Analysis of Missing Health Care Data by Effective Adaptive DASO Based Naive Bayesian Model

[1]Anbumani K, [2]Murali Dhar M S, [3]Subramanian P, [4]Jasmine J, [5]Mahaveerakannan R and [6]John Justin Thangaraj S
[1]Department of EIE, Sri Sairam Engineering College, Chennai, India.
[2]Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, TamilNadu, India.
[4]Department of CSE, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India.
[3,5,6]Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, TamilNadu, India.
[1]anbumani.ei@sairam.edu.in, [2]msmddhar@gmail.com, [3]subramanianp.sse@saveetha.com, [4]jas378@gmail.com, [5]mahaveerakannanr.sse@saveetha.com, [6]johnjustinphd@gmail.com.

Correspondence should be addressed to Anbumani K : anbumani.ei@sairam.edu.in.

Article Info
Journal of Machine and Comp uting (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202303049
Received 28 May 2023; Revised from 18 August 2023; Accepted 20 September 2023.
Available online 05 October 2023.

Abstract – Inevitably, researchers in the field of medicine must deal with the issue of missing data. Imputation is frequently employed as a solution to this issue. Unfortunately, the perfect would overfit the experiential data distribution due to the uncertainty introduced by imputation, which would have a negative effect on the replica's generalisation presentation. It is unclear how machine learning (ML) approaches are applied in medical research despite claims that they can work around lacking data. We hope to learn if and how machine learning prediction model research discuss how they deal with missing data. Information contained in EHRs is evaluated to ensure it is accurate and comprehensive. The missing information is imputed from the recognised EHR record. The Predictive Modelling approach is used for this, and the Naive Bayesian (NB) model is then used to assess the results in terms of performance metrics related to imputation. An adaptive optimisation technique, called the Adaptive Dolphin Atom Search Optimisation (Adaptive DASO) procedure, is used to teach the NB. The created Adaptive DASO method syndicates the DASO procedure with the adaptive idea. Dolphin Echolocation (DE) and Atom Search Optimisation (ASO) come together to form DASO. This indicator of performance metrics verifies imputation's fullness.

Keywords – Missing Data, Electronic Health Records, Naïve Bayesian, Adaptive Dolphin Atom Search Optimization, Machine Learning.

I. INTRODUCTION

Any study that involves or uses clinical data, such as studies of clinical prediction models, must give careful consideration to how missing data will be handled and reported.[1] Diagnostic and prognostic models in clinical prediction employ many input factors (i.e., covariates, predictors) to determine the absolute likelihood of an outcome's existence or occurrence. Most diagnostic and prognostic prediction models in the medical literature use regression modelling methodologies for derivation or validation. Additional work must be done in advance of developing a model when missing data are present in either the development or validation sample [2]. Individuals missing data on any of the predictor or outcome variables are (automatically) removed from the study in the most popular method, known as complete case analysis (CCA) [3]. In general, this approach is inefficient and can result in substantial bias in estimations of the estimated model parameters (for example, regression coefficients), negatively impacting the model's predictive accuracy. However, it is (only) valid under extremely strict circumstances. The loss of many useful observations can result, for instance, from discarding incomplete examples.

Multiple imputations based on additional (seen) patient variables are advised as a result, and multivariable imputation models are commonly used [4]. By using multiple imputation, prediction model coefficients may be computed independently for several, finalised versions of the incomplete datasets. Although multiple probabilities in new patients [5]. Because of this, multiple imputation methods may be used for putting theory into practise and using prediction models in real-world electronic healthcare software [6]. Another strategy takes into account missing information while building, validating, or using a prediction model. Incorporating missing indicator variables, using pattern mixture models, tree-based ensembles, or any number of other machine learning (ML) techniques that avoid missing data imputation are all viable options for this strategy (Box 1) [7].

Existing prediction model reporting guidelines (TRIPOD), in line with the growing body of corroborating literature, advise reporting at the very least on whether or not missing data was a problem in the development and validation sets for the prediction model, how severe the problem was, and what steps were taken to account for it in the analysis [8]. To far, it appears that only a minority of applied prediction studies are following these reporting standards. Many evaluations have indicated that missing data is often improperly handled or neglected [9], even in studies of prediction models that use more conventional (regression-based) methodologies. It is becoming less clear if and how missing data is handled during model construction and validation as ML approaches for predictive modelling emerge that may sidestep the necessity for imputation (for example, random forests with surrogate splits). However, it is unclear how frequently or effectively researchers that use these ML approaches apply appropriate alternative tactics [10].

In order to foretell concrete strength, several ML models have been built recently. In 1998, it was shown that a machine learning (ML) model based on an Network could accurately estimate the compressive strength of High-Performance concrete. The ANN model was found to outperform a regression-based model in terms of prediction [11]. Many different types of ML models, have been presented in the time since. Recent research [12, 13] have shown that expanding the dataset only reduces the volatility of the projected value once the model has mapped the link between the inputs and the output. This is despite the fact that training machine learning models needs enormous quantities of data. This research aims to examine how well prediction models of NB- sets, validation, and (if done) implementation of prediction models. Adaptive DASO optimises the weight of the NB to raise the precision of the classification.

The remaining sections of the paper are as shadows: Following a brief overview of the background literature in Section 2, a explanation of the optional perfect in Section 3, a discussion of the experimental analysis in Section 4, and a summary and conclusion in Section 5.

II. RELATED WORKS

In order to improve performance and generalisation, especially in data-poor scenarios, Hu et al. [14] introduce a novel negative regularisation improved R-Drop approach. When negative regularisation is applied to the R-Drop algorithm's output distributions, the distributions are forced to be inconsistent with one another. To ensure that our model has enough data to work with, we devise a strategy. The negative samples are obtained by taking the maximum value from the in-batch sample and subtracting the maximum value from the mini-batch sample. We use the resulting max-minus negative regularised dropout method to medical forecast datasets with both empty and complete instances to show its efficiency.

Getzen et al. [15] introduce an innovative method for simulating real-world scenarios with missing data in EHR and assessing the effect of such scenarios on predictive models. To do this, they provide a new tier of classification for electronic health record (EHR) data. We use a medical knowledge network to keep track of associations between medical events so that we can create a more plausible framework for missing data. The completion of our realistic missing data architecture is made possible by this. Individuals with lower likelihood of having access to or seeking ICU healthcare were shown to be more negatively impacted by missing data on illness prediction models. In addition, we discovered that the knowledge graph strategy outperforms the elimination of chance occurrences as a means of adding missing data that is representative of the actual world into illness prediction models.

Batra et al. [16] propose an ensemble imputation perfect that learns to use a mix of simple mean imputation, approaches, and then uses them in a way that chooses the best imputation strategy based on attribute correlations on missing value features. In healthcare data where missing values are prevalent, we provide a unique Ensemble Strategy for Missing Value to enable unbiased and reliable predictive statistical modelling. The results were compiled using the eXtreme gradient boosting regressor, the random forest regressor, and the support vector regressor. The suggested method outperforms both conventional missing value imputation techniques and the strategy of simply deleting records with missing values, as demonstrated by experiments and simulations done on real-world healthcare data with various feature-wise missing incidences.

In order to choose the best subset of incomplete features that can improve the learning procedure and maximise the forecast power of the perfect after it has been handled correctly, Awawdeh et al. [17] suggest using evolutionary the imputation for each feature on the performance of the prediction model. The drive of this study is to address the limitations of imputation by developing a new method for dealing with missing data, all while selecting features to improve the representation's learning presentation. The effectiveness of the suggested method was evaluated using a 10-folds cross validation test using ten standard datasets. Common imputation methods used to analyse the results were mean, median, multiple imputation, expectation maximisation, and K- neighbours. The suggested method outperformed the competition in every significant regard, including accuracy. In addition, when compared to three existing evolutionary based imputation strategies, the proposed methodology improved accuracy in 75% of the datasets.

Using a symmetric uncertainty-based L2 regularised regression ensemble and deep learning clustering, Nagarajan et al. [18] provide a novel tactic to lost data imputation in biological datasets. Genomic and non-genomic biological datasets with variable degrees of missing data are used in experiments to model different missingness distributions. We compare our approach to seven classic imputation approaches and two more recent ones. The experimental results show that the projected technique outdoes the other methods we evaluated in terms of computational competence since it preserves the dataset's structure. Therefore, if our proposed method for imputation of lost data works.

To combine the benefits of FIML estimation and self-attention neural networks, [19] develop a new technique they term Full Information Maximum Likelihood (FIML) Optimised Self-attention (FOSA). First, we use FIML to estimate missing values, and then we use the self-attention technique to improve our estimates. Experimental results from both synthetic and real-world datasets consistently show that FOSA outperforms standard FIML approaches in a number of important ways. Even though the Structural Equation Model (SEM) may be mis-specified, leading to subpar FIML estimates, the self-attention component of FOSA's robust architecture effectively corrects and optimises the imputation outputs. Our results demonstrate the robustness and generalizability of FOSA in data imputation by proving its capacity to consistently provide high-quality predictions in the face of up to 40% random missingness.

For missing at random (MAR) type missing data in IoMT, Iris Punitha et al. [20] present a unique Two Tier Missing Data Imputation (TT-MDI) strategy based on an improved linear interpolation method. The cStick IoMT dataset from the Repository was used to evaluate the proposed TT-MDI technique for imputation of MAR missing data. The first level attempts to determine the imputation threshold by using the distances among the class centroids and the associated data instances. The second layer uses the determined cutoff to fill in any blanks in the data. The experimental results show that when using the TT-MDI approach has better accuracy.

III. PROPOSED METHOD

In this section, we first present the imputation procedure.

Imputation For Missing Data

Loss to follow-up, inadequate replies to surveys and questionnaires, and data entry mistakes are all common causes of missing data in medical research. Incorrect treatment of missing data can compromise model generalisation performance by leading to inaccurate estimates. Researchers must carefully choose a strategy for dealing with missing data in command to reduce the possibility of bias and increase the reliability of their results.

Imputation is a common procedure for dealing with instances of missing data. Some characteristics in medical research have continuous values, such as time spent in therapy or patient age. Boolean values are used to express numerous different characteristics, like sex and dizziness. Imputation is thus carried out differently depending on whether the value is continuous or boolean.

In this research, we experimented with the regular imputation method for the continuous missing data and attempted the mode imputation, and Naive Bayes imputation methods for the boolean missing data. Based on the accuracy of their forecasts, we made use of the mode imputation method in our final design. Imputation strategies for missing data in healthcare data are described here, along with their use in the experiments. Assuming a time series with several variables $X = \{x_1,\dots,x_i,\dots,x_t\}^\intercal \in R^{t\times d}$ as a sequence of features. The ith observation x_i consists of d features $\{x_i^1,\dots,x_i^j,\dots,x_i^d\}$ observed features, we introduce a mask matrix of $t\times d$ dimensions, i.e., $M \in R^{\mathrm{t\times d}}$, whose element m_i^j characterizes whether the consistent feature x_i^j is experiential or not, i.e., $m_t^j = 1$ if x_i^j is observed; otherwise, $m_i^j= 0$.

The imputation of a lost feature is distinct as $\delta\left(x_i^j\right) \rightarrow \hat{x}_i^j \in R$, where x_i^j is a missing feature $m_i^j = 0$, and $\hat{x}_i^j$ is a characteristic that has been imputed using a specific approach. Here we have a multivariate period series X, its matrix M, and the resulting imputed series X′, which includes both the original characteristics of X and the features projected using the accusation procedure.

$$X = \begin{pmatrix} x_1^1 & - & x_1^3 & - \\ x_2^1 & x_2^2 & - & x_2^4 \\ x_3^1 & - & x_3^3 & x_4^4 \end{pmatrix}, M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}, X' = \begin{pmatrix} x_1^1 & \hat{x}_1^2 & x_1^3 & \hat{x}_1^4 \\ x_2^1 & x_2^2 & \hat{x}_2^3 & x_2^4 \\ x_3^1 & \hat{x}_3^2 & x_3^3 & x_3^4 \end{pmatrix} \tag{1}$$

Excluding observations (via methods like pairwise and listwise deletion) that have a missing feature is the easiest way to deal with missing features in an incomplete time series. However, these methods may reduce data quality and provide inaccurate estimations. This is why mean replacement and other strategies that require evenly features with a given value are increasingly popular. All the missing parts x_i^j where $m_i^j = 0$, the zero replacement applies zero as the credited feature, i.e., $\hat{x}_i^j = 0$. Likewise, the replaces xij with the mean $\bar{x} = \frac{\sum x_a^b}{n}$, where n is the sum of non-missing topographies in X, and x_a^b is feature with $m_i^j = 1$ therefore, $\hat{x}_i^j = \bar{x}$.

Datasets description

The trials we conducted made use of three different medical prediction databases. The UCI machine learning repository (ASUNCION, 2007) was mined for the open-source Pima Indian Diabetes (PID) and Wisconsin Breast Cancer (WBC) datasets. Our final dataset was produced using information from the TCM Syndrome Biological Technology Platform, a part of the Chinese National Science and Knowledge Chief Project. Table 1 displays summary statistics for all three data sets [14]. The WBC and PID databases are both full, however ours is missing 275 values. In some cases, particularly those involving traditional Chinese medicine treatment, a clear diagnosis may not be possible until further features are collected throughout the diagnostic process. Therefore, we have omitted from our dataset any records that have more than three missing values so that we may do statistical analysis.

Table 1. The Comparison of the Three Datasets.

Statistics	PID Dataset	WBC	Our Dataset
Sum of instances	768	dataset	1920
Sum of categories	2	683	7
Sum of features	8	2	24
Sum of missing values	0	9	275
Sum of training instances	537	0	1,344
Sum of testing instances	231	478	576

Our data collection is comprised of information regarding T2DM syndrome therapies using traditional Chinese medicine. The seven categories into which diabetic symptoms fall. The patient's age and treatment plan are continuous features in our dataset, whereas the remaining characteristics are boolean. Incomplete information is only present in the boolean features. Our data also shows that the majority of patients are male, with the majority of patients falling between the ages of 40 and 80 over the span of 5-20 years. The most prominent manifestations of this illness are a lack of saliva production, increased thirst, and paresthesias and tingling in the feet and legs. In particular, missing data for continuous values are not present in the dataset we have collected.

The WBC dataset contains 683 data points, each of which describes one of nine extracted picture attributes. Features can have values between 1 to 10, where 1 represents a normal or benign case and 10 represents the most aberrant instance possible given the diagnosis. Each of the 768 data points in the Pima Indian Diabetes (PID) dataset has eight different medical characteristics. Two-hundred-seventy-eight of these represent diabetic patients, whereas five-hundred represent those who are not diabetic.

Naive Bayesian Equation:

A Naive Bayes algorithm recognises the occurrence of an event based on the probability of a second event. Bayesian techniques of classification, which Naive Bayes classifiers emphasise, describe the link between numerical value conditional probabilities using an algorithm. The Bayesian classification is of interest because it can be represented as (L—feature) P(L—feature), where L is the label and P is the set of observed features.

The Bayesian theorem puts this in terms of quantifiable terms that are easier to calculate.

$$P(L\ /\ feature) = P(feature\ /\ L)P(L)P(feature)P(L\ /\ feature) = P(feature\ /\ L)P(L)P(feature) \qquad (2)$$

To choose between L1L1 and L2L2, one approach involves calculating the ratio of the posterior probabilities for the two labels.;

$$P(L1\ /\ feature)P(L2\ /\ feature)\ =\ P(feature\ /\ L1)\ P(feature\ /\ L2) \qquad (3)$$

$$P(L1)P(L2)P(L1\ /\ feature)P(L2\ /\ feature)\ =\ P(feature\ /\ L1) \qquad (4)$$

$$P(feature\ /\ L2)\ =\ P(L1)P(L2) \qquad (5)$$

One type of model fits this description; it is termed a generative model since it describes an imaginary random process that generates data. Determining this generative model for each tag is a crucial part of training for a Bayesian classifier. Simplifying the structure of this model can make the generalised form of such a learning phase more manageable, but it is still a difficult undertaking. A approximate estimate of the generating model for each class was evaluated, and the Bayesian classification was proceeded after a very naive assumption was made about the generative model for each label. Various naïve Bayes classifiers make several simplistic assumptions about the data.

Mathematically, the theorem of Bayes may be expressed as

$$P(A\ /\ B)\ =\ (P(B\ /\ A)\ P(A))\ \ /\ P(B) \qquad (6)$$

the two occurrences A and B.

Using a dataset of 1000 patient records as an example, the probability of a certain event can be estimated using the rule R2 as follows: If the main term and sub term2 are known to replace sub term1, then the replacement is performed precisely due to the uniqueness of the structure.

.Let us deliberate the case $(mainterm = "disturbance"\ AND\ subterm2 = "salivary") -> sub - term1 = ??????)$

$According\ to\ the\ rule\ R2,$

$SubTerm1 = "Secretion"\ if\ mainterm = "Disturbance"\ and\ sub -\ term2 = "Salivery"$

$Therefore, the\ probability\ is\ calculated\ as;$

$$P\left(\frac{Secretion}{Disturbance}\right) = \frac{\left(P\left(\frac{Disturbance}{Secretion}\right)*P(Secretion)\right)}{P(Disturbance)} \tag{7}$$

$P(Secretion)$ = 32/1000

$P(Disturbance)$ = 32/1000

$P(Disturbance/Secretion)$ = 32/32

$P(Secretion/Disturbance)$ = ((32/32) ∗ (32/1000)) / (32/1000) = 1

It's also possible to have complete knowledge of the primary word while being unaware of any of the associated sub terms. Furthermore, it is usually preferable to avoid scenarios when the sub term has more than one alternative for filling, as this does not result in the greatest imputation accuracy.

Here's an example in which we have only the main term known and are lacking terms 1 and 2.

.

(MainTerm="disorder" AND SubTerm1=????) − > SubTerm2=????

Subterm1's value must be determined first, then Subterm2's value may be determined from Subterm1. When disorder is the primary term, panic disorder and major depressive disorder are both possible for subterm1. Accuracy for both values must be calculated using RBC and Naive Bayesian approach in order to perform the right imputation..

i) $R6 : (MainTerm = "disorder") -> SubTerm1 = "depression"$

By means of Rule Based Classifier:

$Coverage(R6)$ = 96/1000 = 0.096

$Accuracy(R6)$ = 32/96 = 0.33

By means of Naive Bayesian Equation:

$$P\left(\frac{Depression}{Disorder}\right) = \left(\frac{P\left(\frac{Disorder}{Depression}\right)*P(Depression)}{P(Depression)}\right) \tag{8}$$

$$P(Disorder\ /\ Depression) = ((32\ /\ 32\ )\ *\ (32\ /\ 1000))\ /\ (96/1000) = 32/96 = 0.33$$

ii) R7 : $(MainTerm = "disorder") -> SubTerm1 = "panic"$

$Using\ Rule\ Based\ Classifier$:

Using Naive Bayesian Equation:

$$P\left(\frac{Panic}{Disorder}\right) = \left(\frac{P\left(\frac{Disorder}{Panic}\right)*P(Panic)}{P(Depression)}\right) \tag{9}$$

$$P\left({}^{Disorder}/_{Depression}\right) = \frac{\left(\left({}^{32}/_{32}\right) *\ \left({}^{64}/_{1000}\right)\right)}{(96/1000)\ =\ 64/96} =\ 0.67$$

It is preferable to delete these entries from the dataset rather than impute an erroneous number, given the accuracy level is not 100% in any scenario.

Hyper-parameter tuning of Naïve Bayes

The created Adaptive DASO is used to train the NB classifier. The Adaptive DASO was established by fusing the ideas presented in DE [21] and [22] ASO with the Adaptive notion. Every atom in DASO interacts with every other atom during the startup phase, and then it uses its repulsion qualities to drive away any premature or overconcentrated atoms. Finally, all atoms exhibit attractive behaviour towards one another, guaranteeing full use of the optimisations that have been produced. Coding the Answer: For NIDS with a fitness metric, the proposed optimisation is used to estimate the best solution with a lower error rate. The developed Adaptive RNN involves the following phases during implementation:

Population beginning

Let ν is sum of atoms and the site of dth atom is portrayed as,

$$I_d = [I_d^1, \dots, I_d^a]; d = [1, \dots., l] \tag{10}$$

where I_d^a denotes the ath site constituent of dth atom

Fitness function

The best result is chosen based on its fitness function, which is assessed by calculating the dissimilarity between the anticipated and classifier outputs.,

$$\sigma_d = \frac{1}{\mu}\sum_{s=1}^{\mu} R_s^{(i,j)} - \varepsilon_s \tag{11}$$

where, σ_d designates the fitness rate of dth atom, $R_s^{(i,j)}$ portrays the classifier yield and ε_s specifies the foretold output.

Compute the mass

The mass of the dth atom after f iterations is approximated using the fitness function and is given as,

$$M_d(f) = \frac{e_d(f)}{\sum_{d=1} e_f(f)} \tag{12}$$

where, $M_d(f)$ designates the mass, and the term $e_d(f)$ is uttered as,

$$e_d(f) = \frac{\sigma_d - \sigma_{best}}{e^{\sigma_{worst} - \sigma_{best}}} \tag{13}$$

σ_{best} and σ_{worst} defines a maximum and minimum value, and its expression looks like,

$$\sigma_{best} = \underset{d=1,\ldots,l}{min} \sigma_d \tag{14}$$

$$\sigma_{worst} = \underset{d=1,\ldots,l}{min} \sigma_d \tag{15}$$

Evaluate N Neighbor

The fitness value of interactions between atoms is used to pick the N neighbours, which improves the initial iteration's exploration. N is represented by the following expression,

$$N(f) = l - (l - 2)\sqrt{\frac{f}{d}} \tag{16}$$

Compute the Total force and Constraint force

The total force is defined as the sum of the forces exerted on the dth atoms by their nearest neighbours, and this statement is assessed as,

$$Q_d^a(f) = \sum_{s \in N_{best}} rand_s Q_{ds}^a(f) \tag{17}$$

where, $Q_d^a(f)$ denotes the strength, and the term specifies a random value between 0 and 1, with 0 indicating no force and 1 indicating maximum force. All particles in the population space act optimally, and the pressure exerted by the dth atom's constraints may be written as,

$$\lambda_d^a(f) = H(f)(I_{best}^a(f) - I_d^a(f)) \tag{18}$$

where, H(f) designates multiplier.

Approximation the acceleration

The dth atom hastening at fth phase is intended as,

$$A_d^a(f) = \frac{Q_d^a(f)}{M_d^a(f)} + \frac{\lambda_d^a(f)}{M_d^a(f)} \text{ (19)}$$

where, $Q_d^a(f)$ total force, $\lambda_d^a(f)$ constraint force, $M_d^a(f)$ indicates the mass and $A_d^a(f)$ designates hastening of dth atom at fth period.

Renew the velocity

The velocity of dth atom at f + 1 repetition is articulated as,

$$V_d^a(f+1) = rand_d^a V_d^a(f) + A_d^a(f) \tag{20}$$

Where, $rand_d^a$ indicates the random number, and $A_d^a(f)$ specifies the acceleration

Inform the atom location

As a result, the final DASO method update equation is as follows:.

$$I_d(f+1) = \frac{\omega_{2d} M_d(f)}{\omega_{2d} M_d(f) - ze^{\frac{-20f}{a}}} \begin{bmatrix} I_d(f) + rand_d V_d(f) - \psi\left(2 - \frac{f-1}{a}\right)^3 \\ e^{\frac{-20f}{a}} \sum_{s \in N_{best}} \frac{rand_s[2 \times c_{ds}(f)^{13} - (c_{ds})^7]}{M_d(f)} \\ \frac{(I_s(f) - T_d(f))}{\| I_d(f), I_s(f) \|_2} - Ze^{\frac{-20f}{a}} \\ \frac{I_d(f) + W_d(f) + \omega_{1d} J_d - \omega_{1d} I_d(f)}{\omega_{2d} M_d(f)} \end{bmatrix} \tag{21}$$

where, $M_d(f)$ stipulates the mass of dth atom, $V_d(f)$ is the velocity, Z specifies the multiplier weight , ψ stipulates the depth weight, α demonstrations the extreme repetition, W_d signifies the dimension, J_d depicts the personal greatest solution, ω_{1d} and ω_{2d} are the accidental sum that lies among 0 to 1, correspondingly.

Here, Adaptive concept is presented in ψ from above equation for better presentation of sentiment classification. The appearance ψ is given by,

$$\psi = \psi_{max} - \frac{f(\psi_{max} - \psi_{min})}{a} \tag{22}$$

where, α suggests the depth Adaptive, ψ_{max} and ψ_{min} depicts the maximum and minimum rate of ψ and α indicates the maximum repetition.

Re-compute the fitness:

The goal function, shown in Equation 11, is used to make predictions about the fitness worth, and the solution with the highest fitness is best.

Termination

The process described above is iterated over and over again until the termination conditions are met. The merging of ASO and DE with the adaptive notion yields a more optimum outcome and less computing time.

IV. RESULTS AND DISCUSSION

This section details the investigational evaluation of the projected perfect on three different data sets. The hardware setup for the experiments included a MATLAB 2018a workstation, a Windows 11 desktop, and an Intel(R) Core(TM) i5-12500H mainframe running at 3.1 GHz and 16 GB of RAM. We also used the k-fold cross-validation test in our analysis. Our performance measurements are computed using Equations (23-24) whereas training and prediction times were the primary focus of execution speed measures.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{23}$$

$$F1Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{24}$$

Analysis Of Dropout Rate

Table 2 presents the analysis of dropout rate for three datasets.

Table 2. Experimental Analysis on Three Datasets

Dropout Rate	WBC	PID	Collected dataset
0.1	96.47	93.49	90.28
0.2	97.81	97.22	93.23
0.3	98.54	97.62	91.01
0.4	96.71	97.81	89.38
0.5	97.62	96.71	92.97

In the above **Table 2** characterise that the experimental analysis on three datasets. In the investigation we used different dropout rate to determine the performance. In the analysis of 0.1 dropout rate, the WBC as 96.47 and then PID as 93.49 and also the collected dataset as 90.28 correspondingly. Then the 0.2 dropout rate, the WBC as 97.81 and then PID as 97.22 and also the collected dataset as 93.23 correspondingly. Then the 0.3 dropout rate, the WBC as 98.54 and then PID as 97.62 and also the collected dataset as 91.01 correspondingly. Then the 0.4 dropout rate, the WBC as 96.71 and then PID as 97.81 and also the collected dataset as 89.38 correspondingly. Then the 0.5 dropout rate, the WBC as 97.62 and then PID as 96.71 and also the collected dataset as 92.97 correspondingly. **Fig 1** shows Graphical Representation of proposed model

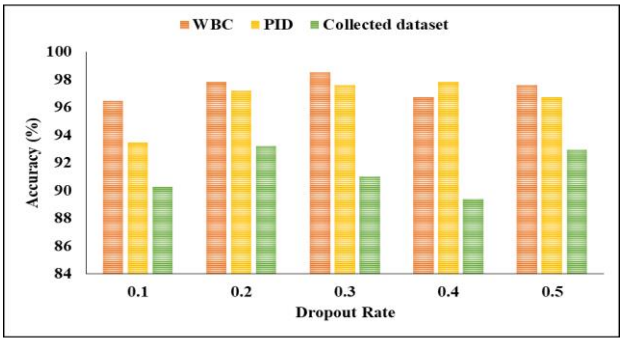

Fig 1. Graphical Representation of Proposed Model

Analysis of Proposed model on missing ratio

Based on the missing ratio, the below **Table 3** presents the validation analysis on three datasets.

Table 3. Experimental Analysis on proposed Model on three datasets

Missing Ratio	WBC		PID		Collected data	
	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
5%	84.27	72.45	76.34	72.5	72.45	74.11
10%	88.67	87.91	78.75	76.25	78.35	77.72
15%	73.55	74.18	79.79	78.75	79.25	79.28
20%	70.78	88.27	82.51	79.44	89.20	87.18
25%	95.87	92.26	93.98	83.05	94.64	90.25

In the above **Table 3** signifies that the Experimental Analysis on projected Model on three datasets. In the evaluation of different missing ratios, in the 5% of missing ratio, the WBC dataset accuracy as 84.27 and F1-score range of 72.45 and another PID dataset, the accuracy value as 76.34 and the F-score range as 72.5 and additionally the collected dataset, the accuracy value as 72.45 and then F1-score rate as 74.11 correspondingly. Then the 10% of missing ratio, the WBC dataset accuracy as 88.67 and F1-score range of 87.91 and another PID dataset, the accuracy value as 78.75 and another PID dataset, the accuracy value as 76.25 and additionally the collected dataset, the accuracy value as 78.35 and F1-score range of and additionally the collected dataset, the accuracy value as 77.72 correspondingly. Then the 15% of missing ratio, the WBC dataset accuracy as 73.55 and F1-score range of 74.18 and another PID dataset, the accuracy value as 79.79 and another PID dataset, the accuracy value as 78.75 and another PID dataset, the accuracy value as 79.25 and additionally the collected dataset, the accuracy value as 79.28 correspondingly. Then the 20% of missing ratio, the WBC dataset accuracy as 70.78 and F1-score range of 88.27 and another PID dataset, the accuracy value as 82.51 and F1-score range of 79.44 and additionally the collected dataset, the accuracy value as 89.20 and F1-score range of 87.18 correspondingly. Then the 25% of missing ratio, the WBC dataset accuracy as 95.87 and F1-score range of 92.26 93.98 and another PID dataset, the accuracy value as 83.05 and additionally the collected dataset, the accuracy value as 94.64 and F1-score range of 90.25 correspondingly. **Fig 2** show in Graphical Comparison on Missing Ratio

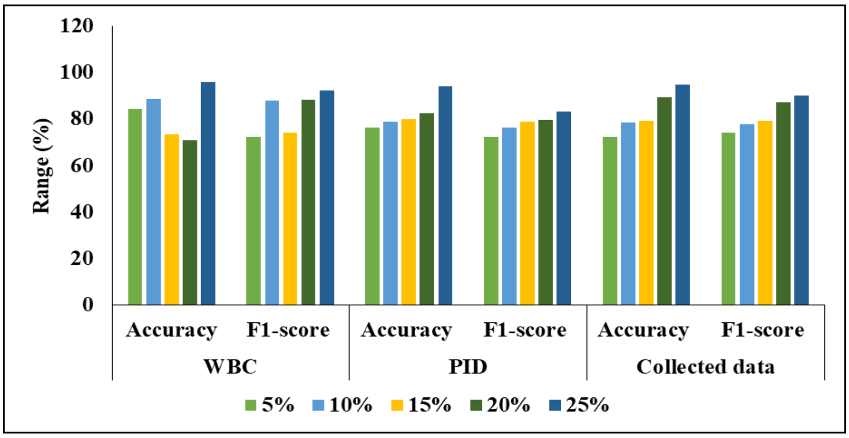

Fig 2. Graphical Comparison on Missing Ratio

V. Conclusion

To further improve presentation and generalisation ability, especially in the case of missing data, we present a straightforward yet effective negative regularisation technique built upon NB. The suggested model preserves both the independence of the deliveries of positive and negative samples at the output level, as well as the distributions derived from the same data sample. Adaptive DASO model is used to fine-tune the NB's hyper-parameters for maximum performance. By combining DA and ASO with an adaptive framework, we get the Adaptive DASO algorithm. In addition, we develop a novel max-minus negative sampling method that is more efficient than the standard in-batch negative example sampling technique and aids in the convergence process. Validation of the usefulness of the suggested strategy, especially in the situation of missing data, is provided by extensive experimental findings on medical datasets containing both full and missing data cases. Using a deep learning architecture and a hybrid optimisation model, it will be possible to forecast the missing healthcare data in the future.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, “A survey on missing data in machine learning,” Journal of Big Data, vol. 8, no. 1, Oct. 2021, doi: 10.1186/s40537-021-00516-9.

[2] Dubey and A. Rasool, “Efficient technique of microarray missing data imputation using clustering and weighted nearest neighbour,” Scientific Reports, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-03438-x.

[3] N. U. Okafor and D. T. Delaney, “Missing Data Imputation on IoT Sensor Networks: Implications for on-Site Sensor Calibration,” IEEE Sensors Journal, vol. 21, no. 20, pp. 22833–22845, Oct. 2021, doi: 10.1109/jsen.2021.3105442.

[4] M. Pazhoohesh, A. Allahham, R. Das, and S. Walker, "Investigating the impact of missing data imputation techniques on battery energy management system," IET Smart Grid, vol. 4, no. 2, pp. 162–175, Feb. 2021, doi: 10.1049/stg2.12011.

[5] C.-Y. Guo, Y.-C. Yang, and Y.-H. Chen, "The Optimal Machine Learning-Based Missing Data Imputation for the Cox Proportional Hazard Model," Frontiers in Public Health, vol. 9, Jul. 2021, doi: 10.3389/fpubh.2021.680054.

[6] F. B. Hamzah, F. Mohd Hamzah, S. F. Mohd Razali, and H. Samad, "A Comparison of Multiple Imputation Methods for Recovering Missing Data in Hydrological Studies," Civil Engineering Journal, vol. 7, no. 9, pp. 1608–1619, Sep. 2021, doi: 10.28991/cej-2021-03091747.

[7] K. Naveen Durai, R. Subha, and A. Haldorai, "Hybrid Invasive Weed Improved Grasshopper Optimization Algorithm for Cloud Load Balancing," Intelligent Automation &amp; Soft Computing, vol. 34, no. 1, pp. 467–483, 2022, doi: 10.32604/iasc.2022.026020.

[8] G. H. Lee, J. Han, and J. K. Choi, "MPdist-based missing data imputation for supporting big data analyses in IoT-based applications," Future Generation Computer Systems, vol. 125, pp. 421–432, Dec. 2021, doi: 10.1016/j.future.2021.06.042.

[9] M. Pazhoohesh, M. S. Javadi, M. Gheisari, S. Aziz, and R. Villa, "Dealing with Missing Data in the Smart Buildings using Innovative Imputation Techniques," IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, Oct. 2021, doi: 10.1109/iecon48115.2021.9612650.

[10] D. Cenitta, R. V. Arjunan, and P. K V, "Missing Data Imputation using Machine Learning Algorithm for Supervised Learning," 2021 International Conference on Computer Communication and Informatics (ICCCI), Jan. 2021, doi: 10.1109/iccci50826.2021.9402558.

[11] Z. Alruhaymi and C. J. Kim, "Study on the Missing Data Mechanisms and Imputation Methods," Open Journal of Statistics, vol. 11, no. 04, pp. 477–492, 2021, doi: 10.4236/ojs.2021.114030.

[12] S. R and A. H, "Improved EPOA clustering protocol for lifetime longevity in wireless sensor network," Sensors International, vol. 3, p. 100199, 2022, doi: 10.1016/j.sintl.2022.100199.

[13] T. Thomas and E. Rajabi, "A systematic review of machine learning-based missing value imputation techniques," Data Technologies and Applications, vol. 55, no. 4, pp. 558–585, Apr. 2021, doi: 10.1108/dta-12-2020-0298.

[14] L. Hu, X. Cheng, C. Wen, and Y. Ren, "Medical prediction from missing data with max-minus negative regularized dropout," Frontiers in Neuroscience, vol. 17, Jul. 2023, doi: 10.3389/fnins.2023.1221970.

[15] E. Getzen, L. Ungar, D. Mowery, X. Jiang, and Q. Long, "Mining for equitable health: Assessing the impact of missing data in electronic health records," Journal of Biomedical Informatics, vol. 139, p. 104269, Mar. 2023, doi: 10.1016/j.jbi.2022.104269.

[16] S. Batra, R. Khurana, M. Z. Khan, W. Boulila, A. Koubaa, and P. Srivastava, "A Pragmatic Ensemble Strategy for Missing Values Imputation in Health Records," Entropy, vol. 24, no. 4, p. 533, Apr. 2022, doi: 10.3390/e24040533.

[17] S. Awawdeh, H. Faris, and H. Hiary, "EvoImputer: An evolutionary approach for Missing Data Imputation and feature selection in the context of supervised learning," Knowledge-Based Systems, vol. 236, p. 107734, Jan. 2022, doi: 10.1016/j.knosys.2021.107734.

[18] G. Nagarajan and L. D. Dhinesh Babu, "Missing data imputation on biomedical data using deeply learned clustering and L2 regularized regression based on symmetric uncertainty," Artificial Intelligence in Medicine, vol. 123, p. 102214, Jan. 2022, doi: 10.1016/j.artmed.2021.102214.

[19] R. Subha and A. Haldorai, "An Efficient Identification of Security Threats in Requirement Engineering Methodology," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–14, Aug. 2022, doi: 10.1155/2022/1872079.

[20] P. I. Punitha and J. G. R. Sathiaseelan, "A Novel Two Tier Missing at Random Type Missing Data Imputation using Enhanced Linear Interpolation Technique on Internet of Medical Things," Indian Journal Of Science And Technology, vol. 16, no. 16, pp. 1192–1204, Apr. 2023, doi: 10.17485/ijst/v16i16.60.

[21] G. M. Borkar and A. R. Mahajan, "A secure and trust based on-demand multipath routing scheme for self-organized mobile ad-hoc networks," Wireless Networks, vol. 23, no. 8, pp. 2455–2472, May 2016, doi: 10.1007/s11276-016-1287-y.

[22] W. Zhao, L. Wang, and Z. Zhang, "A novel atom search optimization for dispersion coefficient estimation in groundwater," Future Generation Computer Systems, vol. 91, pp. 601–610, Feb. 2019, doi: 10.1016/j.future.2018.05.037.