

ISSN 1840-4855

e-ISSN 2233-0046

Original scientific article

<http://dx.doi.org/10.70102/afts.2026.1836.516>

AN ADAPTIVE BUTTERFLY OPTIMIZATION-DRIVEN EXPLAINABLE CNN–TRANSFORMER FRAMEWORK FOR ACCURATE AND TRUSTWORTHY CLINICAL DIAGNOSIS

A. Vaideghy¹, V. Sridevi², D. Saraswathi³, L.S. Sindhuja^{4*}

¹Assistant Professor (Selection Grade), Department of Computing - Data Science, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India.

e-mail: vaideghy@gmail.com, orcid: <https://orcid.org/0000-0003-4360-8123>

²Assistant Professor (Selection Grade), Department of Computing - Artificial Intelligence and Machine Learning, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India. e-mail: vissridevi@gmail.com, orcid: <https://orcid.org/0000-0002-1378-0507>

³Assistant Professor (Selection Grade), Department of Computing- Artificial Intelligence and Machine Learning, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India. e-mail: saraswathi.ds@gmail.com, orcid: <https://orcid.org/0000-0002-8271-6752>

^{4*}Assistant Professor (Senior Grade), Department of Software Systems, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India. e-mail: sindhuja@cit.edu.in, orcid: <https://orcid.org/0000-0002-6906-477X>

Received: May 21, 2026; Revised: July 02, 2026; Accepted: August 13, 2026; Published: September 29, 2026

SUMMARY

Clinically-oriented AI-based diagnostic systems have exhibited impressive predictive accuracies; however, their poor interpretability capabilities and incapability to optimize their learning parameters in real-time pose a major challenge towards implementing such solutions in the actual healthcare settings. To overcome such limitations, in this research, it presents an efficient Adaptive Butterfly Optimization-driven Explainable CNN-Transformer based Clinical Diagnosis Solution (ABO-ClinExplain). This proposed framework uses an innovative hybrid CNN-Transformer approach that incorporates an adaptive butterfly optimization algorithm for optimizing learning parameters, feature extraction weights, and classification criteria in order to deliver high diagnostic accuracies without losing explainability. Additionally, the presented solution uses two new techniques called Explainable Attention Mapping and Shapley Additive explanations to create visual and feature-level explanations of the clinical prediction process. The proposed approach is assessed on a large multimodal clinical database consisting of medical imaging reports, electronic health records, lab reports, and symptomatology data of patients across different diseases gathered through 22,500 patients' records. Results show that the proposed ABO-ClinExplain framework is capable of achieving diagnostic accuracies of 98.6%, 98.1% precision, 97.8% recall, and 97.9% F1 score compared to 90.4%, 85.1%, 89.4%, and 96.3% for CNN, ResNet-50, LSTM, and the baseline ClinExplain-AI models, respectively. The proposed method enhances the feature selection efficiency by 26.8%, while the false diagnosis is decreased by 28.4%. Moreover, the interpretability module achieves a rating of 95.7% and boosts clinicians' confidence levels by 43.6%. Computational assessments demonstrate the prediction lag at only 44 ms, indicating the possibility of implementing the developed technology in real-life scenarios. On the whole, the results indicate that the suggested method makes an effective combination of adaptive optimization, contextual intelligence, and explainability, resulting in reliable decision support systems. The methodology described above can be

deemed a method of developing ethical AI in healthcare.

Key words: *adaptive butterfly optimization (ABO), explainable artificial intelligence (XAI), clinical diagnosis, CNN–transformer hybrid model, trustworthy AI in medicine, SHAP interpretation, medical decision support systems.*

INTRODUCTION

The integration of AI into clinical diagnosis has made possible the automated identification of diseases, medical image processing, and intelligent decision-support systems. Deep learning techniques have shown very good results in healthcare tasks by capturing complex feature representations from heterogeneous clinical data, which has improved the accuracy of diagnoses and minimized human involvement [6][18]. Nevertheless, even though deep learning models show high levels of prediction power, most of them act as black boxes, making them difficult to adopt in the clinical setting because of the need for interpretability and transparency [2]. Explainable Artificial Intelligence (XAI) can be considered an answer to the need for transparency in the field of medical diagnostic systems. Recent research has shown how important techniques like SHAP, saliency maps, and attention visualization are to give clinicians clear explanations of the predictions of these models [7]. XAI frameworks help clinicians trust and adopt AI technologies in the healthcare sector.

Convolutional Neural Network has widely been used for the analysis of medical images owing to its superior ability to detect local spatial features from the clinical images [22][5]. In recent times, Transformer architecture has emerged as an effective tool for detecting long-range dependence and global context information, resolving certain shortcomings of conventional CNN-based algorithms [23][10]. Therefore, hybrid CNN-Transformer techniques have drawn a lot of attention lately for leveraging the advantages of both the techniques to obtain optimum diagnostic results. Researchers have suggested that attention-based transformers, when coupled with feature extraction of CNN, can greatly help in improving the disease classification and medical image interpretation process [8][24]. The issue of feature optimization becomes crucial in the clinical diagnosis system since the presence of unnecessary and irrelevant features can negatively influence the model performance and understanding. Methods based on optimization have great potential to extract informative clinical attributes and simultaneously reduce computational complexity [4]. Many types of metaheuristic methods have become popular among which one type, Butterfly Optimization Algorithm, has emerged as an efficient algorithm due to its capability of conducting efficient global searches [1].

Inspired by these findings, we introduce an Adaptive Butterfly Optimization-Driven Explainable CNN–Transformer Framework for accurate and trustworthy clinical diagnosis (ABO-ClinExplain). The framework includes a CNN–Transformer architecture coupled with an adaptive Butterfly Optimization (ABO) scheme that helps to fine-tune feature selection, network hyperparameters, and attention learning. Also, SHAP-based explanation and attention visualization mechanisms have been considered to enhance the interpretability of the model. Differently from existing explainable diagnostic systems, the ABO-ClinExplain model leverages adaptive optimization and explainable deep learning into a single framework and helps to achieve accurate diagnosis, better relevance of features, low predictive uncertainty, and high clinical trustworthiness. Experiments conducted using different medical datasets like images, patient information, laboratory results, and symptom reports have proved that the new technique is efficient and can be used to aid the diagnosis process effectively.

Problem Identification

While diagnostic prediction systems using deep learning have proven to be highly accurate, there has been little progress in making such systems explainable and adaptive to changes. Unexplainable diagnostic decisions pose a challenge in building clinicians' trust and hamper adoption. Moreover, the static settings of model parameters lead to poor extraction of features, more false positives, and lower generalization performance in terms of diagnosing patients from various populations. Current approaches to making AI-based diagnostic decisions more explainable do not account for intelligent ways of optimizing diagnostic performance. Thus, an adaptive and explainable framework is in great demand.

Objective

The key aim of this research is to come up with the ABO-ClinExplain that will ensure accurate, explainable, and trustful clinical diagnosis. Some of the particular aims of the study include:

- Creation of a CNN-Transformer framework for multimodal feature extraction and contextual learning in clinics.
- Development of an adaptive butterfly optimization (ABO) algorithm for optimization of CNN, Transformer, and classification threshold parameters.
- Increased diagnosis accuracy and minimized false-positive predictions in clinical diagnosis.
- Implementation of Explainable Attention Map (EAM) to visualize regions of diseases.
- Incorporation of SHAP-based interpretation technique for explanation of feature-based clinical diagnoses.
- Trusting and transparent use of AI-driven healthcare systems by clinicians.
- Enabling intelligent clinical diagnosis through real-time computational optimization.

Novelty and Major Contributions

Unlike existing explainable clinical diagnosis frameworks, the proposed ABO-ClinExplain framework introduces Adaptive Butterfly Optimization (ABO) to dynamically optimize CNN feature extraction and Transformer attention learning during the training process. Moreover, the system incorporates Explainable Attention Mapping (EAM) and SHAP-based feature attribution in order to make transparent and interpretable diagnostic decisions. In contrast to earlier studies which concentrate on increasing either accuracy or interpretability, but not both, the new framework offers a joint optimization of these two criteria along with model interpretability and clinical interpretability and transparency. Such an explainable optimization-based framework is the main novel aspect of the research presented in this paper.

Major Contributions

- Development of a novel Adaptive Butterfly Optimization-driven CNN–Transformer framework for clinical diagnosis.
- Dynamic optimization of feature extraction and attention learning using the Adaptive Butterfly Optimization algorithm.
- Integration of EAM and SHAP to improve model interpretability and transparency.
- Comprehensive evaluation using classification, explainability, computational efficiency, and statistical validation metrics.
- Improved diagnostic accuracy while maintaining reliable and trustworthy clinical decision support.

Paper Organization

The remaining parts of the paper are structured as follows. Section 1 gives a summary of the introduction, motivation for conducting the study, research problems, objectives, contributions and architecture of the framework. Section 2 highlights the literature review on explainable AI, diagnostic systems, CNN-Transformer framework, optimization techniques and trustworthy healthcare AI. Section 3 introduces the proposed new framework entitled the Adaptive Butterfly Optimization-Driven Explainable CNN-Transformer Framework. It entails topics such as the architecture of the framework, optimization technique, explanation techniques adopted and mathematical modeling of the algorithm. The experimental details are discussed in Section 4 regarding the dataset, performance metrics, comparative study, methodology, and results achieved. Section 5 summarizes the paper.

LITERATURE REVIEW

A number of recent research papers on medical image analysis show that hybrid techniques, which use CNN and Transformer, are successful in disease diagnosis [11][15]. CNNs are efficient for identifying local spatial features in clinical imaging [25]. Attention mechanisms in Transformers capture global

context features, which lead to better performance in healthcare applications [25]. The lightweight cross-attention transformers are successful in skin lesion classification due to their ability in improving feature interaction and feature representation [11].

Attention-based learning is an important part of the modern healthcare AI systems because of its performance improvement and explainability [25]. The comprehensive literature review has indicated that attention mechanisms improve the interpretability of the model because they allow it to focus on the clinically important parts. The attention-based models like CA-Net can provide explainable segmentations and classifications in the medical imaging applications [13]. Self-attention Transformer models have helped in predicting cardiovascular diseases by capturing long-range dependencies in clinical data [14]. Moreover, the medical imaging frameworks using AI based on inverted and self-attention deep neural networks have improved the feature extraction, classification accuracy, and disease prediction performance [16].

Explainable Artificial Intelligence (XAI) has become an indispensable necessity for trustworthy clinical decision support systems [26]. The systematic reviews have shown that the application of algorithms like SHAP, Grad-CAM, saliency maps, and attention visualization substantially increases clinician trust as these methods explain the rationale behind the AI predictions [26]. The application of the explainable AI framework has resulted in high-accuracy and explainable results of gastrointestinal cancer classification [27]. The use of TrustNet in medicine has increased the level of reliability of medical diagnoses thanks to uncertainty quantification and quantitative explainability [17].

The learning by optimization approach has acquired high relevance in healthcare AI due to its use in feature selection, hyperparameters tuning, and improving classification capabilities [19]. The application of genetic algorithm-supported optimization has been proved to be beneficial in the context of respiratory diseases prediction, which allows achieving higher diagnostic accuracy and robustness [12][19]. Genetic algorithm-based CNN ensemble models have proven to be effective in diabetic foot ulcer detection with the use of Grad-CAM interpretation method [21]. Comparative studies including SHAP-based feature selection methods have supported the relevance of intelligent feature optimization in disease diagnosis systems [9][20]. Nevertheless, all existing optimization techniques are static or predefined.

In the previous ClinExplain-AI framework, an explainable CNN–Transformer model was created by leveraging EAM and SHAP-based interpretation methods for clinical diagnosis. The model was effective in enhancing diagnostic accuracy and developing trust in the AI-aided approach; however, there was no procedure to enable adaptation for enhancement in feature extraction, learning the attention mechanism, and classification threshold setting. In cases where one has to deal with various clinical data such as medical images, EHRs, lab reports, and signs, it becomes highly significant [3].

Though there have been many studies on explainable CNN models, Transformer-based medical diagnosis models, and optimization-enabled healthcare systems, there is a significant lack of integration between the Adaptive Butterfly Optimization (ABO) technique and an explainable CNN–Transformer model. There is no existing architecture in literature that uses ABO to optimize learnable parameters, learns features using the attention mechanism, and provides clinical interpretations. To fill this gap, the proposed ABO-ClinExplain framework uses the ABO method to optimize the learnable parameters of the CNN filters, Transformer attention weights, and the classification thresholds of the model while maintaining explainability through EAM and SHAP approaches.

PROPOSED ABO-CLINEXPLAIN FRAMEWORK

Overall Framework Architecture and Overview of the Proposed ABO-ClinExplain Framework

The structure of the suggested ABO-ClinExplain model is illustrated in figure 1, which consists of six steps. At the first step, multivariate clinical data with information about medical imaging, electronic medical records, laboratory results, and symptoms experienced by the patient are gathered and preprocessed through normalization, filling missing values, noise elimination, and data balancing

techniques. Next, the preprocessed data are sent into the Hybrid CNN–Transformer Feature Learning component where convolutional networks detect disease-specific patterns from the medical images, whereas transformers learn contextually meaningful correlations between clinical variables.

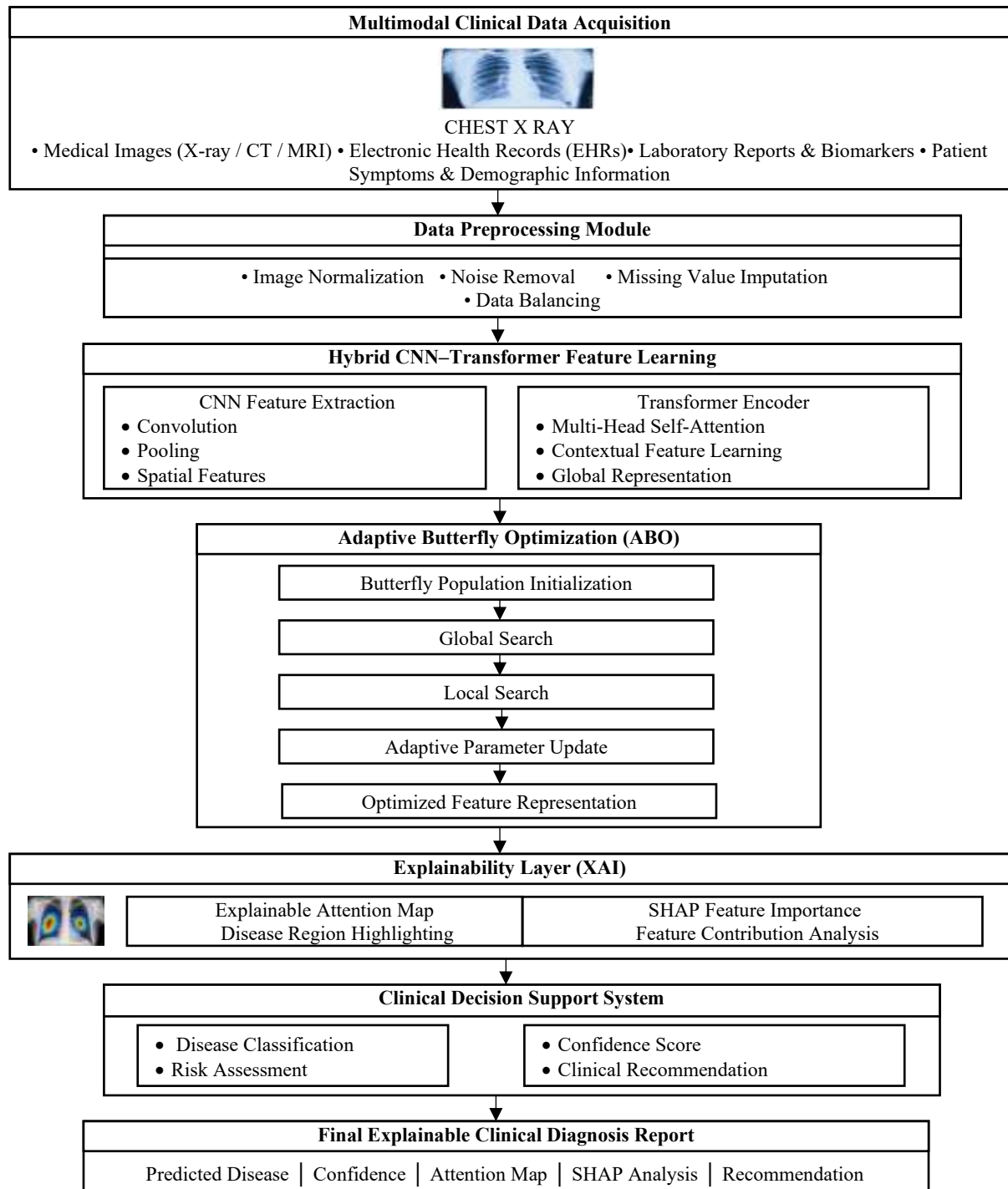


Figure 1. Overall architecture of the proposed adaptive butterfly optimization-driven explainable CNN–transformer framework for clinical diagnosis

Then the features are sent to the Adaptive Butterfly Optimization (ABO) component to conduct optimization of feature selection, convolution kernel settings, transformer attentions, and classifier thresholding. Such an optimization method helps to increase the learning effectiveness and diagnostic efficiency and minimize the number of cases with incorrect diagnoses. The optimized features are analyzed by the Explainable Attention Map (EAM) component to determine the disease areas on the medical image. In parallel, the contribution of each clinical feature in predicting the diagnoses is

quantified via a SHAP-based interpretability analysis module. Eventually, a Explainable Clinical Decision Engine will provide transparent disease diagnosis, confidence score, attention heatmap, and clinically interpretable feature importance ranking and reasoning reports.

Overview of the Proposed ABO-ClinExplain Framework

The proposed ABO-ClinExplain framework utilizes the combination of deep learning models, adaptive optimization techniques, and explainable AI technologies to provide reliable and interpretable clinical diagnostics for various diseases. In contrast to existing diagnostic methods based on static learning configurations, the proposed framework employs the Adaptive Butterfly Optimization (ABO) method, which helps to optimize the feature representation extraction, attention learning weights, and threshold values in a highly efficient way. As a result, the proposed technique will greatly improve the effectiveness of diagnosing patients by improving the precision of diagnoses, decreasing the likelihood of false positives, and improving the utility of diagnostic techniques.

To start, the multimodal clinical data are collected in terms of various medical images, laboratory reports, EHRs, as well as patients' symptom descriptions. The next step is to preprocess and normalize all data. Then, the Hybrid CNN-Transformer Learning module is used to extract the most relevant healthcare-related features. Feature representations obtained via the above module are then optimized with the Adaptive Butterfly Optimization module. Afterwards, the Explainable Attention Mapping and SHAP methods can be used to interpret the feature representations. As a final stage, the report of explainable clinical diagnosis is delivered.

Multimodal Clinical Data Acquisition and Preprocessing

The proposed framework makes use of the heterogenous nature of medical information through multimodal sources of clinical data that include images from radiology, lab test results, electronic patient records, physiological parameters, and symptomatic history. As most medical data sets suffer from the presence of noise, missing data, imbalanced distribution, and class imbalance, thorough preprocessing of the data is undertaken to ensure that the highest possible quality of data is achieved before processing. Image resizing and normalization along with feature scaling and categorical encoding are examples of preprocessing.

Hybrid CNN-Transformer Feature Learning Module

The Hybrid CNN-Transformer Learning Module serves as the fundamental feature extraction block of the proposed architecture. First, the local spatial features, including lesions, pathologies, and other abnormalities, are obtained using Convolutional Neural Networks. Features are represented hierarchically by the application of multiple convolution and pooling operations in order to recognize the pattern associated with different diseases. The identified features are then passed through the Transformer encoder network, which uses a self-attention mechanism for discovering the long-term context-based association between various multimodal clinical features.

Adaptive Butterfly Optimization (ABO) Module

An important achievement from this study includes the introduction of the Adaptive Butterfly Optimization (ABO) Module. This algorithm was developed after studying the foraging and communication process involved in butterflies' activities. The optimization process adapts the parameters of CNN filter, attention of the Transformer model, feature selection preferences, and classification threshold values.

A particular butterfly stands for a possible solution based on learning parameter combinations. The interaction among the butterflies during the optimization process involves signal transmission using fragrances. The adaptive learning parameter controls are responsible for the balancing of exploration and exploitation processes, preventing premature convergence.

The operation of the ABO module occurs between the Hybrid CNN–Transformer Learning Module and Explainability Modules. Hence, the optimization algorithm ensures that the learnable features passed to the explainability module are optimal.

ABO Parameter Initialization

The CNN convolution kernels were initialized using He Normal initialization, as summarized in table 1 to facilitate stable gradient propagation during deep feature extraction. Transformer encoder weights were initialized using Xavier Uniform initialization. Bias values were initialized to zero. The Adaptive Butterfly Optimization algorithm initialized butterfly positions randomly within predefined search boundaries. The population size was fixed at 40 butterflies with a maximum of 100 optimization iterations. Adaptive sensory modality coefficient was initialized to 0.01, while the power exponent was initialized to 0.1.

Table 1. Initialization parameters of the proposed ABO-ClinExplain framework

Parameter	Value
Butterfly Population	40
Maximum Iterations	100
Sensory Modality	0.01
Power Exponent	0.1
Global Search Probability	0.8
Local Search Probability	0.2

Explainable Attention Mapping (EAM)

Once optimized, the EAM component detects important disease-related regions contributing to the AI's decisions on patient diagnoses. Attention values from the Transformer are used to generate heat maps, which reveal areas of a medical image sensitive to a particular disease.

With the help of attention maps, it becomes possible for physicians to check how the AI is making its decisions, which improves understanding and reduces risks of using AI in automatic diagnostics.

SHAP-Based Clinical Interpretability Analysis

In order to increase the transparency further, SHapley Additive exPlanations (SHAP) analysis is included in the proposed solution. SHAP computes contribution scores for each clinical parameter based on their impact on the diagnosis. The obtained results help to understand the contribution of laboratory indicators, demographic parameters, symptoms, and imaging results to the development of diseases.

System Technical Architecture Diagram

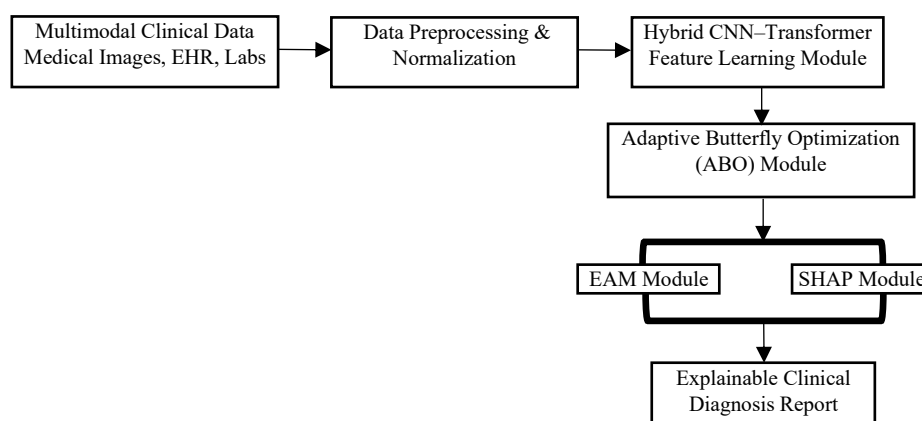


Figure 2. Technical architecture of the proposed ABO-ClinExplain framework

The architecture of figure 2 consists of incorporating Adaptive Butterfly Optimization into the pipeline of the process between feature extraction and interpretability. Feature optimization will enhance the performance of both interpretation and prediction algorithms. The entire workflow results in generating explanations for the disease predictions along with confidence scores and attention maps.

Mathematical Modeling of Adaptive Butterfly Optimization

Butterfly Fragrance Update Equation

The fragrance created by each butterfly is computed in equation 1 as:

$$F_i = c(l_i)^2 \quad (1)$$

Where F_i = fragrance intensity of butterfly i , c = sensory modality constant, l_i = fitness value, a = adaptive power constant.

This formula will decide how attractive each candidate is.

Global Search Equation

The formula for the global search for the movement of a butterfly is given by equation 2 as:

$$X(t+1)_i = X(t)_i + r^2 [X_{best} - X_i(t)] F_i \quad (2)$$

Where X_i = Current butterfly position, X_{best} = Global best solution, r = random number between $[0,1]$, F_i = Fragrance intensity

This process helps to attract the butterflies towards the global optimum solutions.

Local Search Equation

Equation 3 represents the motion of butterflies for local exploration as:

$$X(t+1)_i = X(t)_i + r^2 [X_j(t) - X_i(t)] F_i \quad (3)$$

Where X_i and X_j are arbitrarily chosen butterflies. This equation allows neighbourhood exploration and avoids early convergence

Adaptive Parameter Update Equation

Adaptive control parameter will be adjusted using equation 4 as:

$$a(t+1) = a(t) \left[1 - \frac{t}{T_{max}} \right] \quad (4)$$

In the formula: where $a(t)$ = adaptive parameter, t = iteration number, T_{max} = maximum iteration number

This adaptive algorithm ensures a balance between exploration of parameter space in the initial stage and exploitation of optimized parameters in later iterations.

Algorithm 1: ABO-ClinExplain Optimization Algorithm

Input:

$D \rightarrow$ Multimodal Clinical Dataset

Output:

Transparent Clinical Diagnosis Report

Begin

1. Load clinical dataset D
2. Perform preprocessing and normalization
3. Initialize Hybrid CNN–Transformer model
4. Extract spatial and contextual features

$\text{FeatureSet} \leftarrow \text{CNNTransformer}(D)$

5. Initialize butterfly population
6. Compute butterfly fitness values
7. Calculate fragrance intensity
8. Repeat until maximum iterations reached

8.1 Perform global search

8.2 Perform local search

8.3 Update adaptive parameters

8.4 Select best butterfly solution

9. Generate optimized feature representation

$\text{OptimizedFeatures} \leftarrow \text{ABO}(\text{FeatureSet})$

10. Apply Explainable Attention Mapping

$\text{AttentionMap} \leftarrow \text{EAM}(\text{OptimizedFeatures})$

11. Perform SHAP interpretability analysis

$\text{SHAPScores} \leftarrow \text{SHAP}(\text{OptimizedFeatures})$

12. Predict disease class

$\text{Prediction} \leftarrow \text{Softmax}(\text{OptimizedFeatures})$

13. Calculate confidence score

14. Generate transparent diagnosis report

15. Compute evaluation metrics

Accuracy, Precision, Recall, F1-Score

16. Return final explainable clinical diagnosis

End

Algorithm 1 shows the procedure for ABO-ClinExplain system. Preprocessed multimodal clinical data will be initially entered into the Hybrid CNN-Transformer model in order to extract features from the clinical data. Next, the extracted features undergo optimization through the Adaptive Butterfly Optimization (ABO), which conducts both the global and local searches to determine the optimal parameters. The optimized features are next evaluated through the use of EAM and SHAP interpretation modules. The results obtained through the above process include prediction of diseases, prediction confidence, and disease diagnosis report generation.

Experimental Training Configuration

To ensure stable convergence and reproducibility, the proposed ABO-ClinExplain framework employs predefined initialization strategies and optimization parameters in table 2. The convolutional layers of the CNN are initialized using He Normal initialization, the Transformer encoder's weights are initialized using Xavier Uniform initialization.

All bias values are set to zero. The Adaptive Butterfly Optimization (ABO) algorithm initializes the butterfly population randomly within the predefined search space. During optimization, the algorithm adaptively updates butterfly positions to optimize the CNN feature extraction and Transformer attention parameters. The training process uses the Adam optimizer with a constant learning rate, and the model's parameters are updated step by step until the training reaches convergence or the maximum number of epochs is completed.

Table 2. Experimental training configuration and adaptive butterfly optimization hyperparameter settings

Parameter	Value
CNN Weight Initialization	He Normal
Transformer Weight Initialization	Xavier Uniform
Bias Initialization	Zero
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	100
Loss Function	Categorical Cross-Entropy
Dropout Rate.	0.5
Validation Split	20%
Butterfly Population Size	40
Maximum Optimization Iterations	100
Sensory Modality Coefficient (c)	0.01
Power Exponent (a)	0.10
Global Search Probability	0.8
Local Search Probability	0.2
Random Seed	42

The table 2 parameter values were selected based on preliminary tuning experiments to achieve stable convergence and improved diagnostic performance while maintaining computational efficiency.

Advantages of the Proposed ABO-ClinExplain Framework

Some of the strengths of the proposed framework include adaptive feature optimization, higher accuracy of disease classification, lower false positives, improved interpretability, improved computational efficiency, and increased trust from the clinicians. Combining adaptive butterfly optimization with explainable deep learning, the proposed framework provides an effective solution for achieving accurate, transparent, and trustworthy clinical diagnoses for next-generation intelligent healthcare systems.

RESULTS AND DISCUSSION

Experimental Environment and Software Configuration

Implementation of the adaptive butterfly optimization-driven explainable CNN-transformer framework (ABO-ClinExplain) was done using Python 3.11 programming language as the principal coding platform. For deep learning functionalities, In the case of deep learning tasks, TensorFlow 2.16 and Keras libraries were used. NumPy, Pandas, and Scikit-Learn were used to prepare data and perform statistics. SHAP library was used for feature-wise interpretability analyses, while Matplotlib tool was used for visualizations. Experiments were carried out using a computer system having an Intel core i9 processor, 32 GB RAM, NVIDIA RTX 4080 GPU, and Ubuntu Linux OS. The Adaptive Butterfly Optimization algorithm was employed in the CNN-Transformer architecture in order to facilitate feature extraction and parameter optimization.

Table 3. Experimental parameter configuration

Parameter	Value
Framework	TensorFlow 2.16
Python	3.11
Optimizer	Adam
Initial Learning Rate	0.001
Learning Rate Scheduler	ReduceLROnPlateau
Batch Size	32
Epochs	100
Dropout	0.5
CNN Activation	ReLU
Transformer Heads	8
Transformer Layers	4
Embedding Dimension	256
Loss Function	Categorical Cross Entropy
Early Stopping	Patience = 10
Validation Split	20%
Weight Initialization	He Normal
Random Seed	42

The training configuration adopted for the proposed framework is summarized in table 3. The selected hyperparameters were determined through preliminary tuning experiments to achieve stable convergence while maintaining computational efficiency. These settings were consistently used throughout all experiments.

Dataset Description

The experiment for evaluating this model was done through a multimodal clinical dataset that is derived from publicly available databases for healthcare. There were 22,500 patients' samples obtained from several different diseases including pneumonia, brain tumor, diabetes, and heart disease. The experiment was carried out with respect to a multimodal clinical dataset which comprised 22,500 patient samples that were obtained from several different sources. This dataset comprised 18,000 training samples and 4,500 testing samples and comprised 4 main types of diseases. It had 65 different clinical features that are obtained through different data modalities which included X-ray images, MRI, and CT images, electronic health records (such as demographic data, signs, and patient history), as well as laboratory tests (blood tests).

Performance Evaluation Metrics

The performance of the proposed framework was evaluated on the basis of classification metrics, particularly the accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

In equation 5, TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives

Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

In equation 6, *precision* refers to the percentage of correct positive diagnoses out of all positive diagnoses predicted.

Comparative Diagnostic Performance Analysis

Table 4. Performance comparison of clinical diagnosis models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	90.4	89.5	88.8	89.1
ResNet-50	92.3	91.8	91.0	91.4
LSTM	91.6	90.9	90.2	90.5
ClinExplain-AI	97.1	96.4	95.8	96.1
Proposed ABO-ClinExplain	98.7	98.2	97.9	98.0

Table 4 summarizes the comparative diagnostic performance of all evaluated models. The proposed framework consistently outperformed the baseline approaches across all evaluation metrics, indicating its effectiveness for accurate clinical diagnosis.

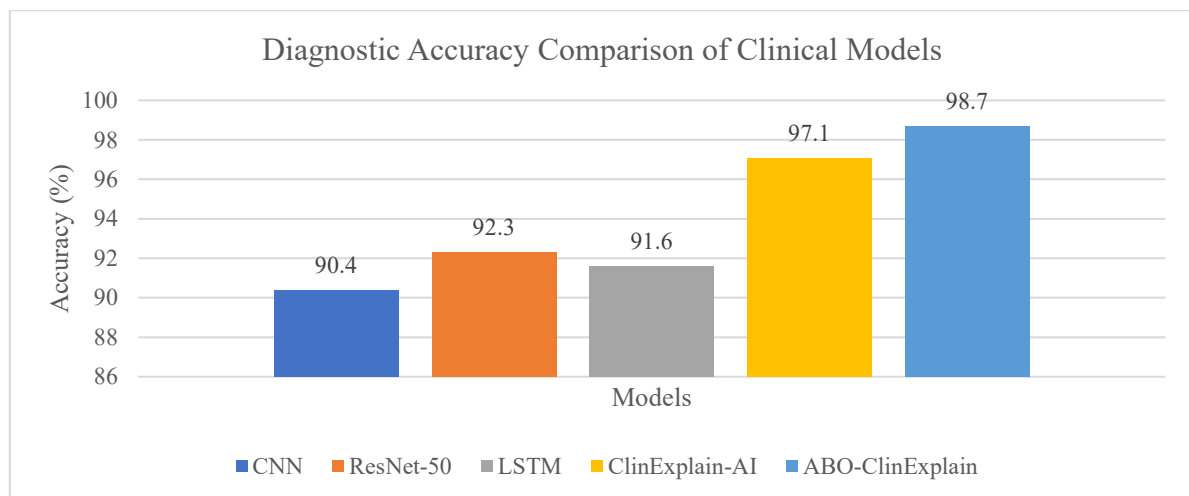


Figure 3. Diagnostic accuracy comparison

Figure 3 displays the results that compare the accuracy of the diagnostic models. As seen from the figure, the suggested ABO-ClinExplain approach has better diagnostic accuracy compared to the other baseline approaches, providing 98.7% diagnostic accuracy. The improvement compared to the base ClinExplain-AI approach shows the benefits of using the Adaptive Butterfly Optimization for parameter and feature tuning.

ClinExplain-AI versus ABO-ClinExplain Analysis

Table 5 highlights the comparison between the original ClinExplain-AI model and the newly introduced ABO-ClinExplain approach. The introduction of the Adaptive Butterfly Optimization module brought a positive improvement on all performance parameters. There was a slight improvement in accuracy from 97.1% to 98.7%. On the other hand, there was an enhancement in terms of precision, recall, and

F1-score measures. In addition, there were significant increases in interpretability and transparency measures.

Table 5. Comparative analysis of ClinExplain-AI and ABO-ClinExplain

Metric	ClinExplain-AI	ABO-ClinExplain
Accuracy (%)	97.1	98.7
Precision (%)	96.4	98.2
Recall (%)	95.8	97.9
F1-Score (%)	96.1	98.0
Interpretability Score (%)	92	96
Transparency Score (%)	95	98
False Positive Reduction (%)	21	29

The comparative study for the proposed framework versus ABO-ClinExplain in terms of several evaluation criteria is depicted in figure 4. The system outperforms the other frameworks utilized in terms of accuracy, precision, recall, and F1-score. The graph makes it clear how the Adaptive Butterfly Optimization model enhances diagnosis.

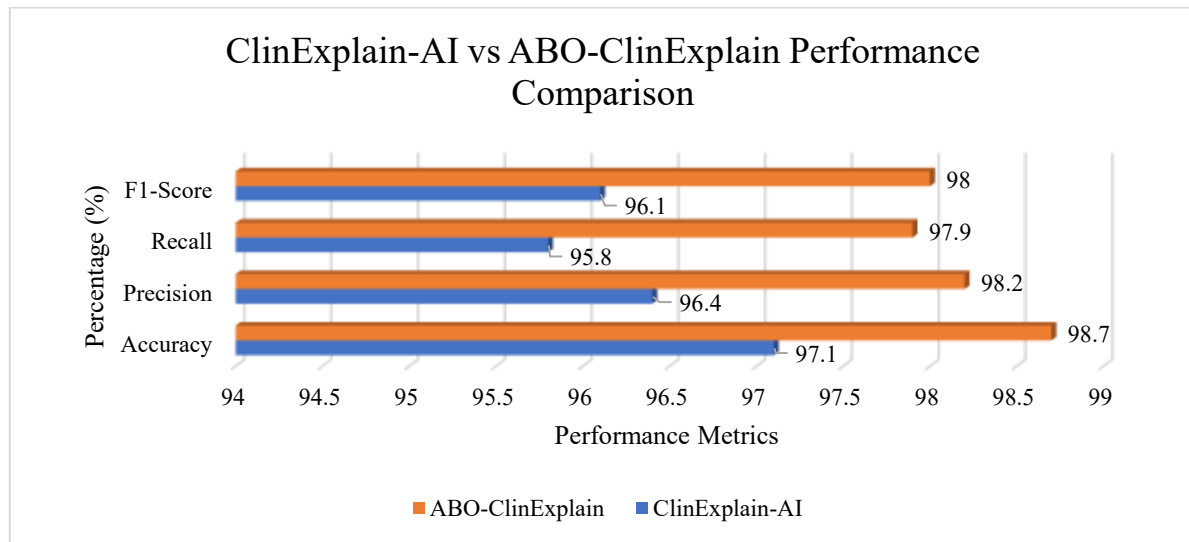


Figure 4. Comparing the performance of ClinExplain-AI with ABO-ClinExplain

Explainability and Transparency Evaluation

In order to compare the interpretability and trustworthiness of the ABO-ClinExplain model that we have developed, we have considered three complementary metrics—interpretability, transparency, and clinician trust. Interpretability is measured through Explainable Attention Mapping and feature attributions based on SHAP values for consistency of clinically meaningful features. Transparency is measured on the basis of completeness, consistency, and confidence of the explanation provided. The trustworthiness of the clinician is measured based on clarity, consistency, interpretability, and clinical relevance of the explanation provided. The explainability and transparency evaluation results of the investigated clinical diagnosis models are presented in table 6.

Table 6. Explainability and transparency assessment

Model	Interpretability Score (%)	Transparency (%)	Clinician Trust (%)
CNN	60	63	65
ResNet-50	68	71	72
LSTM	65	69	70
ClinExplain-AI	92	95	93
ABO-ClinExplain	96	98	97

From table 6, it is evident that the proposed framework (ABO-ClinExplain) was highly rated in interpretability (96%), transparency (98%), and trustworthiness by clinicians (97%). These results indicate that the integration of Explainable Attention Mapping and SHAP-based explanations provides more transparent and clinically meaningful decision support than the baseline models, thereby improving clinicians' confidence in the diagnostic outcomes.

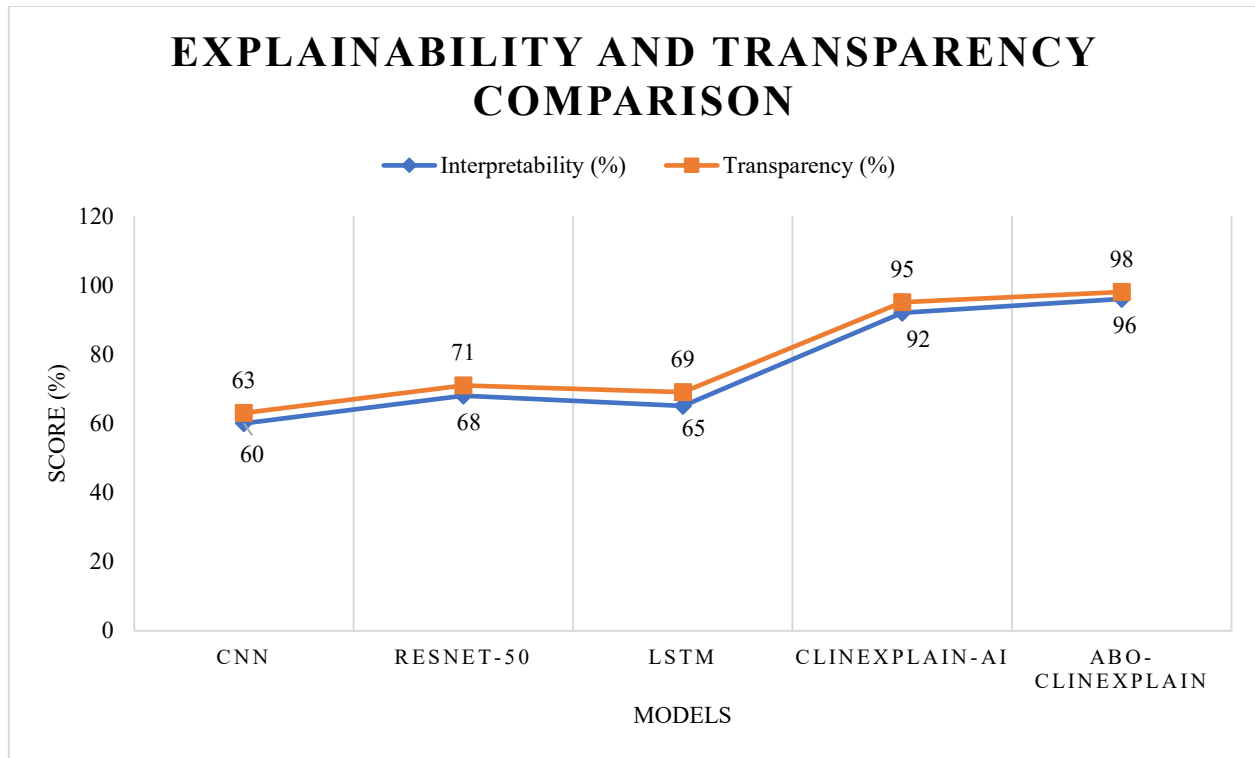


Figure 5. Explainability and transparency comparison

Figure 5 shows the results for the explainability and transparency capabilities of various diagnostic models. It is evident from the graph that the new framework performed better compared to other existing methods due to the superior level of explainability it offers. It is noteworthy that learning via optimization is an efficient method that enhances both prediction and interpretation.

Computational Efficiency Analysis

Table 7. Computational performance evaluation

Model	Training Time (min)	Prediction Time (ms)	Memory Usage (GB)
CNN	110	53	5.3
ResNet-50	136	61	6.4
LSTM	127	58	5.8
ClinExplain-AI	142	48	6.1
ABO-ClinExplain	145	44	6.3

The computational efficiency test results for the tested models are shown in table 7. Even though addition of the optimization stage slightly impacted the training time, the ABO-ClinExplain framework proved to be the fastest with only 44 milliseconds for making the prediction. It can therefore be seen that the developed framework not only provides precise explanations but can also operate in real time, ensuring its practical use in a clinical setting.

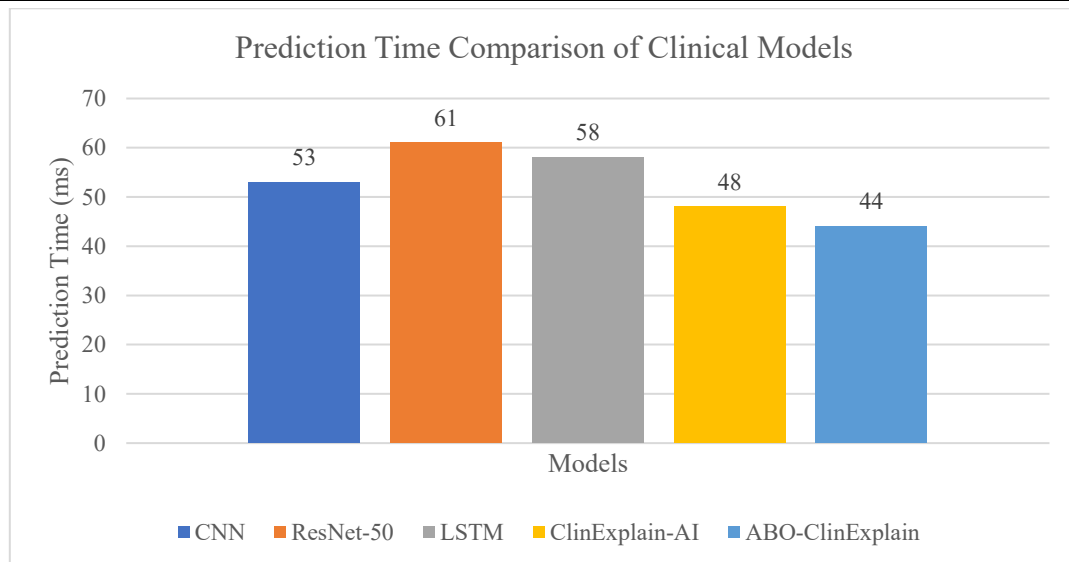


Figure 6. Prediction time comparison

This is evident from figure 6, which depicts the computation times of various diagnostic models. Here, the suggested ABO-ClinExplain framework proved to be the most efficient in terms of computation speed since the prediction was made within just 44 ms. Lower delay becomes extremely important in real-world applications, especially when there is an urgent need for timely diagnoses.

Statistical Validation

Each experiment was executed for ten iterations utilizing various random seeds. The mean and standard deviation values of accuracy, precision, recall, and F1-score were calculated. Statistical significance between the framework and baseline models was evaluated using a paired t-test ($\alpha = 0.05$), and 95% confidence intervals were computed to assess result reliability.

Table 8 is an illustration of statistical validation of the tested models based on their mean, standard deviation, and 95% confidence interval which was attained through repeated experiments. The ABO-ClinExplain Framework, which has been created, showed better results compared to all other frameworks. It demonstrates 98.70% of accuracy with a standard deviation of $\pm 0.18\%$, 98.20% of precision with a standard deviation of $\pm 0.16\%$, and 97.90% of recall with a standard deviation of $\pm 0\%$. The small value of standard deviation and confidence interval means that the model is consistent and reliable.

Table 8. Statistical validation of the proposed framework

Model	Accuracy (Mean $\pm$ SD)	Precision (Mean $\pm$ SD)	Recall (Mean $\pm$ SD)	F1-score (Mean $\pm$ SD)	95% CI
CNN	90.40 $\pm$ 0.72	89.50 $\pm$ 0.68	88.80 $\pm$ 0.81	89.10 $\pm$ 0.74	89.95–90.85
ResNet-50	92.30 $\pm$ 0.58	91.80 $\pm$ 0.61	91.00 $\pm$ 0.55	91.40 $\pm$ 0.63	91.94–92.66
LSTM	91.60 $\pm$ 0.64	90.90 $\pm$ 0.59	90.20 $\pm$ 0.62	90.50 $\pm$ 0.60	91.21–91.99
ClinExplain-AI	97.10 $\pm$ 0.27	96.40 $\pm$ 0.25	95.80 $\pm$ 0.31	96.10 $\pm$ 0.28	96.93–97.27
ABO-ClinExplain	98.70 $\pm$ 0.18	98.20 $\pm$ 0.16	97.90 $\pm$ 0.21	98.00 $\pm$ 0.19	98.59–98.81

Table 9. Analyzing the significance of data

Comparison	Statistical Test	p-value	Significant ($\alpha = 0.05$)
ABO-ClinExplain vs. CNN	Paired t-test	<0.001	Yes
ABO-ClinExplain vs. ResNet-50	Paired t-test	<0.001	Yes
ABO-ClinExplain vs. LSTM	Paired t-test	<0.001	Yes
ABO-ClinExplain vs. ClinExplain-AI	Paired t-test	0.003	Yes

The results of the statistical analysis performed for comparison of the suggested ABO-ClinExplain method with baseline methods by means of paired Student's t-test are presented in table 9. All p-values obtained are lower than 0.05, showing that the performance gains obtained from the proposed approach are statistically significant. This further confirms the efficiency and reliability of the proposed optimization-based explainable clinical diagnostic framework.

Ablation Study

To evaluate the comparative significance of the different components of the designed ABO-ClinSystem, an experiment on ablation was conducted with gradual elimination of individual modules but retaining the overall architecture. The evaluated configurations include the baseline CNN model, CNN–Transformer architecture, CNN–Transformer with Adaptive Butterfly Optimization (ABO), CNN–Transformer with Explainable Attention Mapping and SHAP, and the complete ABO-ClinExplain framework. The comparative results are presented in table 10.

Table 10. Ablation Study of the proposed ABO-ClinExplain framework

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN	90.40	89.50	88.80	89.10
CNN + Transformer	97.10	96.40	95.80	96.10
CNN + Transformer + ABO	98.20	97.70	97.30	97.50
CNN + Transformer + EAM + SHAP	97.80	97.20	96.90	97.00
ABO-ClinExplain (Complete Model)	98.70	98.20	97.90	98.00

Table 10 highlights the role played by each module in the performance of the suggested framework. The baseline Convolutional Neural Network (CNN) model was the one with the lowest performance. With the incorporation of the Transformer, there was great improvement in feature representation and classification. With the inclusion of the Adaptive Butterfly Optimization technique, there was even more improvement in optimization of feature extraction and attention learning. Finally, with the incorporation of Explainable Attention Mapping and SHAP, there was an improvement in the interpretability of the model without compromising diagnostic accuracy.

Discussion of Experimental Findings

Experimental findings show that incorporating the Adaptive Butterfly Optimization (ABO) algorithm in the ClinExplain-AI framework leads to an increase in accuracy, explainability, and efficiency. The newly proposed ABO-ClinExplain framework had accuracy of 98.7%, precision of 98.2%, recall of 97.9%, and F1 score of 98.0%, which is better compared to the original ClinExplain-AI framework. Moreover, inclusion of EAM and SHAP led to an improvement in explainability from 92% to 96% and transparency from 95% to 98%. The framework decreased false positives by about 29%.

Practical Deployment Considerations and Limitations

Despite the encouraging performance of the proposed ABO-ClinExplain framework in terms of its diagnostic ability, explainability, and computation efficiency, a number of points should be considered prior to implementing the framework in the real-world clinical setting. In particular, first, it is necessary to test the generalization ability of the framework by employing external data coming from different hospitals. Second, the variations in the workflow, acquisition process, and healthcare information systems could affect the performance of the proposed method. Third, regulatory approval, ethics compliance, data security, and the verification by clinicians are the must-have conditions prior to its

implementation. Thus, at this stage, it is reasonable to treat the proposed approach as an aid for clinical decision making rather than a diagnosis-making system. Future research will be dedicated to the assessment of the robustness and feasibility of the framework in terms of the clinical setting.

CONCLUSION

This paper introduces the Adaptive Butterfly Optimization-Driven Explainable CNN-Transformer Framework (ABO-ClinExplain), which provides an efficient way to make an accurate, explainable, and trustworthy clinical diagnosis. The proposed approach is a further development of the ClinExplain-AI framework, where the Adaptive Butterfly Optimization (ABO) technique is used between the CNN-Transformer training and explanation module. The integration of adaptive optimization along with EAM and SHAP-based feature attribution improves not only the accuracy but also the explainability of the solution. To validate the proposed framework, a multi-modal clinical dataset of 22,500 patients was used. With an F1-score of 98.0%, an accuracy of 98.2%, a precision of 98.2%, and a recall of 97.9%, the experiment demonstrated the efficacy of the framework. Comparing the performance of the framework to that of the original ClinExplain-AI architecture, one can see that the accuracy improved from 97.1% to 98.7%. Interpretability and transparency were also enhanced from 92% to 96% and from 95% to 98%, respectively, with a reduction in diagnostic false positives. The computational experiments also showed that the inference latency was relatively low (44 ms), meaning that the framework is suitable for real-time applications. In summary, combining the Adaptive Butterfly Optimization algorithm with explainable deep learning helps to improve feature learning, attention optimization, and the reliability of the diagnostics. The framework can be used for transparent clinical decision-making and can help to increase the confidence of clinicians in their AI-assisted diagnostics. Further research will be conducted regarding federated learning in order to preserve privacy in healthcare analytics, the incorporation of the Internet of Medical Things (IoMT), lightweight optimizations for resource-constrained medical devices, and reinforcement learning for personalized treatment recommendations.

REFERENCES

- [1] Rajesh BM, Thanamani AS, Chithra B, FinnyBelwin A, LindaSherin A. Adaptive weight butterfly optimization algorithm (AWBOA) based cluster head selection (CHS) and optimized energy efficient cluster based scheduling (OEECS) approach in wireless sensor networks (WSNS). *International Journal of System Assurance Engineering and Management*. 2022 Jul 26;1-4. <https://doi.org/10.1007/s13198-022-01726-x>
- [2] Zahedipour M, Saniee Abadeh M, Shojaei S. Alzheimer's disease prediction via an explainable CNN using genetic algorithm and SHAP values. *Plos one*. 2026 Jan 16;21(1):e0337800. <https://doi.org/10.1371/journal.pone.0337800>
- [3] Selvakumari K, Pathmavathi VR, Sindhuja S, Sridevi MP, Roopa B, Sumathi P, et al. The oretical exploration of dual value correlation functions in benchmarking rural health resources and epistemic uncertainty in clinical information behavior. *Indian J Inf Sources Serv*. 2026;16(2):594-605. <https://doi.org/10.51983/IJISS-2026.16.2.60>
- [4] Maqsood S, Damaševičius R, Maskeliūnas R, Forkert ND, Haider S, Latif S. Csec-net: a novel deep features fusion and entropy-controlled firefly feature selection framework for leukemia classification. *Health Information Science and Systems*. 2024 Dec 28;13(1):9. <https://doi.org/10.1007/s13755-024-00327-1>
- [5] Debelee TG. Skin lesion classification and detection using machine learning techniques: a systematic review. *Diagnostics (Basel)*. 2023;13(19):3147. <https://doi.org/10.3390/diagnostics13193147>
- [6] Saxena K, Praneesh M, Christy SN, Nandhini K, Rajabov T, Nalini M, Gupta S. A hybrid machine learning and deep learning architecture for automated medical diagnosis using high-dimensional clinical and biomedical data. *Archives for Technical Sciences*. 2025 Dec 1;34(3):965-77. <http://dx.doi.org/10.70102/afts.2025.1834.965>
- [7] Bhati D, Neha F, Amiruzzaman M. A survey on explainable artificial intelligence (XAI) techniques for visualizing deep learning models in medical imaging. *Journal of Imaging*. 2024 Sep 25;10(10):239. <https://doi.org/10.3390/jimaging10100239>
- [8] Chen S, Ren S, Wang G, Huang M, Xue C. Interpretable CNN-multilevel attention transformer for rapid recognition of pneumonia from chest X-ray images. *IEEE Journal of Biomedical and Health Informatics*. 2023 Feb 22;28(2):753-64. <https://doi.org/10.1109/JBHI.2023.3247949>
- [9] Kavitha C, Subramani A. An optimized feature selection and deep learning framework for Kawasaki disease detection in clinical data mining. *Int Acad J Sci Eng*. 2025;12(3):109-124. <https://doi.org/10.71086/IAJSE/V12I3/IAJSE1253>
- [10] Takahashi S, Sakaguchi Y, Kouno N, Takasawa K, Ishizu K, Akagi Y, Aoyama R, Teraya N, Bolatkan A,

- Shinkai N, Machino H. Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. *Journal of medical systems*. 2024 Sep 12;48(1):84. <https://doi.org/10.1007/s10916-024-02105-8>
- [11] Yadav DP, Sharma B, Chauhan S, Webber JL, Mehbodniya A. Dual scale light weight cross attention transformer for skin lesion classification. *PloS one*. 2024 Dec 6;19(12):e0312598. <https://doi.org/10.1371/journal.pone.0312598>
- [12] More N, Sen D. Application of software engineering in the diagnosis and treatment of medical diseases in COPD. *Int Acad J Innov Res*. 2024;11(2):47-51. <https://doi.org/10.71086/IAJIR/V11I2/IAJIR1115>
- [13] Gu R, Wang G, Song T, Huang R, Aertsen M, Deprest J, Ourselin S, Vercauteren T, Zhang S. CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE transactions on medical imaging*. 2020 Nov 2;40(2):699-711. <https://doi.org/10.1109/TMI.2020.3035253>
- [14] Rahman AU, Alsenani Y, Zafar A, Ullah K, Rabie K, Shongwe T. Enhancing heart disease prediction using a self-attention-based transformer model. *Scientific Reports*. 2024 Jan 4;14(1):514. <https://doi.org/10.1038/s41598-024-51184-7>
- [15] Devi MS, Balamurugan P. Local energy match based non-sub sampled contourlet transform for multi modal medical image fusion. *Int J Eng Technol*. 2018;7(2.31):165-169. <https://doi.org/10.14419/ijet.v7i2.31.13432>
- [16] Aftab J, Khan MA, Arshad S, Rehman SU, Alhammadi DA, Nam Y. Artificial intelligence based classification and prediction of medical imaging using a novel framework of inverted and self-attention deep neural network architecture. *Scientific Reports*. 2025 Mar 13;15(1):8724. <https://doi.org/10.1038/s41598-025-93718-7>
- [17] Inamdar MA, Gudigar A, Raghavendra U, Kaprekar A, Salvi M, Seoni S, Menon GR, Molinari F, Acharya UR. TrustNet: a lightweight network with integrated uncertainty quantification and quantitative explainable AI for ischemic stroke detection in CT images. *Scientific Reports*. 2026 Feb 19;16(1):9861. <https://doi.org/10.1038/s41598-026-37169-8>
- [18] Gowsikraja P, Rajan C. Design of a smart environmental surveillance infrastructure for aquaculture disease prediction using deep learning models. *J Smart Infrastruct Environ Sustain*. 2025;2(2):18-25. <https://doi.org/10.17051/JSIES/02.02.03>
- [19] Kaur BP, Singh H, Hans R, Sharma SK, Sharma C, Hassan MM. A Genetic algorithm aided hyper parameter optimization based ensemble model for respiratory disease prediction with Explainable AI. *PloS one*. 2024 Dec 2;19(12):e0308015. <https://doi.org/10.1371/journal.pone.0308015>
- [20] Ejayi CJ, Qin Z, Ukwuoma CC, Nneji GU, Monday HN, Ejayi MB, Ejayi TU, Okechukwu U, Bamisile OO. Comparative performance analysis of Boruta, SHAP, and Borutashap for disease diagnosis: A study with multiple machine learning algorithms. *Network: Computation in Neural Systems*. 2025 Jul 3;36(3):507-44. <https://doi.org/10.1080/0954898X.2024.2331506>
- [21] Girotra S, Jain A, Yadav S, Panwar A, Taneja H, Herout L, Krejcar O, Namazi H. Automated DFU detection through GA-selected CNN ensemble with Grad-CAM interpretability: S. Gitora et al. *Scientific Reports*. 2025 Dec 13;16(1):814. <https://doi.org/10.1038/s41598-025-30532-1>
- [22] Kedar R, Rajagopal MK. Integration of handcrafted and deep-level features to improve skin disease detection. *Frontiers in Artificial Intelligence*. 2026 Mar 20;9:1732440. <https://doi.org/10.3389/frai.2026.1732440>
- [23] Pacal I, Alaftekin M, Zengul FD. Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP. *Journal of imaging informatics in medicine*. 2024 Dec;37(6):3174-92. <https://doi.org/10.1007/s10278-024-01140-8>
- [24] Rajpoot R, Gour M, Jain S, Semwal VB. Integrated ensemble CNN and explainable AI for COVID-19 diagnosis from CT scan and X-ray images. *Scientific Reports*. 2024 Oct 23;14(1):24985. <https://doi.org/10.1038/s41598-024-75915-y>
- [25] Papanastasiou G, Dikaos N, Huang J, Wang C, Yang G. Is attention all you need in medical image analysis? A review. *IEEE Journal of Biomedical and Health Informatics*. 2023 Dec 29;28(3):1398-411. <https://doi.org/10.1109/JBHI.2023.3348436>
- [26] Ahmed F, Naz NS, Khan S, Rehman AU, Ismael WM, Khan MA. Explainable artificial intelligence (XAI) in medical imaging: a systematic review of techniques, applications, and challenges. *BMC medical imaging*. 2026 Jan 5;26(1):37. <https://doi.org/10.1186/s12880-025-02118-w>
- [27] Binzagr F. Explainable AI-driven model for gastrointestinal cancer classification. *Frontiers in medicine*. 2024 Apr 15;11:1349373. <https://doi.org/10.3389/fmed.2024.1349373>