MVIBPM: DESIGN OF A MISSING VALUE IDENTIFICATION TECHNIQUE VIA BIOINSPIRED PREDICTIVE MODELING

Dipalika Das [1*], Maya Nayak[2], Subhendu Kumar Pani[3]

[1] Dipalika Das, Asst. Prof., Trident Academy of Creative Technology, Research Scholar, Department of Computer Science and Engineering, Biju Patnaik University of Technology, Rourkela, Odisha, India.

[2] Dr. Maya Nayak, Professor, Trident Academy of Creative Technology (TACT), Bhubaneswar, Biju Patnaik University of Technology (BPUT) Rourkela, Odisha, India.

[3] Dr. Subhendu Kumar Pani, Professor, Krupajal Engineering College (KEC), Bhubaneswar, Biju Patnaik University of Technology (BPUT), Rourkela, Odisha, India.

Abstract: Detecting absent values in time-series data samples is a challenging signal-processing task that requires pattern analysis, proactive modeling, and regression methods. Researchers propose various models to optimize the efficiency of missing value identification techniques. Most of them remain intricate and unsuitable for extensive information sets. Additionally, the limited effectiveness of basic models when dealing with extensive datasets restricts their applicability for real-time uses. In order to address these challenges, this article introduces a new Elephant Herding Optimization (EHO) Model that aims to enhance an effective ensemble classifier for identifying missing values, particularly suited for feature-based data samples. The proposed model utilizes a combination of Deep Forest (DF), Support Vector Machines (SVM), Naïve Bayes (NB), and k Nearest Neighbour (kNN) classifiers to examine relationships in samples that have missing values. The EHO model optimizes the effectiveness of the proposed classifier by helping to determine hyperparameters for the classifiers, thereby enhancing the performance of the process for identifying missing values. The EHO model employs a practical fitness function that integrates the accuracy, precision, and recall metrics achieved in assessing the efficacy of the process for identifying missing values. In order to assess its effectiveness, the model was applied to several extensive datasets, and an accuracy improvement of 9.5%, with a precision improvement of 8.3%, and a recall improvement of 4.5% was observed when compared with standard regression-based pre-emption models. Due to this, the proposed method was highly scalable and can be applied to multidomain use cases.

Keywords: Missing, Value, NB, kNN, SVM, DF, EHO, Accuracy, Precision, Recall, Optimizations

1. Introduction

A time series generally refers to a collection of measurements obtained at consistent time intervals. The fundamental objective of time series prediction is to foretell future tendencies in the data by examining past data. This is accomplished via the use of historical data. As a result, it plays an essential part in the decision-making process for various applications, including industrial monitoring, business metrics, management of electrical grids, and other applications. The following is a list of probable overarching categories for the challenges with the time series. Suppose we are just concerned with the next one- or two-time steps, as with the overwhelming majority of time series issues. In that case, we will create a one-step or single-step forecast prediction. Predictions that look at the future in multiple steps are called

multistep predictions since they gaze far into the future. When making a prediction involving several stages, you can use either the direct or iterative methods. The direct approach requires the creation of a model that can foretell the outcomes of many stages in the future. At the same time, the iterative method necessitates the creation of a series of predictions for one step at a time up until the relevant step is reached. Time series prediction has seen a recent uptick in using artificial intelligence (AI). The support vector machine is a well-known artificial intelligence technology for time series prediction (SVM). During the 1970s, Vapnik and his fellow employees at AT&T Bell Laboratories made essential strides in the development of SVM. It was first developed to assist with categorization issues and had practical uses such as optical character recognition. In [1, 2, 3], an improvement was made to the support vector machine to address issues with regression. Neural networks must be concerned with local minimums, but support vector machines do not have this worry. However, to tackle quadratic programming issues, a significant amount of computing power is required. The Takagi-Sugeno Modeling (TSM) and Least Squares Support Vector Machine, sometimes known as the LSSVM, were presented in reference [4, 5, 6] as a method for converting constraint problems into a linear system. Although the LSSVM is better at cutting down on computational expenses, the sparsity of the support vectors is lost in the process. The weighted LSSVM is an alternate strategy that was presented to cope with sparsity and also used Local Median-based Gaussian Naive Bayes (LMeGNB) [7, 8, 9]. LSSVM has recently been successful in several domains, including time series prediction and financial forecasting. Since the data for time series are obtained from real-world scenarios, it is common for values to be missing. Failures of the sensors or mistakes made by humans might explain the missing data. [10, 11, 12, 13] Numerous ad hoc approaches have been investigated over the years to address the issue of missing data. These methods encompass different techniques and elimination processes to find a single replacement for each absent value. Ad hoc approaches can potentially affect both standard errors and biases in estimates [14, 15]. Despite this, research demonstrates that they are used regularly [16, 17, 18]. The maximum likelihood method [19, 20] and the multiple imputation methods [21, 22] are two of the most well-known and successful approaches to imputing missing data. When using various imputations, copies of the missing information are first generated, and then those duplicates are separately imputed. The ultimate judgment was arrived at by compiling the results of several parameter estimates and standard errors, one of each kind for every copy examined. Additionally, maximum likelihood approaches using Generative Adversarial Networks (GANs) and bi-directional long short-term memory (Bi-LSTM) [23, 24, 25] take into account all accessible data and produce estimates that indicate the highest likelihood. Since multiple imputations and maximum probability often provide the same results, choosing one over the other is a very subjective decision. Making predictions based on time series is difficult since there is missing information. Compared to different types of data analysis, time series prediction stands out due to the temporal relevance of its predictions.

These findings imply that academics have developed a broad range of models in an effort to improve the efficiency of missing value detection techniques. Despite this, the great majority of these models are very complex to implement and cannot be used for enormous information sets. In addition, the low levels of efficiency that simpler models have when applied to large-scale datasets are a hurdle to the development of real-time applications. This paper provides a unique Elephant Herding Optimization (EHO) Model for optimizing an efficient ensemble classifier for missing value detection that is applicable to feature-based data sets. The goal of this model is to overcome the problems that have been identified. In section 3 of this text, using a wide range of datasets, an evaluation of the usefulness of the model was carried out. This performance was evaluated in comparison to industry standards in order to demonstrate its superiority over contemporary models. The inquiry concludes with some overall

reflections on the provided work, along with suggestions for expanding its relevance across various applications.

2. Proposed Missing Value Identification technique via Bioinspired Predictive Modeling

Upon reviewing the current models for identifying missing values, it becomes evident that many of them are quite intricate and unsuitable for large-scale data sets. Additionally, the more straightforward models that are utilized for extensive datasets often exhibit low efficiency, which restricts their use in real-time applications.

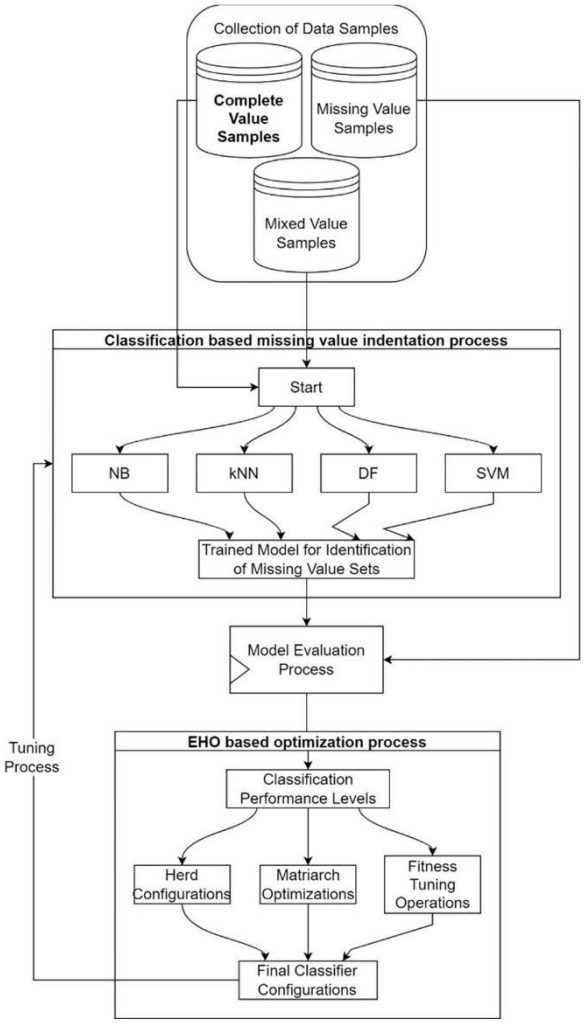

Illustration 1. Sequence of the suggested process for identifying missing values.

To address these challenges, this section presents the development of a new Elephant Herding

Optimization (EHO) Model aimed at optimizing a robust missing value identification ensemble classifier, suitable for feature-based data samples. The entire process of the suggested model is illustrated in Figure 1, which demonstrates that the model utilizes a combination of Deep Forest (DF), Support Vector Machines (SVM), Naïve Bayes (NB), and k Nearest Neighbour (kNN) classifiers to examine the relationship of samples with missing values. The effectiveness of the proposed classifier is enhanced through the EHO model, which helps in determining the hyperparameters of the classifier to boost the efficacy of the missing value identification process. The EHO model incorporates an effective fitness function that integrates accuracy, precision, and recall metrics achieved while assessing the effectiveness of the missing value identification process.

The process indicates that training datasets are utilized to develop the classifiers Naïve Bayes (NB), k Nearest Neighbour (kNN), Deep Forest (DF), and Support Vector Machine (SVM). The classifiers, together with their starting parameter configurations, are presented in the table below.

Classifier	**Parameter Sets**
Naïve Bayes	Priors are determined according to the variance levels present in the training set's samples. The Smoothing Value (Sv) starts at one and is adjusted through the EHO process.
kNN	k = 1, and tuned by the EHO process
Deep Forest	Number of Estimators (N_{est}), initially set as the number of features and tuned by the EHO process Max Depth (M_{dep}) , initially set as 1, and later modified by the EHO process
SVM	Regularization Coefficient (C), initially set as 1, and modified by the EHO process Tolerance (tol), initially set as 0.0001, later modified by the EHO process

Table 1. Classifiers along with their parameter sets

Based on these parameter sets, missing value samples are classified into 1 of N categories. The average value of missing parameters (MPV) is evaluated via equation 1,

$$MPV = \sum_{i=1}^{N_c} \frac{NMVP_I}{N_c} \ldots. (1)$$

Where, NMVP and N_C indicate the values of parameters that are available in the dataset and the overall number of samples found in the specified class. Based on this value of MPV, the accuracy of the classifier is estimated and kept for future reference purposes. If it is found to be beneath a certain threshold, an optimization model based on EHO is triggered, which operates according to the following procedure.

- Optimization framework, organized according to EHO constants,
  - EHO iterations during which Herds will be evaluated and adjusted (Ni)
  - EHO Herds that will be utilized in the optimization process (Nh)
  - The rate at which Herds will learn from one another (Lr)
  - Current parameters for each of the classifiers obtained from Table 1, which has to be optimized by the EHO process
- Once these parameters are set, then generate N_{h} solutions as per the following process,
  - Stochastically modify values for each of the classifier parameters via equations, 2, 3, 4, 5, 6, and 7 as follows,

$$S_v = S_v(Old) \pm STOCH\left(\frac{L_r}{2}, L_r\right) \ldots\ldots. (2)$$

Where STOCH is a Markovian process utilized for generating stochastic number sets.

$$k = k(old) \pm 1 \ldots\ldots (3)$$

Where, increment (+), and decrement (-) operators are selected stochastically for individual solution sets.

$$N_{est} = N_{est}(Old) * STOCH\left(\frac{L_r}{2}, 2 * L_r\right) \ldots\ldots (4)$$

$$M_{depth} = M_{depth}(old) \pm 1 \ldots\ldots. (5)$$

$$C = C(old) * STOCH\left(\frac{L_r}{2}, L_r\right) \ldots\ldots. (6)$$

$$tol = tol(old) \pm STOCH\left(tol * \frac{L_r}{2}, tol * L_r\right) \ldots\ldots. (7)$$

  - Based on these values of classifier parameters, fitness levels are estimated for each Herd via equation 8, $f = \frac{A+P+R}{3} \ldots\ldots. (8)$

Where, A, P, R represent accuracy, precision & recall levels for each of the classifier entities, and are estimated via equations 9, 10, and 11,

$$A = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad \ldots\ldots\ldots (9)$$

$$P = \frac{t_p}{t_p + t_n} \quad \ldots\ldots\ldots (10)$$

$$R = \frac{t_p + f_p}{t_p + t_n + f_p + f_n} \quad \ldots\ldots\ldots (11)$$

Where $t_p t_n$,f_p& f_n denotes the counts of true positives, true negatives, false positives, and false negatives in practical real-time scenarios.

- Repeat this process for all Herds, which assists in the generation of N_h different solution sets.

- After generating the configurations (solution sets), evaluate the fitness threshold of the Herd solution using equation 12.

$$f_{th} = \frac{1}{N_h} \sum_{i=1}^{N_h} f_i * L_r \quad \ldots\ldots\ldots (12)$$

- Herds that showcase $f < f_{th}$ are reconfigured via equations 2, 3, 4, 5, 6, & 7, while other Herds are not modified during consecutive iterations.

- At the end of each iteration, the Herd with the maximum fitness level is marked as the 'Matriarch' Herd and is used to modify the learning rate via equation 13,

$$L_r(New) = L_r(Old) \pm \frac{f(Matriarch)}{\sum_{i=1}^{N_h} f_i} \quad \ldots\ldots\ldots\ldots (13)$$

Where $f(Matriarch)$ represents the highest fitness levels, and the rate is incremented if the current solution is better than the previous. In contrast, the rate is reduced if the current solution has lower performance than the previous one. The existing accuracy rates have led to this enhancement in the proposed model's capability to boost classification outcomes across various applications. This performance will be assessed and contrasted with conventional models in the subsequent section of this document.

3. Outcome & Assessment

A mix of Naïve Bayes (NB), k Nearest Neighbours (kNN), Support Vector Machine (SVM), and Deep Forest (DF) classifiers are used to predict the appropriate class for samples with missing values. The values of this class are averaged to calculate the current missing value sets. The performance of this classifier is improved via an EHO-based optimization process, which assists in identifying optimal hyperparameters that can achieve higher accuracy under different data samples. This accuracy was estimated for the Missing Value Dataset from Kaggle (https://www.kaggle.com/code/alexisbcook/missing-values/data), Brittleness Index Dataset (https://openmv.net/info/brittleness-index), Class Grades Dataset(http://openmv.net/info/class-grades) and Raw Material Properties Dataset (https://openmv.net/info/raw-material-properties) Samples. A total of 500,000 data samples were removed, with 70% utilized for training, and 15% allocated for validation and testing. The accuracy of missing value identification was assessed with respect to the Test Set Samples (TSS) and was compared with the methodologies of TSM [4], LMEGNB [9], and GANBILSTM [25], as presented in Table 2.

TSS	A (%) TSM [4]	A (%) LME GNB [9]	A (%) GAN Bi LSTM [25]	A (%) MVI BPM
833	79.94	79.45	81.15	85.32
1250	81.24	80.64	82.39	86.61
1667	82.29	81.60	83.38	87.66
2500	83.14	82.40	84.20	88.52
2917	83.88	83.13	84.95	89.31
3333	84.63	83.91	85.74	90.14
3750	85.48	84.78	86.62	91.07
4167	86.40	85.69	87.56	92.05
4583	87.33	86.60	88.49	93.02
5000	88.23	87.48	89.38	93.97
5417	89.13	88.36	90.27	94.92
5833	90.09	89.28	91.13	95.86
6250	91.10	90.25	91.95	96.74
6667	92.11	91.21	92.79	97.64
7083	93.07	92.12	93.71	98.60
7500	93.92	92.96	94.64	99.57

Table 2. Accuracy evaluation of different missing value models

The model showed a rise in accuracy of 5.5% compared to TSM [4], 6.4% more than LME GNB [9], and 4.9% above GAN Bi LSTM [25] across different applications. This accuracy enhancement is due to using accuracy during the tuning of the classifier hyperparameter sets.

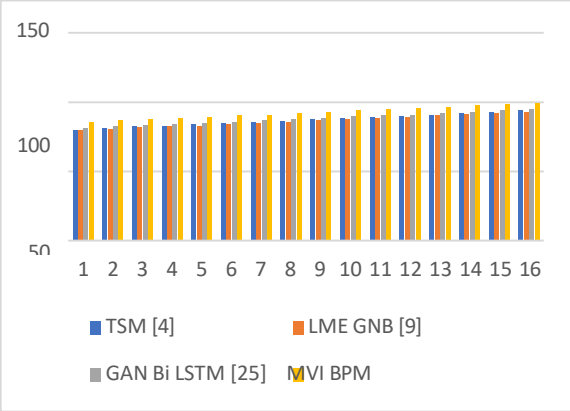

Figure 2. Accuracy evaluation of different missing value models

Similar performance was evaluated for precision levels, and can be observed from table 3 as follows,

TSS	**P (%) TSM [4]**	**P (%) LME GNB [9]**	**P (%) GAN Bi LSTM [25]**	**P (%) MVI BPM**
833	75.90	76.48	78.60	81.23
1250	77.09	77.63	79.82	82.45
1667	78.05	78.56	80.81	83.44
2500	78.83	79.34	81.63	84.27
2917	79.53	80.04	82.36	85.02
3333	80.26	80.79	83.11	85.81
3750	81.07	81.62	83.95	86.70
4167	81.95	82.50	84.86	87.63
4583	82.82	83.37	85.76	88.55

5000	83.67	84.22	86.64	89.46
5417	84.52	85.07	87.52	90.38
5833	85.41	85.96	88.44	91.35
6250	86.36	86.90	89.42	92.36
6667	87.30	87.83	90.40	93.36
7083	88.19	88.71	91.31	94.29
7500	88.99	89.52	92.15	95.15

Table 3. Precision evaluation of different missing value models

According to the estimation and Figure 3, the suggested model demonstrated a precision improvement of 6.5% over TSM [4], 5.5% over LME GNB [9], and 2.9% over GAN Bi LSTM [25] across various use cases. The improvement is a result of utilizing this parameter throughout the EHO tuning procedure, which helps in pinpointing effective parameters for each of the classifiers.

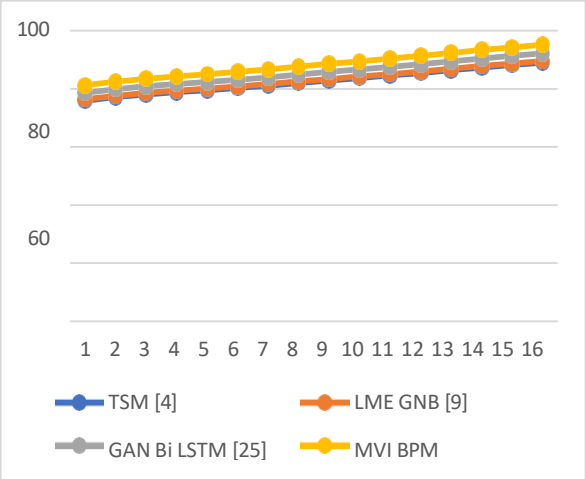

Figure 3. Precision evaluation of different missing value models

The assessment of recall levels for the performance was conducted, and it can be seen in Table 4 as follows,

TSS	R (%) TSM [4]	R (%) LME GNB [9]	R (%) GAN Bi LSTM [25]	R (%) MVI BPM
833	77.92	77.96	79.87	83.27
1250	79.16	79.13	81.10	84.53
1667	80.17	80.08	82.09	85.55
2500	80.99	80.87	82.91	86.39
2917	81.71	81.59	83.66	87.17
3333	82.45	82.35	84.42	87.98
3750	83.28	83.20	85.29	88.88
4167	84.18	84.10	86.21	89.84
4583	85.08	84.99	87.12	90.79
5000	85.95	85.85	88.01	91.72
5417	86.82	86.71	88.90	92.66
5833	87.75	87.62	89.84	93.66
6250	88.73	88.57	90.83	94.70
6667	89.70	89.52	91.81	95.72
7083	90.63	90.42	92.74	96.68
7500	91.45	91.24	93.59	97.56

Table 4. Recall the evaluation of different missing value models

In this assessment and Figure 4, the suggested model demonstrated a recall that was 5.9% greater than TSM [4], 6.2% better than LME GNB [9], and 4.5% higher than GAN Bi LSTM [25] across various use cases. The motivation behind this recall enhancement is the implementation of an ensemble classifier and the incorporation of recall in the EHO tuning process, which aids in pinpointing effective parameters for each classifier.

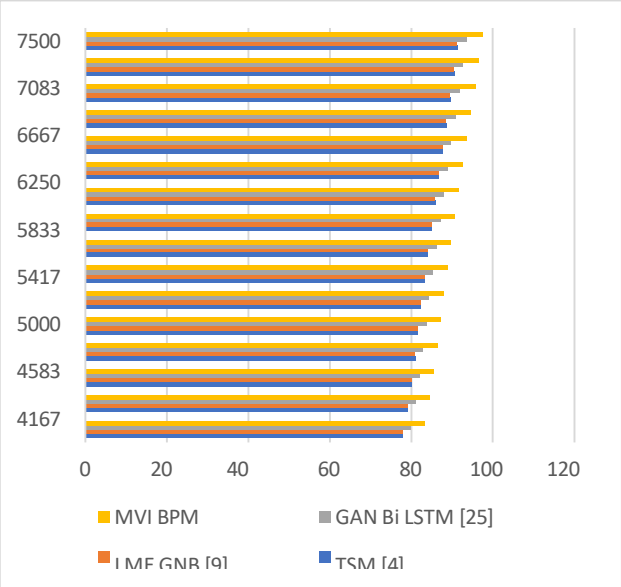

Figure 4. Recall the evaluation of different missing value models

A comparable assessment was conducted regarding the levels of computational delay, which can be seen in Table 5, as detailed below:

TSS	**D (ms) TSM [4]**	**D (ms) LME GNB [9]**	**D (ms) GAN Bi LSTM [25]**	**D (ms) MVI BPM**
833	1.60	1.60	1.56	1.33
1250	1.97	1.97	1.92	1.63
1667	2.33	2.34	2.28	1.92
2500	2.70	2.70	2.64	2.21
2917	3.08	3.08	3.00	2.51

3333	3.49	3.49	3.41	2.85
3750	3.97	3.97	3.88	3.25
4167	4.53	4.53	4.42	3.70
4583	5.15	5.16	5.03	4.21
5000	5.81	5.82	5.68	4.72
5417	6.45	6.46	6.30	5.21
5833	7.01	7.02	6.85	5.63
6250	7.48	7.49	7.30	5.99
6667	7.89	7.90	7.70	6.32
7083	8.31	8.31	8.11	6.66
7500	8.79	8.79	8.58	7.05

Table 5. Delay evaluation for different missing value models

Based on this assessment and Figure 5, the proposed model exhibited a 10.5% enhancement in speed compared to TSM [4] and LME GNB [9] and a 9.5% quicker performance than GAN Bi LSTM [25] across different applications. The reason for this enhancement in delay is the selection of optimal tuning parameters and the application of an ensemble classifier, which helps in determining effective parameters for every classifier.

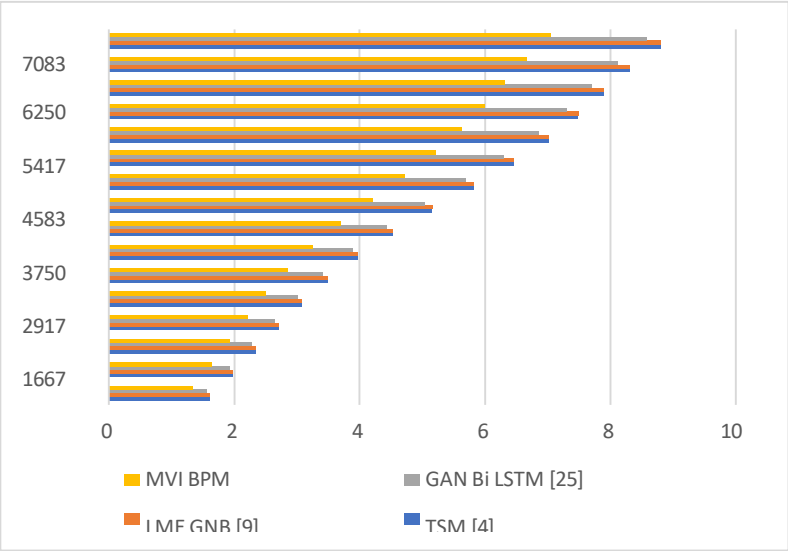

Figure 5. Delay evaluation for different missing value models

As a result of these enhancements, the suggested model achieves low error rates, fast processing speeds, high accuracy, and improved recall performance, making it valuable for a diverse range of missing value identification tasks.

4. Summary & Potential Directions Ahead

The suggested model employs a blend of Naive Bayes (NB), k Nearest Neighbors (kNN), Support Vector Machine (SVM), and Deep Forest (DF) classifiers to determine the appropriate class for samples with missing values. The values of this class are averaged to calculate the current missing value sets. The effectiveness of this classifier is enhanced through an optimization process based on EHO, which helps in determining the best hyperparameters capable of attaining greater accuracy across various data samples. The model's effectiveness was assessed using multiple datasets, revealing that the suggested model demonstrated an accuracy increase of 5.5% compared to TSM [4], 6.4% compared to LME GNB [9], and 4.9% compared to GAN Bi LSTM [25] across various use cases. The improvement in accuracy can be attributed to the use of accuracy metrics when adjusting the hyperparameters of the classifier. Based on precision estimation, the suggested model showed a 6.5% enhancement in precision when compared to TSM [4], a 5.5% gain in precision over LME GNB [9], and a 2.9% increase in precision in relation to GAN Bi LSTM [25] across multiple use cases. The aim of this enhancement in accuracy is to apply this parameter consistently during the EHO tuning procedure, which aids in the identification of optimal parameters for each classifier. The recall evaluation indicated that the suggested model demonstrated a recall that is 5.9% greater than that of TSM [4], 6.2% greater than LME GNB [9], and 4.5% higher than GAN Bi LSTM [25] across various use cases. The purpose of this recall enhancement is the utilization of an ensemble classifier and the incorporation of recall in the EHO tuning process, which helps in identifying effective parameters for each classifier. According to the estimation of computational delays, the proposed model demonstrated a speed increase of 10.5% compared to TSM [4] and LME GNB [9], and a 9.5% faster performance than GAN Bi LSTM [25] across various use cases. The aim of this enhancement in delay is to select the optimal tuning parameters and implement an ensemble classifier that helps in determining effective parameters for each of the classifiers. As a result of these enhancements, the suggested model demonstrates low error rates, quick processing, high

accuracy, and improved recall performance, making it applicable for various applications in identifying missing values.

5. References

[1] Foggo B. and Yu N. published an article titled "Online PMU Missing Value Replacement Through Event-Participation Decomposition" in the IEEE Transactions on Power Systems, volume 37, issue 1, covering pages 488-496 in January 2022, with the doi: 10.1109/TPWRS.2021.3093521.

[2] Chan P. C., Selamat A., Krejcar O., Kuok K.K., Bujang S.D.A., and Fujita H. conducted a systematic review titled "Designs and Methods for Missing Value Imputation Using Nature-Inspired Metaheuristic Techniques," published in IEEE Access, volume 10, on pages 61544-61566, in the year 2022, with the DOI: 10.1109/ACCESS.2022.3172319.

[3] Jena M. and Dehuri S., "A New Comprehensive Framework for Handling Missing Values Imputation and Classification," in IEEE Access, vol. 10, pp. 69373-69387, 2022, doi: 10.1109/ACCESS.2022.3187412.

[4] Zhang D. Li, H. , Li T. , Bouras A., Yu X., and Wang T., "Hybrid Missing Value Imputation Algorithms Using Fuzzy C-Means and Vaguely Quantified Rough Set," in the IEEE Transactions on Fuzzy Systems, volume 30, issue 5, on pages 1396-1408 in May 2022, with the DOI: 10.1109/TFUZZ.2021.3058643.

[5] Razavi-Far R. , Wan D., Saif M. , and Mozafari N. "To Tolerate or To Impute Missing Values in V2X Communications Data?" in IEEE Internet of Things Journal, volume 9, issue 13, pages 11442-11452, on July 1, 2022, with the doi: 10.1109/JIOT.2021.3126749.

[6] Wang T. , Ke H., Jolfaei A. , Wen S. , Haghighi M. S. , and Huang S. , "Missing Value Filling Based on the Collaboration of Cloud and Edge in Artificial Intelligence of Things," in IEEE Transactions on Industrial Informatics, volume 18, issue 8, pages 5394-5402, in August 2022. The DOI for the article is 10.1109/TII.2021.3126110.

[7] Fernstad S. J. and Westberg J. J., “Investigating the Absence--Glyph-Based Visualization for Analyzing Missing Data,” in IEEE Transactions on Visualization and Computer Graphics, vol. 28(10), pp. 3513-3529, 1 Oct. 2022, doi: 10.1109/TVCG.2021.3065124.

[8] Jia L., Wang Z., Lv S., and Xu Z., "PE_DIM: An Effective Probabilistic Ensemble Classification Method for Managing Class Imbalance in Diabetes with Missing Values," in IEEE Access, vol. 10, pp. 107459-107476, 2022, doi: 10.1109/ACCESS.2022.3212067.

[9] Liu X., Li N., Shu G., and Min L., "Creation of High-Quality Spaceborne Interrupted FMCW SAR Images Using Singular Value Threshold-Based Matrix Completion," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Article no. 4505905, doi: 10.1109/LGRS.2022.3157466.

[10] Niemela M., Ayrämö S., and Karkkainen T. authored "Toolkit for Estimating Distance and Validating Clusters on Datasets with Incomplete Data," which was published in IEEE Access, vol. 10, pp. 352-367, in 2022, doi: 10.1109/ACCESS.2021.3136435.

[11] Guo L., Renze L., Xingyu L., Juanjuan T., Lei C., and Yang Z., "Logging Data Completion Based on an MC-GAN-BiLSTM Model," in IEEE Access, volume 10, pages 1810-1822, in the year 2022, with the DOI: 10.1109/ACCESS.2021.3138194.

[12] Zhu X. , Yang J. , Zhang C., and Zhang S., "Effective Use of Incomplete Data in Cost-Sensitive Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 6, pp. 2425-2436, on June 1, 2021, doi: 10.1109/TKDE.2019.2956530.

[13] Wang A., Yang J., and An N., "Regularized Sparse Modeling for Estimating Missing Values in Microarray Data," in IEEE Access, vol. 9, pp. 16899-16913, 2021, doi: 10.1109/ACCESS.2021.3053631.

[14] D. Xu, J. Q. Sheng, P. J. H. Hu, T. -S. Huang, and C. -C. Hsu present "An Unsupervised Method Based on Deep Learning for Filling in Missing Values in Patient Records to Enhance Cardiovascular Patient Management," published in the IEEE Journal of Biomedical and Health Informatics, volume 25, issue 6, pages 2260-2272, in June 2021, doi: 10.1109/JBHI.2020.3033323.

[15] Chen L., Li G., Huang G., and Shi P. presented "An Adaptive Interpolation Framework for Sensor Data That Recognizes Missing Types," published in IEEE Transactions on Instrumentation and Measurement, volume 70, pages 1-15, in 2021, Article number 2510515, doi: 10.1109/TIM.2021.3089783.

[16] Liu A. , Lu J. , and Zhang G. , "Detecting Concept Drift: Addressing Missing Values Through Fuzzy Distance Estimations," in IEEE Transactions on Fuzzy Systems, vol. 29, no. 11, pp. 3219-3233, in November 2021, doi: 10.1109/TFUZZ.2020.3016040.

[17] Mostafa S. M. , Eladimy A. S., Hamad S. , and Amano H. present "CBRG: An Innovative Algorithm for Managing Incomplete Data Utilizing Bayesian Ridge Regression and Feature Selection Driven by Gain Ratio," published in IEEE Access, vol. 8, pp. 216969-216985, 2020, doi: 10.1109/ACCESS.2020.3042119.

[18] Liu Y. , Dillon T., Yu W. , Rahayu W., and Mostafa F. discuss "Missing Value Imputation for Industrial IoT Sensor Data With Large Gaps" in IEEE Internet of Things Journal, volume 7, issue 8, pages 6855-6867, published in August 2020, doi: 10.1109/JIOT.2020.2970467.

[19] Yu Y., Yu J. J. Q., Li V. O. K., and Lam J. C. K., "A Novel Interpolation-SVT Approach for Recovering Missing Low-Rank Air Quality Data," in IEEE Access, vol. 8, pp. 74291- 74305, 2020, doi: 10.1109/ACCESS.2020.2988684.

[20] Li Q., Tan H., Wu Y., Ye L. and Ding F., "Traffic Flow Prediction with Missing Data Imputed by Tensor Completion Methods," in IEEE Access, vol. 8, pp. 63188-63201, 2020, doi: 10.1109/ACCESS.2020.2984588.

[21] Fan J., Zhang P., Chen J., Li B., Han L., and Zhou Y., "Quantitative Assessment of Missing Value Interpolation Techniques for Suomi-NPP VIIRS/DNB Nighttime Light Monthly Composite Images," published in IEEE Access, vol. 8, pp. 199266-199288, 2020, doi: 10.1109/ACCESS.2020.3035408.

[22] Garcia C., Leite D., and Škrjanc I., "Incremental Missing-Data Imputation for Evolving Fuzzy Granular Prediction" in the IEEE Transactions on Fuzzy Systems, volume 28, issue 10, pages 2348 to 2362, in October 2020, with the DOI: 10.1109/TFUZZ.2019.2935688.

[23] Ma Q. et al., “Modeling Incomplete Time-Series from Linear Memory of Latent Variables End-to-End,” in IEEE Transactions on Cybernetics, vol. 50, no. 12, pp. 4908-4920, December 2020, doi: 10.1109/TCYB.2019.2906426.

[24] Lai X., Zhang L., and Liu X., "Takagi-Sugeno Modeling of Incomplete Data for Missing Value Imputation Using Alternate Learning," in IEEE Access, volume 8, pages 83633- 83644, published in 2020, doi: 10.1109/ACCESS.2020.2991669.

[25] Huamin T., Qiuqun D., and Shanzhu X. authored the paper titled "Reconstructing time series with absent values using a 2D representation-based denoising autoencoder," published in the Journal of Systems Engineering and Electronics, volume 31, issue 6, pages 1087-1096, in December 2020, with the DOI: 10.23919/JSEE.2020.000081.