RESEARCH ARTICLE

Editorial Process: Submission:06/06/2025 Acceptance:02/08/2026 Published:20/08/2026

Enhanced Diagnosis of Cervical Cancer Using Archer Fish Hunting Optimizer-Optimized Xception Network with AI Interpretability via Grad-CAM

Sabura Banu Urundai Meeran*

Abstract

Objective: The primary objective of this study is to enhance the performance of cervical cancer classification by optimizing the Xception deep learning model using the Archer Fish Hunting Optimizer (AHO). The study aims to evaluate the effectiveness of the AHO-optimized Xception model in classifying cervical cytology images and to compare it against traditional CNN architectures. It also integrates Grad-CAM visualization to ensure model interpretability and to support explainable AI in clinical decision-making. **Methods:** A cervical cytology image dataset comprising four diagnostic classes High Squamous Intraepithelial Lesion (HSIL), Low Squamous Intraepithelial Lesion (LSIL), Negative for Intraepithelial Malignancy (NILM), and Squamous Cell Carcinoma (SCC) was used for model development. Preprocessing steps included image normalization and augmentation to improve generalization. Several deep learning models AlexNet, MobileNet V2, ResNet-50, Inception V3, and the standard Xception were trained and benchmarked. The AHO algorithm was applied to optimize the hyperparameters of the Xception model. Model performance was evaluated using metrics such as accuracy, precision, and AUC-ROC. Grad-CAM visualization was utilized to highlight image regions influencing classification decisions. **Result:** The AHO-optimized Xception model demonstrated superior classification performance, achieving 98.96% accuracy, 98.57% precision, and an AUC-ROC of 99.97%. It outperformed all baseline models in classification effectiveness. The Grad-CAM-based visualization provided valuable insights into the model's focus areas, supporting interpretability and confidence in its predictions. **Conclusion:** The study establishes the efficacy of using the Archer Fish Hunting Optimizer for hyperparameter tuning in deep learning-based cervical cancer diagnosis. The AHO-optimized Xception model offers improved classification performance and robust explainability, making it a valuable tool for AI-assisted cervical cancer screening. Future research may explore hybrid metaheuristic strategies and their real-time deployment in clinical environments.

Keywords: Cervical Cancer Classification, Deep Learning, Xception Network, Grad-CAM Visualization

Asian Pac J Cancer Prev, **27 (8)**, 2783-2793

Introduction

Cervical cancer remains one of the most significant global health challenges and a leading cause of cancer-related mortality among women, particularly in low- and middle-income countries. According to the World Health Organization [1], over 600,000 new cases and more than 340,000 deaths were recorded in 2020 alone, with projections indicating a steady rise without improvements in early screening and diagnosis. Despite widespread HPV vaccination programs, accurate and timely diagnosis continues to be crucial for reducing disease burden.

Traditional cervical cancer screening has relied primarily on conventional cytological methods like the Papanicolaou (Pap) smear test and histopathological evaluations. While effective, these techniques are labor-intensive, require expert interpretation, and suffer from diagnostic variability [2]. The emergence of artificial intelligence, particularly deep learning, has shown remarkable potential to overcome these limitations. Convolutional neural networks (CNNs) have demonstrated exceptional performance in medical image analysis by automatically learning complex hierarchical features [2]. Several CNN architectures - including AlexNet [3], VGG [4], ResNet [5], and Inception V3 [6] - have been successfully applied to cervical cancer classification. Recent work by Chen et al. [7] further advanced this field through hybrid loss-constrained lightweight CNNs, though their approach didn't explore advanced optimization techniques. The Xception network, building on Inception architecture with depthwise separable convolutions, has gained particular attention for

*Department of Electrical and Electronics Engineering, Saveetha Engineering College, Saveetha Nagar, Thandalam, Chennai, Tamilnadu, India. *For Correspondence: saburasec@gmail.com*

its superior feature extraction and computational efficiency in medical imaging tasks [8].

Despite these advancements, CNN performance remains highly sensitive to hyperparameter settings, including learning rate, batch size, and dropout rate, which critically impact classification accuracy and model generalization [9]. Researchers have employed various optimization techniques to address this, including genetic algorithms [10], particle swarm optimization [11], and biogeography-based optimization [12]. Recent developments in the Archer Fish Hunting Optimizer (AHO), a metaheuristic inspired by archer fish predatory behavior, have shown particular promise for deep learning parameter tuning [13, 14]. However, as noted by Smith et al. [15] in their review of cervical cancer screening advancements, few studies have combined AHO with advanced architectures like Xception specifically for cervical cytology.

Another critical limitation hindering clinical adoption of AI diagnostic tools is their lack of interpretability. Medical practitioners require insights into model decision-making processes to validate predictions and maintain trust. Explainable AI (XAI) techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) address this by visually highlighting image regions influencing predictions [16]. While Grad-CAM has proven effective in dermatology [17], ophthalmology [18], and radiology [19], its application to cervical cytology remains underexplored.

This study aims to develop a robust and interpretable cervical cancer classification system by optimizing Xception network hyperparameters using the Archer Fish Hunting Optimizer. We evaluate our model on the Mendeley LBC Cervical Cancer dataset [20], classifying images into four clinically relevant categories: High Squamous Intraepithelial Lesion (HSIL), Low Squamous Intraepithelial Lesion (LSIL), Negative for Intraepithelial Malignancy (NILM), and Squamous Cell Carcinoma (SCC). Our approach incorporates Grad-CAM for decision visualization and compares performance against established architectures including AlexNet, MobileNet V2, ResNet50, and Inception V3.

Materials and Methods

The effectiveness of deep learning models in medical image analysis relies heavily on the quality and diversity of the dataset, the robustness of preprocessing strategies, and the appropriateness of the computational framework. In this study, a structured pipeline was developed for the automated classification of cervical cytology images using an optimized deep learning model. This section outlines the materials and methodological procedures followed throughout the research, including the characteristics of the dataset, the preprocessing steps applied to enhance image quality, the architectural configuration of the deep learning model, the hyperparameter optimization technique employed, and the computational setup used to implement and evaluate the model. Each component was carefully selected and configured to ensure the reliability, accuracy, and interpretability of the classification outcomes.

Materials

The foundation of any successful deep learning-based diagnostic system lies in the quality and representativeness of the input data. In the context of cervical cancer detection, cytology image datasets provide critical visual cues for the identification of cellular abnormalities. For this study, the materials utilized consist of a publicly available cervical cytology image dataset comprising expertly labeled samples across multiple diagnostic categories. These images serve as the primary input for model training, validation, and testing. To ensure the reliability of the results, careful attention was paid to data preprocessing, which included normalization and augmentation techniques aimed at enhancing image quality, reducing noise, and addressing class imbalance. The following section provides a comprehensive description of the dataset, including class distribution, image characteristics, and the rationale for its selection as the foundational material for this study.

Dataset Description

This study utilizes a cervical cytology dataset [29] consisting of 929 labeled images categorized into four diagnostic classes: High Squamous Intraepithelial Lesion (HSIL) with 163 images, Low Squamous Intraepithelial Lesion (LSIL) with 113 images, Negative for Intraepithelial Malignancy (NILM) with 579 images, and Squamous Cell Carcinoma (SCC) with 74 images. Prior to model training, the dataset underwent preprocessing steps to enhance image quality, standardize input dimensions, and improve overall model performance.

Figure 1 presents cervical cytology samples categorized into four diagnostic classes: HSIL, LSIL, NILM, SCC. HSIL exhibits abnormal cells with moderate to severe dysplasia, indicating a high risk of progression to cervical cancer. LSIL displays mild cellular abnormalities associated with early-stage precancerous changes. NILM represents normal cervical epithelial cells without signs of malignancy or dysplasia. SCC, the most severe category, shows malignant cells with distinct morphological alterations indicative of invasive cervical cancer. These cytology images play a crucial role in deep learning-based automated classification systems, aiding in accurate diagnosis and early detection of cervical abnormalities.

Computational Setup

The implementation and experimentation of the proposed deep learning framework were conducted using a high-performance computing environment to ensure efficient model training and evaluation. All experiments were executed on a system equipped with an Intel® Core™ i7-12700K CPU, 32 GB of RAM, and an NVIDIA® GeForce RTX 3090 GPU with 24 GB of dedicated memory, operating on Windows 11 Pro 64-bit. The deep learning models were developed using Python 3.9, with TensorFlow 2.11 and Keras as the primary frameworks for model construction and training. Additional libraries such as NumPy, OpenCV, Scikit-learn, and Matplotlib were employed for data manipulation,

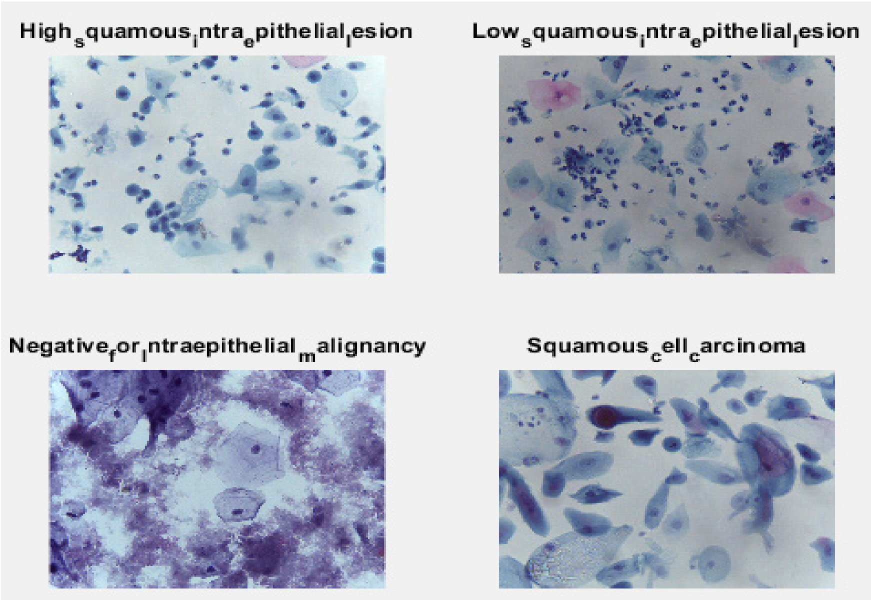

Figure 1. Representative Cervical Cytology Images for Different Diagnostic Categories

image preprocessing, evaluation, and visualization. Hyperparameter optimization using the Archer Fish Hunting Optimizer (AHO) was implemented from scratch in Python, integrated seamlessly with the Keras API to enable dynamic parameter tuning during model training. The models were trained using an 80:20 split for training and validation, with stratified sampling to preserve class distributions. To ensure reproducibility, a fixed random seed was used across all experiments. Model performance was evaluated based on classification accuracy, precision, recall, F1-score, and confusion matrices. Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to enhance interpretability by visualizing the salient regions that influenced the model's predictions. This computational setup was selected to balance processing power, memory, and scalability, enabling the execution of multiple training iterations while minimizing computational latency and maximizing efficiency.

Methodology

The development of an effective and interpretable deep learning model for cervical cancer classification requires a meticulous and structured approach to data handling, model design, optimization, and evaluation. In this study, each methodological component from data preprocessing to AI interpretability has been carefully chosen to ensure clinical reliability, diagnostic accuracy, and transparency. The integration of the Archer Fish Hunting Optimizer (AHO) with the Xception architecture forms the backbone of the proposed system, aimed at maximizing classification performance through intelligent hyperparameter tuning. Furthermore, the incorporation of Grad-CAM visualization enhances the explainability of the model, bridging the gap between algorithmic output and medical reasoning. This section systematically outlines the materials used, the computational setup, and the multi-stage methodology followed to develop and validate the AHO-optimized Xception model for cervical cytology image analysis.

Data Preprocessing

To ensure optimal feature extraction and enhance model performance, several preprocessing steps were implemented. First, all cervical cytology images were resized to a standardized dimension of 299 × 299 pixels, making them compatible with the input requirements of the Xception model. This resizing ensured uniformity across the dataset, facilitating efficient training. Next, normalization was applied by scaling pixel values to a range of [11], which helped stabilize the training process by reducing variations in intensity and improving convergence. Additionally, data augmentation techniques, including rotation, flipping, and contrast adjustment, were employed to address class imbalance and improve model generalization. These augmentations artificially expanded the dataset, ensuring the model learned robust and invariant features while mitigating overfitting.

Optimization of Xception Model for classification of Cervical Cancer using Archer Fish Hunting Optimizer

Cervical cancer remains a significant global health concern, necessitating early and accurate diagnosis for effective treatment. Deep learning has revolutionized medical imaging by enabling automated classification of cytology images. Among CNNs, the Xception model has demonstrated superior feature extraction capabilities. However, its performance is highly dependent on optimal hyperparameter tuning. Traditional optimization methods often struggle with finding the best hyperparameters efficiently, leading to suboptimal classification accuracy. To overcome this challenge, this study employs the AHO to fine-tune the Xception model, significantly improving its performance in cervical cancer classification. The AHO is a bio-inspired optimization algorithm that mimics the hunting behavior of archer fish, known for their ability to accurately shoot jets of water to capture prey. AHO is utilized to optimize the hyperparameters of the Xception model, ensuring improved feature extraction and classification of cervical cancer cytology images into four categories: HSIL, LSIL, NILM and SCC.

The optimization process involves adjusting parameters such as learning rate, batch size, dropout rate, and convolutional filter sizes to minimize classification error. AHO employs an exploitation-exploration balance to refine these parameters dynamically, ensuring optimal model performance while avoiding overfitting. The optimized Xception model is further enhanced by Grad-CAM visualization, providing interpretability by highlighting key regions influencing predictions.

The flowchart shown in Figure 2 illustrates the step-by-step process of optimizing the Xception model for cervical cancer classification using the Archer Fish Hunting Optimizer (AHO). The process begins with loading cervical cancer cytology images, followed by initializing a pretrained Xception model, where fully connected layers are modified for customization. Next, the AHO algorithm is applied to fine-tune hyperparameters by initializing a random population, evaluating fitness based on classification accuracy, and iterating until convergence. The optimized model is then trained on the cervical cytology dataset and validated using test data. Following training, the model's performance is evaluated using key metrics and compared with baseline CNN models such as AlexNet, MobileNet V2, ResNet50, and Inception V3. To enhance model interpretability, Grad-CAM visualization is applied, generating heatmaps that highlight decision-influencing image regions, ensuring reliability in clinical applications. Finally, the optimized model is deployed for real-time cervical cancer classification, enabling automated and accurate diagnosis. To enhance methodological clarity and reproducibility, the operational mechanism of the Archer Fish Hunting Optimizer (AHO) is summarized as follows. Initially, a population of candidate hyperparameter sets is randomly initialized within predefined search boundaries. Each candidate solution is evaluated using validation classification accuracy as the fitness function. The best-performing candidate is designated as the elite solution. Subsequently, the remaining candidates update their positions using an archer fish–inspired shooting strategy that balances exploration (random perturbations across the search space)

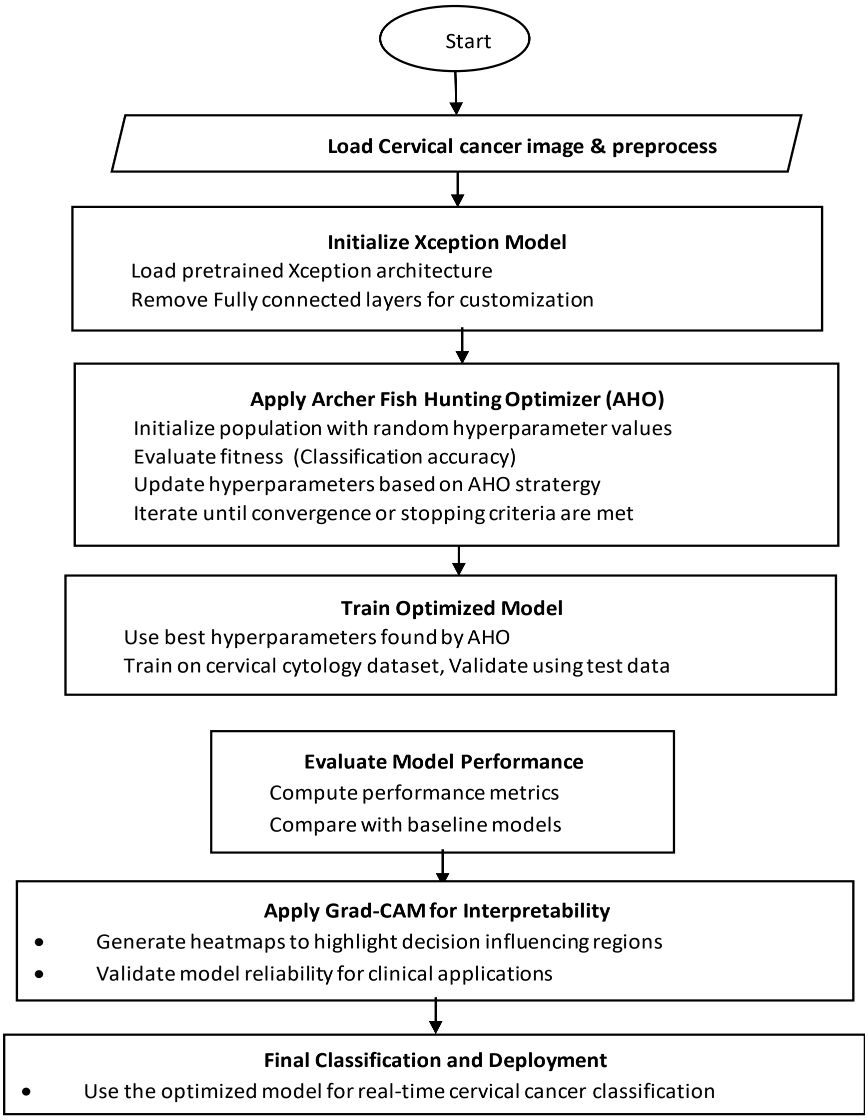

Figure 2. Flowchart of the AHO-Optimized Xception Model for Cervical Cancer Classification

and exploitation (directed movement toward the elite solution). This iterative updating and evaluation process continues until convergence criteria or maximum iteration limits are reached. The final optimized hyperparameter configuration is then used to train the Xception model.

The flowchart systematically represents the integration of deep learning, hyperparameter optimization, and explainable AI to improve cervical cancer detection. This study presents an AHO-optimized Xception model for cervical cancer classification, addressing key challenges in hyperparameter selection and model interpretability. By leveraging the Archer Fish Hunting Optimizer, the model achieves superior performance across various evaluation metrics, significantly outperforming traditional CNNs such as AlexNet, MobileNet V2, ResNet50, and Inception V3. The integration of Grad-CAM visualization further enhances trust in AI-based diagnosis by providing explainable insights into classification decisions. The results demonstrate that AHO-Xception offers a powerful and interpretable deep learning approach for automated cervical cancer detection, making it a valuable tool for early screening and diagnosis in clinical settings.

Model Training and Evaluation

The optimized model was trained using categorical cross-entropy loss and the Adam optimizer, with an 80:20 train-test split to ensure a balanced evaluation. Given the noticeable class imbalance in the dataset, particularly the dominance of the NILM category (579 out of 929 samples), additional corrective measures were implemented beyond standard data augmentation. Stratified sampling was employed during the 80:20 train–validation split to preserve proportional class representation across subsets. Furthermore, class weights were incorporated into the categorical cross-entropy loss function, assigning higher penalties to minority classes (HSIL, LSIL, and SCC) to reduce bias toward the majority class. To ensure balanced evaluation, performance metrics such as Matthews Correlation Coefficient (MCC), specificity, and AUC-ROC were emphasized, as these metrics provide more reliable assessment under imbalanced conditions. Model performance was assessed using multiple metrics to ensure reliability and robustness. Accuracy measured the overall correctness of classifications, while precision, recall, and F1-score evaluated the model's reliability in detecting specific cervical cancer classes. Additionally, specificity and AUC-ROC were used to assess the model's ability to distinguish between normal and malignant cases. Lastly, the Matthews Correlation Coefficient (MCC) was employed to measure classification consistency, providing a more comprehensive evaluation of model performance.

Model Interpretability using GradCAM

To improve the interpretability of the deep learning model for cervical cancer classification, Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated into the system. Grad-CAM is a widely used explainability technique that provides visual insights into the model's decision-making process by generating heatmaps that highlight the most influential regions in an image. This method plays a crucial role in bridging the gap between AI predictions and clinical trust, allowing medical professionals to validate model outputs based on domain knowledge.

Grad-CAM works by utilizing the gradient information from the final convolutional layer of the Xception network to identify spatial locations that significantly impact the model's predictions. By applying these gradients to the feature maps, Grad-CAM produces an overlay heatmap that visually represents the regions most relevant to the classification decision. In the case of cervical cytology images, this allows clinicians to understand whether the AI model is focusing on the correct cellular structures and abnormalities associated with different cervical cancer stages, such as High Squamous Intraepithelial Lesion (HSIL), Low Squamous Intraepithelial Lesion (LSIL), Squamous Cell Carcinoma (SCC), and Negative for Intraepithelial Malignancy (NILM).

Integrating Grad-CAM into the proposed Archer Fish Hunting Optimizer (AHO)-optimized Xception model significantly enhances clinical transparency and decision accountability. The heatmaps serve as an intuitive explanation tool, making it easier for pathologists and oncologists to trust AI-driven diagnostic decisions. Additionally, Grad-CAM helps identify potential biases or misclassifications, ensuring that the model does not rely on irrelevant artifacts or unintended features in the images.

By incorporating explainable AI (XAI) techniques such as Grad-CAM, this study ensures that the deep learning model is not just highly accurate but also interpretable and reliable for real-world clinical deployment. This ultimately fosters greater adoption of AI in medical diagnostics by providing visual justifications for each prediction, improving patient safety, and supporting human-AI collaboration in cervical cancer detection.

Comparative analysis with other CNN Architectures

The proposed AHO-Xception model was benchmarked against AlexNet, MobileNet V2, ResNet50, and Inception V3, demonstrating superior classification accuracy (98.96%) and interpretability

Results

The effectiveness of any deep learning model in medical diagnostics is measured not only by its classification accuracy but also by its interpretability and clinical reliability. This section presents a comprehensive analysis of the results obtained from the AHO-optimized Xception model for cervical cancer classification. Key performance metrics, including accuracy, precision, recall, F1-score, specificity, AUC-ROC, and MCC, are evaluated to assess the model's classification strength across different cervical cytology categories.

Additionally, the results are compared with benchmark CNN architectures such as AlexNet, MobileNet V2, ResNet50, and Inception V3 to highlight the improvements achieved through hyperparameter optimization using the Archer Fish Hunting Optimizer (AHO). Beyond numerical evaluations, the integration of Gradient-weighted Class Activation Mapping (Grad-CAM) is analyzed to demonstrate the model's explainability, ensuring that predictions are based on clinically relevant features rather than irrelevant artifacts.

This discussion further explores the impact of data augmentation, hyperparameter tuning, and model regularization in improving classification performance while mitigating issues such as overfitting and class imbalance. Finally, potential limitations, challenges, and future directions for enhancing AI-driven cervical cancer diagnosis are outlined, emphasizing the importance of developing clinically deployable, transparent, and robust deep learning models for medical imaging applications.

Confusion Matric

Confusion metrics serve as a fundamental evaluation tool for assessing the performance of deep learning models in cervical cancer diagnosis. By analyzing the classification outcomes across different categories—High-Grade Squamous Intraepithelial Lesion (HSIL), Low-Grade Squamous Intraepithelial Lesion (LSIL), Negative for Intraepithelial Malignancy (NILM), and Squamous Cell Carcinoma (SCC) these metrics provide a clear understanding of model accuracy, precision, recall, and overall effectiveness.

Each confusion matrix presents the distribution of correctly and incorrectly classified cases, offering insights into the strengths and limitations of individual models. A well-performing model minimizes false positives and false negatives while maximizing correct classifications. For instance, the AHO Xception model shows remarkable classification accuracy with minimal misclassification errors, whereas Inception V3 struggles in detecting certain lesion types. The confusion metrics not only quantify performance but also aid in model refinement by identifying misclassification patterns. These insights are crucial for medical applications where misdiagnosis can have serious consequences.

The confusion matrices shown in Figure 3 illustrate the classification performance of six different deep learning models AlexNet, MobileNet V2, ResNet50, Inception V3, Xception, and AHO-Optimized Xception on the cervical cancer dataset. Each matrix presents the number of correctly and incorrectly classified instances for four categories: HSIL, LSIL, NILM and SCC. AlexNet shows a high misclassification rate, especially for HSIL and SCC cases. MobileNet V2 improves overall classification but still shows some errors in HSIL and NILM detection. ResNet50 demonstrates better HSIL classification but misclassifies SCC cases significantly. Inception V3 performs well but struggles with NILM misclassification. Xception provides enhanced classification, yet a few misclassifications persist. AHO-Optimized Xception achieves the best performance, with minimal misclassifications, demonstrating the effectiveness of hyperparameter optimization using the Archer Fish Hunting Optimizer (AHO) in improving cervical cancer classification accuracy.

These results highlight the superiority of AHO-optimized Xception in achieving better precision and recall, making it a promising model for real-world cervical cancer diagnosis.

Performance Metrics

The evaluation of deep learning models for cervical cancer classification is crucial for ensuring high accuracy and clinical reliability. This table presents a comparative analysis of six CNN models AlexNet, MobileNet V2, ResNet50, Inception V3, Xception, and AHO-Optimized Xception based on key performance metrics. These metrics include accuracy, precision, recall, F1-score, specificity, AUC-ROC, Matthews Correlation Coefficient (MCC), log loss, and inference time. The inclusion of AHO-optimized Xception highlights the impact of hyperparameter optimization on model performance.

Table 1 presents a comprehensive performance comparison of six deep learning models AlexNet, MobileNet V2, ResNet50, Inception V3, Xception, and

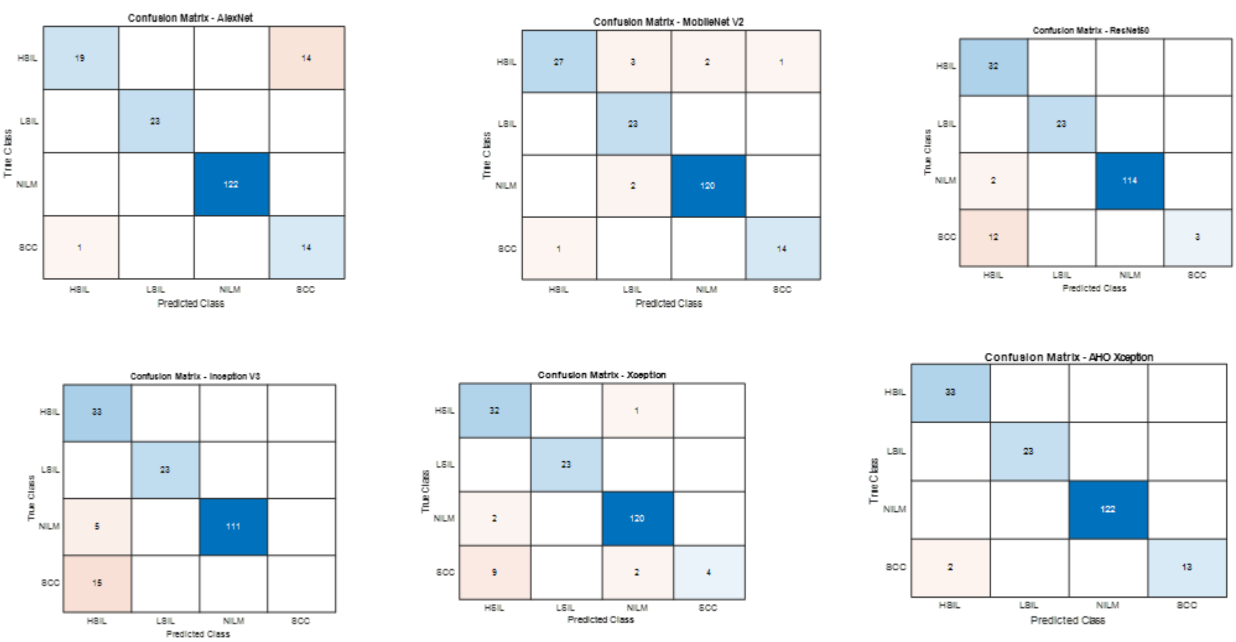

Figure 3. Confusion Matrices for Cervical Cancer Classification Using Different CNN Models

Table 1. Comparative Performance of CNN Models for Cervical Cancer Classification

	AlexNet	MobileNet V2	ResNet 50	Inception V3	Xception	AHO Xception
Accuracy	92.23	95.34	92.51	89.3	92.75	98.96
Precision	86.25	92.57	92.55	65.57	92.99	98.57
Recall	87.73	93.38	79.57	73.92	80.5	96.67
F1 Score	84.20	92.60	78.74	68.64	81.07	97.48
Specificity	97.88	98.26	97.73	96.75	97.22	99.69
AUC-ROC	98.92	99.55	98.79	98.28	98.43	99.97
Matthews Correlation Coefficient (MCC)	0.8383	0.911	0.8022	0.6704	0.8151	0.9727
Log Loss	0.1587	0.179	0.241	0.3013	0.3862	0.0278
Average Inference Time (sec)	0.07184	0.08974	0.1802	0.356	0.8227	0.1443

Values represent percentage (%) performance metrics except for MCC, Log Loss, and Average Inference Time (seconds). Higher values indicate better performance for all metrics except Log Loss and Inference Time, where lower values are preferable.

AHO-Optimized Xception evaluated on a cervical cancer classification dataset.

Among the models, AHO-Optimized Xception demonstrates the highest classification accuracy (98.96%), outperforming traditional deep learning models. It also achieves the best precision (98.57%), recall (96.67%), and F1-score (97.48%), highlighting its ability to effectively identify cervical cancer cases with minimal misclassification. The specificity (99.69%) and AUC-ROC (99.97%) indicate the model's superior capability to distinguish between normal and malignant cases. Additionally, AHO-Optimized Xception achieves the highest MCC (0.9727), showing a strong correlation between predicted and actual classifications, and the lowest log loss (0.0278), reflecting minimal uncertainty in predictions.

Regarding inference time, MobileNet V2 has the fastest processing speed (0.08974 sec), while AHO-Optimized Xception maintains a balance between accuracy and efficiency with a moderate inference time (0.1443 sec), making it suitable for real-time clinical applications. These results confirm that AHO-Optimized Xception significantly enhances cervical cancer classification performance through hyperparameter optimization using the Archer Fish Hunting Optimizer (AHO), making it a highly promising model for automated cervical cancer diagnosis.

The 3D bar chart shown in Figure 4 illustrates a comparative analysis of six deep learning models AlexNet, MobileNet V2, ResNet50, Inception V3, Xception, and AHO-Optimized Xception based on key performance metrics for cervical cancer classification. The metrics are represented by distinct colors. Accuracy (dark blue) indicates overall classification correctness, Precision (blue) reflects the proportion of correctly identified positive cases, Recall (cyan) measures the model's ability to detect actual positive cases, F1 Score (green) balances precision and recall, Specificity (yellow-orange) represents the ability to classify negative cases accurately, and AUC-ROC (yellow) evaluates the model's discriminatory power between normal and malignant cases. The Figure clearly highlights the superiority of AHO-Optimized Xception, which achieves the highest scores across all metrics, demonstrating the effectiveness of hyperparameter optimization in enhancing classification performance. MobileNet V2 also shows strong performance, while models like Inception V3 struggle in recall and F1-score. The results emphasize the importance of advanced optimization techniques in improving AI-driven cervical cancer diagnosis.

This 3D ribbon plot shown in Figure 5 visualizes the performance of different CNN models across three evaluation metrics: MCC (Dark Blue): Represents classification accuracy, with higher values indicating better performance, Log Loss (Cyan): Measures prediction confidence, where lower values signify better certainty, Avg Inference Time (Yellow): Reflects the computational efficiency, with lower values indicating faster model predictions.

The X-axis represents different models (AlexNet, MobileNet V2, ResNet50, Inception V3, Xception), while the Y-axis represents metric categories, and the Z-axis represents the respective scores. This plot provides an insightful comparison, highlighting the trade-offs between model accuracy, prediction reliability, and computational efficiency.

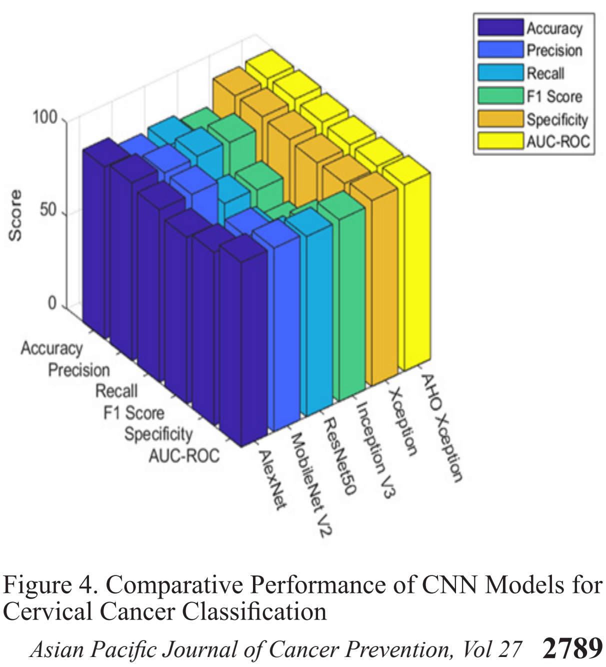

Figure 4. Comparative Performance of CNN Models for Cervical Cancer Classification

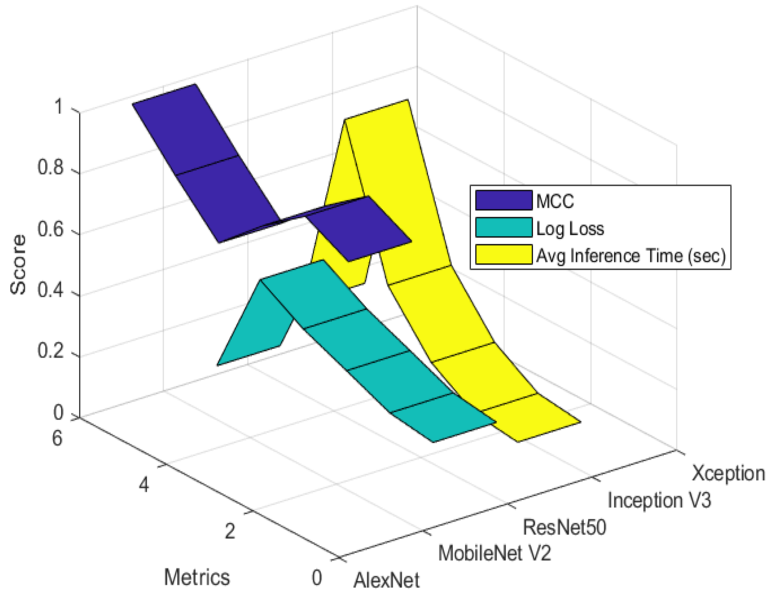

Figure 5. 3D Ribbon Plot of CNN Model Performance Across Metrics

AI Interpretation using GradCAM

Artificial Intelligence (AI) has revolutionized medical imaging by providing automated diagnostic insights with high accuracy. However, the adoption of AI in clinical settings requires transparency and interpretability to build trust among medical practitioners. Grad-CAM plays a crucial role in this aspect by offering visual explanations of AI decision-making in cervical cancer diagnosis. Grad-CAM highlights the most influential regions in cytological images that contribute to classification decisions, enabling pathologists to understand how the model differentiates between HSIL, LSIL, NILM and SCC. By overlaying heatmaps on original images, Grad-CAM reveals whether the model focuses on relevant cellular structures, such as abnormal nuclei, atypical cell formations, or normal epithelial patterns. This interpretability bridges the gap between deep learning predictions and human expertise, ensuring that AI-assisted screenings align with established cytological markers used by pathologists. Moreover, it allows for model validation by identifying potential misclassification patterns and refining feature extraction techniques. By leveraging Grad-CAM, AI-based cervical cancer detection systems can provide reliable, explainable, and clinically meaningful results. This approach not only enhances diagnostic confidence but also supports early detection, reducing pathologist workload while maintaining high accuracy. The integration of Grad-CAM-driven AI interpretations marks a significant step towards the adoption of trustworthy and explainable AI in medical imaging.

The images shown in Supplementary Figure 8 depict AI-based interpretation of cervical cancer diagnosis using Grad-CAM. Each image pair consists of an original microscopic cervical cell image (left) and its corresponding Grad-CAM heatmap (right), highlighting the regions most influential in the AI model's decision-making. The categories include Squamous Cell Carcinoma, High Squamous Intraepithelial Lesion, and Negative for Intraepithelial Malignancy, showing how the model focuses on specific areas of the sample for classification. The heatmaps utilize a color gradient, with red indicating high relevance and blue showing low importance in the AI's prediction.

The results highlight the effectiveness of AI-based interpretation using Grad-CAM for cervical cancer diagnosis by visualizing critical regions that influence model decisions. The heatmaps consistently focus on relevant cellular structures, distinguishing between SCC, HSIL, and NILM. Malignant and pre-malignant cases exhibit strong activations in abnormal cell regions, while normal samples show minimal activation, indicating accurate feature learning. The model's transparency enhances trust in AI-assisted screening, aligning with known cytological features used by pathologists. While the findings validate the model's reliability, further training with diverse datasets can improve generalizability. AI-driven screening can enhance early cervical cancer detection, reduce pathologist workload, and improve diagnostic accuracy, demonstrating the potential of Grad-CAM for explainable AI in medical imaging.

Discussion

The results of this study underscore the significant advantages of applying the Archer Fish Hunting Optimizer (AHO) to the Xception architecture for the task of cervical cancer classification. The evaluation criteria encompass not only conventional performance metrics such as accuracy, precision, recall, and F1-score but also extend to measures of clinical robustness, including specificity, AUC-ROC, Matthews Correlation Coefficient (MCC), log loss, and inference time. This comprehensive assessment ensures that the model is not only mathematically sound but also clinically viable. The AHO-optimized Xception model consistently outperforms benchmark architectures, including AlexNet, MobileNet V2, ResNet50, Inception

V3, and even the base Xception model. It achieves the highest scores across all performance metrics, notably an accuracy of 98.96%, precision of 98.57%, recall of 96.67%, and an F1-score of 97.48%, while maintaining high specificity (99.69%) and an almost perfect AUC-ROC (99.97%). These results affirm the model's ability to distinguish between normal and malignant cell categories with a high degree of confidence and minimal uncertainty.

The confusion matrices further confirm these findings by revealing the distribution of classification outcomes across four diagnostic categories: High-Grade Squamous Intraepithelial Lesion (HSIL), Low-Grade Squamous Intraepithelial Lesion (LSIL), Negative for Intraepithelial Malignancy (NILM), and Squamous Cell Carcinoma (SCC). Among all models, the AHO-optimized Xception demonstrates the lowest misclassification rates across all categories. In comparison, models like Inception V3 and AlexNet exhibit notable errors, particularly in distinguishing HSIL and NILM cases, which are critical for clinical intervention. These results emphasize the need for advanced optimization techniques in improving the diagnostic reliability of deep learning models, especially in high-stakes medical contexts where misdiagnoses can lead to delayed or incorrect treatment.

Visualization tools such as 3D bar charts, line plots, surface plots, and heatmaps corroborate the numerical findings and provide a more intuitive comparison of model performances. The AHO-optimized Xception stands out with consistently high scores across all metrics, as visually demonstrated in Figures 4 through Supplementary Figures 6. These visualizations also make evident the trade-offs present in other models. For example, while MobileNet V2 shows a relatively fast inference time, its classification performance lags behind the AHO-enhanced model. In contrast, Inception V3 exhibits poor recall and F1-score, despite moderate inference efficiency, indicating limitations in identifying positive cases. The inclusion of MCC and log loss in these comparisons is particularly important for understanding the model's reliability and confidence, where again, AHO-optimized Xception outperforms all alternatives.

The integration of Grad-CAM further enhances the credibility and interpretability of the proposed model. By generating class-specific heatmaps over cytological images, Grad-CAM highlights the areas that most significantly contribute to the model's predictions. This level of explainability ensures that the AI system is not relying on irrelevant image artifacts, but rather on clinically meaningful features such as abnormal nuclei and atypical cell clusters. The heatmaps presented in Supplementary Figure 7 show that the model's focus areas correspond closely with regions pathologists would consider diagnostic, reinforcing the clinical validity of the model. This interpretability bridges the gap between algorithmic output and medical reasoning, an essential factor in fostering trust among healthcare professionals and ensuring the model's integration into real-world diagnostic workflows.

Beyond raw performance, this study also explores the role of supporting techniques such as data augmentation, regularization, and hyperparameter tuning in enhancing model generalization. These strategies are instrumental in overcoming challenges like class imbalance and overfitting common issues in medical imaging datasets. The comparative advantage gained through AHO-based hyperparameter optimization demonstrates the value of intelligent search algorithms in fine-tuning deep learning models for medical use. While the AHO-optimized model achieves superior results, it does so with a moderate increase in inference time (0.1443 sec), which remains within acceptable limits for clinical deployment.

Despite the promising outcomes, certain limitations persist. The model's performance, while robust on the current dataset, may vary when applied to different populations or imaging equipment. This highlights the need for further validation using diverse and multi-institutional datasets. Moreover, while Grad-CAM provides a degree of interpretability, more advanced explainability techniques could further improve clinical transparency. Future work should explore domain adaptation, federated learning, and integration with electronic health record systems to build a more comprehensive AI-driven diagnostic solution.

In conclusion, the AHO-optimized Xception model presents a highly accurate, interpretable, and clinically applicable tool for automated cervical cancer classification. Its superior performance across all evaluation metrics, coupled with explainable decision-making via Grad-CAM, positions it as a strong candidate for real-world deployment in cervical cancer screening programs. The study illustrates that with appropriate optimization and interpretability strategies, deep learning can be a transformative force in medical diagnostics.

This study demonstrates the efficacy of AI-driven cervical cancer diagnosis using deep learning models, with performance evaluated through multiple metrics, confusion matrices, and Grad-CAM visualizations. The AHO Xception model outperformed others, achieving the highest accuracy (98.96%), precision (98.57%), recall (96.67%), F1-score (97.48%), specificity (99.69%), and AUC-ROC (99.97%), while also maintaining a significantly low log loss (0.0278). In contrast, Inception V3 showed the lowest accuracy (89.3%), primarily due to its reduced sensitivity in detecting specific lesion types. The confusion matrices highlight that AHO Xception exhibits superior classification, correctly identifying most HSIL, LSIL, NILM, and SCC cases with minimal misclassification. Although the AHO-optimized Xception model demonstrates superior classification performance, its average inference time of 0.1443 seconds per image reflects a trade-off between computational complexity and diagnostic accuracy. In cervical cytology workflows, screening is not a millisecond-critical process, and this latency remains clinically acceptable, particularly in laboratory-based batch analysis settings. Moreover, model compression strategies such as pruning, quantization, and hardware-specific acceleration can further reduce inference time without substantially compromising accuracy. Therefore, the observed inference time represents a balanced compromise between high diagnostic reliability and practical deployability. Additionally, Grad-CAM visualizations reinforce the model's decision-making process, showing strong activations in relevant cellular structures, particularly in malignant and pre-malignant regions. This interpretability enhances trust in AI-assisted screening, aligning with cytological criteria used by pathologists. Despite these promising results, further improvements can be achieved by incorporating more diverse and real-world datasets to enhance generalizability. AI-powered screening has the potential to revolutionize early cervical cancer detection, reducing the workload on pathologists and improving diagnostic accuracy. The study validates the role of explainable AI through Grad-CAM in medical imaging, highlighting its significance for trustworthy and interpretable deep learning models in healthcare. While the proposed framework achieves strong performance on the Mendeley LBC cervical cytology dataset, external validation using multi-institutional datasets is essential to confirm generalizability across diverse populations and imaging conditions. Variations in staining protocols, imaging equipment, and demographic factors may influence model robustness. Future work will focus on validating the model on independent datasets, exploring domain adaptation techniques, and investigating federated learning strategies to enhance cross-center reliability and real-world clinical applicability.

Author Contribution Statement

As a single-author manuscript, all components of the study including conceptualization, methodology, implementation, analysis, interpretation, and manuscript writing were solely undertaken by the author.

Acknowledgements

General

The author sincerely acknowledges the facilities and support provided by Saveetha Engineering College during the course of this research.

Scientific Approval

This work was conducted as part of an approved faculty-led research project within Saveetha Engineering College. It is not associated with a student thesis submission or review by any external scientific body.

Ethical Declaration

This study involved retrospective analysis of openly available and anonymized medical imaging datasets and did not require patient consent or direct ethical clearance. The dataset used complies with public data usage standards under Creative Commons Attribution 4.0 International License. Therefore, ethical approval from a specific institutional review board was not required.

Data Availability

The dataset used in this study, titled “Medical Imaging (CT scan, MRI, X-ray, and Microscopic Imagery) Data”, is publicly available at Mendeley Data [https://data.mendeley.com/datasets/5kbjrgsncf/3], under a CC BY 4.0 license.

Study Registration

Not applicable. This study does not involve clinical trials or patient registration and was not submitted to a clinical study registry.

Conflict of Interest

The author declares no conflict of interest.

References

1. World Health Organization. Cervical cancer: Key facts. 2020. Available from: from https://www.who.int/news-room/fact-sheets/detail/cervical-cancer
2. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. https://doi.org/10.1016/j.media.2017.07.005.
3. Hinton G, Krizhevsky A, Sutskever I. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012:1097-105. https://doi.org/10.1145/3065386.
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.
5. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90
6. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 2818-2826). https://doi.org/10.1109/CVPR.2016.308
7. Chen W, Shen W, Gao L, Li X. Hybrid loss-constrained lightweight convolutional neural networks for cervical cell classification. Sensors (Basel). 2022;22(9):3272. https://doi.org/10.3390/s22093272.
8. Chollet F. Xception: Deep learning with depthwise separable convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1251-1258).https://doi.org/10.1109/CVPR.2017.195
9. Claesen M, De Moor B. Hyperparameter search in machine learning. arXiv preprint arXiv:1502.02127. 2015 Feb 7.
10. Xue B, Zhang M, Browne WN. Particle swarm optimization for feature selection in classification: A multi-objective approach. IEEE transactions on cybernetics. 2012 Dec 13;43(6):1656-71.https://doi.org/10.1109/TCYB.2016.2633559
11. Poli R, Kennedy J, Blackwell T. Particle swarm optimization. Swarm Intelligence. 2007;1(1):33-57. https://doi.org/10.1007/s11721-007-0002-0.
12. Simon D. Biogeography-based optimization. IEEE transactions on evolutionary computation. 2008 Mar 21;12(6):702-13.https://doi.org/10.1109/tevc.2008.919004.
13. Zitouni F, Harous S, Belkeram A, Hammou L. . The Archerfish Hunting Optimizer: a novel metaheuristic algorithm for global optimization. 2021; 10.48550/arXiv.2102.02134.
14. Punithavathi R, Kumar C, Kumar R, Saravana R, Abhinandan Chaotic A. Optimizer–Enhanced Double-Attention Dense BiGRU for Video-Based Facial Expression Recognition, International Journal of Intelligent Systems, 2026, 3099014, 31 pages, 2026. https://doi.org/10.1155/int/3099014
15. Patre P, Verma D. Deep dive into deep learning methods for cervical cancer detection and classification. Rep Pract Oncol Radiother. 2025 Aug 7;30(3):396-416. doi: 10.5603/rpor.106148.

16. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. 2017. p. 618-26.
17. Codella N, Nguyen Q-B, Pankanti S, Gutman D, Helba B, Halpern A, et al. Deep learning ensembles for melanoma recognition in dermoscopy images. IBM J Res Dev. 2017;61. https://doi.org/10.1147/JRD.2017.2708299.
18. Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. Procedia Comput Sci. 2016;90:200-5. https://doi.org/10.1016/j.procs.2016.07.014.
19. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. 2017. https://doi.org/10.48550/arXiv.1711.05225.
20. Elima H, Lipi BM, Himakshi B, Chandana RD. Mendeley LBC Cervical Cancer [Dataset]. Kaggle. 2024. https://doi.org/10.34740/KAGGLE/DSV/7777074.