Jurnal Ilmiah Biologi UMA (JIBIOMA), 8(1) 2026: 28-37

DOI: 10.31289/jibioma.v8i1.7103

Jurnal Ilmiah Biologi UMA (JIBIOMA)

Available online http://jurnalmahasiswa.uma.ac.id/index.php/jibioma

Submitted: 24 Maret 2026; Reviewed: 01 Mei 2026; Accepted: 14 Mei 2026

An Extended Deep Forest Algorithm with Automatic Parameter Selection for Binary Image Classification: Application to Malaria Diagnosis

Jerry Emmanuel* & Temitope Oluwaseun Ogungbesan

Department of Computer and Information Sciences, Covenant University, Nigeria

Abstract

Deep forest has emerged as a lightweight machine-learning approach for image classification, offering lower computational requirements than many deep learning models. However, its classification performance is highly dependent on the selection of the n_estimators parameter, and determining the optimal value remains a challenge. This study aimed to develop an Extended Deep Forest (EDF) algorithm capable of automatically selecting the optimal number of estimators based on training dataset characteristics. The proposed method employs a data-driven interpolation function derived from the empirical relationship between training dataset size and the optimal estimator count. Experiments were conducted using eight binary image datasets, while the performance of EDF and the conventional deep forest was compared on four independent test datasets. The results showed that EDF consistently achieved higher classification performance on datasets requiring more than two estimators while maintaining comparable computational efficiency. Application of the proposed algorithm to thick blood smear malaria diagnosis achieved an accuracy of 94.05% and an F1-score of 94.03%, demonstrating competitive performance while preserving computational efficiency suitable for resource-constrained environments. In conclusion, EDF provides an effective and automated parameter-selection strategy that improves deep forest performance and supports its practical implementation in lightweight image-classification applications, including malaria diagnosis.

Keywords: *Deep Forest; Image Classification; Automatic Parameter Selection; Interpolation Function; Malaria Diagnosis.*

How to Cite: Emmanuel, J., & Ogungbesan, T.O. (2026). An Extended Deep Forest Algorithm with Automatic Parameter Selection for Binary Image Classification: Application to Malaria Diagnosis*. Jurnal Ilmiah Biologi UMA (JIBIOMA), 8(1) 2026: 28-37*

**E-mail:* jerry.emmanuel@covenantuniversity.edu.ng *ISSN 2722-9777 (Online)*

This work is licensed under a Creative Commons Attribution 4.0

INTRODUCTION

Image classification is a fundamental task in computer vision and has been widely applied in medical diagnosis, remote sensing, traffic monitoring, and content moderation. Recent advances in deep learning have significantly improved image classification performance, particularly through Convolutional Neural Networks (CNNs). Architectures such as VGG16 (Simonyan & Zisserman, 2015) and Vision Transformers (Dosovitskiy et al., 2021) have achieved state-of-the-art results across various benchmark datasets. However, these models typically require large training datasets, extensive computational resources, and specialized hardware, making their deployment challenging in resource-constrained environments.

To overcome these limitations, Zhou and Feng (2017) proposed Deep Forest (DF), a deep learning framework based on cascaded decision-tree ensembles rather than neural networks. The framework was later refined as DF21 (Zhou & Feng, 2021), providing a computationally efficient alternative that requires fewer hyperparameters and does not depend on GPU acceleration. Deep Forest can also adapt its complexity according to the characteristics of the training data, making it attractive for practical applications where computational resources are limited.

Despite these advantages, the performance of Deep Forest remains sensitive to hyperparameter selection. Among the available parameters, *n_estimators*, which determines the number of estimator pairs in each cascade level, has a substantial influence on classification accuracy. The official DF21 documentation explicitly recommends parameter tuning to achieve optimal performance. Nevertheless, many implementations continue to use the default value of two estimators regardless of dataset size and complexity. Consequently, model performance may not fully reflect the potential of the Deep Forest architecture.

Previous studies have demonstrated the effectiveness of Deep Forest and related ensemble-learning approaches across various domains. Applications include weighted random forest classification (Utkin & Konstantinov, 2021), cancer survival prediction (Yan et al., 2023), and confidence-enhanced Deep Forest architectures such as DBC-Forest (Ma et al., 2022). Although these studies reported promising results, parameter optimization was generally performed using manual experimentation, grid search, or heuristic approaches. Such methods can be time-consuming, computationally inefficient, and difficult to generalize across datasets. More importantly, no previous study has proposed a systematic and data-driven mechanism for automatically determining the optimal *n_estimators* value in the Deep Forest cascade architecture. This gap limits the practical scalability and usability of the method.

The need for an efficient and easily deployable classification model is particularly evident in automated malaria diagnosis. Microscopic examination of Giemsa-stained blood smears remains the gold standard for malaria detection; however, diagnostic accuracy depends heavily on the expertise of trained microscopists (Arias-Alpízar et al., 2022). According to the *World Malaria Report 2024* (World Health Organization, 2024), approximately 263 million malaria cases were reported worldwide in 2023, with sub-Saharan Africa accounting for the majority of infections and deaths. To address this challenge, numerous computer-aided diagnostic systems based on deep learning have been developed. Studies by Nakasi et al. (2021), Kassim et al. (2021), Abdurahman et al. (2021), and Yang et al. (2020) demonstrated that deep neural networks can achieve high diagnostic accuracy in thick blood smear image analysis. However, recent evaluations indicate that many high-performing models remain difficult to deploy in resource-limited settings because of their computational demands (Mmileng et al., 2025; Ozsahin et al., 2022).

Given these challenges, this study proposes an Extended Deep Forest (EDF) algorithm that automatically determines the optimal number of estimators using a data-driven interpolation function derived from the relationship between training dataset size and the optimal estimator count. Specifically, this study aims to (a) analyze the relationship between training dataset size and the optimal number of estimators in the Deep Forest cascade architecture, (b) develop an interpolation-based mechanism for automatic parameter selection, and (c) evaluate the proposed EDF model through a malaria diagnosis case study using thick blood smear images and its implementation in a mobile-compatible web application. By automating parameter selection while preserving computational efficiency, the proposed approach is expected to improve Deep Forest performance and enhance its applicability in resource-limited environments.

RESEARCH METHOD

Research Design

This study employed a three-phase empirical research design. The first phase focused on an estimator sensitivity analysis to quantify the relationship between training dataset size and the optimal number of *n_estimators* in the Deep Forest model. The second phase involved comparative evaluation between the proposed Extended Deep Forest

(EDF) and the conventional Deep Forest using held-out datasets. The third phase implemented the EDF in a real-world application for malaria detection using thick blood smear images, followed by prototype system development.

All experiments were conducted in a CPU-only environment (Intel Core i7 processor with 16 GB RAM) to simulate realistic resource-constrained conditions. To ensure reproducibility, a fixed random seed (seed = 42) was used across all experiments. A stratified 70:30 train-test split was applied using scikit-learn's train_test_split function to preserve class distribution consistency across datasets. No separate validation set was used during the estimator sensitivity phase; evaluation was performed exclusively on held-out test sets.

Datasets

Eight publicly available binary image datasets were used in the estimator sensitivity experiment. These datasets were selected to ensure variability in dataset size, class distribution, and application domain. The datasets were sourced from Kaggle and ranged from 0.53 MB to 3,604 MB, enabling the analysis of scalability effects in relation to model parameter behavior.

In addition, four independent held-out datasets (*good_bad*, *run_walk*, *cat_rabbit*, and *Hymenoptera*) were used exclusively to evaluate the generalization performance of the proposed EDF model. For the malaria application, the thick blood smear dataset provided by Kassim et al. (2021) was used, consisting of 3,024 images collected from 350 patients at Chittagong Medical College Hospital.

Data Preprocessing

All images were resized to a uniform dimension based on the median width and height of each dataset. This adaptive resizing strategy was chosen to minimize distortion caused by fixed-size resizing (e.g., 224×224), which may negatively affect feature representation in heterogeneous image datasets (Bradski & Kaehler, 2023).

Image loading, resizing, and conversion to array format were performed using OpenCV (Bradski & Kaehler, 2023). Label encoding was conducted using scikit-learn's LabelEncoder. Each dataset was split into training and testing sets using a stratified 70:30 ratio to maintain class distribution consistency. This split ratio is consistent with common practice in Deep Forest-based studies, balancing training sufficiency and evaluation reliability (Zhou & Feng, 2017; Yan et al., 2023).

Estimator Sensitivity Experiment

To evaluate the effect of *n_estimators* on classification performance, each dataset was trained using the Deep Forest classifier with *n_estimators* values ranging from 1 to 6. All other hyperparameters were kept at their default configuration to isolate the influence of this parameter.

Model performance was evaluated using six metrics: Accuracy, F1-score, Macro Precision, Weighted Precision, Macro Recall, and Weighted Recall. This resulted in a total of 48 experimental runs across all datasets.

For each dataset and metric, the optimal *n_estimators* value was recorded. A monotonic relationship between dataset size (MB) and optimal *n_estimators* was then analyzed visually and empirically. Only metrics demonstrating consistent monotonic behavior were considered suitable for constructing the interpolation function. Weighted metrics were prioritized due to their robustness in handling class imbalance, which is common in image classification datasets (Utkin & Konstantinov, 2021).

Extended Deep Forest (EDF) Algorithm

The experimental results indicated a consistent positive correlation between dataset size and optimal *n_estimators* for Weighted Precision and Weighted Recall. Based on these findings, a data-driven interpolation function was constructed using SciPy's interp1d (univariate interpolation) method (Virtanen et al., 2020).

The interpolation function maps dataset size to the predicted optimal number of estimators: f: dataset size → n_estimators. The function is constructed using empirically derived (dataset size, optimal estimator) pairs obtained from the sensitivity experiment.

This interpolation function was integrated into the initialization phase of the Deep Forest pipeline, resulting in the Extended Deep Forest (EDF). The EDF retains the original architecture of Deep Forest, including cascade construction and multi-grained scanning, with the only modification being automatic determination of *n_estimators* prior to training (Zhou & Feng, 2017; Zhou & Feng, 2021).

Malaria Application and Prototype Development

The proposed EDF was applied to the thick blood smear malaria dataset. Based on the interpolation function, the optimal *n_estimators* was determined as 4 for the infected-class training partition.

The trained model was serialized using the pickle format for deployment. A three-tier web application was developed, consisting of a client interface, a Flask-based application server, and local storage.

To ensure input validity, an image verification module based on the ORB (Oriented FAST and Rotated BRIEF) algorithm was implemented (Guo et al., 2023). This module performs feature matching between user-submitted images and reference blood smear images to filter out invalid inputs such as non-cell images. Only validated images are forwarded to the classifier for prediction.

RESULT AND DISCUSSION

Correlation Between Dataset Size and Optimal Estimators

Table 2 presents the optimal *n_estimators* values for each dataset across different evaluation metrics. The results indicate a clear trend: smaller datasets such as Breast Cancer (0.005 MB), Brain Tumor (69.2 MB), and Elephant (103.5 MB) generally achieve optimal performance using fewer estimators (1–2). In contrast, larger datasets such as Pizza vs Not Pizza (1,032.0 MB) and Organic vs Recycle (2,523.5 MB) tend to perform best with higher estimator values (3–4).

A consistent positive monotonic relationship between dataset size and optimal *n_estimators* was observed particularly for Weighted Precision and Weighted Recall. This finding supports the use of a data-driven interpolation approach as the basis for the Extended Deep Forest (EDF), since these metrics provide stable behavior under class imbalance conditions (Utkin & Konstantinov, 2021).

Table 2. Optimal *n_estimators* per Dataset Across Evaluation Metrics

Dataset	Train MB	Accuracy	F1 Score	Macro Prec.	Wtd. Prec.	Wtd. Recall
Breast Cancer	0.005	1	1	3	1	1
FaceMask	24.2	4	4	1	1	1
Brain Tumor	69.2	2	2	2	2	2
Elephant	103.5	2	2	2	2	2
Human vs Horses	231.2	3	3	3	3	3
HotDog	831.1	4	5	3	3	3
Pizza	1,032.0	4	4	4	4	4
Organic vs Recycle	2,523.5	4	4	4	4	4

Extended Deep Forest vs Traditional Deep Forest

Table 3 compares the performance of the Extended Deep Forest (EDF) with the traditional Deep Forest (DF) on four held-out datasets. The results show that EDF improves performance on datasets where the optimal estimator value is greater than two. For example, accuracy increased from 86.35% to 87.14% on the good_bad dataset

(+0.79%) and from 88.62% to 89.58% on the cat_rabbit dataset (+0.96%). In contrast, performance remained identical for run_walk and Hymenoptera, where the interpolation function correctly selected the default value of two estimators.

In terms of computational cost, EDF requires longer training time on larger datasets due to the increased number of estimators. For instance, training time on good_bad increased from 22,461 seconds (DF) to 46,389 seconds (EDF), approximately a twofold increase. This is expected because doubling the number of estimators increases cascade computations proportionally.

However, this increase only affects training time. Inference time remains nearly unchanged (EDF: 0.0298 s; DF: 0.030 s on the malaria dataset). Therefore, EDF remains suitable for deployment in resource-constrained environments, particularly when training is performed offline. The slight increase in computational cost is justified by the consistent improvement in classification performance for larger datasets.

Table 3. Performance Comparison: Extended Deep Forest vs Traditional Deep Forest

Dataset	Model	Est.	Acc (%)	F1 (%)	Mac. P (%)	Wtd. P (%)	Wtd. R (%)	Time (s)
good_bad	EDF	4	87.14	87.14	87	87	87	46,389
good_bad	DF	2	86.35	86.35	86	86	86	22,461
run_walk	EDF	2	65.02	65.17	65	65	65	698
run_walk	DF	2	65.02	65.17	65	65	65	659
cat_rabbit	EDF	3	89.58	89.58	90	90	90	3,589
cat_rabbit	DF	2	88.62	88.62	89	89	89	2,492
Hymenoptera	EDF	2	71.67	71.67	72	72	72	910
Hymenoptera	DF	2	71.67	71.67	72	72	72	882

Malaria Classification and Comparison with Literature

The proposed EDF achieved an accuracy of 94.05% and an F1-score of 94.03% on the thick blood smear malaria dataset. In comparison, the traditional Deep Forest achieved 93.72% accuracy and 93.70% F1-score. This represents a consistent improvement of 0.33 percentage points, which is attributed to the automatic selection of the optimal number of estimators.

Table 4 compares the proposed method with existing studies in malaria parasite classification. EDF outperforms traditional Deep Forest (Zhou & Feng, 2017), Mask R-CNN (Kassim et al., 2021), and DeepMCNN (Manescu et al., 2020) in terms of accuracy. It also achieves significantly lower inference time (0.0298 s), making it faster than YOLOV4-based approaches and comparable lightweight models.

However, the highest-performing neural network models, such as CNN-based methods reported by Kassim et al. (2021) and Faster R-CNN models (Yang et al., 2019), still achieve higher accuracy (up to 96.89%). This difference reflects the trade-off between model complexity and computational efficiency. These models require extensive computational resources and GPU support, which limits their use in mobile or low-resource environments (Ozsahin et al., 2022; Mmileng et al., 2025).

Although EDF shows slightly lower accuracy compared to the best-performing deep learning models (approximately 1.84% lower than 95.89% reported by Ramarolahy et al., 2021), it offers a substantial advantage in terms of deployment feasibility. The inference time of less than 30 milliseconds makes EDF suitable for real-time clinical decision support systems in resource-limited settings.

Table 4. Comparison with State-of-the-Art Methods for Malaria Detection

Authors	Model	Accuracy (%)	F1 Score (%)
Yang et al. (2020)	VGG19 CNN	93.46	93.40
Kassim et al. (2021)	Mask R-CNN	90.00	79.97
Kassim et al. (2021)	CNN	96.89	81.80
Manescu et al. (2020)	DeepMCNN	91.00	—
Yang et al. (2019)	Faster R-CNN	96.84	—
Ramarolahy et al. (2021)	CNN	95.30	95.30
Meng et al. (2022)	NCGCN	94.17	94.20
Nakasi et al. (2021)	Faster R-CNN	—	79.97
Nakasi et al. (2021)	SSD MobileNet v2	—	60.38
Traditional DF (Zhou & Feng, 2017)	Deep Forest	93.72	93.70
Proposed EDF	Extended Deep Forest	94.05	94.03

NCGCN = Neighbour Correlated Graph Convolutional Network

Limitations

Several limitations should be acknowledged in this study. First, the interpolation function is based on a limited number of datasets, which may affect generalization. Expanding the dataset coverage, particularly within the 300–800 MB range, could improve interpolation accuracy.

Second, this study focuses only on the *n_estimators* parameter. Other Deep Forest hyperparameters such as *n_trees*, *max_depth*, and *n_bins* may also influence performance and could be incorporated in future work.

Third, the current evaluation is limited to binary classification tasks. Extending the proposed method to multi-class classification may require re-evaluation of the relationship between dataset size and optimal estimator configuration.

Finally, although EDF increases training time for larger datasets, this trade-off remains acceptable in offline training scenarios. However, it may be less suitable for applications requiring frequent model retraining in time-sensitive environments.

CONCLUSION

This study proposes the Extended Deep Forest (EDF), a novel approach that addresses a long-standing limitation in the Deep Forest framework: the absence of a systematic method for selecting the *n_estimators* hyperparameter. By identifying an empirical relationship between training dataset size and the optimal number of estimators, and formalising it through a SciPy-based interpolation function, the EDF enables automatic parameter selection without requiring additional labelled data or cross-validation overhead. This data-driven mechanism is fully integrable into existing Deep Forest pipelines and consistently improves classification performance over the traditional baseline.

In the context of automated malaria diagnosis using thick blood smear images, the proposed EDF achieved 94.05% accuracy and 94.03% F1-score, outperforming the traditional Deep Forest model and several conventional CNN-based approaches reported in the literature. Importantly, the EDF maintains low computational requirements and is fully operable on CPU-only environments, making it suitable for deployment in resource-constrained settings where GPU-based deep learning models are often impractical. The developed prototype system, incorporating ORB-based image validation, further demonstrates the feasibility of an end-to-end lightweight diagnostic pipeline with automated input quality control.

Despite these advantages, several limitations remain. Future research should expand the dataset range used for interpolation, particularly within mid-scale datasets, to improve prediction stability. Additional work is also needed to explore relationships between other Deep Forest hyperparameters and dataset characteristics, which may further enhance model performance. Moreover, extending the EDF framework to multi-class classification problems is an important direction, as the current formulation is limited to binary tasks. Finally, prospective field validation in real clinical environments is necessary to confirm the practical effectiveness of the proposed system, alongside comparisons with newer lightweight neural architectures to better understand the trade-off between accuracy and deployability.

REFERENCES

Abdurahman, F., et al. (2021). Malaria parasite detection in thick blood smear images using modified YOLOV3 and YOLOV4 models. BMC Bioinformatics, 22(1), 1–17.

Arias-Alpízar, K., et al. (2022). Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: A review. PLOS Neglected Tropical Diseases, 16(11), e0010795. https://doi.org/10.1371/journal.pntd.0010795

Bradski, G., & Kaehler, A. (2023). Learning OpenCV 5: Computer vision with Python. O'Reilly Media.

Deep Forest Documentation. (n.d.). Parameters tuning — DF21. https://deep-forest.readthedocs.io/en/latest/parameters_tunning.html

Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR).

Guo, X., et al. (2023). A light field full-focus image feature point matching method with an improved ORB algorithm. Sensors, 23(1), 123. https://doi.org/10.3390/s23010123

Kassim, Y. M., et al. (2021). Diagnosing malaria patients with Plasmodium falciparum and Plasmodium vivax using deep learning for thick smear images. Diagnostics, 11(11), 1–19.

Ma, P., et al. (2022). DBC-Forest: Deep forest with binning confidence screening. arXiv preprint arXiv:2112.13182. https://doi.org/10.48550/arXiv.2112.13182

Mmileng, O. P., et al. (2025). Application of ConvNeXt with transfer learning and data augmentation for malaria parasite detection in resource-limited settings using microscopic images. PLOS ONE, 20(6), e0313734. https://doi.org/10.1371/journal.pone.0313734

Nakasi, R., Mwebaze, E., & Zawedde, A. (2021). Mobile-aware deep learning algorithms for malaria parasites localization. Algorithms, 14(1), 1–16.

Ozsahin, D. U., et al. (2022). Evaluating the performance of deep learning frameworks for malaria parasite detection using microscopic images of peripheral blood smears. Diagnostics, 12(11), 2702. https://doi.org/10.3390/diagnostics12112702

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).

Utkin, L. V., & Konstantinov, A. V. (2021). A weighted random forest classifier by using a ranking loss function. In Proceedings of the 29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2021) (pp. 253–258).

Virtanen, P., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

World Health Organization. (2024). World malaria report 2024. WHO Press.

Yan, W., et al. (2023). Multiview deep forest for overall survival prediction in cancer. Diagnostics, 13(3), 512. https://doi.org/10.3390/diagnostics13030512

Yang, F., et al. (2020). Deep learning for smartphone-based malaria parasite detection in thick blood smears. IEEE Journal of Biomedical and Health Informatics, 24(5), 1427–1438.

Zhou, Z. H., & Feng, J. (2017). Deep forest: Towards an alternative to deep neural networks. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), 3553–3559.

Zhou, Z. H., & Feng, J. (2021). Deep forest. National Science Review, 6(1), 74–86. https://doi.org/10.1093/nsr/nwy108