



ARTICLE

Integration of YOLOv11 and Histogram Equalization for Fire and Smoke-Based Detection of Forest and Land Fires

Christine Dewi^{1,2}, Melati Viaeritas Vitrieco Santoso¹, Hanna Prillysca Chernovita³, Evangs Mailoa¹, Stephen Abednego Philemon¹ and Abbott Po Shun Chen^{4,*}

¹Department of Information Technology, Satya Wacana Christian University, Jalan Diponegoro No. 52-60, Salatiga, 50711, Indonesia

²School of Information Technology, Deakin University, 221 Burwood Highway, Burwood, VIC 3125, Australia

³Department of Information Systems, Satya Wacana Christian University, Jalan Diponegoro No. 52-60, Salatiga, 50711, Indonesia

⁴Department of Marketing and Logistics Management, Chaoyang University of Technology, 168 Jifeng East Road, Taichung City, 413310, Taiwan

*Corresponding Author: Abbott Po Shun Chen. Email: chprosen@gm.cyut.edu.tw

Received: 01 May 2025; Accepted: 18 June 2025; Published: 30 July 2025

ABSTRACT: Early detection of Forest and Land Fires (FLF) is essential to prevent the rapid spread of fire as well as minimize environmental damage. However, accurate detection under real-world conditions, such as low light, haze, and complex backgrounds, remains a challenge for computer vision systems. This study evaluates the impact of three image enhancement techniques—Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), and a hybrid method called DBST-LCM CLAHE—on the performance of the YOLOv11 object detection model in identifying fires and smoke. The D-Fire dataset, consisting of 21,527 annotated images captured under diverse environmental scenarios and illumination levels, was used to train and evaluate the model. Each enhancement method was applied to the dataset before training. Model performance was assessed using multiple metrics, including Precision, Recall, mean Average Precision at 50% IoU (mAP50), F1-score, and visual inspection through bounding box results. Experimental results show that all three enhancement techniques improved detection performance. HE yielded the highest mAP50 score of 0.771, along with a balanced precision of 0.784 and recall of 0.703, demonstrating strong generalization across different conditions. DBST-LCM CLAHE achieved the highest Precision score of 79%, effectively reducing false positives, particularly in scenes with dispersed smoke or complex textures. CLAHE, with slightly lower overall metrics, contributed to improved local feature detection. Each technique showed distinct advantages: HE enhanced global contrast; CLAHE improved local structure visibility; and DBST-LCM CLAHE provided an optimal balance through dynamic block sizing and local contrast preservation. These results underline the importance of selecting preprocessing methods according to detection priorities, such as minimizing false alarms or maximizing completeness. This research does not propose a new model architecture but rather benchmarks a recent lightweight detector, YOLOv11, combined with image enhancement strategies for practical deployment in FLF monitoring. The findings support the integration of preprocessing techniques to improve detection accuracy, offering a foundation for real-time FLF detection systems on edge devices or drones, particularly in regions like Indonesia.

KEYWORDS: Histogram equalization; YOLO; forest and land fire detection; deep learning

1 Introduction

The Eaton Fire, more widely known as the LA Wildfire—a forest fire incident that occurred in early 2025—has so far claimed 30 lives and destroyed up to 15,000 buildings. This incident highlights the



importance of early detection and mitigation of forest and land fires (hereafter referred to as FLF). As of this year, FLF has burned 126 hectares of land across Indonesia, causing significant ecological and economic impacts [1]. Whether caused by extreme weather and drought or by human activities like uncontrolled burning to clear land, FLF can result in environmental degradation, increased greenhouse gas emissions, and health issues due to the resulting smoke.

Early detection of forest and land fires is crucial for mitigating and preventing such disasters. Small sources of fire must be detected quickly before they spread and turn into large-scale wildfires. Delays in detection can accelerate fire spread. An effective monitoring system for FLF can enhance awareness and preparation in reducing post-disaster impacts. In turn, this will also support the sustainability of ecosystems for the future.

By utilizing deep learning models, Table 1 shows that various studies have developed early detection systems for fire and smoke, for example, using YOLOv4 to YOLOv8, as well as other variants. However, among these systems, there are still two main aspects that need improvement—detection speed and accuracy, especially in recognizing fire and smoke. The optimization of YOLOv5 up to YOLOv11 for real-time detection has been explored in several studies [2–4]. Some researchers have also combined the YOLO model with other techniques (such as filter pruning [5]) to reduce computational load on low-power devices.

Table 1: Summary of FNF detection methods in the literature

Author	Method	Contribution	Limitation
Goncalves et al. [2], 2024	YOLOv5, 7, and 8 models	Detection speed & smoke localization	Detection of small areas with visual disturbances (fog, clouds, sun)
Wicaksono [3], 2024	YOLOv8 model	Real-time detection efficiency & accuracy	Effectiveness in real-time field conditions
Venâncio et al. [5], 2022	Combination of YOLOv4 model and pruning filter	Computational load (83.6%) & memory (83.86%) efficiency on low-power devices	Accuracy of detecting multiple fire points in complex scenarios
Pan et al. [6], 2021	Distillation approach in CNN	Lightweight Faster R-CNN framework	Segmentation accuracy compared to pixel labels
Perrolas et al. [7], 2022	Segmentation-classification with SqueezeNet + U-Net and quad-tree	Accuracy in segmenting small fire areas & fire localization	Effectiveness on other image types & accuracy in large fire areas being roughly segmented
Tingting Li et al. [8], 2022	Adversarial fusion network (AFN) & domain-adversarial neural network (DANN)	Optimized detection of small smoke in complex scenarios	Detection accuracy in complex scenarios & data variability due to lack of feature detail
Li et al. [9], 2025	YOLOv8 model optimization with SE-ResNeXt & Focal-SIoU	Optimized detection of varying fire sizes	High computational cost for low-power devices

(Continued)

Table 1 (continued)

Author	Method	Contribution	Limitation
Muksimova et al. [10], 2024	Miti-DTER with modified AlexNet backbone & HE data	Effective feature extraction & training stability	Real-time detection effectiveness & non-fire scenarios
Farhan et al. [11], 2022	YOLOv4 model & HE data	Accuracy & speed of fire detection	Detection stability is affected by input resolution
Ayumi et al. [12], 2024	Xception & MobileNet models and CLAHE data	Fire and smoke detection accuracy	Detection of data variations in complex real-time scenarios

The novelty of this research does not lie in proposing a fundamentally new detection architecture or image enhancement technique, but rather in conducting a comprehensive empirical evaluation of the recently launched YOLOv11 model combined with selected image enhancement techniques such as Histogram Equalization (HE) [11], Contrast Limited Adaptive Histogram Equalization (CLAHE) [12], and the DBST-LCM CLAHE method proposed by Chakraverti et al. [13].

Prior works in forest fire detection have focused on various improvements across model design, training strategies, and preprocessing methods. For example, Pan et al. [6] developed a lightweight Faster R-CNN via distillation, Venâncio et al. [5] optimized YOLOv4 pruning for low-power devices, and Muksimova et al. [10] integrated Miti-DETR with HE to improve feature extraction and training stability. Other studies enhanced detection speed and real-time efficiency using YOLOv5, YOLOv7, YOLOv8 [2,3], and SE-ResNeXt backbones [9].

However, challenges remain in balancing computational efficiency, detection accuracy, and robustness under complex real-world conditions, especially with smoke and small fire regions amid environmental disturbances like fog and varying illumination. The recent work of [14] introduces an adaptive hierarchical multi-headed CNN with a modified convolutional block attention mechanism to improve aerial forest fire detection precision. Although this method shows promising accuracy, its high computational complexity may limit practical real-time deployment on resource-constrained platforms.

In contrast, this study prioritizes evaluating YOLOv11's performance when combined with different image enhancement methods, including HE [11], CLAHE [12], and DBST-LCM CLAHE [13], to improve detection precision and robustness while maintaining computational efficiency suitable for real-time wildfire detection. The empirical results highlight YOLOv11's strong potential as an efficient single-shot detector, with DBST-LCM CLAHE showing superior precision by adaptively enhancing local contrast without excessive noise amplification, advancing beyond traditional HE and CLAHE techniques.

Thus, this work contributes valuable benchmarking and practical insights into how state-of-the-art detection models can be effectively paired with image enhancement preprocessing to address forest and land fire detection challenges in realistic environments, complementing existing research focused on architectural innovations. Unlike recent methods that focus primarily on model complexity and attention mechanisms [10–12], our study explores a lightweight yet effective combination of modern object detection and preprocessing techniques to achieve a balance between accuracy and real-time feasibility.

The structure of this report is as follows: the theoretical foundation of this study will be explained in Related Work. The materials and methods used will then be discussed in [Section 3](#), Methodology. [Section 4](#)—Results and Discussions—will present the experimental results, analysis, comparisons, and insights related to FNF detection. Finally, [Section 5](#), Conclusions, will summarize the findings and explore potential directions for future research.

2 Related Work

2.1 Forest and Land Fires (FLF) Detection

The process of identifying and recognizing early indicators of FLF, such as localized temperature spikes and the appearance of smoke from fire sources, through the use of visual analysis technology is commonly referred to as FLF detection. By analyzing images to assess the size and density of fire and smoke in forest and land areas, FLF detection has become a critical component of early disaster warning systems and global forest conservation. Indonesian FLF Prevention Patrol System [15] serves as a real-world example of early FLF detection implementation.

Li et al., in their study, combined AFN and DANN models to improve the detection accuracy of small-sized smoke in complex forest scenarios [8]. This method showed significant improvements in both detection accuracy and generalization ability, particularly in reducing false alarms. Such capability is critical to ensure detection systems can identify potential fires even from the early appearance of smoke before it spreads and becomes uncontrollable. This is especially crucial in tropical forests, like those in Indonesia, where vegetation is dense.

In addition, the YOLO model has become one of the most widely adopted techniques for developing real-time FLF detection systems [16]. Utilizing YOLOv5, v6, and v8 for smoke localization, Goncalves et al. [2] still encountered challenges in detecting small fires under visually disturbed conditions caused by haze and sunlight. Caixiong Li et al. [9] further explored YOLOv8's capability in recognizing varying fire sizes, although the method required high computational resources. This can hinder FLF detection efficiency in real-time scenarios with limited computing power.

On the other hand, Venâncio et al. [5] addressed this issue by combining YOLOv4 with pruning filters, which demonstrated YOLO's strong potential in FLF detection. However, this still needs further testing under underexplored data conditions. A clear example would be maximizing early-stage fire detection when smoke or fire visuals are faint and scattered, while simultaneously avoiding false alarms. To support effective early warning systems and FLF mitigation, developing an approach that can maintain accuracy under such conditions becomes highly crucial.

2.2 Histogram Equalization (HE)

Histogram Equalization (HE) is an image enhancement technique that improves visual quality by equalizing contrast through the redistribution of pixel intensity values. It has been successfully applied in medical and forestry imaging, particularly in improving Signal-to-Noise Ratio (SNR) metrics [11,17]. However, HE has limitations in scenes with uneven lighting, where global adjustments may fail to enhance local details.

To overcome this, Adaptive Histogram Equalization (AHE) was introduced, operating locally on sub-regions of the image. Yet, AHE often excessively amplifies noise. CLAHE (Contrast Limited Adaptive Histogram Equalization) improves upon AHE by applying a clip limit to the histogram, thus preserving visual stability while enhancing local contrast [18]. CLAHE is particularly useful in handling varying illumination

levels and textured backgrounds, making it more suitable for complex forest scenes compared to traditional HE or Retinex-based methods.

Recent studies have also explored hybrid techniques such as Fuzzy Contrast Enhancement (FCE) and learning-based histogram models to further improve detail visibility in low-resolution or noisy images [19]. Applications in plant disease detection have demonstrated that HE and CLAHE can improve classification accuracy by clarifying subtle patterns and textures during preprocessing [20,21].

To address dynamic scene complexity and preserve fine details, DBST-LCM (Dynamic Block Size Technique-Local Contrast Modification) was developed. This method adapts enhancement parameters by dynamically selecting block sizes based on image features and applying localized adjustments. It then performs CLAHE, followed by a feedback-driven quality check to ensure contrast clarity and sharpness in complex backgrounds [13]. Unlike conventional methods, DBST-LCM provides both adaptability and structure-aware enhancement, making it especially effective for detecting subtle smoke or fire signatures in challenging FLF conditions.

These three techniques—HE, CLAHE, and DBST-LCM—were chosen for this study due to their progressive improvements in enhancing visual cues critical for fire and smoke detection under varying illumination and environmental conditions. Their performance will be evaluated comprehensively to determine their suitability for real-time FLF detection.

2.3 YOLOv11

YOLOv11 is the latest evolution in the YOLO (You Only Look Once) architecture series, developed to improve real-time object detection by enhancing both accuracy and computational efficiency. Key innovations in this version include the C3K2 block and the C2PSA attention module, which are designed to optimize feature extraction and attention to critical image regions. These improvements allow YOLOv11 to achieve better inference speed and precision compared to its predecessors—YOLOv8, YOLOv9, and YOLOv10—without significantly increasing model complexity. In this research, the official open-source implementation of YOLOv11 provided by Ultralytics (available at <https://docs.ultralytics.com/models/yolo11/>, accessed on 17 June 2024) has been adopted to ensure reproducibility and consistency with the original architecture.

Here are the key features of YOLOv11: (1) Efficient Feature Extraction with the C3K2 Block: The C3K2 block is an enhancement of the Cross Stage Partial (CSP) architecture used in previous versions. By using two small convolutions (3×3 kernel) instead of one large convolution, the C3K2 block maintains feature extraction performance while reducing the number of parameters and computational load [22]. (2) Improved Spatial Attention with C2PSA: C2PSA (Cross Stage Partial with Spatial Attention) is a new module that introduces a spatial attention mechanism to help the model focus more on major areas in the image, such as small objects or partially occluded objects. This improves the model's sensitivity to spatial variations in the image [23]. (3) Multi-Scale Feature Combination through SPPF: Like previous versions, YOLOv11 retains the Spatial Pyramid Pooling Fast (SPPF) module, which combines features from various scales to enhance the detection of both small and large objects [24]. (4) CBS Blocks for Inference Stability: YOLOv11 also uses a Convolution-BatchNorm-SiLU (CBS) arrangement in the head section to ensure stable and effective data flow, supporting more accurate bounding box prediction and classification [23].

As part of the Single Shot Detector (SSD) architecture family, YOLOv11 performs object detection in a single forward pass, eliminating the need for region proposal stages found in two-stage detectors like Faster R-CNN. This makes YOLOv11 highly efficient and well-suited for real-time applications, including wildfire detection, where rapid response is critical. Combined with its enhanced modules—such as C3K2, C2PSA, and SPPF—YOLOv11 achieves robust performance in detecting fire and smoke under challenging

conditions like low light, varied object scales, and partial occlusion. These improvements make it a robust and practical solution for early forest and land fire monitoring systems, particularly when deployed in real-time surveillance setups using UAVs (drones) or edge devices in high-risk areas such as peatlands or remote conservation forests.

3 Methodology

3.1 D-Fire Dataset

Focused on effective early detection of fire and smoke in real-time conditions, this experiment is based on real images representing fire and smoke events from various environmental conditions. These include scenarios with only fire, only smoke, a combination of both, and negative examples without fire or smoke but with visual elements that might be misinterpreted. Based on these four scenarios, Venancio et al. [25] developed the D-Fire dataset, as outlined in Table 2. D-Fire consists of 21,527 labeled images categorized accordingly. Although the fire category contains fewer images, it has a higher annotation density, with an average of 2.52 fire objects per image. In contrast, smoke objects in the smoke and fire-and-smoke categories have an average of 1.13 annotations per image. In total, the dataset includes 26,557 bounding boxes: 14,692 labeled as fire and 11,865 as smoke, as shown in Table 3.

Table 2: The distribution of images per scenario category in the D-Fire dataset

Scenario	Description of scenario	Image
Fire	Images containing only fire	1164
Smoke	Images containing only smoke	5867
Fire and smoke	Images containing both fire and smoke	4658
None	Images containing neither fire nor smoke	9838
Total of images		21,527

Table 3: Number of annotations and density per image in the D-Fire dataset

Object category	Bounding box	Average per-image
Fire	14,692	2.52
Smoke	11,865	1.13
Total	26,557	

To support variation, the images were collected from several sources, including internet searches, legal fire simulations at the Technological Park of Belo Horizonte (Brazil), surveillance camera footage from Universidade Federal de Minas Gerais (UFMG), and Serra Verde State Park [26]. Additionally, some synthetic images were generated using montage techniques by overlaying artificial smoke onto green landscape backgrounds with photo editing software to simulate real forest conditions. Fig. 1 visually represents the diversity within the dataset, highlighting different instances of fire and smoke along with their corresponding ground-truth labels. The dataset captures a broad spectrum of scene types—such as forests, parks, and semi-urban environments—as well as variations in camera angles, smoke density, and lighting conditions.

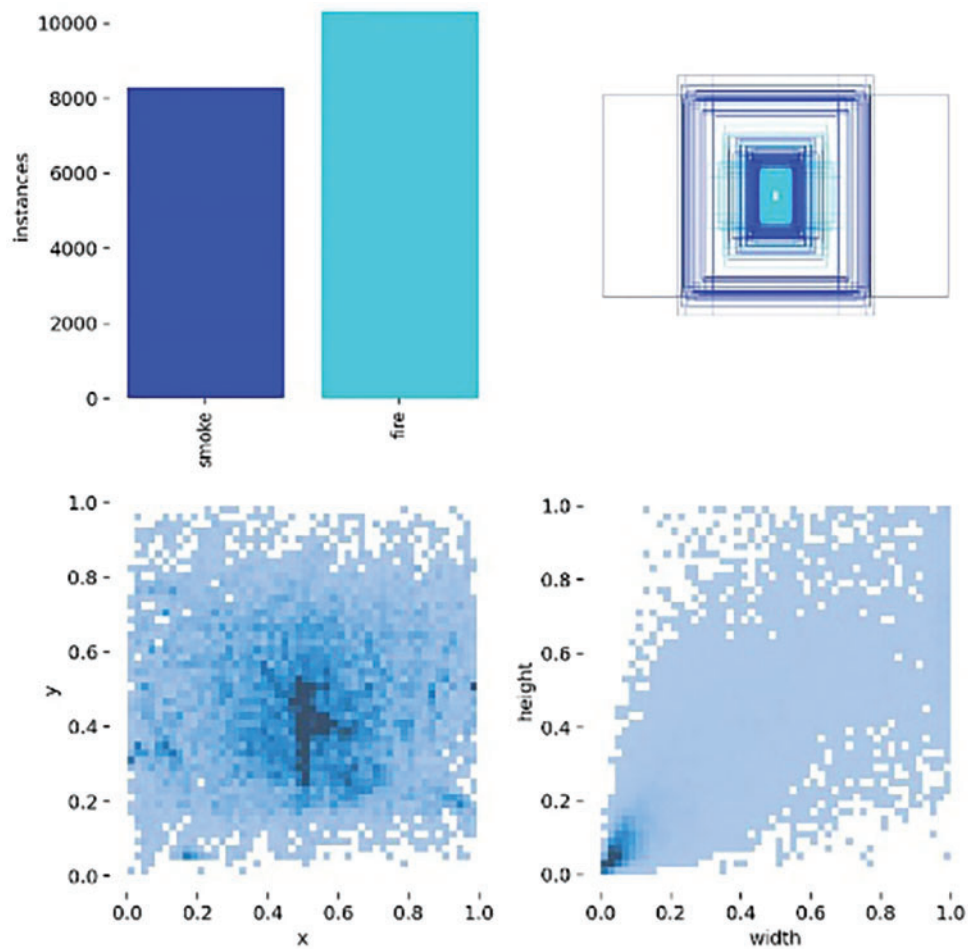


Figure 1: D-Fire dataset instances

This diversity allows for a comprehensive evaluation of image enhancement methods, especially in challenging scenarios such as diffused smoke, reduced visibility, and fluctuating lighting conditions. As a result, D-Fire is particularly well-suited for testing contrast-based techniques aimed at enhancing feature clarity in complex visual environments.

3.2 Experiment Setting

The D-Fire dataset is initially divided into two parts with an 80:20 ratio for training and testing. We then combined these two folders before randomly splitting the dataset again into a 70:20:10 ratio for training, validating, and testing. Afterward, we applied 3 image enhancement methods and categorized our dataset into 3 groups: X Dataset, Y Dataset, and Z Dataset. X Dataset consists of images that were applied to HE (Histogram Equalization), which works by equalizing the pixel intensity distribution in the image to flatten the image contrast. The Y dataset contains images that have been applied to CLAHE (Contrast Limited Adaptive HE). Improving upon HE, CLAHE divides the image into smaller blocks and limits contrast amplification to avoid excessive noise in the same areas. Finally, the Z Dataset consists of images that have been applied DBST-LCM CLAHE (Dynamic Block Size Technique-Local Contrast Modification), where, before CLAHE, a combination of noise reduction based on shift transformation (DBST) and local contrast

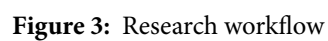
modification (LCM) is applied. Fig. 2 shows examples of images that have undergone image enhancement with (a) HE, (b) CLAHE, and (c) DBST-LCM CLAHE.



Figure 2: Examples of image enhancement methods with (a) HE, (b) CLAHE, and (c) DBST-LCM CLAHE

We implemented the YOLOv11x variant for forest and land fire detection, utilizing a configuration of $\text{depth_multiple} = 1.00$, $\text{width_multiple} = 1.50$, and $\text{max_channels} = 512$. These values were not the result of empirical tuning but were adopted directly from the official Ultralytics YOLOv11 model configuration as the standard settings for the YOLOv11x variant. This ensures consistency with the original reference implementation and benchmark results. The model processes $640 \times 640 \times 3$ input images through a deep convolutional pipeline. It features down-sampling stages via alternating Conv and C3 modules, which progressively reduce spatial resolution while increasing feature depth. The SPPF block aggregates multi-scale context before passing features to the neck, where they are unsampled and concatenated to enhance semantic richness. Final detection heads operate at resolutions of 80×80 , 40×40 , and 20×20 to detect small, medium, and large-scale objects, respectively. The architecture's combination of C3K2, SPPF, and C2PSA modules improves feature extraction efficiency and attention precision, capabilities particularly valuable for detecting subtle visual cues such as smoke. Fig. 3 illustrates the research workflow that is based on YOLOv11 architecture with customized depth, width, and max channels [22].

Model training was conducted on Google Collaboratory using an NVIDIA A100-SXM4-40GB GPU. The training process spanned 100 epochs with a batch size of 6 and an input resolution of 640×640 pixels. The selection of these hyperparameters was based on hardware availability and training stability considerations, in line with practices found in related studies [25].



3.3 Training Result

Table 4 shows the results of training the YOLOv11x model on the D-Fire dataset, which has been processed using three different image enhancement techniques: Histogram Equalization (HE), CLAHE, and DBST-LCM CLAHE. During the training process, X Dataset (HE) achieved the best performance with an mAP50 score of 0.771, followed by Z Dataset (DBST-LCM CLAHE) with 0.770, and Y Dataset (CLAHE) with 0.759. These results provide an initial indication that the HE method offers strong support in the object detection model training process, even when compared to more complex methods like DBST-LCM CLAHE.

Table 4: Training results of the D-Fire dataset with YOLOv11x

Class	Images	Instances	HE (X Dataset)			CLAHE (Y Dataset)			DBST-LCM CLAHE (Z Dataset)		
			P	R	mAP50	P	R	mAP50	P	R	mAP50
Smoke	2127	2419	0.834	0.791	0.84	0.822	0.79	0.833	0.839	0.788	0.846
Fire	1167	2993	0.735	0.615	0.703	0.704	0.6	0.685	0.741	0.605	0.694
All	4305	5412	0.784	0.703	0.771	0.763	0.695	0.759	0.79	0.697	0.77

The evaluation of detection results is based on the Intersection over Union (IoU) value, which calculates the overlap ratio between the predicted box and the ground truth box. The formula can be seen in Eq. (1):

$$IoU = \frac{Area_{pred} \cap Area_{gt}}{Area_{pred} \cup Area_{gt}} \quad (1)$$

The prediction outputs are then classified into True Positive (TP), False Positive (FP), and False Negative (FN). This classification serves as the basis for calculating Precision and Recall, as shown in Eqs. (2) and (3) [27]. Another evaluation indicator is the F1 Score, which is the harmonic mean of Precision and Recall, as shown in Eq. (4).

$$Precision (P) = \frac{TP}{TP + FP} \quad (2)$$

$$Recall (R) = \frac{TP}{TP + FN} \quad (3)$$

Based on the training results on the X, Y, and Z datasets, the model showed consistent detection performance, especially for the Smoke class. The highest mAP50 score for this class was achieved with the Z Dataset, reaching 0.846. In terms of Precision and Recall, the Smoke class in the Z Dataset recorded the highest Precision (0.839), while the highest Recall (0.791) was obtained from the X Dataset. On the other hand, for the Fire class, the highest mAP50 score was achieved with X Dataset (0.703), with the highest Precision (0.741) recorded in Z Dataset and the highest Recall (0.615) in X Dataset. Another evaluation index, F1, is shown in Eq. (4).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

The training processes for each enhanced dataset are visualized in Fig. 4 (X Dataset for HE), Fig. 5 (Y Dataset for CLAHE), and Fig. 6 (Z Dataset for DBST-LCM CLAHE). Across more than 30 epochs,

the loss values for all datasets consistently declined, indicating successful convergence of the model parameters. Throughout training, the key components of the YOLOv11x loss function—namely, box loss (localization error), classification loss, and distribution focal loss (bounding box refinement)—exhibited a steady downward trend, reflecting stable and effective optimization during training. These loss components correspond to the model's efforts in improving bounding box localization, object confidence prediction, and class classification performance, which are essential for accurate fire and smoke detection in various environmental conditions.

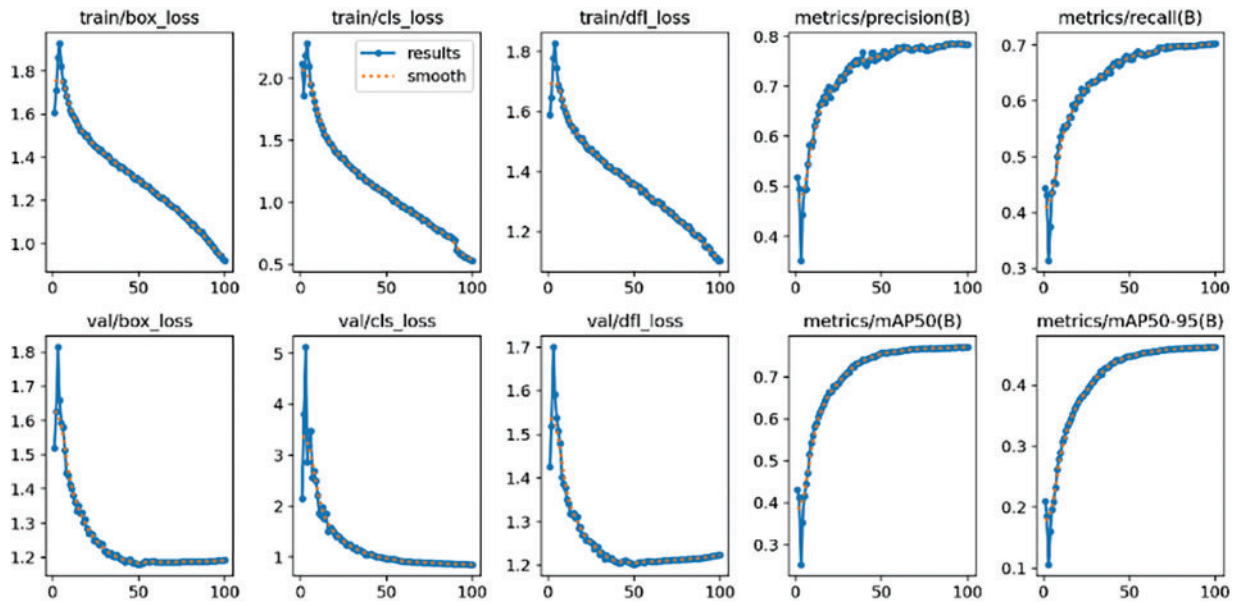


Figure 4: Training process YOLOv11x with X Dataset (HE)

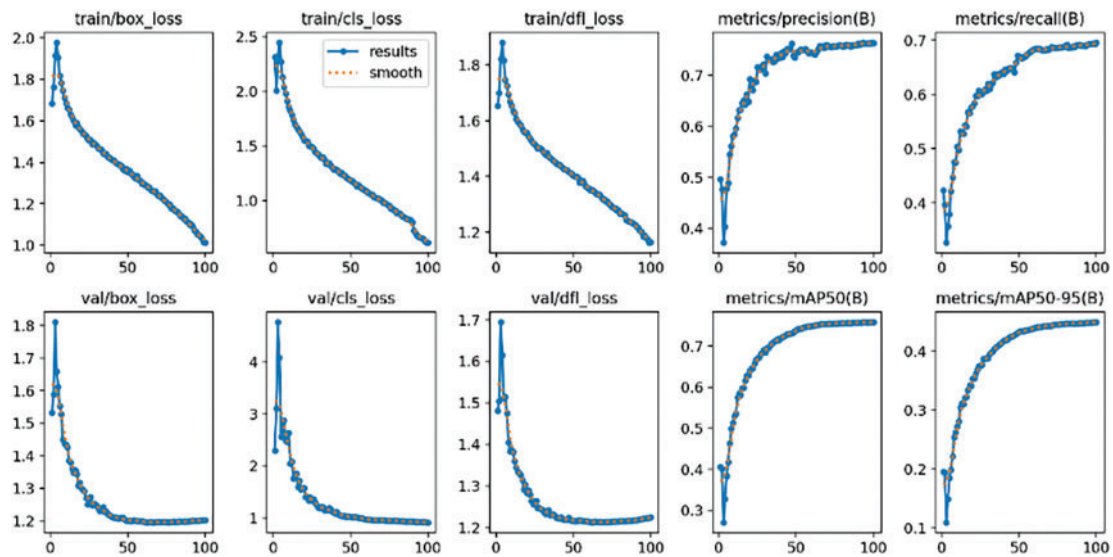


Figure 5: Training process YOLOv11x with Y Dataset (CLAHE)

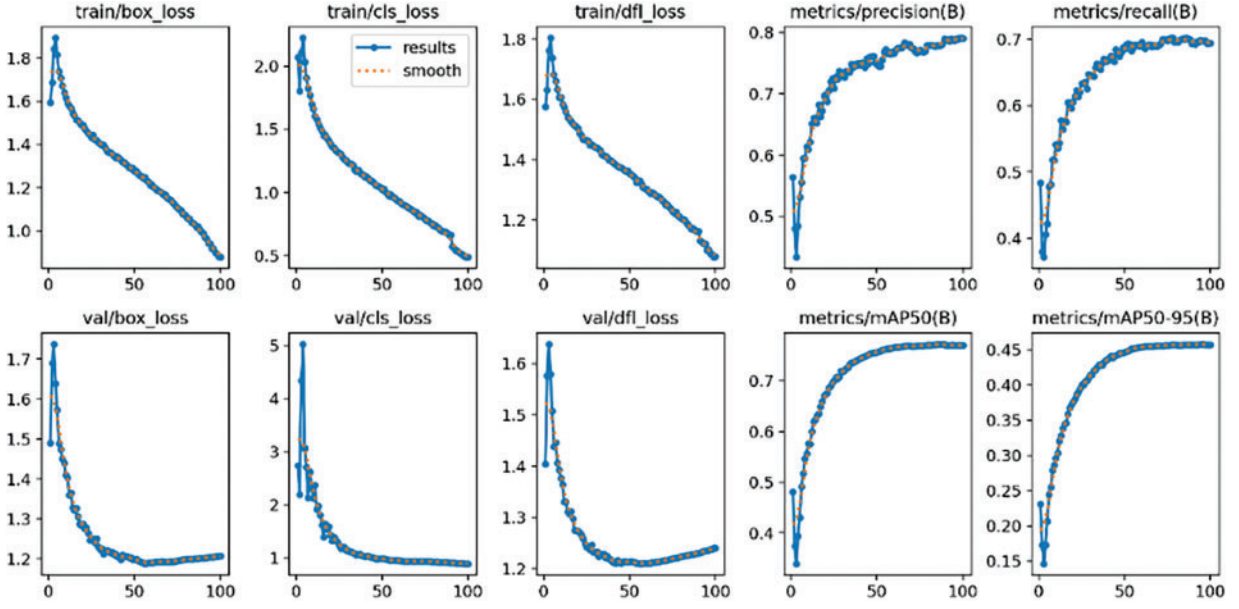


Figure 6: Training process YOLOv11x with Z Dataset (DBST-LCM CLAHE)

While the overall trends were similar, minor differences in the loss values were observed across the three datasets. The model trained with the Z Dataset showed slightly lower overall loss values, particularly in the validation box and classification losses, suggesting better optimization. The Y Dataset achieved moderate loss reductions, whereas the X Dataset maintained slightly higher losses throughout training. These differences highlight the potential benefits of advanced image enhancement techniques in improving model performance.

$$\begin{aligned}
 \mathcal{L} = & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B t_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (\psi_i - \hat{\psi}_i)^2 \right] \\
 & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B t_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] + \sum_{i=0}^{S^2} \sum_{j=0}^B t_{ij}^{obj} (c_i - \hat{c}_i)^2 \\
 & + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B t_{ij}^{noobj} (c_i - \hat{c}_i)^2 + \sum_{i=0}^{S^2} t_{ij}^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2
 \end{aligned} \quad (5)$$

As described in Eq. (5), the YOLOv11x model adopts a loss (\mathcal{L}) function structure based on the basic YOLO architecture. This loss function minimizes errors in predicting bounding box parameters. (x, ψ, w, h) , object confidence \mathcal{C} , and class probabilities $p(c)$ [27]. Here, the symbols $\hat{x}, \hat{\psi}, \hat{w}, \hat{h}, \hat{c}, \hat{p}$ represent the predicted values, while the corresponding ground truth labels are denoted without a hat. The indicator t_{ij}^{obj} identifies whether an object exists within a cell i and whether the j -th bounding box predictor is responsible for that detection. In the training configuration, the hyperparameter λ_{coord} was set to 0.5, assigning moderate weight to localization errors. Similarly, λ_{noobj} was set to 0.5 to reduce the impact of prediction in grid cells without objects, preventing distraction from large background regions.

YOLOv11x introduces significant advancements in real-time object detection, emphasizing both efficiency and precision. Its architecture integrates modules such as the C3K2 block for lightweight feature extraction and the C2PSA module for enhanced spatial attention. During training, YOLOv11x employed a combination of data augmentation techniques, including HSV augmentation, random rotation, translation, perspective transformation, scaling, and both vertical and horizontal flipping. Additionally, advanced augmentations such as Mosaic and MixUp were applied to improve generalization across diverse environmental conditions, with optional use of CutMix further enhancing training robustness.

4 Results and Discussions

To visualize the performance of the YOLOv11x model across different image enhancement techniques, Figs. 7–9 present the validation results for datasets enhanced using Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), and DBST-LCM CLAHE, respectively. These figures demonstrate the model's ability to detect fire and smoke under various visual conditions. The HE-enhanced dataset (Fig. 7) shows consistent detection, but with relatively moderate confidence scores. In contrast, the CLAHE-enhanced dataset (Fig. 8) reveals slightly improved detection clarity in several images but also includes some lower confidence values and misidentifications. The DBST-LCM CLAHE dataset (Fig. 9) displays the highest consistency in detection with generally stronger confidence levels, especially for both smoke and fire, indicating better contrast and feature enhancement.



Figure 7: Validation Batch 0 with YOLOv11x and X Dataset (HE)

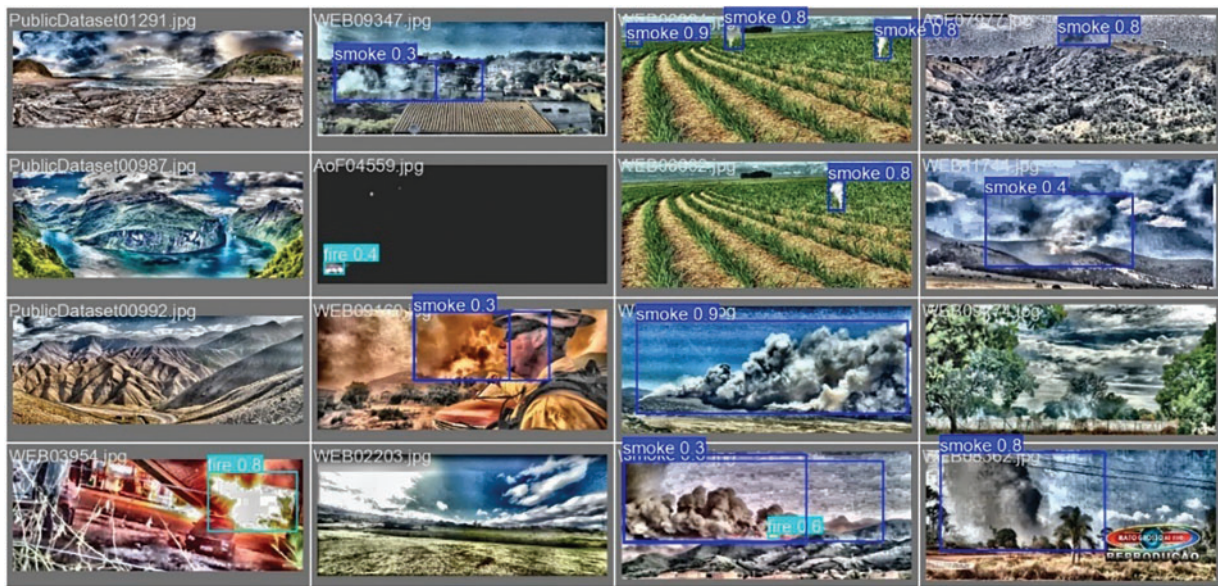


Figure 8: Validation Batch 0 with YOLOv11x and Y Dataset (CLAHE)

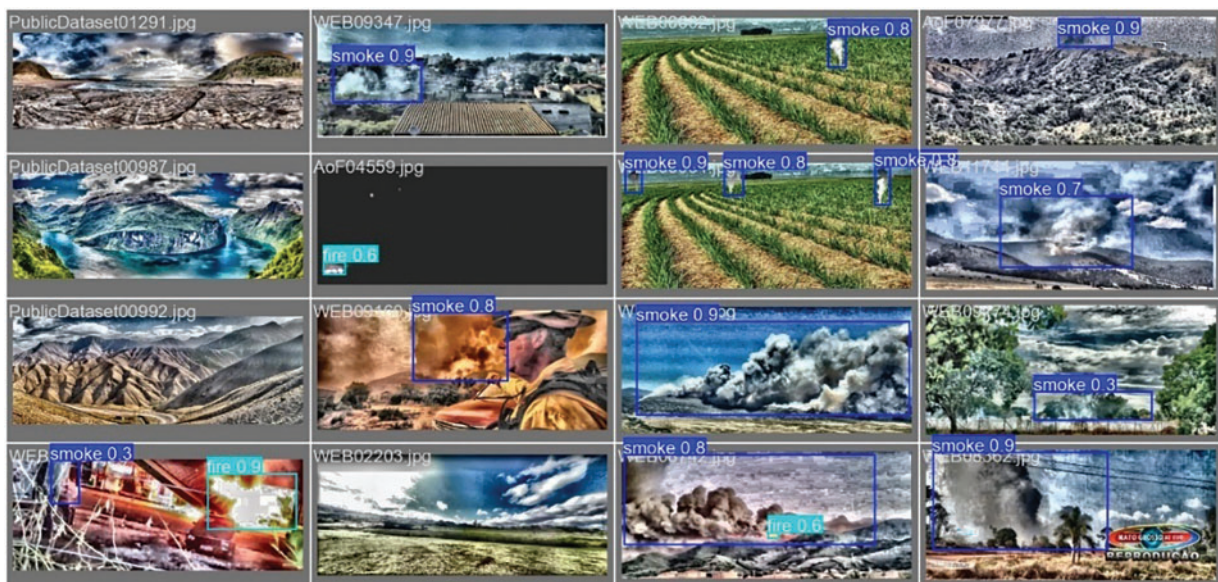


Figure 9: Validation Batch 0 with YOLOv11x and Z Dataset (DBST-LCM CLAHE)

To further support this observation, [Fig. 10](#) presents key evaluation metrics for the DBST-LCM CLAHE-enhanced dataset. The precision-recall (PR) curve shows high precision values for smoke (0.864) and fire (0.694), contributing to a mean Average Precision at 50% Intersection over Union (mAP@50) of 0.770 across all classes. The F1 curve peaks at 0.74 at a confidence threshold of 0.353, indicating a good balance between precision and recall. The confusion matrix supports this finding, with 2016 out of 2047 smoke instances correctly identified (98.49% true positive rate) and 2071 out of 2179 fire instances accurately detected (95.63%). Although 905 background instances were misclassified as fire, the primary object classes remained

distinguishable. These results highlight the capability of DBST-LCM CLAHE in improving both detection accuracy and object localization, especially under challenging visual conditions involving dispersed smoke.

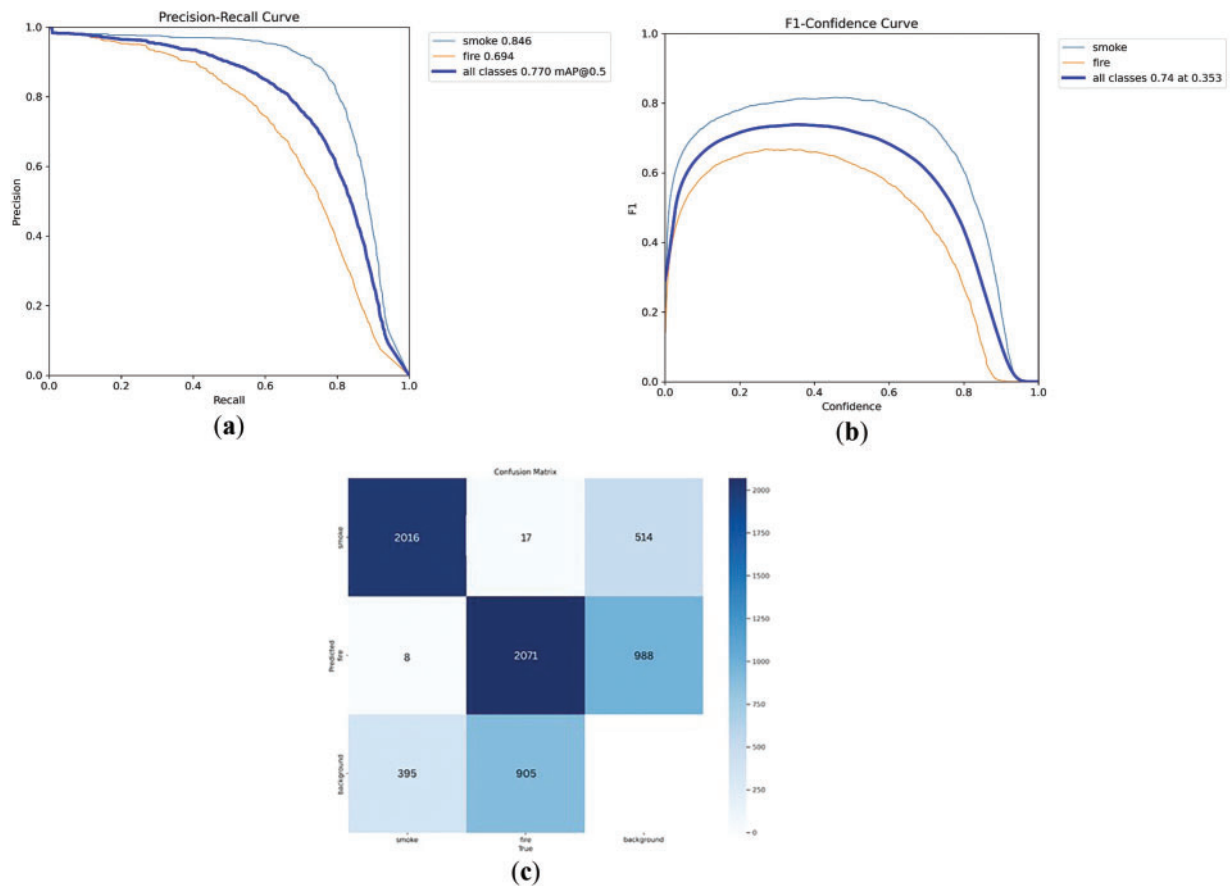


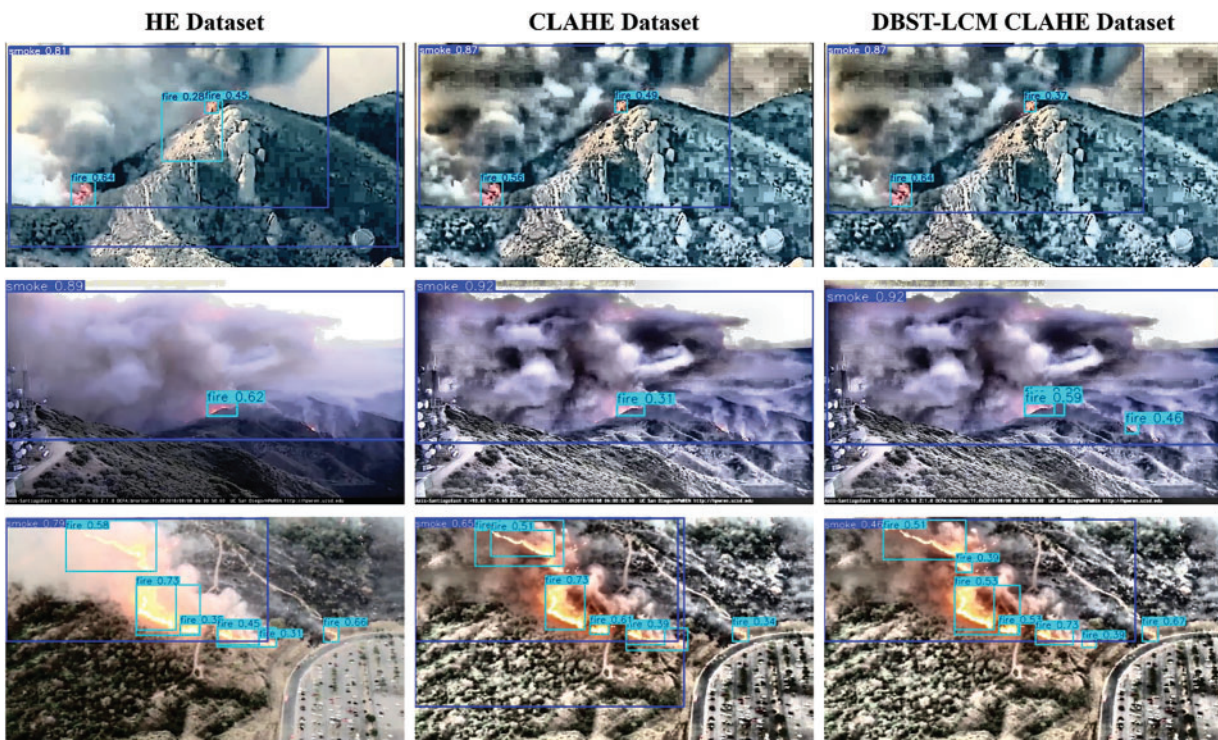
Figure 10: Performance of DBST-LCM CLAHE based on precision-recall (PR) curves (a), per-class F1-scores (b), and confusion matrix (c)

Table 5 presents the validation or testing results of the YOLOv11x model on the three image-enhanced datasets. Z Dataset achieved the highest Precision score of 0.79, indicating its strong ability to correctly identify fire and smoke with fewer false positives. Meanwhile, X Dataset maintained the highest mAP50 score of 0.771 and the highest Recall at 0.703, showing better overall detection accuracy and generalization. This suggests that while DBST-LCM CLAHE (applied on Z Dataset) is highly precise, HE (applied on X Dataset) remains more balanced and effective in capturing a broader range of fire and smoke instances, especially in varied or low-light conditions. CLAHE, although beneficial for enhancing local contrast, showed slightly lower performance in comparison to the other two techniques.

Table 5: Validating the result of the D-Fire dataset with YOLOv11x

Class	Images	Instances	HE (X Dataset)			CLAHE (Y Dataset)			DBST-LCM CLAHE (Z Dataset)		
			P	R	mAP50	P	R	mAP50	P	R	mAP50
Smoke	2127	2419	0.834	0.791	0.84	0.821	0.791	0.833	0.839	0.788	0.846
Fire	1167	2993	0.735	0.615	0.703	0.706	0.601	0.685	0.741	0.605	0.694
All	4305	5412	0.784	0.703	0.771	0.763	0.695	0.759	0.79	0.697	0.77

In addition, the detection performance of YOLOv11x on each dataset is visualized in Fig. 11. When comparing the different image enhancement techniques, DBST-LCM CLAHE provides better object localization for complex smoke features, especially under varying lighting and contrast conditions. This is noticeable in the second and third rows of predictions, where smoke areas are more comprehensively mapped. However, in some cases, such as those shown in the first column of Fig. 11, X Dataset allows the model to detect fire and smoke more completely and with higher confidence scores, likely due to the global contrast enhancement provided by HE.

**Figure 11:** Recognition results with YOLOv11x

Each image enhancement method contributes uniquely to the detection process. (1) Histogram Equalization (HE) improves global contrast, making image features easier to recognize in dark or low-light areas. (2) CLAHE enhances local contrast and preserves excellent details, supporting more precise detection in regions with small intensity variations. (3) DBST-LCM CLAHE, as a hybrid method, balances both local and

global contrast enhancement, making it effective for handling complex scenarios such as smoke dispersion. (4) Additionally, the DBST preprocessing helps suppress background interference, while the LCM operation helps maintain object structure. (5) Although more computationally complex, the contribution of DBST-LCM CLAHE, particularly its highest precision score, demonstrates its strength in reducing false positives and improving detection robustness under real-world conditions.

Overall, all three enhancement techniques have proven to improve detection quality in diverse ways. Histogram Equalization remains an efficient and practical approach with the highest overall detection score. However, for applications that require high sensitivity and precision, such as early smoke detection, DBST-LCM CLAHE shows promising results and potential for further development.

5 Conclusions

This study aimed to empirically evaluate the effectiveness of three image enhancement techniques—Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), and DBST-LCM CLAHE—on the object detection performance of the YOLOv11x model for early Forest and Land Fire (FLF) detection. Using the D-Fire dataset, which includes over 21,000 annotated images representing varied times and weather conditions, the model was trained and tested to reflect realistic environmental challenges.

Results show that all three enhancement techniques positively impact detection accuracy, especially in low-illumination and high-noise scenarios. HE demonstrated the highest mean Average Precision at 50% IoU (mAP50) of 0.771, along with a balanced performance in Precision (0.784) and Recall (0.703), indicating strong generalization across fire and smoke cases. In contrast, DBST-LCM CLAHE achieved the highest Precision score (0.790), reducing false positives and demonstrating superior robustness in complex scenes, such as dispersed smoke. CLAHE performed slightly lower but remained valuable in enhancing local feature details.

The detection visualizations further confirm that each enhancement method contributes uniquely: HE improves global contrast, aiding overall detection; CLAHE accentuates local details; while DBST-LCM CLAHE successfully combines both strategies, with added benefits from DBST in noise suppression and LCM in structure preservation. These differences suggest that the enhancement method can be selected based on detection priority—completeness, precision, or clarity in complex backgrounds.

Importantly, this study contributes not by proposing new architectures but by benchmarking YOLOv11x—one of the most recent lightweight detectors—under realistic FLF conditions enhanced by proven preprocessing strategies. The inclusion of diverse evaluation metrics such as F1-score, precision-recall curves, and qualitative assessments (e.g., bounding box visualizations) strengthens the credibility of these findings.

For future development, one valuable direction would be the deployment of the system on edge computing platforms, evaluating real-time inference speed, power consumption, and detection latency. Additionally, training a localized model using region-specific data, such as forest imagery from Indonesia, could further improve detection accuracy by incorporating native vegetation types, fire patterns, and regional atmospheric conditions.

Ultimately, this research provides practical insights into how advanced image enhancement techniques, when combined with modern detectors like YOLOv11x, can deliver high-performance FLF detection systems suited for early warning applications in real-world environments.

Acknowledgement: The authors thank the Vice-Rector of Research, Innovation, and Entrepreneurship at Satya Wacana Christian University.

Funding Statement: This research was funded by the Directorate of Research, Technology, and Community Service, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia the Regular Fundamental Research scheme, with grant numbers 001/LL6/PL/AL.04/2025, 011/SPK-PFR/RIK/05/2025.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Christine Dewi; Melati Viaeritas Vitrieco Santoso; Abbott Po Shun Chen; Hanna Prillysca Chernovita; data collection: Melati Viaeritas Vitrieco Santoso; Stephen Abednego Philemon; analysis and interpretation of results: Evangs Mailoa; Christine Dewi; Hanna Prillysca Chernovita; Abbott Po Shun Chen; draft manuscript preparation: Abbott Po Shun Chen; Christine Dewi; Hanna Prillysca Chernovita; Evangs Mailoa; Stephen Abednego Philemon. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: D-Fire: an image dataset for fire and smoke detection (<https://github.com/gaias/dFireDataset>, accessed on 1 December 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Akter S, Grafton RQ. Do fires discriminate? Socio-economic disadvantage, wildfire hazard exposure and the Australian 2019–20 ‘Black Summer’ fires. *Clim Change*. 2021;165(3):53. doi:10.1007/s10584-021-03064-6.
2. Gonçalves LAO, Ghali R, Akhloufi MA. YOLO-based models for smoke and wildfire detection in ground and aerial images. *Fire*. 2024;7(4):140. doi:10.3390/fire7040140.
3. Wicaksono P. Deep learning wildfire detection to increase fire safety with YOLOv8. *Int J Intell Syst Appl Eng*. 2024;12(3):4383.
4. Alkhamash EH. A comparative analysis of YOLOv9, YOLOv10, YOLOv11 for smoke and fire detection. *Fire*. 2025;8(1):26. doi:10.3390/fire8010026.
5. de Venâncio PVAB, Lisboa AC, Barbosa AV. An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Comput Appl*. 2022;34(18):15349–68. doi:10.1007/s00521-022-07467-z.
6. Pan J, Ou X, Xu L. A collaborative region detection and grading framework for forest fire smoke using weakly supervised fine segmentation and lightweight faster-RCNN. *Forests*. 2021;12(6):768. doi:10.3390/f12060768.
7. Perrolas G, Niknejad M, Ribeiro R, Bernardino A. Scalable fire and smoke segmentation from aerial images using convolutional neural networks and quad-tree search. *Sensors*. 2022;22(5):1701. doi:10.3390/s22051701.
8. Li T, Zhang C, Zhu H, Zhang J. Adversarial fusion network for forest fire smoke detection. *Forests*. 2022;13(3):366. doi:10.3390/f13030366.
9. Li C, Du Y, Zhang X, Wu P. YOLOGX: an improved forest fire detection algorithm based on YOLOv8. *Front Environ Sci*. 2025;12:1486212. doi:10.3389/fenvs.2024.1486212.
10. Muksimova S, Umirzakova S, Mardieva S, Abdullaev M, Cho YI. Revolutionizing wildfire detection through UAV-driven fire monitoring with a transformer-based approach. *Fire*. 2024;7(12):443. doi:10.3390/fire7120443.
11. Farhan MS, Sthevanie F, Ramadhani KN. Video based fire detection method using CNN and YOLO Version 4. *Indones J Comput*. 2022;7(2):65–78. doi:10.34818/INDOJC.2022.7.2.654.
12. Ayumi V, Noprisson H, Ani N. Forest fire detection using transfer learning model with contrast enhancement and data augmentation. *J Nas Pendidik Teknik Inform*. 2024;13(1):1–10. doi:10.23887/janapati.v13i1.75692.
13. Chakraverti S, Agarwal P, Pattanayak HS, Chauhan SPS, Chakraverti AK, Kumar M. De-noising the image using DBST-LCM-CLAHE: a deep learning approach. *Multimed Tools Appl*. 2024;83(4):11017–42. doi:10.1007/s11042-023-16016-2.
14. Mowla MN, Asadi D, Tekeoglu KN, Masum S, Rabie K. UAVs-FFDB: a high-resolution dataset for advancing forest fire detection and monitoring using unmanned aerial vehicles (UAVs). *Data Brief*. 2024;55(1):110706. doi:10.1016/j.dib.2024.110706.

15. Sitanggang IS, Syaufina L, Trisminingsih R, Ramdhany D, Nuradi E, Hidayat MFA, et al. Indonesian forest and land fire prevention patrol system. *Fire*. 2022;5(5):136. doi:10.3390/fire5050136.
16. Saleh A, Zulkifley MA, Harun HH, Gaudreault F, Davison I, Spraggon M. Forest fire surveillance systems: a review of deep learning methods. *Heliyon*. 2024;10(1):e23127. doi:10.1016/j.heliyon.2023.e23127.
17. Winarno G, Irsal M, Karenina CA, Sari G, Hidayati RN. Metode histogram equalization untuk peningkatan kualitas citra dengan menggunakan studi phantom lumbosacral. *J Kesehat*. 2022;7(2):104. doi:10.22146/jkesvo.71469.
18. Härtinger P, Steger C. Adaptive histogram equalization in constant time. *J Real Time Image Process*. 2024;21(3):93. doi:10.1007/s11554-024-01465-1.
19. Singh P, Ganotra D. Histogram based resolution enhancement of an image by using artificial neural network. In: *Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*; 2021 Apr 8–10; Erode, India. doi:10.1109/ICCMC51019.2021.9418295.
20. Sai VT, Sai Akhil NE, Jashnavi TJM, Kanakala NVK. Image quality enhancement for wheat rust diseased leaf image using histogram equalization & CLAHE. *E3S Web Conf*. 2023;391:01029. doi:10.1051/e3sconf/202339101029.
21. Sayyid MFN. Klasifikasi penyakit daun jagung menggunakan metode CNN dengan image processing HE Dan CLAHE. *J Tek Inform Dan Teknol Inf*. 2024;4(1):86–95. (In Indonesian). doi:10.55606/jutiti.v4i1.3425.
22. Ghosh A. YOLO11: redefining real-time object detection [Online]. [cited 2025 Jun 17]. Available from: <https://learnopencv.com/yolol1/>.
23. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements. *arXiv:2410.17725v1*. 2024.
24. Hidayatullah P, Syakrani N, Sholahuddin MR, Gelar T, Tubagus R. YOLOv8 to YOLO11: a comprehensive architecture in-depth comparative review. *arXiv:2504.12345*. 2025.
25. De Venâncio PVAB, Rezende TM, Lisboa AC, Barbosa AV. Fire detection based on a two-dimensional convolutional neural network and temporal analysis. In: *Proceedings of the 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*; 2021 Nov 2–4; Temuco, Chile. doi:10.1109/LA-CCI48322.2021.9769824.
26. Boroujeni SPH, Mehrabi N, Afghah F, McGrath CP, Bhatkar D, Biradar MA, et al. Fire and smoke datasets in 20 years: an in-depth review. *arXiv:2503.14552*. 2025.
27. Terven J, Cordova-Esparza DM, Romero-González JA, Ramírez-Pedraza A, Chávez-Urbiola EA. A comprehensive survey of loss functions and metrics in deep learning. *Artif Intell Rev*. 2025;58(7):195. doi:10.1007/s10462-025-11198-7.