced one. In these cases, it may be possible to misinterpret the classification model developed using traditional learning algorithms. This is because the ML algorithms are usually constructed by reducing errors to boost performance (https://www.analyticsvidhya.com/blog/2017/03/imbalanced-data-classification/). Consequently, the distribution of classes is not taken into account. There are several approaches [10] for handling imbalanced datasets, such as over-sampling, under-sampling, bagging based methods, etc. We have used SMOTE (Synthetic Minority Oversampling Technique) for balancing our referenced dataset and the number of samples after SMOTE is 51598. SMOTE [11] is a very popular solution for tackling the problem of imbalance. This helps to balance the distribution of classes by replicating them randomly through instances of minority classes. SMOTE synthesizes new instances of minorities as shown in FIGURE. 2 within existing instances of minorities. It produces records of virtual training for the minority class by linear interpolation. In each case, a random collection of one or more of the k-nearest minority class neighbours can produce these synthetic training records. Many classifiers may be used to process the results following the over-sampling procedure.

B. MULTI-OBJECTIVE OPTIMIZATION (MOO)

MOO [12] means that more than one objective must be optimized concurrently in a multiple-criterion decision-taking environment. There is no specific solution to the problem of nontrivial multi-target optimization that can optimize each target in parallel since the goal functions are contradictory.
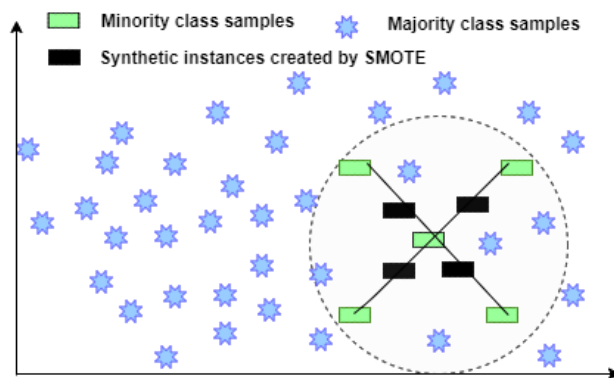

FIGURE 2. SMOTE: A dataset balancing technique.

Algorithm 1 SMOTE Algorithm

1: Set A = minority class set
2: **for each** $x \in \mathcal{A}$ **do**
3:     Calculate nearest k-neighbors of x by measurement of the Euclidean distance of x from each other in set A.
4: **end for**
5: Fix N % The sampling rate as per the unbalanced ratio.
6: **for each** $x \in \mathcal{A}$ **do**
7:     Create a set $A1$ by choosing N instances (i.e. $x1, x2, \ldots xn$) randomly from its k-nearest neighbors.
8:     **for each** example $x_k \in \mathcal{A}1(k = 1, 2, \ldots., N)$ **do**
9:         Generate new example by using the following formula:

$$x' = x + rand(0, 1) \times \|x - x_k\| \qquad (1)$$

10:     **end for**
11: **end for**

So, finally, a non-dominated solution set is achieved as the outcome. There are 2 solutions to MOO problems.

1) To acquire the optimal results, it first discovers several non-dominated solutions and then employs advanced understanding.
2) A composite function is formulated by using the chosen preference vector. Then any single objective optimization algorithm is applied to the developed composite function to discover the best one.

The definition of ''Pareto superiority/dominance'' is used as follows when finding optimal solutions in multi-objective optimization, meaning that all objectives can be taken into consideration at the same time. A solution c1 dominates another solution c2 if and only if:

- c1 is not weaker than c2 in each objective.
- c1 is specifically superior to c2 for at least one of the objectives.

Pareto optimal options are solutions not dominated by any member in the field of search. Let the solutions to the hypothetical problem given in FIGURE. 3 are to be minimized with two goals: f1 and f2. Solution P has little f1 but a large f2,

while solution Q contains large f1 and low f2. When it is necessary to minimize all objectives, we cannot conclude that solution P is better than solution Q or vice versa. So the two solutions, P and Q, are Pareto optimal or non-dominated solutions because none of them is dominated by other solutions. A Pareto front is obtained by joining all the Pareto optimal solutions in objective space.
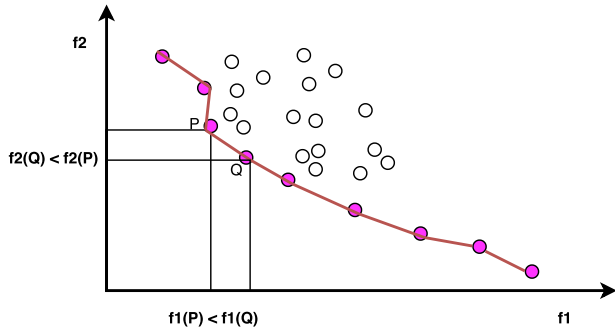

FIGURE 3. Non-dominance principle.

C. PROPOSED METHOD FOR FEATURE SELECTION (FS)

Mirjalili suggests an ant lion optimizer (ALO) [13] based on ant lion hunting in nature. Ant-lions chase for larvae, and the larva excavates a cone-formed hole (plot) in the sand by walking along a circular road. The larva hides and waits until the ants are trapped in the ditch. The ant lions seek to capture the prey while it is in the pit and throw sand at the pen. The victim is sucked into the ground, where it is then devoured. Ant-lions throw out the remnants of the trap and amend it to continue hunting.

In this study, a binary version of the ant lion optimization [14] is used because in the FS problem the solutions are restricted to {0,1}. ALO is very famous for preventing local optimums, exploitation, refined exploration, and convergence. Flexible border reduction technology and ALO elitism lead to high-speed exploitation and convergence. The random walk and roulette wheel selection techniques of ALO maintain population diversity. Again, the multi-objective version of ALO is used for finding the best feature subsets of the referenced dataset with regard to fewer attributes and maximal classification efficiency. ALO proved itself as an excellent optimizer in several problems, including: Optimal Power Flow, Load Dispatch, Feature Selection, Image Processing, Computer Vision, Neural Networks, and Medical Applications, etc. ALO provides massive benefits, including: easy conception, simplicity of implementation, and a small number of controlling variables [15]. This provoked us to explore a multi-objective variant of ALO in the era of FS tasks. The working principle of MOALO-CD includes the following functions:

1) **Encoding**
   Each ant and ant lion location is encoded as an L-bit pattern in which L is the number of characteristics present in the original dataset. The appearance or

missing of a particular variable in the competitor substring is determined by the bit position (1: present and 0: absent). The pictorial representation of the encoding scheme is shown in FIGURE 1.

2) **Fitness Evaluation**
   Each ant and ant lion is evaluated by using two fitness criteria/objectives: #features (Fitness1) and the accuracy of the wrapper classifier (Fitness2).

$$Fitness1(X) = \sum_{i=1}^{L} x_i \quad (2)$$

where X is pattern of L bit length.

$$Fitness2(X) = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

where *TP*, *FN*, *TN*, and *FP* show true positive, false negative, true negative, and false positive, respectively. To determine the Fitness2 for each ant and ant lion, the following processes are adopted:

- Initially, a compact dataset is created by extracting the properties from the relevant positional vector cells containing 1.
- The Fitness2 is then determined using the KNN algorithm with a 10-fold CV. The total samples are arbitrarily separated into ten batches for training and testing.

3) **Repository Maintenance**
   The archive will be updated by the best ant lions for any population shift. An individual ant lion can reside in the repository if it is not dominated by the ant lions already present in the archive. Former archive members should be erased if they are dominated by this inclusion. The external repository may overflow as its size is fixed by the user. If overflow emerges, individuals are organized in decreasing order [16] of their CD values. From the bottom 5% of the archive, pick any alternative that includes the most crowded archive substitutes [17] and remove it.

4) **Position Update of Ants**
   To carry on with the upcoming repetition, each ant picks an ant lion haphazardly by the roulette wheel principle and twists it by equation 4 (RW1) [18].

$$x_{out}^d = \begin{cases} x_{in}^d & \text{if } r1 \geq r \\ round(r2) & \text{otherwise} \end{cases} \quad (4)$$

where $r1$ and $r2$ are two randomly [0, 1] generated numbers, x_{out}^d is the d dimension value of the output vector from mutation, x_{in}^d is the d dimension value of the input vector for mutation (shown in FIGURE 4), and r is the mutation rate computed by equation 5.

$$r = 0.9 + \frac{-0.9 \times (i-1)}{GEN - 1} \quad (5)$$

Here, *i*: current generation and *GEN*: the total generations to complete. Afresh, the best solution (ant lion)
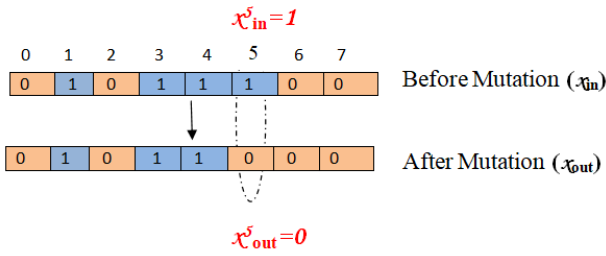
FIGURE 4. Mutation.

from the repository is picked by the roulette wheel and mutation using equation 4 is carried out on it, resulting in RW2. Then the new position of the ant is set by the outcome of the crossover among RW1 and RW2, which is done using the equation 6.

$$x^d = \begin{cases} x_1^d & \text{if } rand \geq 0.5 \\ x_2^d & \text{otherwise} \end{cases} \quad (6)$$

where x^d is the crossover result between vectors x_1^d and x_2^d at dimension d.

5) **Catching Prey/Replacement of Ant lion with Fitter Ant**

After updating the position of each ant, an ant lion can be replaced by its respective ant if it is suitable.

The details of feature selection using binary multi-objective ant lion optimization are given in Algorithm 2.

D. FS METHODS USED FOR COMPARISON

The following is a list of FS approaches used here to compare the efficiency of the proposed FS mechanism:

1) **MOGA-CD [19]** GA is an important natural selection-inspired evolutionary search algorithm. The working procedure of GA is to create a group of individuals where each individual is the nominee solution to a specific problem. Every individual is assessed by a fitness function to determine the quality of the solution. In each generation, the best or fittest individuals are selected and recombined to create offspring so that the population progressively finds better and better alternatives. The initialization of chromosomes in MOGA is structurally the same as the encoding of ants and ant lions in MOALO. In this case, we used the same fitness functions to evaluate each chromosome. At each iteration, the best individuals are stored in an external repository, and the repository is updated as given in section II-C. A detailed algorithm for MOGA-CD based FS is given below:

2) **MOGA-CD and MOALO-CD Filter Version**

To compare the performance of the proposed wrapper-based FS mechanism MOALO-CD, we have implemented the filter versions of the two multi-objective evolutionary algorithms, MOGA and

---

Algorithm 2 MOALO-CD for Feature Selection

Input:
A – External archive
N - No of ants
M – No of ant lions
GEN – No of generations (100)
Fitness1 – No of features in the candidate feature subset.
Fitness2 – Classification accuracy of the wrapper classifier

1: **for** $i \leftarrow 1$ *to N* **do**
2:     Initialize $ant[i] \in \{0, 1\}$ randomly
3:     Calculate Fitness1 and Fitness2 for $ant[i]$
4: **end for**
5: **for** $i \leftarrow 1$ *to M* **do**
6:     Initialize $ant\_lion[i] \in \{0, 1\}$ randomly
7:     Calculate Fitness1 and Fitness2 for $ant\_lion[i]$
8: **end for**
9: Store the fittest ant_lions (non dominated solutions) into external achieve A.
10: **for** $COUNT \leftarrow 1$ *to GEN* **do**
11:     Calculate mutation rate r using equation 5.
12:     **for each** ant_j **do**
13:         Select an ant_lion (RW) at random using Roulette Wheel selection.
14:         Apply mutation on RW with mutation rate r, resulting in RW1.
15:         Select a fittest ant_lion from A by roulette wheel and mutate it to get RW2.
16:         Set new position of ant_j to the output of CROSSOVER(RW1, RW2).
17:     **end for**
18:     Evaluate all ants (calculate Fitness1 and Fitness2).
19:     Restore an ant_lion with its respective ant if it is fitter(catching prey).
20:     Update the repository A by newly found fitter ant_lions
21: **end for**
22: Use CD measure to obtain best out of bests from the repository.

MOALO, by taking the following two fitness functions:

$$Minimize\ Fitness1(X) = \sum_{i=1}^{L} x_i \quad (7)$$

where X is pattern of L bit length.

$$Maximize\ Fitness2(X) = \frac{1}{|B|} \sum_{i=1}^{|B|} |corr(b_i, y)| \quad (8)$$

where Fitness2 (X) is the average correlation between the selected feature subset B (consists of all those features of X where $x_i = 1$) and the corresponding class attribute y [20].

Algorithm 3 MOGA-CD for Feature Selection

    Input:
A – External archive
N – Population size
GEN – No of generations (100)
Fitness1 – No of features in the candidate feature subset.
Fitness2 – Classification accuracy of the wrapper classifier

1: **for** $i \leftarrow 1$ to N **do**
2:     Initialize P[i] randomly
3: **end for**
4: **for** $i \leftarrow 1$ to N **do**
5:     Calculate Fitness1 and Fitness2
6: **end for**
7: Store the non dominated solutions found in P into external achieve A.
8: **for** $COUNT \leftarrow 1$ to GEN **do**
9:     Perform Roulette Wheel selection to choose fittest individual to recombine.
10:    Perform two – point crossover
11:    Perform mutation
12:    Evaluate updated population P (calculate Fitness1 and Fitness2)
13:    Update the repository A
14: **end for**
15: Use CD measure to obtain best out of bests from the repository.

---

3) Recursive Feature Elimination (RFE)

The RFE [21] selection process is fundamentally a recursive process, which classifies features according to their significance. The importance value of each feature is computed, and an irrelevant one is discarded at each iteration. Another option, not used here, is to each time delete a group of features to speed up the operation. The recursion is required because, for such steps, when measured using another sub-set of features during the elimination period, the relative value of each feature will alter substantially. To create a final ranking, the (inverse) order where features are eliminated is used. The feature selection method only takes into account the first n attributes in this ranking.

E. CLASSIFICATION

The next task, after getting important features from the infant dataset, is to build classification models to train the machine so that it can predict whether an infant will survive or not, depending upon the selected features. In this research, we have used six state-of-the-art classification techniques, namely Logistic Regression [22], KNN [22], Decision Tree [23], Random Forest (RF) [24], Gaussian

Naive Bayes,[3] and Support Vector Machine [5], [24] to classify the reduced datasets. During the analysis of the infant data set, every algorithm was checked with ten-fold cross-validation. Algorithm performance was recorded and subsequently tabled.

F. EXPERIMENTAL SETUPS

In this study, we used the 2018 birth data file of the US Territories recorded by the CDC and the NCHS. For more details about the dataset, refer to section II-A. To reduce the dimensionality of the dataset and to get the most influential factors causing infant death, we used MOALO-CD as a feature selection mechanism. All the experiments are executed on Python 3.7 using a PC with an Intel Core i3-7020U CPU @ 2.30 GHz and 4.00 GB RAM. The parameter settings of feature selection techniques are given in Table 1.

To contrast, the proposed method for selecting features from infant data with the benchmark methods discussed in section II-D, the below mentioned multi-objective performance indicators [25] are used.

1) GENERATIONAL DISTANCE (GD)

$$GD(A, PF) = \frac{1}{|A|} \left(\sum_{a \in A} \min_{r \in PF} \|F(a) - F(r)\|^p \right)^{\frac{1}{p}} \quad (9)$$

where A is the content of the external archive after 100th generations. The F(X) is a point in the objective space whose co-ordinate values are (*Fitness1(X), Fitness2(X)*) and X is any solution (*r or a*) lying in the true pareto front (PF) or calculated pareto front (A) respectively. PF is the representative of the True Pareto front. As in most practical problems, the true PF is unknown, so the following steps have been followed here to find an approximation of the true PF [26].

a) Run MOALO-CD and MOGA-CD on the problem 150 times.

b) All the non-dominated solution sets obtained in the above step are combined into a single solution set S. Then the front containing only the solutions from S is used as an approximation of the true Pareto front.

2) INVERTED GENERATIONAL DISTANCE (IGD)

$$IGD(A, PF) = \frac{1}{|PF|} \left(\sum_{i=1}^{|PF|} \min_{a \in A} \|F(a) - F(i)\|^p \right)^{\frac{1}{p}} \quad (10)$$

For both GD and IGD calculation $p = 2$ is taken.

3) HYPERVOLUME (HV)

The HyperVolume measure can be explained as the volume of the area in the objective space dominated by the calculated PF, A, and delimited from above by a reference point $rp \in R^m$ such that for all $a \in A$, a dominates rp.

---

[3]https://medium.com/hackernoon/implementation-of-gaussian-naive-bayes-in-python-from-scratch-c4ea64e3944d

Methods	Parameters	Values
MOGA-CD(FILTER and WRAPPER)	Population size, N	50
	Maximum number of generations, GEN	100
	External Archive size	50
	Crossover rate, CR	0.8
	Mutation rate, MR	0.01
	No of ants	50
MOALO-CD(FILTER and WRAPPER)	No of ant lions	50
	Maximum number of generations, GEN	100
	External Archive size	50

4) SPREAD

Spread (*S*) can be computed as:

$$S = \sqrt{\sum_{j=1}^{O} \left(max_{i=1}^{|A|} Fitness_j^i - min_{i=1}^{|A|} Fitness_j^i \right)^2} \quad (11)$$

Here, *O* is the number of fitness functions (here *O* = 2). A bigger value of this measure indicates a better spread of solutions.

In this article, we have used the classification performance assessment measures [27] [4] listed in Table 2 to evaluate the performance of all classifiers.

III. PERFORMANCE ANALYSIS

First, wrapper-based MOALO-CD and MOGA-CD optimization algorithms are executed to find the best feature subsets by using the 10-fold CV with the KNN method to get the second fitness function (Accuracy) of each candidate feature subset. During the optimization process, cross-validation is used to avoid the over-fitting problem.

The True Pareto front along with the same obtained from wrapper based MOALO-CD and MOGA-CD are shown in FIGURE. 5. There are 5 non-dominated solutions present in the true PF, whereas both PFs obtained from MOALO-CD and MOGA-CD contain 4 solutions each. In the case of MOALO-CD, there are two non-dominated solutions present in the same location [42, 99.5] but having a different subset of features. The performance of the two above discussed methods is evaluated by using four multi-objective performance indicators, explained in section II-F and values are given in Table 3. The non dominated set output from MOALO-CD is better as the GD and IGD values are less and the HV value is more as compared to the same obtained from MOGA-CD. But the portion of the objective space covered by the MOGA-CD Pareto front is more than that of the MOALO-CD Pareto front.

Secondly, filter based MOALO-CD and MOGA-CD methods are implemented using the fitness functions described by equations 7 and 8. The resulting Pareto fronts, along with the true PFs are shown in FIGURE. 6. The True PF has only one solution and that is far away from the fronts obtained
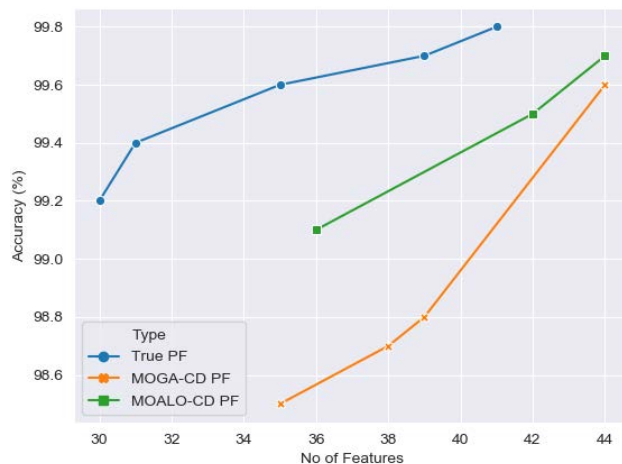
[4]https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

FIGURE 5. Pareto fronts obtained from wrapper based methods.

from the above-mentioned filter methods. The cardinality of the solutions present in the Pareto fronts is less (2) in MOALO-CD as compared to MOGA-CD (4). The values of the multi-objective performance indicators for the filter methods are given in Table 4. The MOALO-CD Filter version also outperforms in terms of GD, IGD, and HV values.

Here, we have used the Student t-Test to examine the IGD figures of the four approaches (MOALO-CD Wrapper vs MOGA-CD Wrapper and MOALO-CD Filter vs MOGA-CD Filter) acquired from 30 different trials to identify the optimal wrapper and filter method, as well as for solid affirmation of convergence demonstration. Table 5 shows the statistical outcome of a two-tailed t-test at a level of significance of 0.05. The outcomes of '*Technique*1 *vs Technique*2' are expressed as ">>" or "≪" or "==", indicating whether the first technique is substantially superior than, markedly inferior than, or substantially equivalent to Technique 2. The MOALO-CD exceeds the MOGA-CD in both the filter and wrapper versions, according to the findings. Hence, we may say that the suggested MOALO-CD has superior optimization ability for the FS issue in the US Territories Birth data.

The number of features in the best solutions picked by MOALO-CD and MOGA-CD wrapper versions is 44 and 44, respectively, by using the crowding distance measure. So, while implementing the non-evolutionary Recursive Feature Elimination (RFE) algorithm, we have supplied the value for

TABLE 2. List of classification performance measures.

Performance Measures	Formula
Accuracy	$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$
Sensitivity	$Sensitivity = \frac{TP}{TP+FN}$
Specificity	$Specificity = \frac{TN}{TN+FP}$
Gmean	$Gmean = \sqrt{Sensitivity \times Specificity}$
F-score	$F-score = \frac{(2 \times Sensitivity \times Specificity)}{(Sensitivity+Specificity)}$
ROC	ROC curve is a graphical representation of sensitivity against (1-specificity).
Precision	$Precision = \frac{TP}{TP+FP}$
FPR (False Positive Rate)	$FPR = \frac{FP}{FP+TN}$
MCC (Matthews Correlation Coefficient)	MCC$=\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (FN+TN) \times (FP+TN) \times (TP+FN)}}$
ERR (Error Rate)	$ERR = \frac{FP+FN}{TP+FP+FN+TN}$

TABLE 3. Performance comparison of multi-objective wrapper methods.

Methods/Measures	GD	IGD	Spread	HV
MOGA-CD WRAPPER	0.105	0.101	**0.34**	0.21
MOALO-CD WRAPPER	**0.104**	**0.09**	0.32	**0.22**

TABLE 4. Performance comparison of multi-objective filter methods.

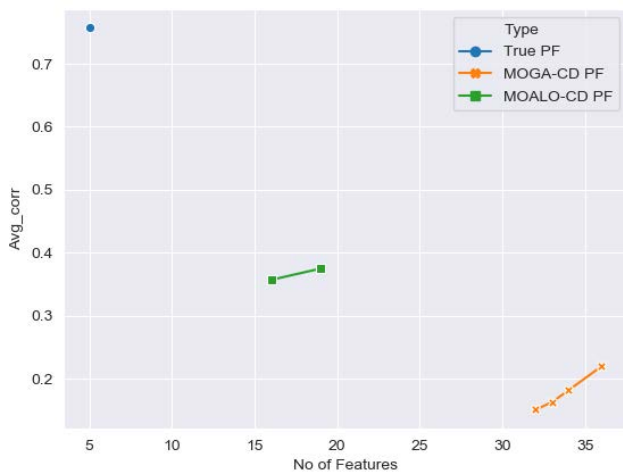| Methods/Measures	GD	IGD	Spread	HV
MOGA-CD FILTER	1.06	1.008	**0.50**	0.26
MOALO-CD FILTER	**0.7**	**0.66**	0.47	**0.32**

FIGURE 6. Pareto fronts obtained from filter based methods.

the required number of features, which is equal to 44, which is the average number of features resulting from the above two methods.

After getting the five sets of features by MOGA-CD WRAPPER, MOALO-CD WRAPPER, MOGA-CD FILTER, MOALO-CD FILTER, and RFE methods, we performed our next experiment of classification with these five reduced datasets along with the original dataset. Here we have subjected the above six (including original) datasets to six different states of art classifiers. The experimental results along with the performance measures after repeating the 10 fold cross-validation 3 times are outlined in Table 6, 7, 8, 9, 10, and 11.

By observing Table 9, and 10, we can infer that most of the classification models (LR, RF, DT, and KNN) performed well on the reduced dataset produced by **MOGA-CD Filter**

approach. Although the number of features present in the best solution of the MOALO-CD Filter approach is less (16) as compared to the MOGA-CD Filter method (36), it is missing some very important features like breastfeed_at_discharge, mother's_edu, cigaratte_before_pregnancy, mother's_height, and pre_preg_diabetes, etc. The feature set selected by the MOGA-CD Filter contains these features, and they influence the infant's health status.

Similarly, a performance comparison is made by considering the datasets obtained from the MOALO-CD Wrapper and MOGA-CD Wrapper using Table 7, and 8. Experimental results reveal that all the classifiers, LR, SVM, RF, DT, KNN, and GNB, showed superior performance when they considered the reduced dataset generated by **MOALO-CD Wrapper** version. So, for further experimental comparisons, 3 datasets are taken into consideration:

1) RDS1: Reduced Dataset obtained From RFE.
2) RDS2: Reduced Dataset obtained From MOGA-CD Filter.
3) RDS3: Reduced Dataset obtained From MOALO-CD Wrapper.

Various plots for the performance metrics values of the above three datasets after applying the said classifiers are given in FIGURE. 7, 9, 10, 8, 12, 11, 13, and 14. As can be observed, the accuracy values for the RDS3 are very attractive concerning all the classifiers as compared to the RDS2 and RDS1. Only in the case of RDS1, the SVM classifier has given better accuracy than other datasets. RF, DT, and KNN defeated others in classifying infant states, thus producing higher classification accuracy, sensitivity, specificity, precision, MCC, F1-score, and lower error rate. As the sensitivity and specificity values of RF, DT, and KNN are very high and their gap is smaller, these three classifiers can be treated as

TABLE 5. Findings of student t-test.

Dataset	MOALO-CD Wrapper vs MOGA-CD Wrapper		MOALO-CD Filter vs MOGA-CD Filter	
	p-value	Sig.diff	p-value	Sig.diff
US Territories Birth Data	2.748E-4	>>	1.452E-3	>>

TABLE 6. Classification performance results: Whole dataset.

Measures/Methods	LR	SVM	RF	DT	KNN	GNB
Accuracy	82.40	82.21	99.90	99.82	98.11	94.50
Sensitivity	0.70	0.67	1	0.99	0.99	0.94
Specificity	0.94	0.96	0.99	0.99	0.96	0.94
TP	4545	4381	6443	6440	6430	6063
TN	6094	6224	6449	6441	6229	6130
FP	1898	2062	0	3	13	380
FN	363	233	8	16	228	327
Gmean	0.81	0.80	0.99	0.99	0.98	0.94
Precision	0.92	0.94	0.99	0.99	0.96	0.94
FPR	0.05	0.036	0.001	0.002	0.035	0.05
MCC	0.66	0.67	0.99	0.99	0.96	0.89
F1-Score	0.8	0.79	0.99	0.99	0.98	0.94
Error Rate	0.17	0.17	0.0006	0.001	0.018	0.054

TABLE 7. Classification performance results: Reduced dataset by MOALO-CD wrapper.

Measures/Methods	LR	SVM	RF	DT	KNN	GNB
Accuracy	90.31	77.85	99.90	99.85	99.80	96.27
Sensitivity	0.81	0.56	0.99	0.99	1	0.97
Specificity	0.99	0.99	0.99	0.99	0.99	0.94
TP	5231	3632	6442	6437	6443	6297
TN	6419	6411	6446	6444	6429	6123
FP	1212	2811	1	6	0	146
FN	38	46	11	13	28	334
Gmean	0.89	0.74	0.99	0.99	0.99	0.96
Precision	0.99	0.98	0.99	0.99	0.99	0.94
FPR	0.005	0.007	0.001	0.002	0.004	0.05
MCC	0.81	0.61	0.99	0.99	0.99	0.92
F1-Score	0.89	0.71	0.99	0.99	0.99	0.96
Error Rate	0.09	0.22	0.0009	0.001	0.002	0.03

TABLE 8. Classification performance results: Reduced dataset by MOGA-CD wrapper.

Measures/Methods	LR	SVM	RF	DT	KNN	GNB
Accuracy	90.27	77.8	99.90	99.72	99.6	96.13
Sensitivity	0.81	0.56	1	0.99	1	0.97
Specificity	0.98	0.99	0.99	0.99	0.99	0.94
TP	5281	3635	6443	6435	6443	6284
TN	6364	6413	6445	6429	6412	6117
FP	1162	2808	0	8	0	159
FN	93	44	12	28	45	340
Gmean	0.89	0.74	0.99	0.99	0.99	0.96
Precision	0.98	0.98	0.99	0.99	0.99	0.94
FPR	0.014	0.006	0.001	0.004	0.006	0.05
MCC	0.81	0.61	0.99	0.99	0.99	0.92
F1-Score	0.89	0.71	0.99	0.99	0.99	0.96
Error Rate	0.09	0.22	0.0009	0.002	0.003	0.03

TABLE 9. Classification performance results: Reduced dataset by MOALO-CD filter.

Measures/Methods	LR	SVM	RF	DT	KNN	GNB
Accuracy	80.27	78.48	99.72	99.67	97.87	85.37
Sensitivity	0.65	0.60	1	0.99	0.99	0.74
Specificity	0.94	0.96	0.99	0.99	0.96	0.96
TP	4248	3886	6443	6431	6406	4796
TN	6108	6238	6422	6427	6220	6217
FP	2195	2557	0	12	37	1647
FN	349	219	35	30	237	240
Gmean	0.78	0.76	0.99	0.99	0.97	0.84
Precision	0.92	0.94	0.99	0.99	0.96	0.95
FPR	0.05	0.03	0.005	0.004	0.03	0.03
MCC	0.63	0.61	0.99	0.99	0.95	0.72
F1-Score	0.76	0.73	0.99	0.99	0.97	0.83
Error Rate	0.19	0.21	0.002	0.003	0.02	0.14

TABLE 10. Classification performance results: Reduced dataset by MOGA-CD filter.

Measures/Methods	LR	SVM	RF	DT	KNN	GNB
Accuracy	80.75	74.87	99.89	99.72	99.26	82.29
Sensitivity	0.63	0.5	0.99	0.99	0.99	0.67
Specificity	0.98	0.99	0.99	0.99	0.98	0.96
TP	4087	3223	6442	6433	6441	4359
TN	6331	6436	6444	6432	6364	6257
FP	2356	3220	1	10	2	2084
FN	126	21	13	25	93	200
Gmean	0.78	0.7	0.99	0.99	0.99	0.80
Precision	0.97	0.99	0.99	0.99	0.98	0.95
FPR	0.01	0.003	0.002	0.003	0.014	0.03
MCC	0.65	0.57	0.99	0.99	0.98	0.67
F1-Score	0.76	0.66	0.99	0.99	0.99	0.79
Error Rate	0.19	0.25	0.001	0.002	0.007	0.17

TABLE 11. Classification performance results: Reduced dataset by RFE.

Measures/Methods	LR	SVM	RF	DT	KNN	GNB
Accuracy	81.90	80.41	99.90	99.82	98.41	86.60
Sensitivity	0.68	0.64	1	0.99	0.99	0.76
Specificity	0.95	0.96	0.99	0.99	0.96	0.96
TP	4408	4153	6443	6439	6424	4942
TN	6161	6224	6449	6440	6274	6230
FP	2035	2290	0	4	19	1501
FN	296	233	8	17	183	227
Gmean	0.80	0.78	0.99	0.99	0.98	0.86
Precision	0.93	0.94	0.99	0.99	0.97	0.95
FPR	0.04	0.03	0.001	0.002	0.03	0.03
MCC	0.66	0.64	0.99	0.99	0.96	0.74
F1-Score	0.79	0.76	0.99	0.99	0.98	0.85
Error Rate	0.18	0.19	0.0006	0.001	0.01	0.13

the best model for predicting whether an infant will survive or not.

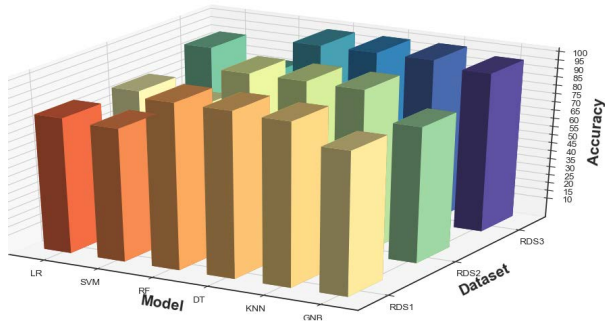The precision value of RF and DT for all the datasets is 0.99, which is an indication of more positive outcomes with
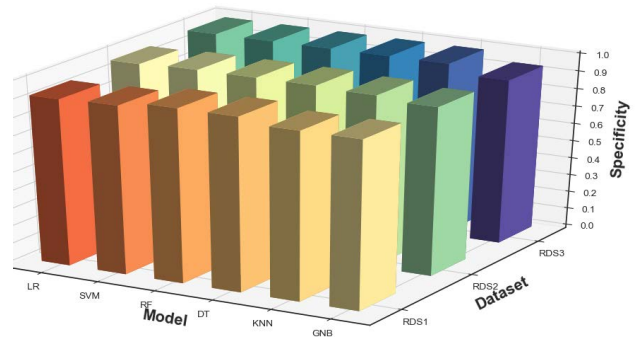
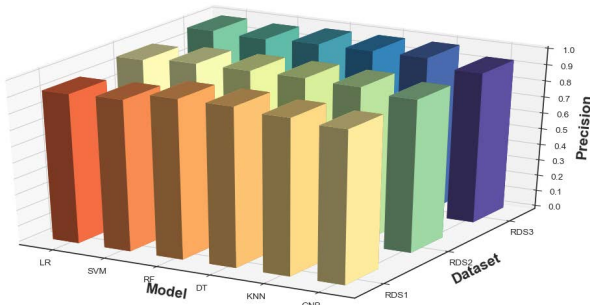FIGURE 7. Accuracy comparison.

FIGURE 8. Precision comparison.
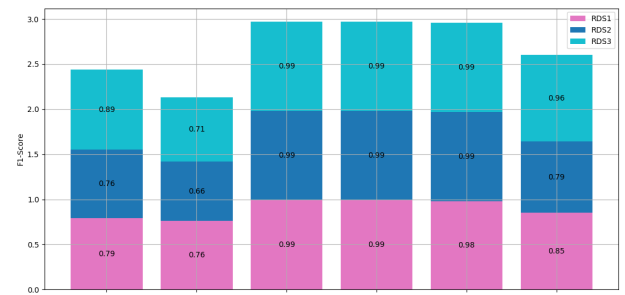

FIGURE 9. Sensitivity comparison.

FIGURE 10. Specificity comparison.
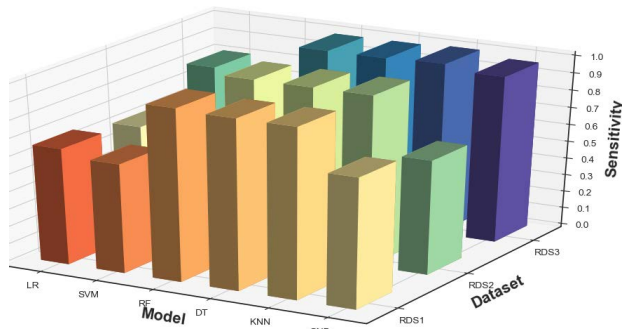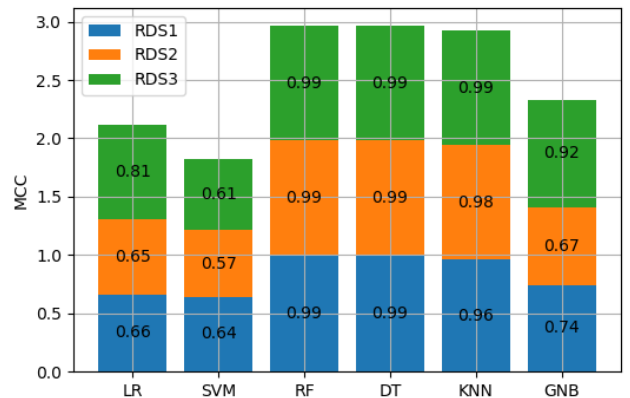

FIGURE 11. F1-score comparison.

FIGURE 12. MCC comparison.

respect to the number of positive outcomes forecasted by the above two classifiers.

ROC is a probability curve, and AUC is a degree or separability measure. This shows how much one model can differentiate between classes. From FIGURE 15, FIGURE 16, and FIGURE 17 it can be concluded that for all three datasets, as AUC of RF and DT is equal to 1, these two models are the best ones to differentiate between alive and dead infants. As for the RDS3, the AUC value of RF, DT, and KNN is equal to 1, and the same for GNB is also very close to 1, so it can be concluded that the feature subset produced by **MOALO-CD Wrapper** is more relevant in classifying the infant's health status.

The proportion of samples that have been wrongly classified (ERR) is higher in SVM and LR as compared to others.

As the Gmean is very close to 1, the combined efficiency of two classes (alive or dead) is higher when applying RF, DT, and KNN to classify the infant state for RDS1, RDS2, and RDS3.

The MCC value of 0.99 for RF, DT, and KNN for RDS3 indicates that these are perfectly correct binary classifiers for the RDS3.

The FPR value of the random forest model is much smaller. So the number of wrongly categorized negative events as positive (false positives) concerning the total number of actual negative events is much less as compared to other models.

The combined performance of both the alive and dead classes is also very attractive and eye-catching when the dataset is trained and tested under RF, DT, and KNN. It is found that RF proves itself as a very fast and accurate classifier in classifying infants as alive or dead. The performances of DT and KNN are also remarkable as they are very close to RF. After comparing the performance results of all the six classifiers with respect to the three chosen datasets, it is observed that all the classification models proved their
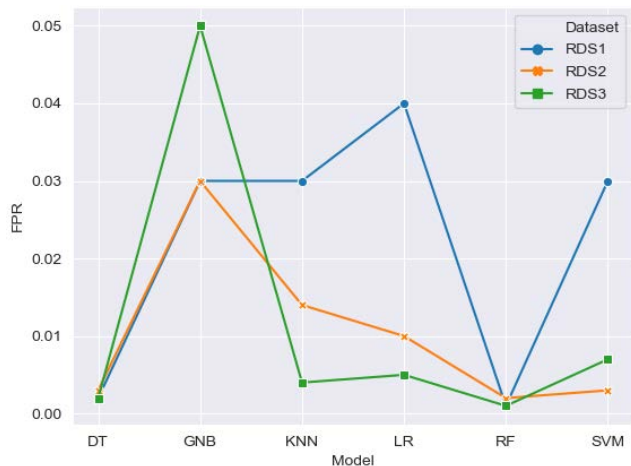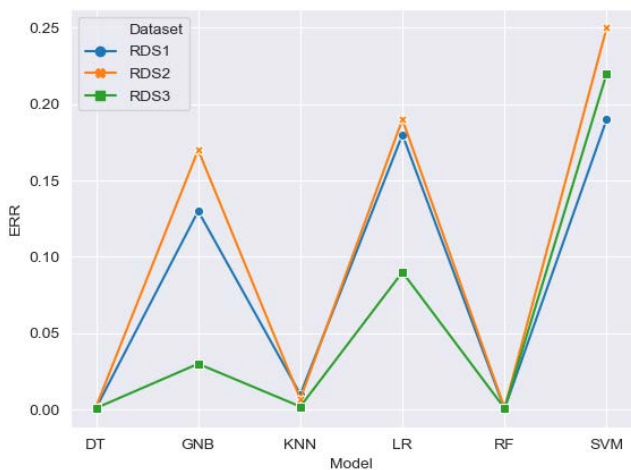
FIGURE 13. FPR comparison.
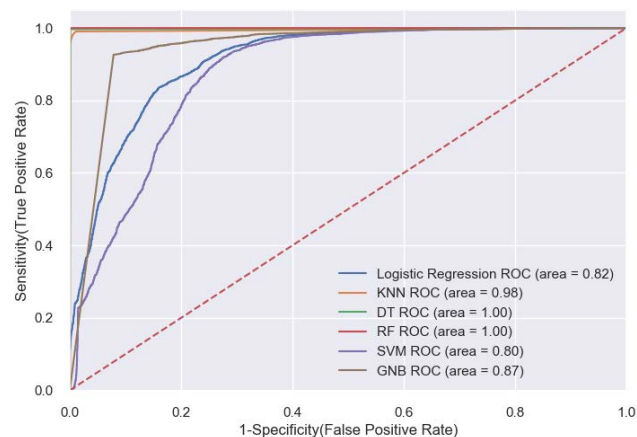

FIGURE 14. ERR comparison.

FIGURE 15. ROC of RDS1.

efficiency in classifying whether an infant is alive or not when applied to RDS3 (the dataset obtained from MOALO-CD Wrapper). So, at the end of 100 iterations, a set of features selected by the MOALO-CD Wrapper, MOGA-CD Filter, and RFE approach are listed in Table 12 for easy comparison.
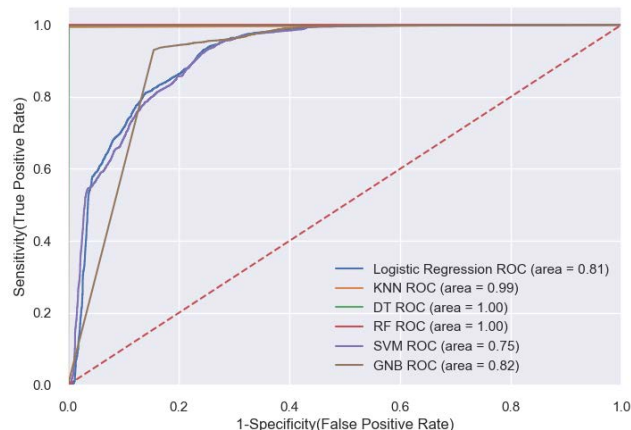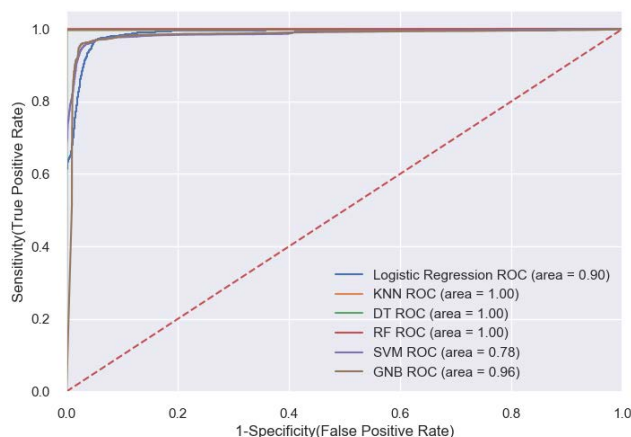

FIGURE 16. ROC of RDS2.

FIGURE 17. ROC of RDS3.

The P-values (with a 0.05 significance level) and absolute correlation values (with target attribute) of all the features selected by the above three methods are listed in the Table 12. The P-value gives us the likelihood of finding an observation on the assumption that a given hypothesis is valid. This likelihood is used to accept the hypothesis or deny it. In feature selection, removing various features from the dataset will have different effects on the dataset's p-value. If a particular feature has a p-value > threshold (here 0.05), then this feature does not provide any noticeable change to the output and can be easily removed without consequences. However, a feature with a p-value < threshold provides a very significant change to the output, and it cannot be removed.

IV. DISCUSSION

Many studies have been performed over the last few years to accurately classify infants' health. Singha *et al.* [28] have analyzed different attributes of pregnant women to identify risky maternal factors that are responsible for neonatal infant mortality. Based on a statistically significant pattern, an ML model is suggested to predict the key factors and high-risk mothers for proper medical care. The birth data of the U.S. Territories (2013) was trained and tested by using LR, NB, and LSVM. Rinta-Koski *et al.* [29] have proposed Gaussian

TABLE 12. P-values and correlation (Corr.) values of the features chosen by MOALO-CD wrapper, MOGA-CD filter and RFE.

SNO	MOALO-CD wrapper	p-value	Corr.	MOGA-CD filter	p-value	Corr.	RFE	p-value	Corr.
1	mothers_hispanic_origin	2.09e-32	0.052	marital_status	6.144e-16	0.036	birth_month	0.205	0.005
2	marital_status	6.144e-16	0.036	facility	1.691e-33	0.053	time_of_birth	1.092e-25	0.046
3	facility	1.691e-33	0.053	mothers_height	1.66e-68	0.077	birth_day_of_week	8.961e-30	0.049
4	mothers_edu	3.61E-166	0.121	births_now_living	0.003	0.052	mothers_age	1.225e-33	0.053
5	births_now_dead	2.00E-04	0.016	cigaratte_1st_trimster	2.501e-05	0.019	residence_status	7.767e-25	0.045
6	interval_since_last_live_birth	3.84e-10	0.028	cigaratte_3rd_trimster	0.0	0.587	mothers_hispanic_origin	2.09e-32	0.052
7	cigarette_before_preg	3.35e-06	0.02	mothers_race_hispanic	2.659e-34	0.037	mothers_race_hispanic	2.659e-34	0.037
8	mothers_height	1.66e-68	0.077	BMI_recode	3.112e-5	0.010	paternity_ack	1.584e-13	0.032
9	#prenatal_visits	0.00e+00	0.186	gestational_hypertension	1.418e-61	0.073	marital_status	6.144e-16	0.036
10	cigaratte_1st_trimster	2.50e-05	0.019	paternity_ack	1.584e-13	0.032	mothers_edu	3.61e-166	0.121
11	cigaratte_3rd_trimster	0.00e+00	0.587	fathers_race	0.007	0.011	fathers_race	0.007	0.011
12	BMI	0.00e+00	0.12	fathers_race_hispanic	0.328	0.004	fathers_hispanic_origin	1.246e-45	0.062
13	gestational_hypertension	1.42e-61	0.073	pre_preg_wt_recode	1.027e-07	0.011	fathers_edu	4.472e-34	0.053
14	prev_preterm_birth	3.281e-98	0.092	pre_preg_diabetes	7.00e-236	0.121	prior_termination	1.757e-06	0.021
15	#prev_ceaserian	3.374e-23	0.011	gonorrhea	2.640e-58	0.010	interval_since_last_live_birth	3.84E-10	0.028
16	pre_preg_diabetes	7.003e-236	0.144	syphilis	1.029e-157	0.001	cigaratte_3rd_trimster	0.0	0.587
17	hepatites_B	7.836e-47	0.063	mothers_edu	3.61e-166	0.121	mothers_height	1.66e-68	0.077
18	infertility_treat_used	1.842e-05	0.019	fetal_presentation_at_delivery	0.0	0.021	BMI	0.00e+00	0.12
19	chlamydia	1.167e-42	0.06	route_method of_delivery	1.789e-209	0.135	delivery_wt_recode	9.690e-36	0.012
20	hepatites_C	1.823e-71	0.078	delivery_method	9.326e-228	0.141	gestational_diabetes	0.014	0.002
21	succ_ext_cephalic_version	7.779e-64	0.072	reptured_uterus	0.013	0.011	pre_preg_hypertension	8.748e-23	0.013
22	fail_ext_cephalic_version	2.678e-62	0.072	unplanned_hysterectomy	0.003	0.002	gestational_hypertension	1.418e-61	0.073
23	induction_of_labour	2.531e-78	0.082	admit_intensiv_care	8.580e-75	0.032	prev_preterm_birth	3.281e-98	0.092
24	steroids	0.0	0.406	attendent_at_birth	2.01e-99	0.093	infertility_treat_used	1.842e-05	
25	chorioamnionitis	1.883e-265	0.152	mother_transferred	5.067e-224	0.139	induction_of_labour	2.531e-78	0.082
26	route_method of_delivery	1.789e-209	0.135	5_min_APGAR	1.956e-33	0.053	chorioamnionitis	1.883e-265	0.152
27	delivery_method	9.326e-228	0.141	10_min_APGAR	0.0	0.493	fetal_presentation_at_delivery	0.0	0.021
28	reptured_uterus	0.013	0.011	plurality	1.907e-299	0.162	trial_of_labor_attempted	0.0	0.041
29	attendent_at_birth	2.01e-99	0.093	immediate_ventilation	0.0	0.818	mother_transferred	5.067e-224	0.139
30	plurality	1.907e-299	0.162	>6hrs_ventilation	0.0	0.706	5_min_APGAR	1.956e-33	0.053
31	10_min_APGAR	0.0	0.493	antibiotics_for_newborn	0.0	0.604	10_min_APGAR	0.0	0.493
32	last_menses_month	4.862e-20	0.04	seizures	0.0	0.235	plurality	1.907e-299	0.162
33	surfactant	0.0	0.65	anencephaly	0.0	0.194	last_menses_month	4.862e-20	0.04
34	seizures	0.0	0.235	sus_chromosomal_disorder	8.275e-226	0.140	birth_weight	0.0	0.592
35	>6hrs_ventilation	0.0	0.706	infant_transferred	0.0	0.32	immediate_ventilation	0.0	0.818
36	admission_NICU	0.0	0.716	breastfeed_at_discharge	0.0	0.894	antibiotics_for_newborn	0.0	0.604
37	antibiotics_for_newborn	0.0	0.604				anencephaly	0.0	0.194
38	spina_bifida	9.039e-222	0.139				heart_disease	0.0	0.214
39	heart_disease	0.0	0.214				gastroschisis	0.0	0.171
40	cleft_lip_w/wo_palate	4.699e-206	0.134				cleft_lip_w/wo_palate	4.699e-206	0.134
41	hypospadias	8.238e-175	0.124				limb_defect	2.098e-210	0.135
42	infant_transferred	0.0	0.32				infant_transferred	0.0	0.32
43	down_syndrome	0.0	0.189				down_syndrome	0.0	0.189
44	breastfeed_at_discharge	0.0	0.894				breastfeed_at_discharge	0.0	0.894

Process Classification to predict preterm infant in-hospital–mortality where, in addition to clinical ratings, gestational age at birth, and birth weight, the characteristics are derived from direct cardiac, arterial, and oximeter sensor measurements. Thompson and Steele [30] have identified the use of ML-based analysis of the cost and usage data set of a healthcare project for the development of long-stay newborn models. In this research, the 2014 SID (State Inpatient Databases) for Florida was used, which contains 230 attributes per patient admission record. Implementation results showed that RF was the best performing model, with a ROC score of 0.877. Podda *et al.* [22] have estimated preterm neonatal mortality risk by developing ML-based predictive models. The developed models use Italian neonatal data (2008 - 2014) for training and (2015-2016) for testing. Out of six ML predictors, this study chose ANN and compared it with LR, where ANN had better discrimination than LR with a p-value < 0.002. Preterm birth and neonatal death are strongly associated with each other. So there is a need for early and accurate prediction of preterm infants for appropriate clinical intervention. Li *et al.* [31] have made a comparison of five feature selection methods, namely the chi-square test, information gain, minimum redundancy maximum relevance (MRMR), stepwise logistic regression, and Gini index in RF, to identify the risk factors for SGA infants. They have evaluated their work by applying 4 classifiers (LR, NB, SVM, and RF) while taking precision and AUC as evaluation criteria. Again, to construct powerful SGA prediction models, they proposed an ensemble method based on feature subsets that performed better than individual ones. Gismondi *et al.* [32] attempted ANN, while Semenova *et al.* [33] proposed a boosted decision tree method for analyzing infant mortality and exploring the association between heart rate variability (HRV) and blood pressure (BP) in preterm babies, respectively. In the study performed by Lee *et al.* [24], 275 infants born before 32 weeks were taken with an average birth weight of 929g and 49% of them were female. They found that the RF model had excellent performance (sensitivity-88%, AUC-0.93) for the prediction of preterm newborn mortality.

However, the prediction of infant health status using the factors selected by evolutionary approaches is less explored. Also, extracting the crucial factors causing infant deaths from the USA birth data file is still an unexplored area. So, in this article, an evolutionary wrapper-based multi-objective FS technique called MOALO-CD is attempted to mine the valuable aspects of causing infant death. By observing the selected features, their p-values, and correlation values, we found some of the most interesting and influential causes

of infant death in the USA are: mothers' education, the number of prenatal visits, pre-pregnancy diabetes, smoking during pregnancy, BMI, birth weight, spina bifida, 5-min, and 10-min APGAR scores, antibiotics for newborns, breastfeeding, ventilation requirement, admission to NICU, steroids, plurality, etc. Maternal education has shown that increased education provides mothers with more links to children's health services and a sense of healthy behaviours (including exercise and not smoking). Education could also improve the skills needed to access the health care system and make effective use of it. In general, the mortality rates of infants born to more trained mothers are lower. Low birth weight is the result of insufficient growth in the fetus, and the lower the birth weight, the greater the immaturity and death risk. Numerous factors lead to low birthweight: low socioeconomic status, low rates of employment, very late or very early puberty, inadequate diet, health problems, and substance misuse. Prenatal visits play a significant role in reducing risk factors and optimizing pregnancy outcomes, particularly if treatment is sufficient and early, by delivering effective healthcare and helping pregnant women improve their overall health. Infant mortality risk among normal-weight women was lowest and increased with increasing category of BMI. Maternal weight and weight gain extremes can contribute to the high infant mortality rates [34] in the US. The Body Mass Index (BMI) of approx. $30kg/m^2$ is nearly 40% more likely to die for babies born to women who have obese pregnancies than for infants born to mothers of average weight. Preconception weight loss therapy may not be a realistic preventive technique, however, as half of the pregnancies are unplanned and significant reductions in body weight during pregnancy (20%–30%) may be needed to substantially reduce the risk of infant mortality. Spina bifida [35] occurs when the spinal cord of a developing baby fails to develop or close properly while it is in the womb. The risk of infant mortality in children with spina bifida is 4.4%. The risk of premature delivery, low weight, and sudden baby death syndrome, both caused by child mortality, were enhanced by maternal smoking. Studies have shown higher mortality among diabetic mothers' infants (DMIs) [36] compared with controls. The rate of neonatal mortality is more than five times that of non-diabetic mothers' babies and is higher for gestational age (GA) groups at all gestational ages and birth weight. These risks can be reduced by optimal glycaemic regulation before and during the entire pregnancy. The APGAR score [37] is a way of easily summing up newborn children's health against infant mortality. For babies with 5 minute (1-3) extremely low APGAR, neonatal and post-neonatal death rates have remained high (37 weeks). Only in babies with high APGAR ratings (almost 7) did neonatal and post-neonatal deaths gradually decrease with gestational age. In a dose-response way, breastfeeding was associated with the lowest risk of mortality among exclusively breastfed infants, the intermediate-risk among partially breastfed infants, and the highest risk among non-breastfed infants. Children's mortality most commonly happens in infants without maternal antibiotic use (11.9%)

and is lower for all prenatally infected babies independent of maternal suggestion [38]. Antenatal steroid [39] intervention in neonatal mortality and morbidity is very successful, but it tends to have poor penetration in low- and middle-income countries. If this technique is greatly extended, up to 500 000 neonatal lives may be saved annually. Respiratory distress syndrome (RDS), sepsis, and birth asphyxia are the most common signs of ventilation for newborns [40]. Relevant mortality predictors for ventilated neonates were <2500 g, <34 weeks of gestation, <7.1 of initial pH, pain, pulmonary hemorrhage, apnea, hypoglycemia, neutropenia, and thrombocytopenia. As the risk of accident-induced baby mortality is associated with low birth weight and multiples are of lower weight than singletons, the indirect impact of plurality [41] on a child's death thus remains. Because of outside causes, direct influences on infant mortality may include heightened workload and parental care heterogeneity, when parents take care of more than one child at a time. As a family grows in size, parental care per child decreases—twins may represent a severe form of family development. Additionally, in the first year of a child's life, twin parents face more fear, tension, and depression than single parents do. Pregnancy-induced hypertension (PIH) is a type of hypertension disorder during pregnancy that includes gestational hypertension, pre-eclampsia, and eclampsia [42]. The effect of gestational hypertension on the neonatal outcomes of twins is even more positive than the effect of singleton pregnancies. The correlation of PIH with child mortality depends on the disparity in birth weight, pregnancy, and death age.

V. CONCLUSION AND FUTURE WORK

The concept of infant mortality is examined in this article. Because of the high imbalance in the data, this is a difficult binary classification task. So, SMOTE is applied as the preprocessing technique for balancing the above-discussed infant mortality dataset. In this article, an evolutionary wrapper-based multi-objective FS technique called MOALO-CD is attempted to mine the valuable aspects of causing infant death. Smoking during pregnancy, birth weight, breastfeeding, BMI, prenatal visits, use of steroids, APGAR score, etc., are some of the important factors affecting infant health selected by the above-discussed FS mechanism. The proposed methodology is compared with one non-evolutionary FS method (RFE), two filter-based multi-objective approaches (MOALO-CD Filter and MOGA-CD Filter), and one wrapper-based multi-objective algorithm (MOGA-CD Wrapper) by considering four multi-objective performance indicators. After getting five sets of the reduced dataset, six standard classifiers are applied to them to verify their classification performance in predicting the infant's health state (alive or dead). In the case of a reduced dataset obtained from the MOALO-CD wrapper approach, all the classifiers showed their best performance. Experimental findings reveal that RF can easily and reliably identify an infant as dead or alive. DT and KNN's performance is also excellent as it is very similar to the RF. In the future, the proposed model

could also be tried on different infant mortality datasets from different countries in different years. Some improved evolutionary feature selection algorithms can also be attempted to mine more relevant factors causing infant death in our society.

REFERENCES

[1] S. Mani, A. Ozdas, C. Aliferis, H. A. Varol, Q. Chen, R. Carnevale, Y. Chen, J. Romano-Keeler, H. Nian, and J.-H. Weitkamp, "Medical decision support using machine learning for early detection of late-onset neonatal sepsis," *J. Amer. Med. Inform. Assoc.*, vol. 21, no. 2, pp. 326–336, Mar. 2014.

[2] K. Wilson, D. Wallace, S. Goudar, D. Theriaque, and E. McClure, "Identifying causes of neonatal mortality from observational data: A Bayesian network approach," in *Proc. Int. Conf. Data Sci. (ICDATA)*, 2015, p. 132.

[3] K. J. Rittenhouse, B. Vwalika, A. Keil, J. Winston, M. Stoner, J. T. Price, M. Kapasa, M. Mubambe, V. Banda, W. Muunga, and J. S. A. Stringer, "Improving preterm newborn identification in low-resource settings with machine learning," *PLoS ONE*, vol. 14, no. 2, Feb. 2019, Art. no. e0198919.

[4] A. Saravanou, C. Noelke, N. Huntington, D. Acevedo-Garcia, and D. Gunopulos, "Infant mortality prediction using birth certificate data," 2019, *arXiv:1907.08968*.

[5] Z. Kefi, K. Aloui, and M. Saber, "New approach based on machine learning for short-term mortality prediction in neonatal intensive care unit," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 1–9, 2019.

[6] J. Liu, C. Sun, and Y. Lai, "A data transmission approach based on ant colony optimization and threshold proxy re-encryption in WSNs," *J. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 23–31, Dec. 2021.

[7] Y. Tadepalli, M. Kollati, S. Kuraparthi, P. Kora, A. K. Budati, and L. K. Pampana, "Content-based image retrieval using Gaussian–Hermite moments and firefly and grey wolf optimization," *CAAI Trans. Intell. Technol.*, vol. 6, no. 2, pp. 135–146, Jun. 2021.

[8] A. Rezaeipanah and M. Mojarad, "Modeling the scheduling problem in cellular manufacturing systems using genetic algorithm as an efficient meta-heuristic approach," *J. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 228–234, Oct. 2021.

[9] H. Ullah, B. Ahmad, I. Sana, A. Sattar, A. Khan, S. Akbar, and M. Z. Asghar, "Comparative study for machine learning classifier recommendation to predict political affiliation based on online reviews," *CAAI Trans. Intell. Technol.*, vol. 6, no. 3, pp. 251–264, Sep. 2021.

[10] G. Mohindru, K. Mondal, and H. Banka, "Different hybrid machine intelligence techniques for handling IoT-based imbalanced data," *CAAI Trans. Intell. Technol.*, vol. 6, no. 4, pp. 405–416, Dec. 2021.

[11] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, Apr. 2018.

[12] M. T. M. Emmerich and A. H. Deutz, "A tutorial on multiobjective optimization: Fundamentals and evolutionary methods," *Natural Comput.*, vol. 17, no. 3, pp. 585–609, Sep. 2018.

[13] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, May 2015.

[14] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.

[15] A. S. Assiri, A. G. Hussien, and M. Amin, "Ant lion optimization: Variants, hybrids, and applications," *IEEE Access*, vol. 8, pp. 77746–77764, 2020.

[16] S. Bai, S. Song, S. Liang, J. Wang, B. Li, and E. Neretin, "UAV maneuvering decision-making algorithm based on twin delayed deep deterministic policy gradient algorithm," *J. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 16–22, Dec. 2021.

[17] J. Piri and P. Mohapatra, "An analytical study of modified multi-objective Harris hawk optimizer towards medical data feature selection," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104558.

[18] J. Piri, P. Mohapatra, and R. Dey, "Multi-objective ant lion optimization based feature retrieval methodology for investigation of fetal wellbeing," in *Proc. 3rd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Sep. 2021, pp. 1732–1737.

[19] J. Piri, P. Mohapatra, and R. Dey, "Fetal health status classification using MOGA—CD based feature selection approach," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2020, pp. 1–6.

[20] J. Ahuja and S. D. Ratnoo, "Feature selection using multi-objective genetic algorith m: A hybrid approach," *INFOCOMP J. Comput. Sci.*, vol. 14, no. 1, pp. 26–37, Jun. 2015.

[21] S. Chatterjee, D. Dey, and S. Munshi, "Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification," *Comput. Methods Programs Biomed.*, vol. 178, pp. 201–218, Sep. 2019.

[22] M. Podda, D. Bacciu, A. Micheli, R. Bellù, G. Placidi, and L. Gagliardi, "A machine learning approach to estimating preterm infants survival: Development of the preterm infants survival assessment (PISA) predictor," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, Dec. 2018.

[23] M. Z. Arif, R. Ahmed, U. H. Sadia, M. S. I. Tultul, and R. Chakma, "Decision tree method using for fetal state classification from cardiotography data," *J. Adv. Eng. Comput.*, vol. 4, no. 1, pp. 64–73, 2020.

[24] J. Lee, J. Cai, F. Li, and Z. A. Vesoulis, "Predicting mortality risk for preterm infants using random forest," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, Dec. 2021.

[25] C. Audet, J. Bigeon, D. Cartier, S. Le Digabel, and L. Salomon, "Performance indicators in multiobjective optimization," *Eur. J. Oper. Res.*, vol. 292, no. 2, pp. 397–422, Jul. 2021.

[26] H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "Reference point specification in inverted generational distance for triangular linear Pareto front," *IEEE Trans. Evol. Comput.*, vol. 22, no. 6, pp. 961–975, Dec. 2018.

[27] P. Mohapatra, S. Chakravarty, and P. K. Dash, "Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system," *Swarm Evol. Comput.*, vol. 28, pp. 144–160, Jun. 2016.

[28] A. K. Singha, D. Phukan, S. Bhasin, and R. Santhanam, "Application of machine learning in analysis of infant mortality and its factors," Working Paper, Tech. Rep., 2016.

[29] O.-P. Rinta-Koski, S. Särkkä, J. Hollmén, M. Leskinen, and S. Andersson, "Prediction of preterm infant mortality with Gaussian process classification," in *Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, 2017, pp. 193–198.

[30] B. Thompson, K. O. Elish, and R. Steele, "Machine learning-based prediction of prolonged length of stay in newborns," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1454–1459.

[31] J. Li, L. Liu, M. Zhou, J.-J. Yang, S. Chen, H. Liu, Q. Wang, H. Pan, Z. Sun, and F. Tan, "Feature selection and prediction of small-for-gestational-age infants," *J. Ambient Intell. Humanized Comput.*, pp. 1–15, Jun. 2018, doi: 10.1007/s12652-018-0892-2.

[32] R. C. Gismondi, R. M. V. R. Almeida, and A. F. C. Infantosi, "Artificial neural networks for infant mortality modelling," *Comput. Methods Programs Biomed.*, vol. 69, no. 3, pp. 237–247, Nov. 2002.

[33] O. Semenova, G. Carra, G. Lightbody, G. Boylan, E. Dempsey, and A. Temko, "Prediction of short-term health outcomes in preterm neonates from heart-rate variability and blood pressure using boosted decision trees," *Comput. Methods Programs Biomed.*, vol. 180, Oct. 2019, Art. no. 104996.

[34] J. Ozawa, K. Tanaka, K. Kabe, and F. Namba, "Impact of being large-for-gestational-age on neonatal mortality and morbidities in extremely premature infants," *Pediatric Res.*, vol. 90, no. 4, pp. 910–916, 2021.

[35] N. D. Pace, A. M. Siega-Riz, A. F. Olshan, N. C. Chescheir, S. R. Cole, T. A. Desrosiers, S. C. Tinker, A. T. Hoyt, M. A. Canfield, S. L. Carmichael, and R. E. Meyer, "Survival of infants with spina bifida and the role of maternal prepregnancy body mass index," *Birth Defects Res.*, vol. 111, no. 16, pp. 1205–1216, Oct. 2019.

[36] D. Care, "Medical care in diabetesd2021," *Diabetes Care*, vol. 44, no. 1, pp. S125–S150, 2021.

[37] S. Cnattingius, S. Johansson, and N. Razaz, "Apgar score and risk of neonatal death among preterm infants," *New England J. Med.*, vol. 383, no. 1, pp. 49–57, Jul. 2020.

[38] B. D. Reed, K. R. Schibler, H. Deshmukh, N. Ambalavanan, and A. L. Morrow, "The impact of maternal antibiotics on neonatal disease," *J. Pediatrics*, vol. 197, pp. 97–103, Jun. 2018.

[39] I. Ali, R. I. Batta, R. M. Yaseen, and J. Hasson, "Antenatal corticosteroids and fetal lung immaturity in preterm birth," *Heliyon*, vol. 6, no. 6, Jun. 2020, Art. no. e04116.

[40] A. A. Chakkarapani, R. Adappa, S. K. M. Ali, S. Gupta, N. B. Soni, L. Chicoine, and H. D. Hummler, "'Current concepts of mechanical ventilation in neonate'—Part 1: Basics," *Int. J. Pediatrics Adolescent Med.*, vol. 7, no. 1, pp. 15–20, Mar. 2020.

[41] K. A. Ahrens, M. E. Thoma, L. M. Rossen, M. Warner, and A. E. Simon, "Plurality of birth and infant mortality due to external causes in the United States, 2000–2010," *Amer. J. Epidemiol.*, vol. 185, no. 5, pp. 335–344, 2017.

[42] A. C. O. Gynecologists, "Gestational hypertension and preeclampsia: ACOG practice bulletin, number 222," *Obstetrics Gynecol.*, vol. 135, no. 6, pp. e237–e260, 2020.

JAYASHREE PIRI received the B.Tech. degree in information technology from the BPUT, India, in 2007, and the M.E. degree in information technology from Jadavpur University, Kolkata, India, in 2012. She is currently pursuing the Ph.D. degree with the Department of Computer Science, IIIT Bhubaneswar, India. She has more than ten years of teaching experience. She has published more than ten research articles in peer-reviewed journals, such as *Computers in Biology and Medicine*, IEEE Access, and presented various papers in international conferences. Her main research interests include multi-objective optimization, evolutionary computing, medical data mining, and machine learning.

PUSPANJALI MOHAPATRA (Member, IEEE) received the B.E. degree in electrical engineering from IGIT Sarang, Odisha, in 1999, the M.Tech. degree in computer science from Utkal University, Bhubaneswar, in 2002, the Ph.D. degree in CSE from Siksha 'O' Anusandhan University, India, and the M.B.A. degree in systems, IT & HR from BPUT Rourkela, India, in 2009. She is currently an Assistant Professor at the International Institute of Information Technology, Bhubaneswar, and has more than 18 years of teaching experience. More than 35 papers in reputed journals, conferences, and book chapters are to her credit. Her research interests include biomedical data mining, time series data analysis, and different fields like machine learning, deep learning, and evolutionary computing. She is also associated with various educational and research societies like IEEE, ACM, IAENG, and OITS.

DEBABRATA SINGH received the Ph.D. degree in computer science and engineering from the Siksha 'O' Anusandhan (SOA) Deemed to be University, Bhubaneswar, Odisha, India. He is currently working as an Associate Professor with the Department of Computer Application, ITER, SOA Deemed to be University. He is keenly interested in interdisciplinary research and development and has experience spanning fields of WSN, bioinformatics, wireless ad hoc networks, network security, cloud computing, cloud native applications, and the IoT applications. In his past 15 years of teaching experience, he has authored or coauthored over 73 research papers in international journal (SCI/SCIE/ESCI/Scopus) and conferences, including IEEE, Springer, and Elsevier conference proceeding. He also serves as an acquisition editorial/reviewing board member in reputed journals. He has received the "Bentham Ambassador" Award in 2019 and the Best Outstanding Researcher Award 2020 from the Kamarajar Institute of Education and Research (KIER).

DEBABRATA SAMANTA (Member, IEEE) received the bachelor's degree (Hons.) in physics from Calcutta University, Kolkata, India, the M.C.A. degree from the Academy of Technology, under WBUT, West Bengal, and the Ph.D. degree in computer science and engineering from the National Institute of Technology, Durgapur, India, in the area of SAR image processing. He is currently working as an Assistant Professor with the Department of Computer Science, CHRIST (Deemed to be University), Bengaluru, India. He has successfully completed six consultancy projects. He has received funding 5250 USD under Open Access, Publication Fund. He has received funding under the International Travel Support Scheme, in 2019, for attending conference in Thailand. He has received Travel Grant for speaker in Conference and Seminar for two years, since July 2019. He is the owner of 21 patents (three Design Indian Patent and two Australian patent granted, and 16 Indian patents published) and two copyrights. He has authored or coauthored over 184 research papers in international journals (SCI/SCIE/ESCI/Scopus) and conferences, including IEEE, Springer, and Elsevier conference proceeding. He is the coauthor of 12 books and the co-editor of 11 books, available for sale on Amazon and Flipkart. He has presented various papers at international conferences and received best paper awards. He has authored or co-authored of eight book chapters. He also serves as an acquisition editor for Springer, Wiley, CRC, Scrivener Publishing LLC, Beverly, USA, and Elsevier. His research interests include interdisciplinary research and development and has experience spanning fields of SAR image analysis, video surveillance, heuristic algorithm for image classification, deep learning framework for detection and classification, blockchain, statistical modeling, wireless ad hoc networks, natural language processing, and V2I communication. He is a Professional IEEE Member, an Associate Life Member of the Computer Society of India (CSI), and a Life Member of the Indian Society for Technical Education (ISTE). He has received the "Scholastic Award" at the 2nd International conference on Computer Science and IT application, CSIT-2011, Delhi, India. He is a convener, a keynote speaker, a session chair, a co-chair, a publicity chair, a publication chair, an advisory board, and a technical program committee members in many prestigious international and national conferences. He was an invited speaker at several institutions.

DILBAG SINGH (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from the Thapar Institute of Engineering and Technology, Patiala, India, in 2019. He worked as an Assistant Professor in three well-known universities of India, such as Chandigarh University, Mohali, India, Manipal University Jaipur, Jaipur, India, and Bennett University, Greater Noida, India. In 2021, he moved to the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea, where he is currently affiliated. He has published more than 70 research articles in SCI/SCIE indexed journals. He has also submitted five patents and has published three books and two book chapters. His H-index is 31. His research interests include image processing, computer vision, deep learning, metaheuristic techniques, and information security. He has acted as a lead guest editor/an editorial board member of many SCI/SCIE indexed journals, such as *Journal of Healthcare Engineering*, *Mathematical Problems in Engineering*, and *Journal of Intelligent & Fuzzy Systems*. He was in the top 2% list issues by "World Ranking of Top 2% Scientists," in 2021. He was part of the 11 Web of Science/Scopus indexed conferences.

MANJIT KAUR (Member, IEEE) received the Master of Engineering degree in information technology from Punjab University, Chandigarh, India, in 2011, and the Ph.D. degree from the Thapar Institute of Engineering and Technology, Patiala, India, in 2019. She worked as an Assistant Professor in three well-known universities of India, such as Chandigarh University, Mohali, India, Manipal University Jaipur, Jaipur, India, and Bennett University, Greater Noida, India. In 2021, she moved to the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea, where she is currently affiliated. She has published more than 60 SCI/SCIE indexed articles so far. Her research interests include wireless sensor networks, digital image processing, and metaheuristic techniques. She was in the top 2% list issues by "World Ranking of Top 2% Scientists," in 2021. She was part of the 11 Web of Science/Scopus indexed conferences.

HEUNG-NO LEE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California at Los Angeles, Los Angeles, CA, USA, in 1993, 1994, and 1999, respectively. He was with HRL Laboratories, LLC, Malibu, CA, USA, as a Research Staff Member, from 1999 to 2002. From 2002 to 2008, he was an Assistant Professor with the University of Pittsburgh, Pittsburgh, PA, USA. In 2009, he moved to the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea, where he is currently affiliated. His research interests include information theory, signal processing theory, blockchain, communications/networking theory, and their application to wireless communications and networking, compressive sensing, future internet, and brain–computer interface. He has received several prestigious national awards, including the Top 100 National Research and Development Award in 2012, the Top 50 Achievements of Fundamental Research Award in 2013, and the Science/Engineer of the Month (January 2014).

● ● ●