Received: July 31, 2024. Revised: September 20, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Explainable Artificial Intelligence (XAI) Framework for Deep Learning Based Classification to Generate Textual Explanations on Predicted Images

Sheela B. P[1]* **H. Girisha[1]**

[1]*Department of Computer Science and Engineering, Rao Bahadur Y. Mahabaleswarappa Engineering College, Bellary, Visvesvaraya Technological University, Belagavi, 590018, India.*
* Corresponding author's Email: sheelabpbphd@gmail.com

Abstract: The Explainable Artificial Intelligence (XAI) is a set of techniques and methods designed to make machine learning models and AI systems more transparent and interpretable. The goal of XAI is to enable humans to understand and trust the decisions made by AI systems, especially in critical or sensitive applications. In the context of generating explanations, XAI aims to provide human-understandable reasons for making a particular prediction or decision based on Artificial Intelligence (AI) model. XAI is an evolving field where the researchers have different ideas. Shapley values, which help explain the predictions, can be slow in calculation because they need to be figured out in many ways for each prediction. The current techniques for explaining AI are not fast and can be costly. This means they might not work well when trying to explain a lot of predictions at once. To overcome these issues, there developed a novel Convolution Bat Optimization based SHapley Additive exPlanations (CBO-SHAP) algorithm. Initially, a dataset contains a large number of images that are collected and pre-processed, then the data got divided into training, testing sets, and validation. Bat optimization is used for the classification and segmentation of the images which is performed in the pooling layer. To create the textual representation during testing, the trained images were applied to XAI with SHAP algorithm. Later, the image explanations were translated into the textual explanation that is readable by humans. This process enhances the whole operation with very less execution time of 78.3427sec. Comparative analyses against existing methods confirm the superior performance of our model, boasting high accuracy, f-measure, recall, and precision rates of approximately 99%, 99.28%, 99.22%, and 99.1% respectively.

Keywords: Image classification, Explainable artificial intelligence, Convolution bat optimization based SHAP, Machine learning.

1. Introduction

The rising dependence on AI systems in crucial domains such as security, autonomous driving, and healthcare necessitates the need for picture categorization to produce explanations of expected images using ML models. Comprehending the logic underlying artificial intelligence judgements is essential for accountability, transparency, and trust. While traditional picture classification offers predictions, it is not interpretable, which makes it challenging for users to comprehend or verify results. This gap is filled by Explainable AI (XAI) techniques, which offer concise, comprehensible justifications for every prediction, boost user confidence, guarantee ethical AI use, and make it easier to debug and refine AI models.

The Artificial Intelligence (AI) algorithms are particularly termed as Deep Neural Networks (DNN) that are changing the human's way to approach real-world problems. Moreover, the application of Machine Learning (ML) strategies is increasingly automating different aspects of scientific, commercial, and social processes in recent years. The increase is partially due to growing research in the branch of ML known as Deep Learning (DL), in which thousands of neural parameters are studied to generalize on how to perform a specific task [1]. The successful application of DL models in healthcare [2], developmental disorders [3], autonomous robotics system [4], ophthalmology [5], image processing [6],

speech and audio processing [7], cyber-security [8], and other fields demonstrate the reach of DL algorithms in daily life. Explanation Artificial Intelligence (XAI) is an important field that focuses on various strategies for breaking the black-box character of ML models and producing human-range explanations [8]. Moreover, this black box which depicts is too sophisticated to interpret, or opaque models like the popular DL models; also ML models are overly complex; for example, linear regression tree and logistic decision trees are transparent models [9]. Therefore, these models can provide some insight into the relationship between the feature value and the desired results, which are creating them informal to interpret [10]. Nevertheless, this is not the case with sophisticated strategies [11].

Each and every explanation is consistent across the same data points and yields stable or related explanations over the time period [12]. Furthermore, explanations should create the AI strategy sensitive in order to enable human understanding, decision-making confidence, and just decisions and encourage impartial judgement [13]. Therefore, in order to retain openness, confidence, and equality in the ML decision-making development, ML schemes must provide a suitable explanation or an interpretable answer [14]. An explanation is a method of validating an AI agent or strategies output decision. In a medical application XAI is efficiently utilized for a cancer detection system. It is utilized to employ microscopic pictures which could be a map of input pixels that contribute to the model outcome [15]. An explanation for a voice recognition model could be the power spectrum information at a certain time that contributed more to the present output result. Moreover, explanation has certain parameters mainly based on the activations of the trained models, which can be explained using substitutes such as decision trees, gradients, or other ways [16]. Consequently, reinforcement learning models are used for XAI option to provide the best solution for all the system. However, XAI and interpretable AI models are frequently misleading and generic, hence, it should integrate certain forms of reasoning [17].

In past, a lot of XAI merged models were developed such as Gradient based class activation mapping replica [18], DL-based multi-label classification model [19], ML based cognitive impairment diagnosis [20], etc. Nevertheless, some drawbacks were not accurately evaluated and also new problems were generated with this technique. Therefore, to address such issues XAI application is merged with ML with optimization algorithms. This combination can give the finest result and SHAP model is adapted to enhance the textual representation of each collected dataset. The main contributions of this work are:

- Our model provides the use of efficient bat optimization for effective classification and segmentation of images in the pooling layer.
- Generation of clear, human-readable textual explanations from trained images using the SHAP method during testing.
- Our proposed worm Convolution Bat Optimization based SHAP (CBO-S) technique, which integrates Bat Optimization and the SHAP algorithm within a Convolution Neural Networks (CNN) model to enhance the textual explanation of input images
- The proposed algorithm is evaluated using Python programming language to demonstrate significant efficiency with a minimal execution time by achieving high accuracy, f-measure, recall, and precision rates compared to existing methods.

The rest of the sections are organized into Literature works in section 2, System Model and Problem Statement covered in section 3, proposed methodology in section 4 discusses on CBO-S model, Results and discussion covered in section 5, and finally, the research work has been concluded in section 6.

2. Related works

A novel lightweight single CNN model for COVID-19, pneumonia, and tuberculosis classification in CXR images has been created by Bhandari et al. [21] this is supported by an explanation generation (XAI) framework. A publicly available dataset of 7132 chest x-ray (CXR) pictures is used to validate the proposed model. Gradient-weighted Class Activation Mapping (Grad-CAM), Local Interpretable Modelagnostic Explanation (LIME), and Shapley Additive Explanation (SHAP) are also used to analyse and clarify the results in order to make them more understandable. The suggested model demonstrates that XAI and CNN models may produce believable and coherent results for classifying and identifying lung diseases.

Deep learning models, particularly ResNet50V2 [22], are effective in animal species classification. The model uses visual cues to identify five species, and after fine-tuning, it achieves high precision in identifying species. This study highlights the potential of deep learning techniques in visual recognition tasks, suggesting potential applications in biodiversity monitoring and wildlife management.

In [23], explores the application of MobileNet architecture in animal image classification,

demonstrating its effectiveness in real-world applications. The study uses an Animal Image Dataset of 800 images, demonstrating a classification accuracy of 90.7%. The combination of MobileNet and traditional classifiers, such as Support Vector Machines, enhances performance and improves the accuracy of animal classification.

In [24], presents a deep learning model for recognizing rare species in Vietnam, such as Panthera pardus, Dalbergia cochinchinensis, and Macaca mulatta. The model uses convolutional neural networks and Inception residual structures for lightweight classification, fine-tuning MobileNetV2 and InceptionV3, and achieves high accuracy of 95.8% and 113 FPS CPU inference speed.

Wild cat populations like Caracals and African Leopards are facing a decline due to human activities like hunting, poaching, and habitat destruction. To conserve these species, a deep learning approach was developed to identify and classify them from images. The Xception transfer learning model, optimized for this task, achieved 96% accuracy, making it effective for monitoring and managing these species [25].

Deep learning (DL) [26] has shown promise in manufacturing for defect identification, classification, and localization. However, accuracy and interpretability must be compromised. XAI has emerged as an appropriate model for producing human-understandable explanations of "black-box" approaches. Muddamsetty et al. [27] proposed the Similarity Difference and Uniqueness (SIDU) technique, an advanced XAI visual explanation system that successfully localizes object areas responsible for prediction. The study explores the integration of hybridized metaheuristics, XGBoost, and SHAP to provide explainable insights into toluene behavior in the atmosphere [28].

The critical contribution of the proposed CBO-S algorithm is described below:

- Many CNN-based XAI model that was only trained on a small number of datasets. This highlights the need for studies that train on larger, more diverse datasets to improve the generalizability of the results.
- Many are addressed the accuracy of explaining spam images using CNN-based LIME and SHAP but faced challenges with multi-classification. Future research could focus on enhancing multi-classification capabilities within these frameworks.
- Many researchers are utilized SVM-based synthetic oversampling to enhance model interpretability but suffered from poor overall risk estimation rates. Research could aim to develop methods to improve risk estimation accuracy.
- Many researchers are faced difficulties in object localization across various appearances and locations within an image. Studies could focus on improving object localization techniques, potentially through advanced localization algorithms or combined approaches.
- Increasing the reliability and consistency of identification methods and developing robust methods for multiple object classification.

3. System model and problem statement

Due to a lack of standardization, XAI (Explainable Artificial Intelligence) has emerged as a developing field in which academia has contributed many definitions and perspectives. Additionally, generating Shapley values for multiple aspects of a prediction instance is computationally expensive and slow, making existing techniques unsuitable for providing global explanations when dealing with a large number of prediction instances. This represents a significant challenge in the field. To address these challenges, machine learning (ML) models combined with optimization algorithms have been developed [35-38]. Fig. 1 illustrates the system model and problem statement for the XAI application.

4. Proposed methodology

The present work aims to develop a novel convolution-based Bat Optimization [28] based CBO-S algorithm. Moreover, Explainable AI (XAI) models, such as the SHAP algorithm, are integrated with machine learning models and applied to processed images to provide textual explanations of the given input images. Additionally, image data is converted into textual representations. Initially, the input datasets are fed into the input layer of the CNN

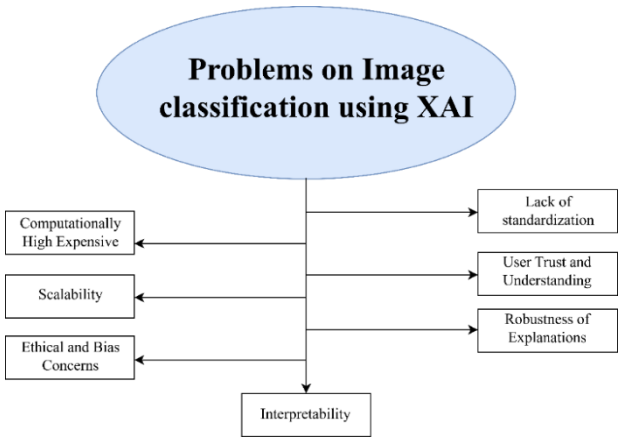

Figure. 1 System model and problem statement

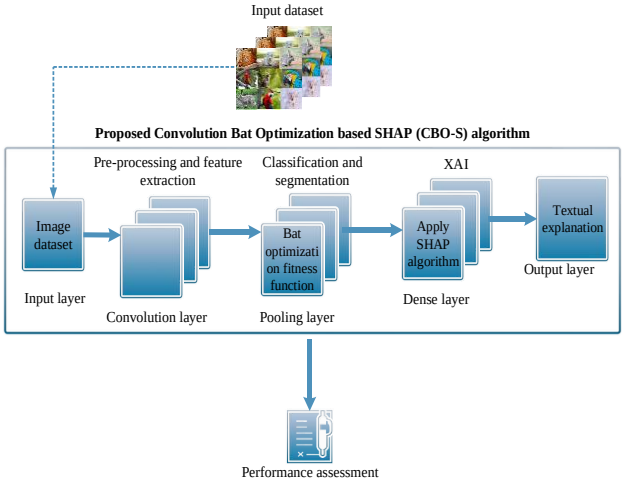

Figure. 2 Proposed methodology

model [29]. Pre-processing and feature extraction are performed in the convolution layer. Subsequently, pre-processed images are fed into the pooling layer, where the bat optimization [32] fitness function is initiated to classify and segment the pre-processed images.

The proposed methodology uses the SHAP algorithm to generate image explanations using the XAI application illustrated in Fig. 2. The algorithm is designed to be efficient and accelerates the process with minimal execution time. The CBO-S algorithm is used to train various types of images from online sources, with unwanted information removed using a convolution layer. Feature extraction is performed in the convolution layer, and the extracted features are passed through a pooling layer. The trained images are then applied to the XAI framework using the SHAP algorithm to generate textual representations. The evaluated metrics are then validated against other models for accuracy, precision, recall, and other measures.

4.1 Design of proposed CBO-S

4.1.1. Input layer:

The aim of the proposed framework is to give the textual explanation of the given input image and to enhance the whole process with very less execution time. Initially the different types of image datasets are gathered from the standard web sources and it is trained in MATLAB. Inside the system, the dataset is initialized using the data initialization function. It is expressed in Eq. (1).

$$I_L(pd^*) = (l_1, l_2, l_3 \ldots \ldots l_n) \quad (1)$$

Where, I_L represents the dataset initialization function, pd^*represents the gathered image dataset, l represents the data present in the image dataset and n represents the total number of data in the dataset.

4.1.2. Convolution layer

Data pre-processing is done before the image data is passed through the convolution layer, where the noisy features and error values of the input images are eliminated in order to improve the quality of the data, then the normalization of the pixel values of the images are performed to bring them within the common range. The model's training speed is enhanced, preventing gradients from becoming too large or small, and the pre-processing function also improves system efficiency and reduces computation time. It is expressed in Eq. (2).

$$P_L^* = \{I_L(pd^*) - \tau(pd)\} \quad (2)$$

Where, P_L^* signifies the pre-processing function and$\tau(pd)$ denotes the noisy features and error values of the input images. Afterwards, the feature extraction of the processed images is done within the convolution layer. Here, some of the significant features like important pixels, textural patterns, edges, boundaries and relevant regions such as object, shapes, and textures of the images are extracted and it is expressed in below Eq. (3).

$$F_L^* = \{P_L^* - \eta(pd)\} \quad (3)$$

Where, the F_L^* implies the feature extraction function, $\eta(pd)$ denotes the unnecessary features.

4.1.3. Pooling layer

The primary purpose of a pooling layer is to reduce the computational complexity of the model and control over fitting by retaining the most relevant information while discarding the less important details. The classification and segmentation of the images are performed in this layer using Bat Optimization (BO) Algorithm.

The BO algorithm is based on the echolocation behaviour of bats and it addresses the practical problem by simulating virtual bats. The bat employs Sound Navigation and Ranging (SONAR) echoes to find obstacles and avoid them. These sound waves are translated into frequencies, and if there are obstructions, the obstacles will reflect the waves back to the source. The amount of time between the wave's emission and reflection has an impact on the bat's movement. After receiving the reflected signal, the bat uses its own pulse to calculate the prey's distance. The range of the pulse rate is 0 to 1, with 0 denoting

no emission and 1 denoting maximum emission. The loudness is greatest when the bat is farthest away from its prey. If the pulse rate rises, loudness decreases. Bats randomly fly with a velocity V_j of at a position of P_j while emitting sounds with an initial frequency of $f_{\min}$, a wavelength of λ and the loudness of L_0. The bat thereafter determines the target's proximity and modifies the pulse' frequency or wavelength and rate of emission P_{ER}, if necessary. The steps for BO algorithm are given below:

Step 1: Initialization is done for the bats' position P_j, velocity V_j, frequency f_j, pulse emission rate P_{ER}, and loudness L_j. Both the population size and iteration count are initialized. The fitness function is represented as $F(P_j)$.

Step 2: The following equations are updated to get new solutions by updating the bats' position, velocity, and frequency.

$$P_j^{t+1} = P_j^t + V_j^{t+1} \tag{4}$$

$$V_j^{t+1} = V_j^t + \left(P_j^{t+1} - P_j^*\right) f_j \tag{5}$$

$$f_j = f_{min} + (f_{max} - f_{min})\gamma \tag{6}$$

Where, $y \in [0, 1]$ is a uniformly generated random vector, and P_j^* is the present global best location or solution, as determined by comparing all solutions among all n bats. The minimum frequency is denoted as f_{min}, while the maximum frequency is represented as f_{max}.

Step 3: A random number is created following the update of the bats' positions. A new solution is generated around the present global best solution utilising a local random walk if the random number produced is higher than the pulse emission rate, its equation is given by,

$$P_{\text{new}} = P_{\text{old}} + \mu L^{t+1} \tag{7}$$

Where, $\mu \in [-1,1]$ is denoted as the random number. L^{t+1} is an average bat volume at this time period. Then, the new solution is evaluated using fitness function $F(P_{new})$.

Step 4: The new solution or location has been accepted, if $\gamma < L_j$ and $F(P_{new}) > F(P_j)$ .Then, using the below equation, P_{ER} and L_j are updated.

$$L_j^{t+1} = \beta L_j^t \tag{8}$$

$$P_{ER}^{t+1} = P_{ER}^0\left[1 - e^{(-\psi t)}\right] \tag{9}$$

Where, β and η are constant. When $t \to \infty$, for any $0 < \beta < 1$ and $0 < \psi$, we have

$$L_j^t \to 0$$

$$P_{ER}^t \to P_{ER}^0$$

Loudness reduces as the number of iterations rises and eventually becomes zeros. On the other hand, when the number of iterations increases, the pulse rate emission rises and gradually goes towards the original value.

Step 5: The bats are classified based on their fitness function and the optimum solution or position is determined.

Step 6: It is suggested to return to step 2 if the number of iterations is less than the maximum number. If not, the process can be concluded.

4.1.4. Dense layer

The dense layer in neural networks is vital for learning complex patterns from input data. Explainable AI (XAI) techniques aim to provide human-readable explanations for neural network predictions, analyzing dense layer activations for insights into decision-making. The SHAP algorithm, applied in XAI, generates textual explanations for images by averaging feature values' marginal contributions. SHAP scores use Shapley values to rank model features' influences, showing each pixel's contribution to predicted images and justifying classifications. Red pixels increase class likelihood, while blue decrease it. Eq. (10) [21] calculates Shapley values when pixelated.

$$\phi_j = \sum_{A \subseteq F\{j\}} \frac{|A|!(F-|A|-1)!}{F!}\left[f_y(A \cup j) - f_y(A)\right] \tag{10}$$

Where, F refers to all feature sets, A is a subset of F, $f_y(A \cup j)$ is the trained model using A and the j^{th} feature, $f_y(A)$ denotes the trained model without that feature, and $\left[f_y(A \cup j) - f_y(A)\right]$ is the value determined on any feasible subset $A \subseteq F\{j\}$. Finally, the weighted average of all potential differences makes up the computed Shapley values. Additionally, feature attributions are made using this representation. The SHAP algorithm belongs to the additive feature

attribution class, and models with a linear function of binary variables are used as explanations in this class. Each original feature (y_j) is replaced with a binary variable (d_j^*) in SHAP that indicates whether the characteristic is existing or not that is stated in Eq. (11).

$$h(d^*) = \phi_0 + \sum_{j=1}^{F} \phi_j d_j^* \quad (11)$$

Where, $h(d^*)$is the explanation model, the extent to which the presence of feature j contributes to the final outcome, and ϕ_j contributes to help the understanding of the original model.

4.1.5. Output layer

The output layer of the system is designed to present the textual interpretation of the image in a format easily understandable by humans. By providing a clear and concise description of the image content, it enables users to comprehend the information conveyed by the AI system with ease. The flowchart of the proposed CBO-S is represented in Fig. 3.

5. Results and discussion

The effectiveness of the suggested approach is covered in this sector. With the help of Bat optimization and SHAP algorithm, evaluate the system's performance using various measures, including accuracy, precision, recall and F-Measure. The experimental setup is tabulated in table 1.

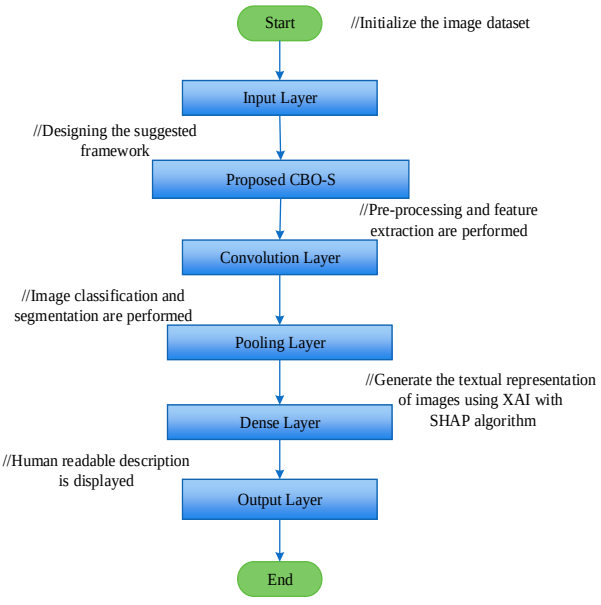

Figure. 3 Flowchart of the proposed framework

Table 1: Experimental Setup

Processor	Intel(R) Core(TM) i5-3570, @3.40GHz
RAM	8.00 GB (7.88 GB usable)
System Type	64-bit operating system
Edition	Windows 10 Pro
Version	22H2
Tool	Python

5.1 Dataset description

The dataset comprises of 5400 colour images with height of 240 * 240 pixels from kaggle repository of animals, spanning across 90 distinct categories or classes [39]. These include a wide variety of creatures such as antelope, badger, bat, bear, bee, beetle, and many others. From mammals like chimpanzees, dogs, and elephants, to birds like eagles, flamingos, and owls, and even marine life like dolphins, sharks, and seals, the dataset offers a comprehensive representation of the animal kingdom. This diverse collection is a valuable resource for training and testing in the field of computer vision and image recognition. For the experimental purpose the dataset is divided into 70:30 as training (3780) and testing datasets (1620) with the epochs ranging between 0-100.

5.2 Performance matrices

To validate our proposed Convolution Bat Optimization based SHAP (CBO-S) algorithm, the performance metrics are evaluated in terms of accuracy, precision, recall, f-measure, error rate, execution time and AUC are shown in Figs. 4 to 10 respectively.

5.2.1. Accuracy:

The percentage of all accurate predictions is an extremely efficient and commonly used measure for the evaluation of models. For ML to evaluate an algorithm's effectiveness in classification, accuracy is crucial. The percentage of perfectly predicted images out of all of the provided images is called accuracy. It is determined by using the below equation,

$$A_T^* = \frac{T_{tp}^* + T_{tn}^*}{T_{tp}^* + T_{tn}^* + T_{fp}^* + T_{fn}^*} \quad (12)$$

Where, A_T^* indicates the accuracy, T_{tp}^*denoted as true positive, $T_{tn}^*$is true negative, T_{fp}^* indicates the false positive, $T_{fn}^*$is false negative.

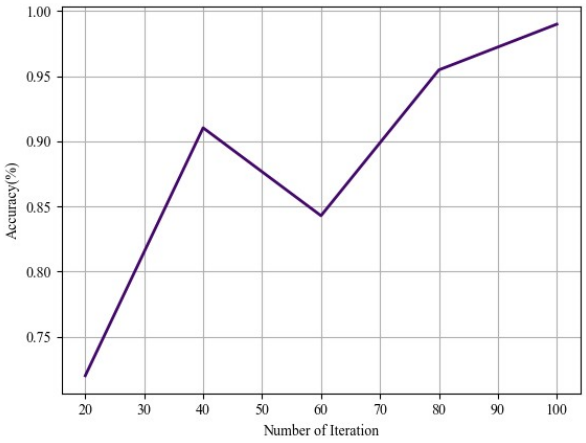

Figure. 4 Accuracy of proposed CBO-S

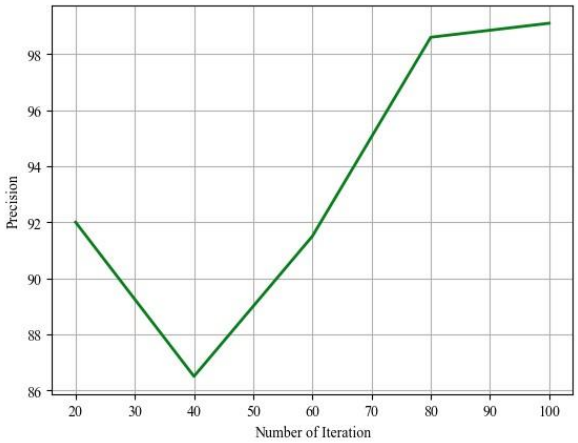

Figure. 5 Precision of proposed CBO-S

5.2.2. Precision:

The amount of the relevant information of images is precision. Precision is the percentage of true predicted positive outcomes to all predicted positive outcomes, or the proportion of accurately identified positive items to all positive items. The equation is,

$$P_T^* = \frac{T_{tp}^*}{T_{tp}^* + T_{fp}^*} \quad (13)$$

5.2.3. Recall:

Recall is a metric used to assess a model's ability to recognize typical data aspects; it represents the percentage of true positive outcomes that are correctly recorded. The formula for recall is

$$R_T^* = \frac{T_{tp}^*}{T_{tp}^* + T_{fn}^*} \quad (14)$$

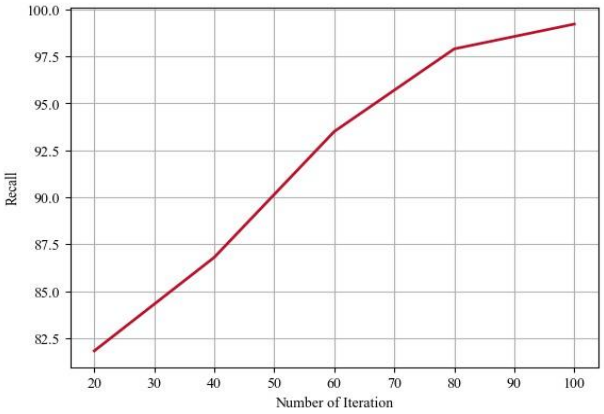

Figure. 6 Recall of proposed CBO-S

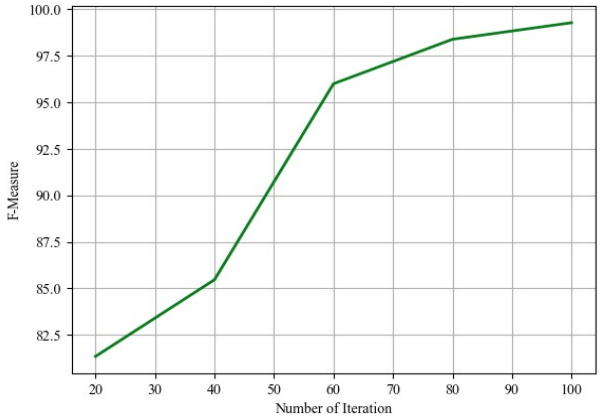

Figure. 7 F-Measure of proposed CBO-S

5.2.4. F-Measure

The precision-to-recall ratio is expressed by the F1-score. The F1- score represents the harmonic average for recall and accuracy levels. It is expressed in below equation.

$$F_1^* = \frac{2(P_T^* * R_T^*)}{P_T^* + R_T^*} \quad (15)$$

5.2.5. Error rate

The error rate is the proportion of incorrect outcomes compared to total attempts or observations. It is used to assess accuracy and performance in various fields. Lower error rates indicate higher precision and effectiveness in the respective domain. In our model obtained very less error rate of 0.015.

5.2.6. Execution time

Execution time, or runtime, is the duration a program takes to complete its tasks on a computing system. It encompasses all processing steps from start to finish. Execution time can vary based on factors

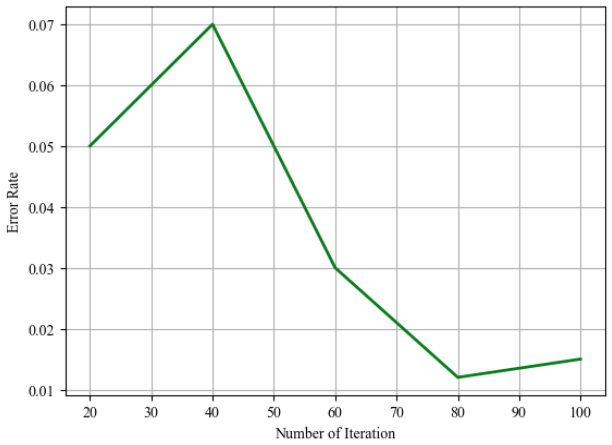

Figure. 8 Error rate of proposed CBO-S

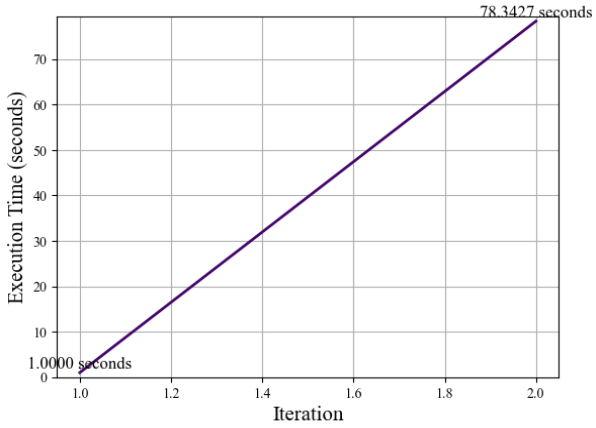

Figure. 9 Execution time of the proposed CBO-S

like the complexity of the program, the efficiency of the algorithm, and the size of the input data. It is a crucial metric in evaluating the performance and efficiency of software, especially in scenarios where processing speed is a critical factor. In this work, here it achieved a better execution time of 78.3427 seconds compared to ResNet50, Xception, InceptionRenNetV3, MobileNet, and EfficientNet.

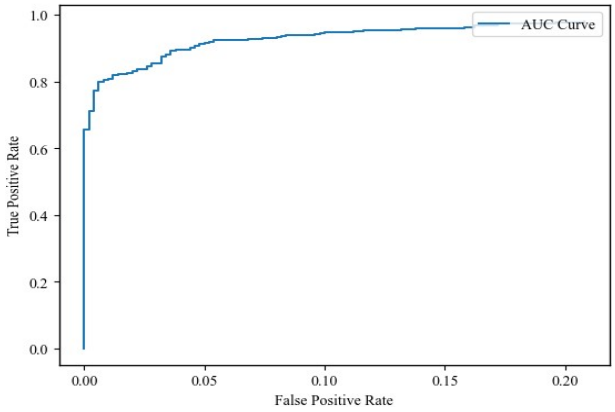

Figure. 10 ROC curve of the proposed CBO-S

5.2.7. Area under the Receiver Operating Characteristic Curve (AUC-ROC) Curve

It provides a quantifiable assessment of different threshold levels, which is crucial in binary classification tasks. The ROC curve represents a probability curve, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). Meanwhile, the AUC calculates the degree of separability, indicating how effectively the model can differentiate between classes. A higher AUC value signifies a better-performing model, showcasing its ability to make accurate classifications. Here, the proposed method CBO-S achieves the better values of 0.98.

5.3 Comparative analysis

The created model has been implemented in Python framework, and the effectiveness of the model was confirmed by comparing its metrics to those of the other models in terms of accuracy, precision, recall, f-measure, specificity, AUC with the existing techniques like ResNet50, Xception,

Table. 2 Comparison values of existing methods

Method	Accuracy	Recall	Precision	F-Measure
ResNet50 [22]	88.5	84.3	84.7	84.5
Xception [25]	96	86.1	86.7	86.4
InceptionRenNetV3 [24]	95.3	87.5	87.8	87.5
MobileNet [23]	90.7	86.9	88.2	87.5
EfficitientNet [27]	83.3	76.9	83.3	80.0
Explainable artificial intelligence model [23]	93	93.3	92.6	93.3
Explainable potential [28]	87.5	89.5	85	87.1
Proposed CBO-S	99	99.22	99.1	99.28

InceptionRenNetV3, MobileNet, EfficitientNetB3 [30] as mentioned in table 2.

5.3.1. Comparison of proposed with other existing methods in terms of accuracy:

The obtained results are compared with the other methods to check the performance level. Here, we extract the accuracy level. ResNet50 has 88.5, very low accuracy level when compared to all the other methods. Xception has 96 level higher than ResNet50 but lower than all the other methods. MobileNet has 90.7 higher than ResNet50 and Xception but lower than the other methods. InceptionRenNetV3 has 95.3 of accuracy level, EfficitientNetB has 83.3 of accuracy level. Our proposed method CBO-S had score 99 of accuracy higher than all other methods as shown in Fig. 11.

5.3.2. Comparison of proposed with other existing methods in terms of recall:

The CBO-S method was compared to existing methods in recall, with ResNet50 having the lowest recall value of 84.3. Xception had a higher recall value of 86.1 but lower than other methods. MobileNet had a higher recall value of 86.9 but lower than ResNet50 and Xception. EfficitientNetB had a higher recall value of 76.9, while InceptionRenNetV3 had a higher recall value of 87.5.

5.3.3. Comparison of proposed with other existing methods in terms of precision:

The precision of the proposed method CBO-S is compared with the other methods and its graph is shown in below Fig. 11. ResNet50 has 84.7 of precision ranking very low when compared to the other methods. Xception has the value 86.7 of precision higher than the ResNet50 but lower than the remaining methods. InceptionRenNetV3 has 87.8 which is higher than ResNet50 and Xception but lower than the other methods. MobileNet has 88.2 of precision, EfficitientNetB has 83.3 of precision. CBO-S scores 99.1, a high precision value when compared to the existing methods.

5.3.4. Comparison of the proposed method with the other existing methods in terms of F-Measure:

The CBO-S method is compared to existing F-Measure methods, with ResNet50 having the lowest F-Measure value of 84.5. Xception has a higher F-Measure value of 86.4, followed by InceptionRenNetV3 and MobileNet at 87.5 and 87.5 respectively. EfficitientNetB has an F-Measure value

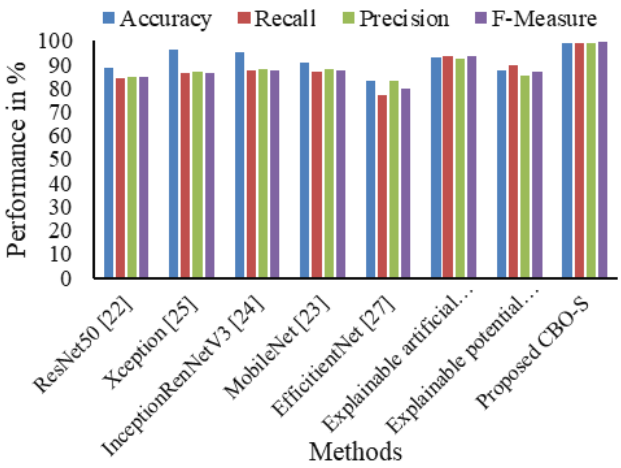

Figure. 11 Comparison of accuracy, Recall, Precision, F-Measure

of 80.0, while CBO-S scores 99.28, indicating a high F-Measure value compared to the existing method.

5.3.5. Comparison of the proposed with the other existing methods in terms of AUC

The proposed method of CBO-S is compared with the existing methods in AUC and it is shown in table 3. Deep Spatial Fusion CNN (DSF-CNN) has the least value of 83.89 in AUC when compared to all the other methods. Hybrid-LSTM has the value 90.1 of AUC higher than the DSF-CNN but lower than the remaining methods. Two-stage CNN has 90.21 which is higher than DSF-CNN and Hybrid-LSTM but lower than the other methods. EMS-Net has 91.36 of AUC. CBO-S scores 98 a high AUC value when compared to the existing method [31].

The CBO-S model outperforms other models like ResNet50, Xception, InceptionResNetV2, MobileNet, and EfficientNet due to its optimized architecture, attention processes, better data augmentation methods, and better learning rate schedules. It can identify intricate patterns in data and reduce noise, enhancing focus on essential characteristics. Additionally, using a larger, more diversified pre-trained model or integrating transfer learning can improve accuracy. Advancements in optimization methods tailored to the CBO-S

Table. 3 Comparison values of existing methods in terms of AUC

Method	AUC
Two-stage CNN	90.21
DSF-CNN	83.89
Hybrid-LSTM	90.1
EMS-Net	91.36
Proposed CBO-S	98

architecture also enhance model performance and convergence speed, resulting in significant accuracy gains. Overall, the CBO-S model offers a more efficient and effective approach to data analysis.

6. Conclusions

The study developed a novel technique called CBO-S using Bat Optimization and the SHAP algorithm in a CNN model to improve textual explanation of input images. The dataset was pre-processed, divided into training, testing, and validation sets, and the images were categorized and segmented using bat optimization. The trained images were then fed into XAI using the SHAP method to generate textual representations. The model effectively converted image explanations into easily understandable textual representations, improving efficiency with a minimal execution time of 78.3427sec. Comparative analyses confirmed the model's superior performance, with high accuracy, f-measure, recall, and precision rates of approximately 99%, 99.28%, 99.22%, and 99.1%, respectively. Future research should investigate additional optimization techniques or algorithmic enhancements to further improve computational efficiency, especially for large-scale applications.

Conflicts of Interest

No author has disclosed any conflicts of interest.

Author Contributions

Sheela B. P was held responsible for identifying the initial problem, algorithm write-up, analysis, drafting of the manuscript, and simulation and was made responsible for the figures, final formatting and for applying for publishing in the journal. Girisha H was responsible for the Literature survey and helped in the initial review process, complexity analysis of the research and the evaluation of the research work. All the authors worked together to implement and evaluate the integrated system, and approve the final version of the paper.

Acknowledgments

Authors acknowledge the support from Rao Bahadur Y, Mahabaleswarappa Engineering College for the facilities provided to carry out the present research.

Funding Statement

There are no particular grants from the funding organizations for this research.

References

[1] B. H. M. Van der Velden, I. P. Jansen, L. M. Steens, R. H. G. Oude Elferink, R. J. P. van Heeswijk, M. B. Van Leeuwen, and W. B. Nagengast, “Explainable artificial intelligence (XAI) in deep learning-based medical image analysis”, *Medical Image Analysis*, Vol. 81, p. 102470, 2022, doi: 10.1016/j.media.2022.102470.

[2] K. Hauser, E. Weninger, L. Scholler, and D. P. Neureiter, “Explainable artificial intelligence in skin cancer recognition: A systematic review”, *European Journal of Cancer*, Vol. 167, pp. 54–69, 2022, doi: 10.1016/j.ejca.2022.02.022.

[3] S. O’Sullivan, C. Herrler, K. Eck, and H. U. Kauczor, “Explainable artificial intelligence (XAI): closing the gap between image analysis and navigation in complex invasive diagnostic procedures”, *World Journal of Urology*, Vol. 40, No. 5, pp. 1125–1134, 2022, doi: 10.1007/s00345-021-03833-6.

[4] C. C. Yang, “Explainable artificial intelligence for predictive modeling in healthcare”, *Journal of Healthcare Informatics Research*, Vol. 6, No. 2, pp. 228–239, 2022, doi: 10.1007/s41666-021-00111-1.

[5] M. S. Veldhuis, R. M. van Oorschot, L. A. L. Verheij, and J. M. Kerkhoff, “Explainable artificial intelligence in forensics: Realistic explanations for number of contributor predictions of DNA profiles”, *Forensic Science International: Genetics*, Vol. 56, p. 102632, 2022, doi: 10.1016/j.fsigen.2021.102632.

[6] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”, *Information Fusion*, Vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[7] C. Mencar and J. M. Alonso, “Paving the way to explainable artificial intelligence with fuzzy modeling: tutorial”, *Fuzzy Logic and Applications: 12th International Workshop, WILF 2018, Genoa, Italy, September 6–7, 2018, Revised Selected Papers*, pp. 69–86, 2019, doi: 10.1007/978-3-030-20245-3_6.

[8] E. Holder and N. Wang, “Explainable artificial intelligence (XAI) interactively working with humans as a junior cyber analyst”, *Human-Intelligent Systems Integration*, Vol. 3, No. 2, pp. 139–153, 2021, doi: 10.1007/s42454-020-00024-4.

[9] H. Lundberg, M. Ekstedt, P. Davidsson, M. Almgren, and R. Jardler, "Experimental Analysis of Trustworthy In-Vehicle Intrusion Detection System Using eXplainable Artificial Intelligence (XAI)", *IEEE Access*, Vol. 10, pp. 102831–102841, 2022, doi: 10.1109/ACCESS.2022.3213502.

[10] N. Capuano, M. Gaeta, and V. Salerno, "Explainable Artificial Intelligence in CyberSecurity: A Survey", *IEEE Access*, Vol. 10, pp. 93575–93600, 2022, doi: 10.1109/ACCESS.2022.3202811.

[11] A. K. M. B. Haque, A. K. M. N. Islam, and P. Mikalef, "Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research", *Technological Forecasting and Social Change*, Vol. 186, p. 122120, 2023, doi: 10.1016/j.techfore.2022.122120.

[12] J. Vice and M. M. Khan, "Toward accountable and explainable artificial intelligence part two: The framework implementation", *IEEE Access*, Vol. 10, pp. 36091–36105, 2022, doi: 10.1109/ACCESS.2022.3162369.

[13] A. Rawal, A. K. Gautam, A. Sharma, and V. S. Dhaka, "Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges, and Perspectives", *IEEE Transactions on Artificial Intelligence*, Vol. 3, No. 6, pp. 852–866, 2021, doi: 10.1109/TAI.2021.3103995.

[14] C. Zednik, "Solving the black box problem: A normative framework for explainable artificial intelligence", *Philosophy & Technology*, Vol. 34, No. 2, pp. 265–288, 2021, doi: 10.1007/s13347-020-00401-4.

[15] J. Borrego-Díaz and J. Galán-Páez, "Explainable Artificial Intelligence in Data Science: From Foundational Issues towards Socio-technical Considerations", *Minds and Machines*, Vol. 32, No. 3, pp. 485–531, 2022, doi: 10.1007/s11023-021-09602-3.

[16] H. Nasiri, A. Homafar, and S. C. Chelgani, "Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using an explainable artificial intelligence", *Results in Geophysical Sciences*, Vol. 8, p. 100034, 2021, doi: 10.1016/j.ringps.2021.100034.

[17] R. Guidotti, A. Monreale, and D. Pedreschi, "Principles of explainable artificial intelligence", in *Explainable AI Within the Digital Transformation and Cyber Physical Systems: XAI Methods and Applications*, pp. 9–31, 2021, doi: 10.1007/978-3-030-57369-0_2.

[18] R. Schönhof, C. Bil, and M. Dankel, "Feature visualization within an automated design assessment leveraging explainable artificial intelligence methods", *Procedia CIRP*, Vol. 100, pp. 331–336, 2021, doi: 10.1016/j.procir.2021.05.076.

[19] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 103, p. 102520, 2021, doi: 10.1016/j.jag.2021.102520.

[20] M. Sidulova, N. Nehme, and C. H. Park, "Towards Explainable Image Analysis for Alzheimer's Disease and Mild Cognitive Impairment Diagnosis", n: *Proc. of 2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1-5, 2021, doi: 10.1109/AIPR53593.2021.9681235.

[21] M. Bhandari, T. B. Shahi, B. Siku, and A. Neupane, "Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI", *Computers in Biology and Medicine*, Vol. 150, p. 106156, 2022, doi: 10.1016/j.compbiomed.2022.106156.

[22] D. Hindarto, "Use ResNet50V2 Deep Learning Model to Classify Five Animal Species", *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, Vol. 7, No. 4, pp. 758–768, 2023, doi: 10.35870/jtik.v7i4.1845.

[23] M. Sowmya, M. Balasubramanian, and K. Vaidehi, "Classification of Animals Using MobileNet with SVM Classifier", *Lecture Notes on Data Engineering and Communications Technologies*, pp. 347–358, 2022, doi: 10.1007/978-981-19-3015-7_25.

[24] T. T. A. Loan, P. T. Anh, L. V. Nam, and H. V. Dung, "An Effective Deep Learning Model for Recognition of Animals and Plants", *Journal of Computer Science and Cybernetics*, Vol. 38, No. 1, pp. 15–29, 2022, doi: 10.15625/1813-9663/38/1/16309.

[25] R. Pillai, N. Sharma, and R. Gupta, "Classification of Endangered Species of Wild Cats through Xception Transfer Learning Model", In: *Proc. of* in *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, 2023, doi: 10.1109/APSIT58554.2023.10201732.

[26] M. Lee, J. Jeon, and H. Lee, "Explainable AI for domain experts: a post Hoc analysis of deep learning for defect classification of TFT–LCD panels", *Journal of Intelligent Manufacturing*, pp. 1–13, 2021, doi: 10.1007/s10845-021-01808-8.

[27] S. M. Muddamsetty, M. N. Jahromi, A. E. Ciontos, L. M. Fenoy, and T. B. Moeslund, "Visual explanation of black-box model: Similarity Difference and Uniqueness (SIDU) method", *Pattern Recognition*, Vol. 127, p. 108604, 2022, doi: 10.1016/j.patcog.2022.108604.

[28] N. Bacanin, S. Husić, M. Tuba, P. Jovanovic, and Z. Papić, "The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere", *Science of The Total Environment*, Vol. 929, p. 172195, 2024, doi: 10.1016/j.scitotenv.2024.172195.

[29] M. Pujar, M. R. Mundada, B. J. Sowmya, S. Supreeth, and G. Shruthi, "An Efficient Framework for Web Content Mining Systems Using Improved CD-PAM Clustering and the A-CNN Technique", *SN Computer Science*, Vol. 4, No. 5, 2023, doi: 10.1007/s42979-023-02137-w.

[30] S. I. Nafisah and G. Muhammad, "Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence", *Neural Computing and Applications*, pp. 1–21, 2022, doi: 10.1007/s00521-022-07117-0.

[31] Z. Senousy, M. M. Gaber, and M. M. Abdelsamea, "AUQantO: Actionable Uncertainty Quantification Optimization in deep learning architectures for medical image classification", *Applied Soft Computing*, p. 110666, 2023, doi: 10.1016/j.asoc.2023.110666.

[32] R. Govardanagiri and V. Sanjeevulu, "Hybrid Grasshopper and Improved Bat Optimization Algorithms-based clustering scheme for maximizing lifetime of Wireless Sensor Networks (WSNs)", *International Journal of Intelligent Engineering & Systems*, Vol. 15, No. 3, 2022, doi: 10.22266/ijies2022.0630.45.

[33] S. I. Saheb, K. U. R. Khan, and C. S. Bindu, "A Multi-Objective Sensor Deployment using BAT Algorithm to Improve Coverage, Lifetime and Energy Consumption in Wireless Sensor Networks", *International Journal of Intelligent Engineering & Systems*, Vol. 15, No. 3, 2022, doi: 10.22266/ijies2022.0630.39.

[34] J. R. Saini, S. Vaidya, I. Dhulekar, and K. Pal, "GRAINY-XAI: A Domain-wise Granular Analysis of Sarcastic Text Using Explainable-AI Techniques", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 4, pp. 981–993, 2024, doi: 10.22266/ijies2024.0831.74.

[35] Y. T and G. Murtugudde, "Study of Evolutionary Techniques in the field of Network Security", In: *Proc. of 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Vol. 4, pp. 594–598, 2020, doi: 10.1109/icstcee49637.2020.9277082.

[36] T. Yerriswamy and M. Gururaj, "An Efficient Hybrid Protocol Framework for DDoS Attack Detection and Mitigation Using Evolutionary Technique", *Journal of Telecommunications and Information Technology*, Vol. 4, No. 2022, pp. 77–83, 2022, doi: 10.26636/jtit.2022.165122.

[37] H. V. Ramachandra et al., "Secured Wireless Network Based on a Novel Dual Integrated Neural Network Architecture", *Journal of Electrical and Computer Engineering*, Vol. 2023, pp. 1–11, 2023, doi: 10.1155/2023/9390660.

[38] S. B. et al., "Machine learning model for emotion detection and recognition using an enhanced Convolutional Neural Network", *Journal of Integrated Science and Technology*, Vol. 12, No. 4, 2024, doi: 10.62110/sciencein.jist.2024.v12.786.

[39] Animal Image Dataset (90 different animals)," *Kaggle*, [Online]. Available: https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals. [Accessed: Aug. 30, 2024].