Journal of Intelligent Systems and Internet of Things
Vol. 13, No. 01, PP. 46-58, 2024
DOI: https://doi.org/10.54216/JISIoT.130104

Enhancing Air Pollution Monitoring and Prediction using African Vulture Optimization Algorithm with Machine Learning Model on Internet of Things Environment

Naresh Sharma[1], Rohit Sharma[*2]

[1] Department of Computer Science and Engineering, SRM Institute of Science and Technology, NCR Campus, Ghaziabad, India
[2] Department of Electronics and Communication Engineering, ABES Engineering College, Ghaziabad, UP, India
Emails: nrssharma@gmail.com; rohitapece@gmail.com

Abstract

An optimal solution for monitoring air pollution, the Internet of Things (IoT)-enabled system delivers real-time data and insights on the air quality within a specific location. Air pollution poses a substantial risk to human health worldwide, with pollutants like nitrogen dioxide, particulate matter, ozone, and sulfur dioxide contributing to a range of cardiovascular and respiratory ailments. Monitoring air pollution levels is critical to understand the effect on public health and the environment. Air Pollution Monitoring includes the systematic analysis and measurement of pollutant concentration in the air, through a network of monitoring stations equipped with instruments and sensors. This station provides real-time data on air quality, allowing authorities to evaluate issue warnings, and pollution levels, and implement strategies to alleviate its negative impact. Machine learning (ML) approaches are becoming more integrated into air pollution monitoring systems for enhancing efficiency and accuracy. By analyzing vast quantities of information gathered from satellite imagery, monitoring stations, and other sources, ML approaches could detect patterns, forecast pollution levels, and pinpoint sources of pollution. This study introduces Air Pollution Monitoring and Prediction using African Vulture Optimization Algorithm with Machine Learning (APMP-AVOAML) model in IoT environment. The drive of the APMP-AVOAML methodology is to recognize and classify the air quality levels in the IoT environment. In the APMP-AVOAML technique, a four stage process is encompassed. Firstly, min-max normalization is applied for scaling the input data. Secondly, a harmony search algorithm (HSA) based feature selection process is executed. Thirdly, the extreme gradient boosting (XGBoost) model is utilized for air pollution prediction. Finally, AVOA based parameter selection process is exploited for the XGBoost model. To illustrate the performance of the APMP-AVOAML algorithm, a brief experimental study is made. The resultant outcomes inferred that the APMP-AVOAML methodology has resulted in effectual outcome.

Received: December 14, 2023 Revised: February 08, 2024, Accepted: April 18, 2024

Keywords: Air Pollution Monitoring; Air Quality Index; African Vulture Optimization Algorithm; Machine Learning; Internet of Things

1. Introduction

Environmental pollution has transformed the conditions of weather and climate. Furthermore, the earth's system has water, air, and land, is changed by pollution [1]. When compared to others, severe pollution is air pollution that disturbs each creature on the earth. Also, the air functions as a cover of elements that defend the globe [2]. Due to this cause, air pollution is perceived to be one of the foremost focuses and extreme tasks the globe has met so far. Air pollution is hazardous for human health and must decline fast in rural and urban areas so it is essential to forecast the excellence of air precisely [3]. Generally, numerous kinds of pollution are accessible such as water

pollution, soil pollution air pollution, etc. but the most significant of these is air pollution which must be measured instantly as humans inhale oxygen over the air. There are numerous reasons for air pollution [4]. Outdoor air pollution is affected by factories, industries, and vehicles and Indoor air pollution is affected if the air inside the house is polluted by chemicals, smoke, and smell.

The Air Quality Index (AQI) is an evaluation parameter, which linked to public health openly [5]. A high level of AQI specifies unsafe contact for the human populace. So, the need to forecast the AQI in progress was inspired by the experts to observe and perfect the quality of airy [6]. Frequently, air quality-based research works aimed the emerging countries, while the attention to the most deadly chemicals such as PM2.5 is initiated in manifold folds in emerging countries [7]. Some researchers tried to start the research of air quality forecast. After going over the obtainable work, a sturdy necessity was to be handled to fill this crack by diagnosis and forecast of AQI for Indian cities. Numerous traditional techniques are there to extend it but outcomes are not precise and it includes a lot of mathematical calculations [8]. Many techniques were used in the works to forecast AQI such as deterministic, physical, statistical, and Machine Learning (ML) methods [9]. The standard approaches depend upon statistics and probability, which are highly compound and less effective. The ML-based AQI forecast techniques were more consistent and reliable. Expanded sensors and technologies formed data collection simple and exact. The precise and reliable forecasts over large environmental data need hard analysis, so only the ML technique can able to deal with it well [10].

This study introduces Air Pollution Monitoring and Prediction using African Vulture Optimization Algorithm with Machine Learning (APMP-AVOAML) model in IoT environment. The drive of the APMP-AVOAML methodology is to recognize and classify the air quality levels in the IoT environment. In the APMP-AVOAML technique, a four stage process is encompassed. Firstly, min-max normalization is applied for scaling the input data. Secondly, harmony search algorithm (HSA) based feature selection process is executed. Thirdly, extreme gradient boosting (XGBoost) model is utilized for air pollution prediction. Finally, the AVOA based parameter selection process is exploited for the XGBoost model. The resultant outcomes implied that the APMP-AVOAML algorithm has resulted in effectual outcome.

2. Related Works

Ansari and Alam [11] projected a new BO-HyTS technique that unites seasonal auto-regressive integrated moving average (SARIMA) and LSTM and modified it by utilizing Bayesian optimizer in order to forecast levels of air pollution. The projected BO-HyTS system can take either linear or non-linear features of the time-series data, hence enlarging the accuracy of the predicting procedure. Moreover, numerous AQI predicting methods such as classical time-series, ML, and DL, are used in order to estimate air quality from time-series data. Hardini et al. [12] projected an air quality valuation model to make simpler future forecasts. The Data Preparation Module contains real-time data collection and organizing to certify compatibility with subsequent components. In this paper, the Sparse Spectrum GPR (SSGPR) model is utilized for AQI predicting, while the cloud method is implemented for evaluation of air quality. In [13], a new Hybrid Interpretable Predictive ML technique is projected for the PM 2.5 forecast, which has dual innovations. At first, a hybrid method structure is built with DNN and Non-linear Auto Regressive Moving Average with Exogenous Input technique. Next, automated feature generation and FS processes are united into this hybrid method.

Abbas and Raina [14] projected a novel Ensemble Empirical Mode Decomposition (EEMD) multi-stage predicting technique. Firstly, input data were decayed by enhanced empirical wavelet transform (EEWT) to upsurge the dimensional of data. The 3 single forecast methods are combined into united weighted predicting system by weight assignment. Liao et al. [15] examined a novel AQP system that depends on Dynamic Multi-granularity Spatio-temporal GNN (DM_STGNN) approach. DM_STGNN was dependent upon an elegant encoding-decoding structure. Likewise, the method constructed a multi-granularity graph framework renowned as the HYSPLIT method and utilized LSTM. This work also constructed an attention mechanism based LSTM for decoder and AQP. Moreover, an unsupervised pre-trained technique was employed in order to improve DM_STGNN.

Ahmed et al. [16] introduced a new hybrid DL system. This method stated as CLSTM-BiGRU, integrates a CNN, LSTM, and a Bi-GRU system. The projected CLSTM-BiGRU system also executed 19 remotely detected predictor variables and attaching the GWO method. Mehrabi et al. [17] use Sentinel 5P image and a new AI technique for forecasting air pollution. An ML system multi-layer perceptron neural network (MLPNN), is united with electromagnetic field optimizer (EFO) technique in order to forecast the daily attention of PM 2.5. Firstly, a dataset was ready. Then, principal component analysis (PCA) was employed to determine the contributing factors and generate a condensed dataset.

3. Proposed Methodology

In this study, we have introduced an APMP-AVOAML model. The drive of the APMP-AVOAML approach is to recognize and classify the air quality levels. In the APMP-AVOAML methodology, a four-stage process is encompassed. Fig. 1 depicts the workflow of the APMP-AVOAML model.

A. Min-Max Normalization

Firstly, min-max normalization was applied for scaling the input data. Min-max normalization is a data pre-processing approach utilized to normalize mathematical characteristics to a certain range within (0,1) [18]. It functions by subtracting the lowest value of the feature and dividing it by the variance among the highest and lowest rates. This assured that all the features are proportionally mapped to the appropriate range, which makes them comparable and prevents specific features from dominating others because of variances in their scales. Min-max normalization is paramount in ML and data analysis to enhance the convergence and performance of models, especially while handling features with differing scales.

B. HSA-based Feature Selection

Secondly, the HSA-based feature selection process is executed. At this phase, an optimizer algorithm dependent upon the HS development (AHHS) has been utilized for FS [19]. By vigorously altering the hyperparameters PAR and HMCR in the iteration process, the arbitrarily produced solution vector is permitted to consume an opportunity to differ even if it is similar to the present optimum result, thus enhancing the variety of solutions and skipping out of the local optimal as possible.

Figure 1: Workflow of APMP-AVOAML model

Exactly, we develop a cosine resemblance feature to define the track of hyperparameter alteration by equating the resemblance among the presently produced harmonic vector and the optimum vector, and once the resemblance is greater, the search probability has properly decreased. The definite stages of the process are:

1. Initialize parameters: Dimensionality harmony vector (N), Harmony Memory Space Size (HMS), i.e., amount of original features, Pitch Adjustment Rate (PAR), HM Consideration Rate (HMCR), the highest amount of iterations (T), and amount of base learners (M).

2. Set the harmonic memory space, and produce HMS harmony vector randomly. The formula is as below:

$$HM = \begin{bmatrix} X^1 \\ X^2 \\ \vdots \\ X^{HMS} \end{bmatrix} = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_N^1 & |d(X^1) \\ x_1^2 & x_2^2 & \dots & x_N^2 & |d(X^2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{HMS} & x_2^{HMS} & \dots & x_N^{HMS} & |d(X^{HMS}) \end{bmatrix} \quad (1)$$

Where HMS denotes the dimension of harmony space, N refers to the dimension of the harmony vector, and $d(X)$ signifies the fitness function rate.

The merged determining performance indicator d is determined as below:

$$d(X) = \lambda \times \frac{\sum_{i=1}^{M}(1 - Acc_i)}{M} + (1 - \lambda) \times \frac{\sum_{i=1}^{N} X_i}{N} \quad (2)$$

Whereas Acc_i denotes the accuracy of ith base learner, M refers the amount of base learners, X_i is the acoustic vector module range from (0 or 1), N signifies the harmony vector length, and λ refers the multiple-objective coefficient.

3. Produce a novel harmony vector. If the random number (r1) is lesser than or equivalent to HMCR, then hunt for the present module of optimum performance vector and alter HMCR affording to Eq. (3). If r2 is lesser than or equivalent to PAR; then reverse the present module variation and regulate PAR utilizing Eq. (4). When r1 is larger than HMCR, then produce the present module at random.

$$hmcr_j = HMCR - \theta \times cosine\_similarity_i \quad (3)$$

where $cosine\_similarity_i$ denotes the cosine resemblance of the i modules of the freshly produced harmony vector to the initial i modules of the optimum solution vector from the existing HM space.

$$par_i = PAR + \theta \times \left| |x_i^{best} - x_i^{worst}| - 1 \right| \quad (4)$$

where x_i^{best} signifies the ith module of the best solution vector and x_i^{worst} denotes the ith module of the worse solution vector.

4. Upgrade the harmony vector. If the fitness function value of produced harmony vector is lesser than the worse vector, it will be substituted.

5. Repeat 3-4 constantly till the greatest amount of iteration T has been attained and the optimum feature subset is result.

The fitness function (FF) employed in the HSA has been planned to take a balance among the elected feature counts from every result (lower) and classifier accuracy (superior) realized with utilizing these elected features.

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|C|} \quad (5)$$

whereas $\gamma_R(D)$ illustrates the classifier error values of offered classifier. $|R|$ represents the cardinality of elected subset and $|C|$ demonstrates the entire feature counts within data, α and β imply the 2 parameters equal to the effect of classifier quality and subset length. $\in$ [1,0] and $\beta = 1 - \alpha$.

C. Air Pollution Prediction using XGBoost

Thirdly, the XGBoost model is utilized for air pollution prediction. XGboost: Gradient-boosted-DT is used in XGboost [20]. The method continuously produces DT, and weights must be important in this architecture. The values of independent variable have been provided as weights and then transferred to the DT for outcome prediction. When the weight of variables is estimated inaccurately by a tree improvement, such variables have been provided to the following DT. Finally, numerous predictors or classifiers should be integrated to make an accurate and robust model. This can overcome the problems like modified predictions, regression, classification, and rankings. XGBoost has a supervised ML technique dependent upon ensemble trees. Fig. 2 depicts the infrastructure of XGBoost model. It targets enhancing a cost objective function comprising a regularization term (β) and loss function (d):

$$\Omega(\theta) = \sum_{m=1}^{n} b\,(x_i, \hat{x}_i) + \sum_{j=1}^{JSS} \beta\,(p_j), \quad (6)$$

Now, n describes the instances values at the training set, $\hat{x}_m$ refers to the prediction values, J means the number of trees that will be generated and f_k refers to a tree in the ensemble trees. The normalization formula will be represented as:

$$\beta(f_u) = \gamma L + \frac{1}{2}[\alpha \sum_{i=1}^{L} | c_i | + \lambda \sum_{i=1}^{L} c_i^2], \quad (7)$$

Here c denotes the weight comprised in every leaf, λ means a regularization value provided at weights, and. Assume $f_u(x_i) = c_{q(x_i)}$, and γ defines the minimum loss split decreased value L is the number of leaf values and q exists in $[1, L]$. A greedy technique will be implemented for choosing the separated values for increasing the gain values.

Algorithm 1: XGboost Pseudocode

Initialize the ensemble system E_i
 	
for $K = 1$ to N trees do
 	
Pseudo residuals must be computed for every training model: Set the pseudo-residuals for all the samples utilized for training as the negative gradients of loss function about the evaluations (Negative-Gradients)
 	
end for
 	
Set the pseudo-residuals in a regression tree: By utilizing (X_train, Negative_Gradients), maximum_features, maximum_depth.
 	
Upgrade the E_i
 	
Upgrade the training predictions: all the training Samples: Compute tree.predict() algorithm, upgrade the training forecastes by multiplying the rate of learning through the existing prediction tree.
 	
To make predictions under the testing dataset employing tree predict, test_predictions.

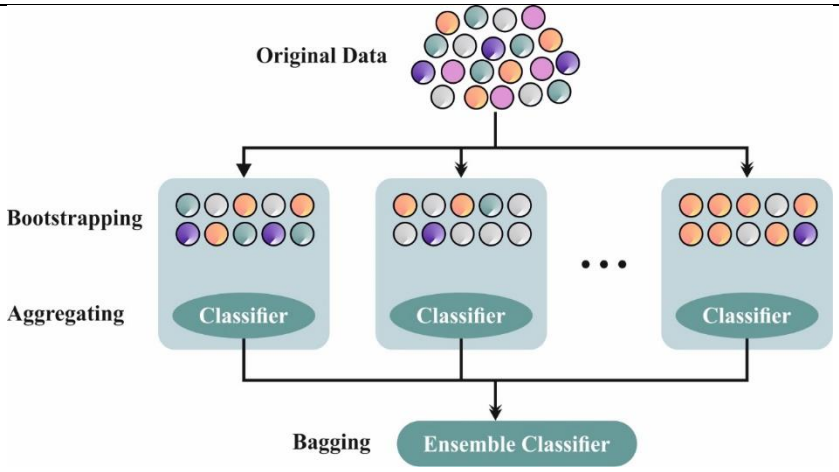

Figure 2: Structure of XGBoost model

D. Parameter Selection

Finally, the AVOA-based parameter selection process is exploited for the XGBoost model. The AVOA gets stimulated by the performances of predatory animals such as leopards, lions, and wolves, and their prey namely stags and gazelles [21]. The following are the steps of novel AVOA.

Determining the best vulture in any group

The chance of choosing the chosen vultures to guide the other vultures nearby most top performances in every group is defined utilizing the equation:

$$R(i) = \begin{cases} B_1 \; if \; p_i = L_1 \\ B_2 \; if \; p_i = L_2 \end{cases}, \quad (8)$$

$$p_i = \frac{F_i}{\sum_{i=1}^{n} F_i}, \tag{9}$$

whereas B_1 and B_2 signifies the 1st-best and 2nd-best solutions, correspondingly, L_1 and & are the parameters that evaluted before the searching function. p_i defines the probability of electing the optimum performance.

The rate of starvation of vultures

The mathematical process of the rate of being replete takes a declining trend, and to model.

$$t = h \times \left(\sin^w\left(\frac{\pi}{2} \times \frac{it}{\text{Max}it}\right) + \cos\left(\frac{\pi}{2} \times \frac{it}{\text{Max}it}\right) - 1\right), \tag{10}$$

$$F = (2 \times r_1 + 1) \times z \times \left(1 - \frac{it}{\text{Max } it}\right) + t, \tag{11}$$

whereas F implies the vultures are satiated. it and $Maxit$ demonstrate the existing iteration number and maximal iterations, correspondingly. Z illustrates the random number from the range of -1 to 1. h stands for the random number among -2 and 2. r_1 represents the random number among zero and one.

Searching prey (exploration)

During the exploration phase, the vultures search for food in distinct regions. This procedure is illustrated as:

$$P(i+1) = R(i) - D(i) \times F, \tag{12}$$

$$D(i) = |X \times R(i) - P(i)|, \tag{13}$$

$$P(i+1) = R(i) - F + r_2 \times \big((ub - lb) \times r_3 + lb\big), \tag{14}$$

Whereas, $P(i+1)$ and $P(i)$ imply the vulture position vector under the next and existing iteration, correspondingly. X is utilized as a co-efficient vector that enhances the random.

Exploitation phase

During the exploitation stage, there are 2 stages with 2 various approaches:

1) Exploitation (first phase)

$$P(i+1) = D(i) \times (F + r_4) - d(t), \tag{15}$$

$$d(t) = R(i) - P(i), \tag{16}$$

$$P(i+1) = R(i) - (S_1 + S_2), \tag{17}$$

$$S_1 = R(i) \times \left(\frac{r_5 \times P(i)}{2\pi}\right) \times \cos\big(P(i)\big), \tag{18}$$

$$S_2 = R(i) \times \left(\frac{r_6 \times P(i)}{2\pi}\right) \times \sin\big(p(i)\big). \tag{19}$$

2) Exploitation (second phase)

$$P(i+1) = \frac{A_1 + A_2}{2}, \tag{20}$$

$$A_1 = B_1(i) - \frac{B_1(i) \times P(i)}{B_1(i) - P(i)^2} \times F, \tag{21}$$

$$A_2 = B_2(i) - \frac{B_2(i) \times P(i)}{B_2(i) - P(i)^2} \times F, \tag{22}$$

$$P(i+1) = R(i) - |d(t)| \times F \times Levy(d). \tag{23}$$

The AVOA grows an FF to reach a superior classier solution. It explains a positive integer to imply a good solution of candidate results. In this case, the decreasing classifier error rate can be supposed that FF.

$$fitness(x_i) = ClassifierErrorRate(x_i)$$

$$= \frac{No.of\ misclassified\ samples}{Total\ no.of\ samples} * 100 \quad (24)$$

4. Performance Validation

The performance evaluation of the APMP-AVOAML methodology was examined deploying the AQI dataset containing 2000 instances with 4 classes as portrayed in Table 1.

Table 1: Details of the dataset

Classes	No. of Instances
Good	500
Satisfactory	500
Moderate	500
Poor	500
Total Instances	2000

Fig. 3 establishes the confusion matrices formed by the APMP-AVOAML system under 80:20 and 70:30 of TRAS/TESS. The results state that the APMP-AVOAML approach has effective detection under all classes.

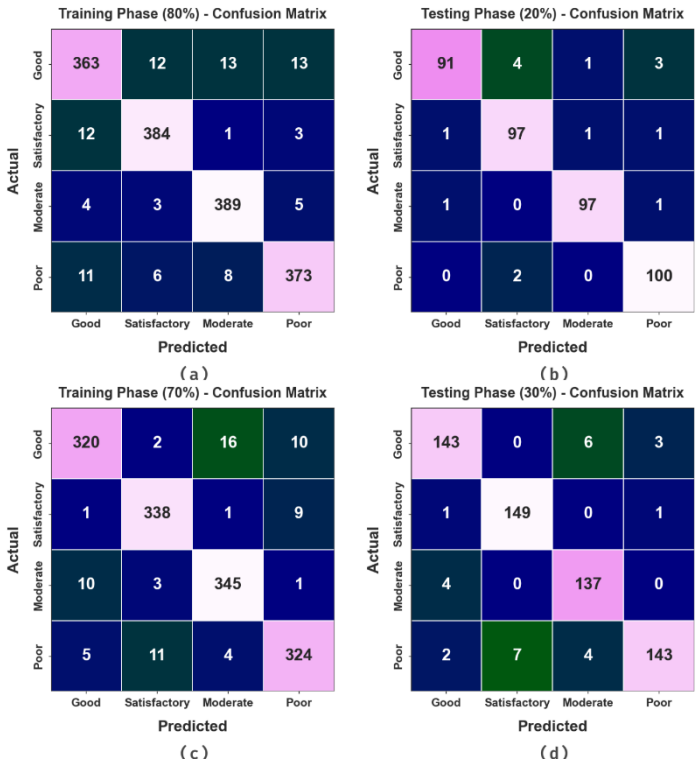

Figure 3: Confusion matrices of (a-b) 80% and 20% of TRAS/TESS and (c-d) 70% and 30% of TRAS/TESS

In Table 2 and Fig. 4, the detection output of the APMP-AVOAML system is clearly demonstrated. The results ensured the ability of the APMP-AVOAML technique to detect the classes. With 80%TRAS, the APMP-AVOAML technique offers average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} of 97.16%, 94.30%, 94.31%, 94.30%, and 96.21%, respectively. Additionally, With 20%TESS, the APMP-AVOAML system provides average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} of 98.12%, 96.31%, 96.23%, 96.24%, and 97.49%, respectively. Besides, With 70%TRAS, the APMP-AVOAML method delivers average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} of 97.39%, 94.79%, 94.77%, 94.77%, and 96.52%, correspondingly. Moreover, With 30%TESS, the APMP-AVOAML approach offers average $accu_y$, $prec_n$, $reca_l$, F_{score}, and AUC_{score} of 97.67%, 95.33%, 95.40%, 95.32%, and 96.92%, respectively.

Table 2: Detection outcome of APMP-AVOAML technique under 80% and 70% of TRAS and 20% and 30% of TESS

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$	AUC_{score}
TRAS (80%)					
Good	95.94	93.08	90.52	91.78	94.14
Satisfactory	97.69	94.81	96.00	95.40	97.12
Moderate	97.88	94.65	97.01	95.81	97.59
Poor	97.12	94.67	93.72	94.19	95.99
Average	97.16	94.30	94.31	94.30	96.21
TESS (20%)					
Good	97.50	97.85	91.92	94.79	95.63
Satisfactory	97.75	94.17	97.00	95.57	97.50
Moderate	99.00	97.98	97.98	97.98	98.66
Poor	98.25	95.24	98.04	96.62	98.18
Average	98.12	96.31	96.23	96.24	97.49
TRAS (70%)					
Good	96.86	95.24	91.95	93.57	95.22
Satisfactory	98.07	95.48	96.85	96.16	97.66
Moderate	97.50	94.26	96.10	95.17	97.04
Poor	97.14	94.19	94.19	94.19	96.15
Average	97.39	94.79	94.77	94.77	96.52
TESS (30%)					
Good	97.33	95.33	94.08	94.70	96.26
Satisfactory	98.50	95.51	98.68	97.07	98.56
Moderate	97.67	93.20	97.16	95.14	97.49
Poor	97.17	97.28	91.67	94.39	95.38
Average	97.67	95.33	95.40	95.32	96.92

The performance of the APMP-AVOAML technique is offered in Fig. 5 in the form of validation accuracy (VALAC) and training accuracy (TRAAC) curves under 80:20 of TRAS/TESS. The figure shows valuable clarification into the behavior of the APMP-AVOAML technique over abundant count of epochs, demonstrating its learning model and generalized proficiencies. Mainly, the outcome determines a stable development in the VALAC and TRAAC with a growth in epochs. It safeguards the adaptive nature of the APMP-AVOAML system from the pattern detection procedure on both the data. The rising trend in VALAC summarizes the ability of the APMP-AVOAML system to regulate to the TRA data and is also best in providing precise classification of hidden data, indicating robust generalization skills.

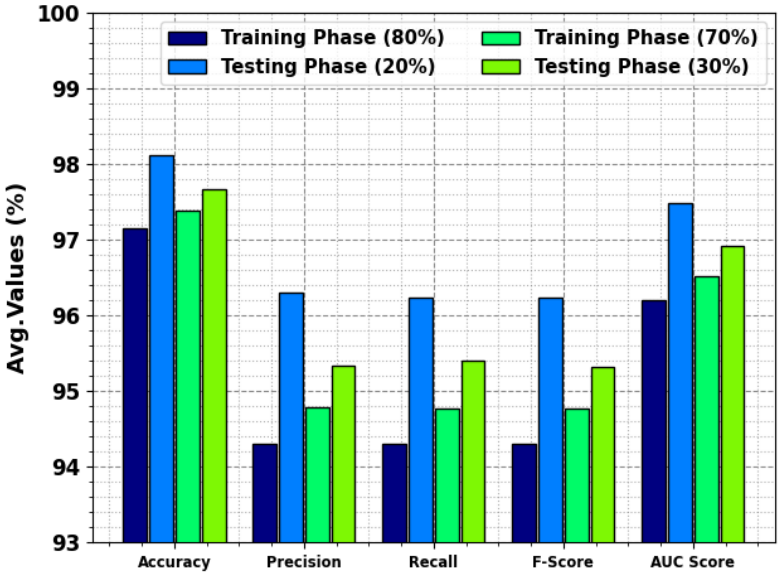

Figure 4: Average outcome of APMP-AVOAML technique

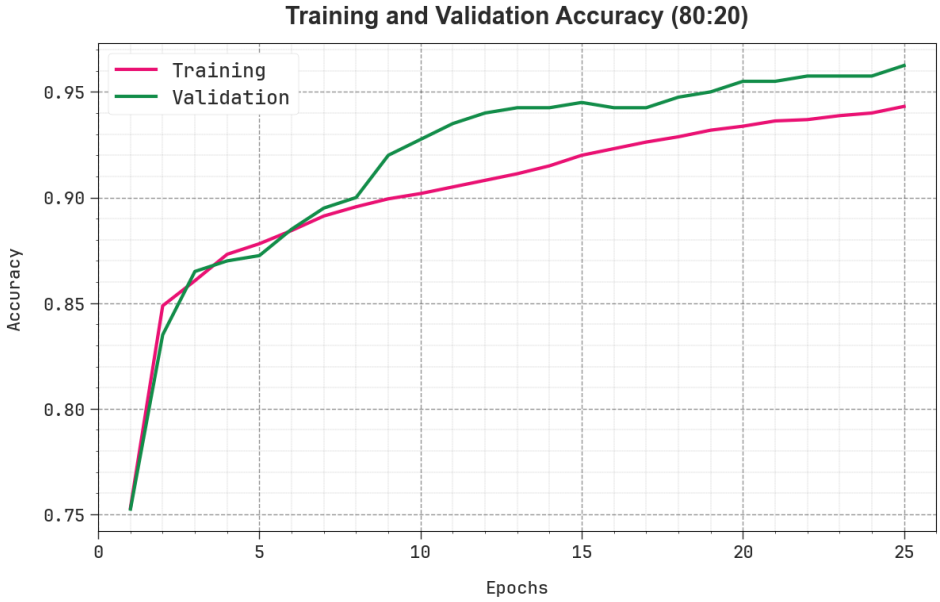

Figure 5: $Accu_y$ curve of APMP-AVOAML technique under 80:20 of TRAS/TESS

Fig. 6 exhibits an overall representation of the validation loss (VALLS) and training loss (TRALS) outcomes of the APMP-AVOAML system over separate epochs under 80:20 of TRAS/TESS. The advanced decrease in TRALS emphasises the APMP-AVOAML approach improving the weights and minimizing the classification error on the TRA and TES data. The outcome states a clear data of the APMP-AVOAML model's linked with the TRA data, emphasizing its ability to take patterns within both datasets. Remarkably, the APMP-AVOAML system recurrently improves its parameters in declining the differences among the prediction and actual TRA class labels.

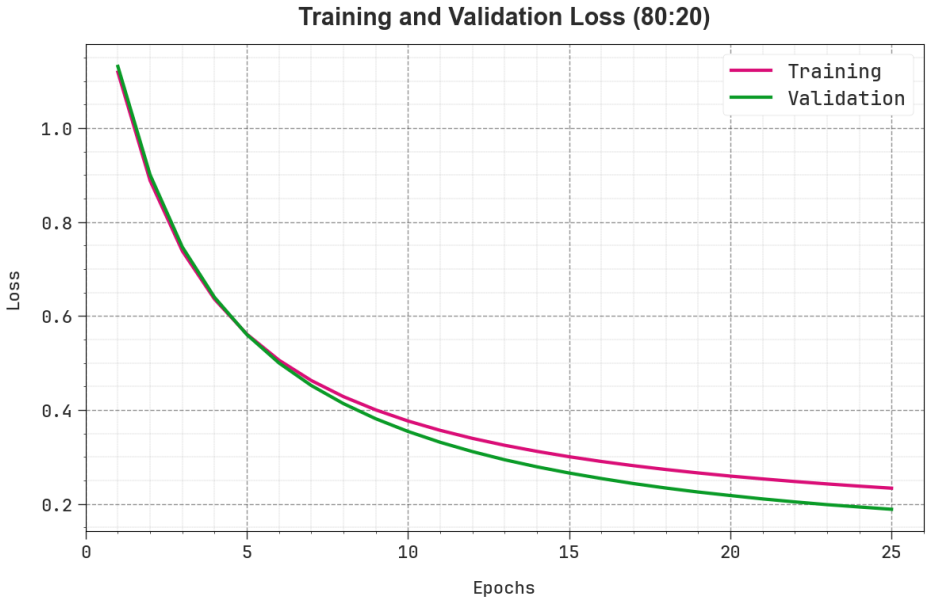

Figure 6: Loss curve of APMP-AVOAML technique under 80:20 of TRAS/TESS

Inspecting the PR outcome, as shown in Fig. 7, the results assured that the APMP-AVOAML system increasingly attains enhanced PR values over each under 80:20 of TRAS/TESS classes. It confirms the improved skills of the APMP-AVOAML technique in the classification of different classes, representing proficiency in the recognition of classes.

Moreover, in Fig. 8, ROC curves molded by the APMP-AVOAML technique outperformed the identification of different labels under 80:20 of TRAS/TESS. It delivers a complete understanding of the trade-off amongst TPR and FRP over dissimilar detect threshold rates and count of epochs. The outcome emphasized the improved classifier outcomes of the APMP-AVOAML system at 2 classes, demonstrating the performance in dealing with many classification problems.

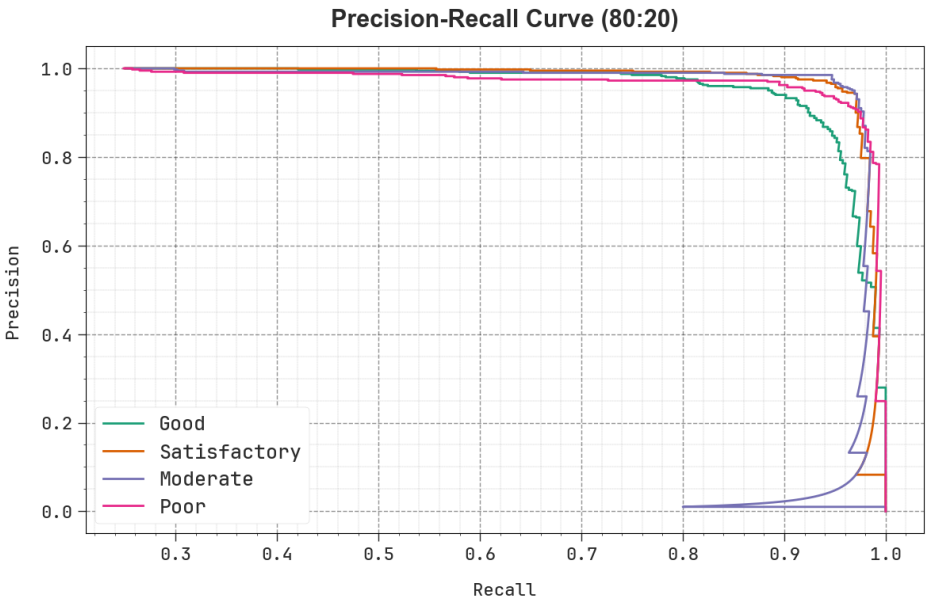

Figure 7: PR curve of the APMP-AVOAML technique under 80:20 of TRAS/TESS

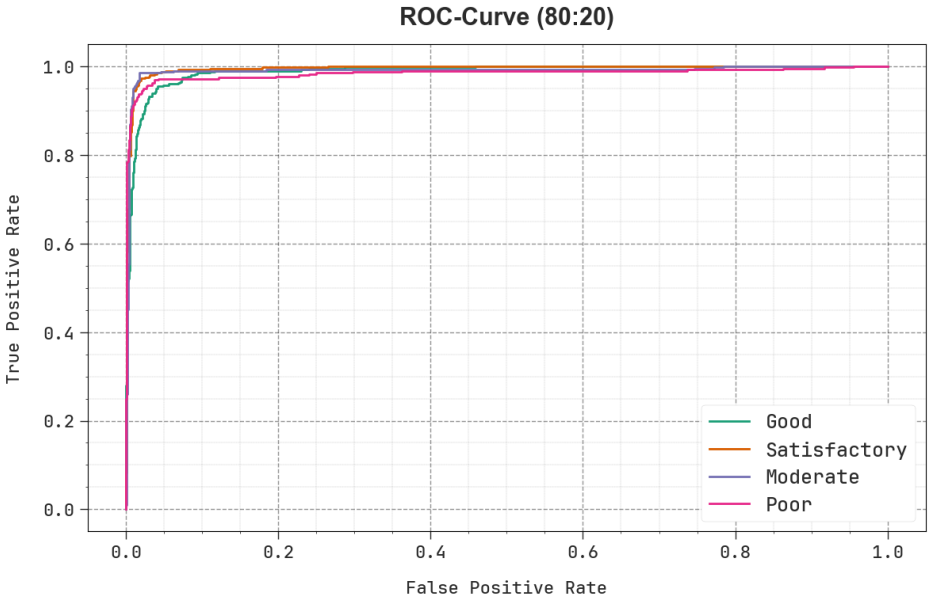

Figure 8: ROC curve of the APMP-AVOAML technique under 80%TRAS:20%TESS

Table 3: Comparative outcome of APMP-AVOAML technique with other approaches

Methods	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{Score}$
RNN Algorithm	95.55	91.33	94.54	90.97
SVR+wavelet Ensemble	96.66	94.00	90.44	93.97
Logistic Regression	93.71	92.15	91.64	94.43
Random forest	86.70	91.70	92.18	94.37
BOSDL-AQIP	97.45	94.78	90.50	93.67
APMP-AVOAML	98.12	96.31	96.23	96.24

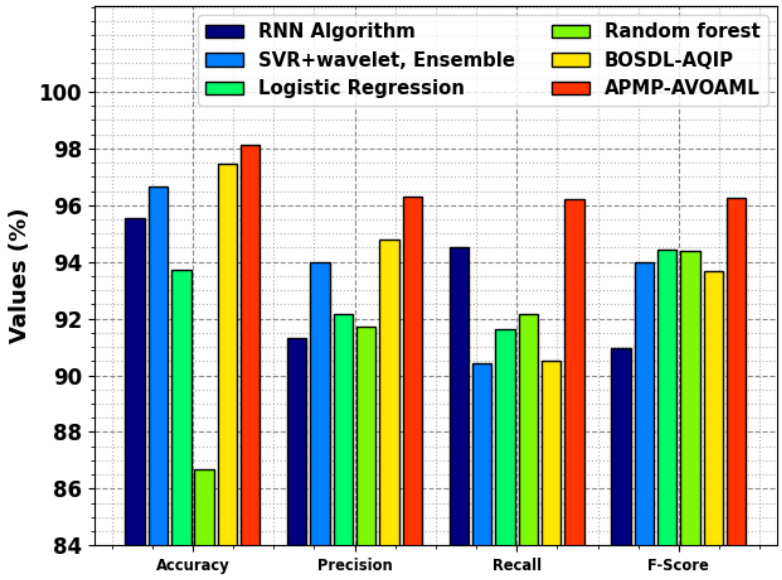

Figure 9: Comparative analysis of APMP-AVOAML technique with other models

In Table 3 and Fig. 9, the overall results of the APMP-AVOAML technique undergo comparison with other models [22]. Based on $accu_y$, the APMP-AVOAML technique reaches an increased $accu_y$ of 98.12% while the RNN,

SVR+WE, LR, RF, and BOSDL-AQIP techniques obtain decreased $accu_y$ of 95.55%, 96.66%, 93.71%, 86.70%, and 97.45%, correspondingly. At the same time, Based on $prec_n$, the APMP-AVOAML technique reaches increased $prec_n$ of 96.31% while the RNN, SVR+WE, LR, RF, and BOSDL-AQIP approaches get decreased $prec_n$ of 91.33%, 94.00%, 92.15%, 91.70%, and 94.78%, correspondingly.

Finally, Based on $reca_l$ the APMP-AVOAML technique reaches an increased $reca_l$of 96.23% while the RNN, SVR+WE, LR, RF, and BOSDL-AQIP techniques obtain decreased $reca_l$ of 94.54%, 90.44%, 91.64%, 92.18%, and 90.50%, respectively. Therefore, the APMP-AVOAML technique can be used for an automated detection method.

5. Conclusion

In this study, we have introduced an APMP-AVOAML model. The drive of the APMP-AVOAML approach is to recognize and classify the air quality levels. In the APMP-AVOAML system, a four-stage process is encompassed. Firstly, min-max normalization is applied for scaling the input data. Secondly, the HSA-based feature selection process is executed. Thirdly, the XGBoost model is utilized for air pollution prediction. Finally, the AVOA based parameter selection process is exploited for the XGBoost model. To illustrate the performance of the APMP-AVOAML methodology, a brief experimental study is made. The resultant outcomes inferred that the APMP-AVOAML methodology has resulted in effectual solution

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Wardana I, Gardner JW, Fahmy SA. Optimising deep learning at the edge for accurate hourly air quality prediction. Sensors. 2021;21(4):1064.
[2] Naranjo, F.V., Vivar, S.M., Arias, E.J. and Atassi, R., 2023. Early Energy Consumption Prediction as a Key Element in Smart City Sustainability. Journal of Intelligent Systems and Internet of Things, 11(1), pp.12-2.
[3] Patil RM, Dinde HT, Powar SK (2020) A literature review on prediction of air quality index and forecasting ambient air pollutants using machine learning algorithms 5(8):1148–1152.
[4] Sweileh WM, Al-Jabi SW, Zyoud SH, Sawalha AF (2018) Outdoor air pollution and respiratory health: a bibliometric analysis of publications in peer-reviewed journals (1900–2017). Multidiscip Respiratory Med.
[5] Nahar K, Ottom MA, Alshibli F, Shquier MA (2020) Air quality index using machine learning—a jordan case study. COMPUSOFT, Int J Adv Comput Technol 9(9):3831–3840
[6] Afotey, B. and Lovely-Quao, C., 2023. Ambient air pollution monitoring and health studies using low-cost Internet-of-things (IoT) monitor within KNUST Community. Journal of Intelligent Systems and Internet of Things, 10(2), pp.49-9.
[7] Marcelo Y. Villacis, Oswaldo T. Merlo, Diego P. Rivero, S. K. Towfek. "Optimizing Sustainable Inventory Management using An Improved Big Data Analytics Approach." Journal of Intelligent Systems and Internet of Things, Vol. 11, No. 1, 2024 ,PP. 55-64.
[8] Tao Q, Liu F, Li Y, Sidorov D. Air pollution forecasting using a deep learning model based on 1d convents and bidirectional gru. IEEE Access. 2019;7:76690–8.
[9] Ruiz, D.P., Vasquez, R.A.D. and Jadan, B.V., 2023. Predictive Energy Management in Internet of Things: Optimization of Smart Buildings for Energy Efficiency. Journal of Intelligent Systems and Internet of Things, 10(2), pp.08-8.
[10] Jeya S, Sankari L. Air pollution prediction by deep learning model. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE; 2020, p. 736–41.
[11] Ansari, M. and Alam, M., 2024. An intelligent IoT-cloud-based air pollution forecasting model using univariate time-series analysis. Arabian Journal for Science and Engineering, 49(3), pp.3135-3162.
[12] Hardini, M., Sunarjo, R.A., Asfi, M., Chakim, M.H.R. and Sanjaya, Y.P.A., 2023. Predicting air quality index using ensemble machine learning. ADI Journal on Recent Innovation, 5(1Sp), pp.78-86.
[13] Gu, Y., Li, B. and Meng, Q., 2022. Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing, 468, pp.123-136.
[14] Abbas, Z. and Raina, P., 2024. A wavelet enhanced approach with ensemble based deep learning approach to detect air pollution. Multimedia Tools and Applications, 83(6), pp.17531-17555.
[15] Liao, H., Yuan, L., Wu, M. and Chen, H., 2023. Air quality prediction by integrating mechanism model and machine learning model. Science of The Total Environment, 899, p.165646.
[16] Ahmed, A.A.M., Jui, S.J.J., Sharma, E., Ahmed, M.H., Raj, N. and Bose, A., 2024. An advanced deep learning predictive model for air quality index forecasting with remote satellite-derived hydro-climatological variables. Science of The Total Environment, 906, p.167234.

[17] Mehrabi, M., Scaioni, M. and Previtali, M., 2023. Forecasting air quality in kiev during 2022 military conflict using sentinel 5P and optimized machine learning. IEEE Transactions on Geoscience and Remote Sensing.
[18] Gajera, V., Gupta, R. and Jana, P.K., 2016, July. An effective multi-objective task scheduling algorithm using min-max normalization in cloud computing. In 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 812-816). IEEE.
[19] Zhang, Z., Lu, Y., Ye, M., Huang, W., Jin, L., Zhang, G., Ge, Y., Baghban, A., Zhang, Q., Wang, H. and Zhu, W., 2023. A novel evolutionary ensemble prediction model using harmony search and stacking for diabetes diagnosis. Journal of King Saud University-Computer and Information Sciences, p.101873
[20] Dasari, S. and Kaluri, R., 2024. An Effective Classification of DDoS Attacks in a Distributed Network by Adopting Hierarchical Machine Learning and Hyperparameters Optimization Techniques. IEEE Access.
[21] Mohamed, A.A., Kamel, S., Hassan, M.H. and Zeinoddini-Meymand, H., 2024. CAVOA: A chaotic optimization algorithm for optimal power flow with facts devices and stochastic wind power generation. IET Generation, Transmission & Distribution, 18(1), pp.121-144.
[22] K. Manikandan, R. Rajkumar, P. Jayanthi, S. Jothimani, R. Venkatesh and S. Govinda Rao, "An Intelligent Bayesian Optimization with Stacked BiLSTM Model for Air Quality Index Prediction," *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2023, pp. 1699-1704, doi: 10.1109/ICIRCA57980.2023.10220650.