RESEARCH ARTICLE

Automatic Pollution Detection on High-Voltage Isolators Using a Two-Phase Approach

Vladimir Maksimović[1,*] **and Branimir Jakšić**[1]

[1] Faculty of Technical Sciences, University of Pristina in Kosovska Mitrovica, Kneza Miloša 7, Kosovska Mitrovica 38220, Serbia

Abstract

The accumulation of atmospheric and industrial pollution of high-voltage insulators is one of the frequent problems and a pattern of failures in transmission systems. In this paper, a two-phase approach based on deep learning is proposed for the detection of pollution of high-voltage insulators. The proposed approach automatically detects three types of pollution (salt, soot and excrement) based on UAV images. In addition to detection, the classification of pollution is automatically done into three levels (low, medium and high pollution). In the first phase, the You Only Look Once (YOLO) detector is used for precise detection and isolation of insulators, where an average accuracy of mAP@0.5 of 93.75% is achieved. The model was trained on a Merged Public Insulator Dataset (MPID) database containing over 5000 insulators. The second phase utilizes the Zenodo dataset, which contains over 14,000 synthetic insulator images. In the second phase, the model was trained using the EfficientNet-B0 convolutional network to classify the type and level of pollution. In order to solve the problem of real data, fine-tuning was done for all 10 classes. The results show that the accuracy is 88% on a partial classification of 10 levels of pollution. When using 4 levels of pollution, the model achieves an accuracy of 91%. Additional automation was achieved with priority metrics, which by analyzing 100 images showed 24% of critical cases. The system determines the cleaning priority based on the pollution intensity, ensuring that critical cases are addressed first. A comparative analysis was performed when the model was trained with MobileNetV2, ResNet16 and VGG. The results show that the proposed model with the highest recall minimizes the risk of missed critical insulators. For example, in real-world applications, MobileNetV2 has a larger difference, which means too many false positives, while ResNet18 has a smaller difference, which means more false negatives which is a security risk.

Keywords: insulator detection, insulator pollution, deep learning, neural networks, multi-phase approach.

1 Introduction

Today, power systems are a key infrastructure for reliable transmission of electricity. The importance of these systems is reflected in the efforts of researchers to solve problems in such systems. One of the very common patterns on high-voltage lines is flashover on insulators caused by excessive accumulation of pollution [1].

Submitted: 28 January 2026
Accepted: 20 February 2026
Published: 27 February 2026

Vol. 2, **No.** 1, 2026.
10.62762/TEPNS.2026.648298

***Corresponding author:**
✉ Vladimir Maksimović
vladimir.maksimovic@pr.ac.rs

Citation
Maksimović, V., & Jakšić, B. (2026). Automatic Pollution Detection on High-Voltage Isolators Using a Two-Phase Approach. *ICCK Transactions on Electric Power Networks and Systems*, 2(1), 7–21.

© 2026 ICCK (Institute of Central Computation and Knowledge)

The standard approach to inspecting lines and insulators requires monitoring from the ground, i.e., patrolling or using helicopters, which is very expensive and demanding. Advances in technology have made it faster and easier to inspect energy systems, but also more efficient by inspecting unmanned aerial vehicles (UAVs) equipped with high-resolution cameras and artificial intelligence algorithms [2, 3]. This approach has made it possible to largely automate processes and find problems, such as inspecting power lines using a drone [4]. By using UAVs for inspection, costs are significantly reduced, and on the other hand, productivity and efficiency are increased. In the paper [5], a deep learning method for detecting insulator defects with an accuracy of over 90% is presented, while in the paper [6] it is shown that this approach can increase the detection speed with a reduction in the number of parameters by 46.9% and a reduction in floating point operations per second (FLOPs) by 43.0%, while maintaining a frame per second (FPS) of 63.6. Inspection has been revolutionized by the use of convolutional neural networks (CNNs), and in particular real-time object detectors such as YOLO. The authors in [7] and [8] successfully demonstrate the application of YOLO algorithms for the detection of physical damage and defects on insulators where they are located. However, such solutions are based on fault detection, but not on problems such as layers of pollution that can make causing much more difficult. In the paper [9], the analysis of power lines is presented, i.e. a post-processing algorithm is developed in order to create a metric of vegetation encroachment in power lines, enabling a quantitative estimation of the distribution of surrounding vegetation. However, the problem is also the pollution of the insulators, as in this paper proposed an approach to solve the mentioned problem. In [1], a multi-modal data fusion is shown that uses UAV footage to detect the state of the pollution. Stains such as salt, dust, water and snow were analysed, but the intensity of pollution was not analysed. The authors in [10] developed a drone robot for autonomous cleaning that is designed to identify an insulator using a camera and clean it using a robotic module. With the help of artificial intelligence, the state of cleanliness of the insulator is proposed in the paper [11], but the paper only discusses the type of pollution, not the intensity. Since there is no base with a real insulator pollution, the authors in [12] also used a synthetic base to estimate the purity of the insulator, where they achieved a 0.99% improvement over the original model [13]. Due to the exposure of the insulator, layers of pollution such as salt, dust, excrement or some other organic matter are formed. These layers of pollution form a conductive surface, especially in the presence of moisture, which significantly reduces the electrical strength of the insulator and increases the risk of flashover [14]. Using these approaches has shown that costs are reduced by 60-80% compared to traditional inspection methods and speed up the process by 3 to 5 times with a great reduction in risk [15, 16].

Analyzing the literature, the lack of fine-grained classification of type and level of pollution is clearly evident because existing systems are mainly based on binary classification or on defects. Also, it is necessary to form an end-to-end approach that will determine the priority of cleaning. Without automatic prioritization, human operators must manually review thousands of detections. Validation of existing results is based on standard metrics such as F1, Accuracy, Precision, Recall, so it is important to analyze how many images have at least one detection and whether the system gives false alarms. This paper proposes an approach to solve key problems related to insulator pollution as well as an end-to-end approach for fine-grained automatic detection, classification and prioritization of pollution on high-voltage insulators.

2 System Model

The proposed approach for automatic detection and classification of pollution on insulators is based on a two-phase architecture that sequentially processes UAV footage. The MPID [17] was used to train the system, which contains 5,000 images with real-world and multi-type images of the insulators taken by UAVs. The MPID database was used to train the YOLO model [18]. In addition, a Zenodo [19] database containing 14,432 images was used to train the EfficientB0 model for the classification and pollution level of insulators. Zenodo is a synthetically generated dataset, created with the help of a tool to simulate pollution in images of insulators. The Zenodo dataset contains 3,608 synthetic images per pollution category (clean, soot, salt and excrement) which is in total 14,432 images. Within each pollution category, the images are generated for three insulator materials (glass, porcelain and polymer), with approximately 1,202 images per material [19]. The aim of the Zenodo database is to enable the initial training of models under controlled conditions, where the classes and levels of pollution (clean, low, medium, high) are clearly marked for different types of pollution such

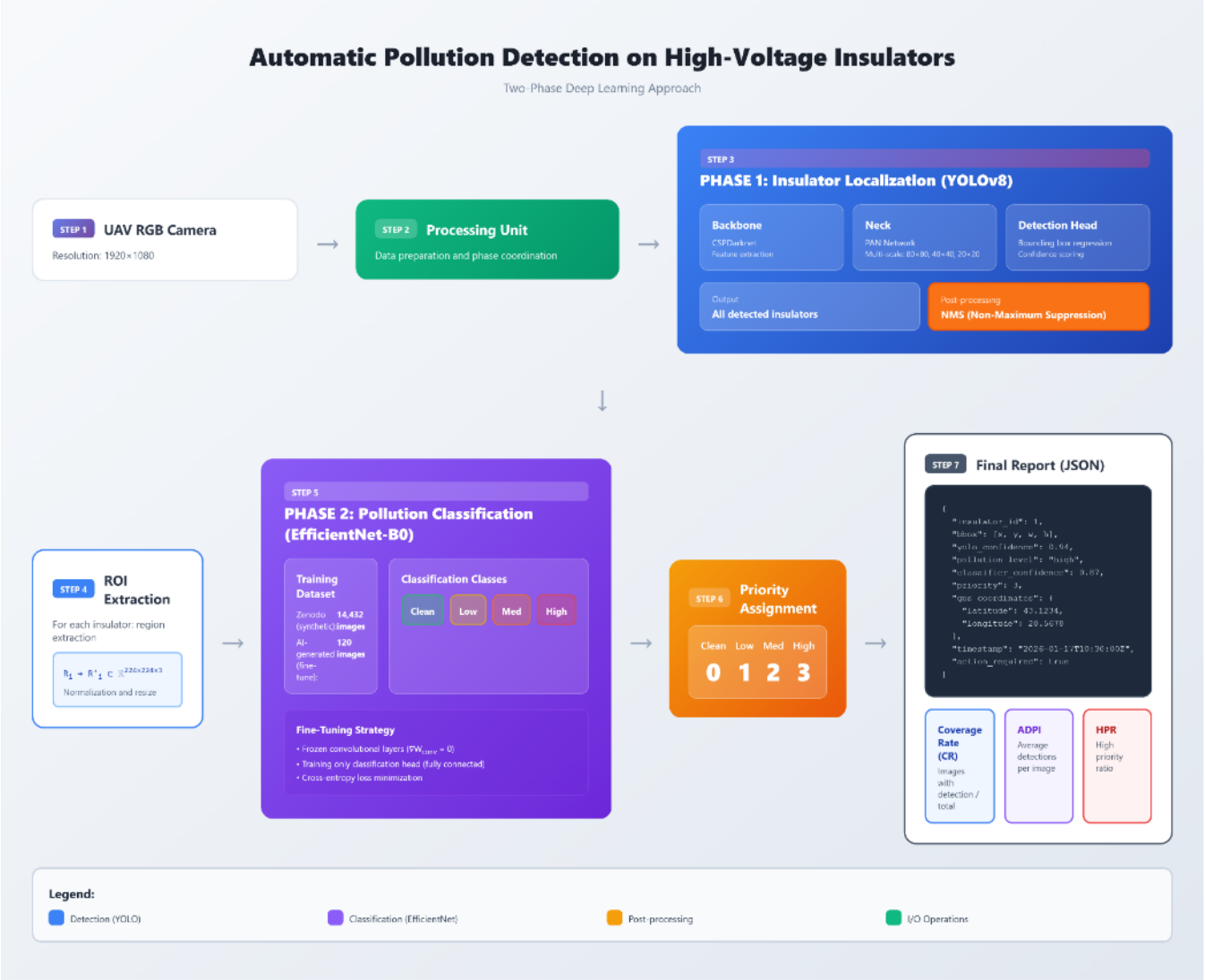


Figure 1. Architecture of the proposed approach.

as soot, salt and excrement.

Given that Zenodo is a synthetically generated database, the problem of real-world images appears, so for this purpose, fine-tuning of the model with real images generated by AI was done [20]. There are 30 images for each class.

The proposed approach is implemented as a processing chain where each stage has a specialized role and an optimized architecture for that specific task. In the first phase, the object detector finds all the insulators in the image and generates their locations as bounding boxes. In the second phase, the trained network simultaneously determines the type and intensity of pollution, which allows the system to generate a prioritized report. Figure 1 shows the architecture of the proposed approach.

In the first phase, insulator detection is done, where the system must find all the insulators in the image regardless of their size, position or angle of the image. In real conditions, the images on which the insulators are located can be small, e.g., 50x50 pixels, or larger than 300x300 pixels. In addition to resolution, they can be hidden by other infrastructure, atmospheric conditions, and the like. Treating object detection as a single regression problem using convolutional networks that simultaneously predicts a location and a given class solves this problem by the YOLO method [21]. The model generates a set of predictions that include bounding box positions, insulators, and the corresponding confidence score. The YOLO model is very fast because it does not treat the image as a sliding window or region proposals, but sees it all at once [21]. By treating object detection as a regression problem, the model directly predicts bounding boxes and classes for the input image. The YOLO model

approximates the function in a single propagation:

$$f_{\text{det}} : I \rightarrow \{(b_i, s_i)\}_{i=1}^{N} \tag{1}$$

where $b_i = (x_i, y_i, w_i, h_i)$ denotes the coordinates of the i-th insulator; $s_i \in [0, 1]$ — reliability; N — number of detected insulators.

In the proposed approach, the loss function used by YOLOv8 is the multi-part loss function, which combines three components:

$$\mathcal{L}_{\text{det}} = \lambda_{\text{box}}\mathcal{L}_{\text{CIOU}} + \lambda_{\text{obj}}\mathcal{L}_{\text{Obj}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} \tag{2}$$

$\mathcal{L}_{\text{CIOU}}$ measures localization error, $\mathcal{L}_{\text{Obj}}$ measures the presence of an object, $\mathcal{L}_{\text{cls}}$ in this paper has a trivial value (one class — insulator). Standard YOLOv8 default values were used in all experiments. Coefficients λ affect optimization and the weighting coefficients $\lambda_{\text{box}}, \lambda_{\text{obj}}$, and λ_{cls} correspond to localization, objectness and classification losses, respectively.

In order to prepare the input image for the phase of the system where it is classified, a region of interest (ROI) extraction is performed for each detected insulator:

$$R_i = I\big|_{b_i} \tag{3}$$

where b_i denotes the bounding box of the i-th detected insulator and $R_i \in \mathbb{R}^{H\times W\times 3}$ represents the RGB region of interest extracted from the input image, which is resized and normalized to $R_i \in \mathbb{R}^{224\times 224\times 3}$.

After the localization of the insulator has been carried out, it is necessary to determine what pollution is present and to what level. The system needs to understand the intensity of pollution, because the priority depends on the intensity. After the initial detection of objects, the YOLO model automatically applies NMS (Non-Maximum Suppression) to remove multiple detections of the same instance of the object.

In the second phase, the type of pollution is classified using EfficientNet-B0 [22], which approximates:

$$f_{\text{type}} : R_i \rightarrow \hat{p}_i \tag{4}$$

where $\hat{p}_i$ is the output probability of the distribution over the pollution class:

$$\hat{p}_i = \left[p_i^{(1)}, p_i^{(2)}, \ldots, p_i^{(|C|)}\right] \tag{5}$$

The prediction is made as:

$$\hat{c}_i = \arg\max_{c\in C} p_i^{(c)} \tag{6}$$

The model is trained by minimizing cross-entropy losses where y_k corresponds to the ground truth notation:

$$L_{\text{type}} = -\sum_{k=1}^{|C|} y_k \log(\hat{p}_k) \tag{7}$$

EfficientNet uses an approach where it scales all the dimensions of the network proportionally, which reduces the number of parameters, and thus we have a model with better performance compared to standard approaches. Each scaled ROI is forwarded to EfficientNet, which classifies the type and injector into one of the pollution classes.

The model in phase two is first trained with the Zenodo dataset and during the training, augmentation is applied over the images. It is trained by minimizing the cross-entropy error where $y_{i,c} \in \{0, 1\}$:

$$L_{CE} = \sum_{i=1}^{N}\sum_{c\in C} y_{i,c} \log \hat{p}_{i,c} \tag{8}$$

Regardless of the fact that the model is synthetic, additional fine-tuning has been done with additional synthetic-real life (generated) 30 images per class where the convolutional layers freeze ($\nabla W_{conv} = 0$) and only the classification "head" is trained:

$$W_{fc}^{(t+1)} = W_{fc}^{(t)} - \eta\nabla L(W_{fc}) \tag{9}$$

The symbol ∇ denotes the gradient of the loss function with respect to the classification head weights. This achieves the adaptation of the model to the domain shift without over-adaptation. Thus, for each detected insulator, a corresponding region (patch) is extracted from the image, which is then prepared for classification using a series of transformations (resize to 224×224 pixels, normalization according to ImageNet statistics). The sample prepared in this way is forwarded to the EfficientNet-B0 model, which has been fine-tuned to classify the four classes: clean, low, med and high. At the output, a prediction of the class is obtained, along with the probability, which allows for an additional assessment of the certainty of the classification.

After phase 2, a template (form) is defined based on the intended class and a report is generated for each one.

$$\pi(\hat{c}_i) = \begin{cases} 0, & \text{clean} \\ 1, & \text{low} \\ 2, & \text{med} \\ 3, & \text{high} \end{cases} \tag{10}$$

The end system can be described as a composition of functions where such architecture allows complete system autonomy, scalability and easy expansion. The classification levels are defined by visual intensity parameters. Unlike IEC standards that rely on conductivity, this RGB-based approach utilizes surface coverage density and opacity as the primary metrics. Low intensity represents sparse contamination with high surface visibility. Medium intensity indicates a semi-opaque layer with significant coverage and High intensity corresponds to a dense, opaque layer shows contamination that completely cover significant portions of the insulator. These visual definitions allow the EfficientNet-B0 model to quantify pollution based on the textural and color features present in the UAV imagery.

This modular approach has several advantages: if a better detector is encountered, only phase 1 can be replaced; Each network is specialized for its task and the system is easier to debug, test and maintain.

In training, standard measures for performance evaluation were used, namely accuracy, precision, recall, F1 and confusion matrix [23]:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

where:

- TP (True Positive) – accurately classified positive examples,
- TN (True Negative) – accurately classified negative examples,
- FP (False Positive) – negative examples classified as positive,
- FN (False Negative) – positive examples classified as negative.

Accuracy (Acc) measures the proportion and proportion of correctly classified examples in relation to the total number.

Precision, recall, and F1 are counted for each class. Precision stands for the accuracy of positive predictions and is calculated as

$$\frac{TP}{TP + FP},$$

recall measures the model's ability to recognize all relevant instances and is defined as

$$\frac{TP}{TP + FN},$$

and F1-score is the harmonic mean of precision and recall:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

A confusion matrix is used to show how the examples of each class are classified as a training appendix:

$$M_{ij} = |\{x : y = i, \hat{y} = j\}| \tag{13}$$

M_{ij} is the number of instances that belong to class III but are classified as class J.

In addition to the standard object measures, the final report consolidates and calculates the following measures.

Coverage Rate (*CR*): The ratio of the number of images containing at least one detection to the total number of images.

$$\text{CR} = \frac{\text{images with detection}}{\text{total number of images}} \tag{14}$$

Also, detections per image (ADPI) which characterizes the behavior of the first phase of the proposed system and calculate how many insulators are detected on average per UAV image. Mean YOLO confidence (MYC) represents the average confidence produced by the detector and provides an overall measure of reliability of the insulator localization stage. Mean phase confidence (MCC) corresponds to the average confidence score generated by the second-phase pollution classifier, indicating the pollution type and severity estimation across all detected insulators. High priority ratio (HPR) supports the practical objective of the proposed approach by measuring the proportion of images that contain at least one insulator classified as highly polluted, enabling automated prioritization of critical cases for maintenance planning.

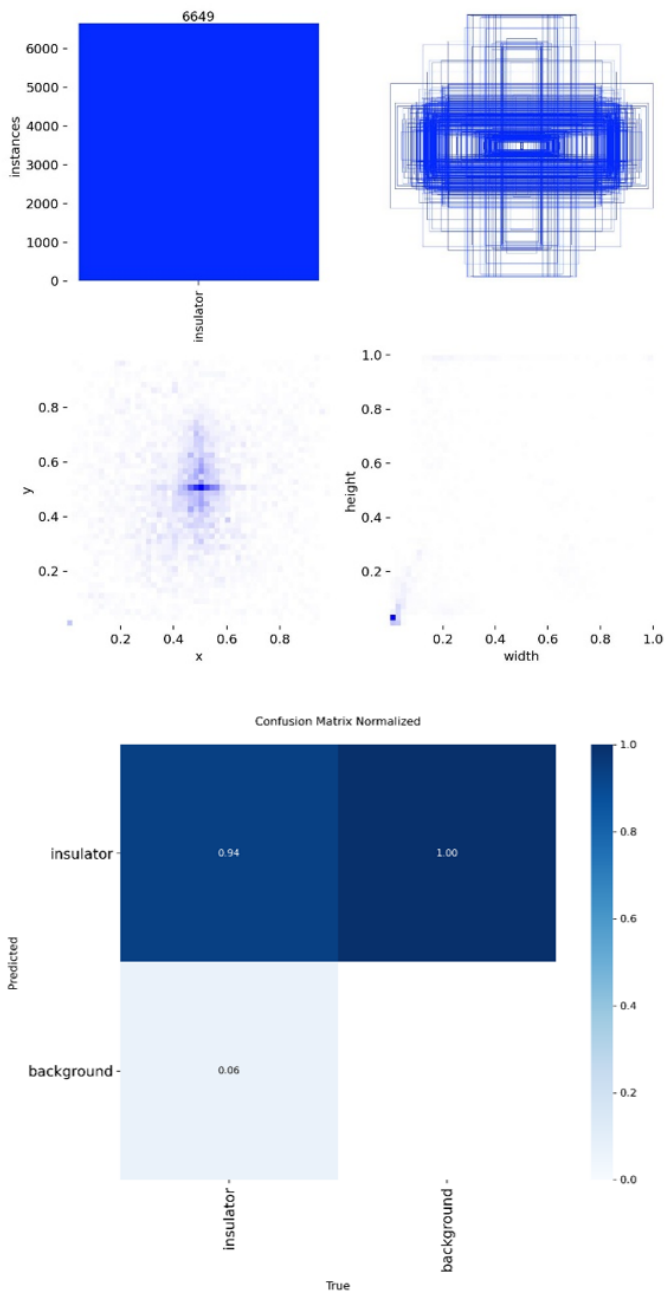


Figure 2. Confusion matrix for phase 1.

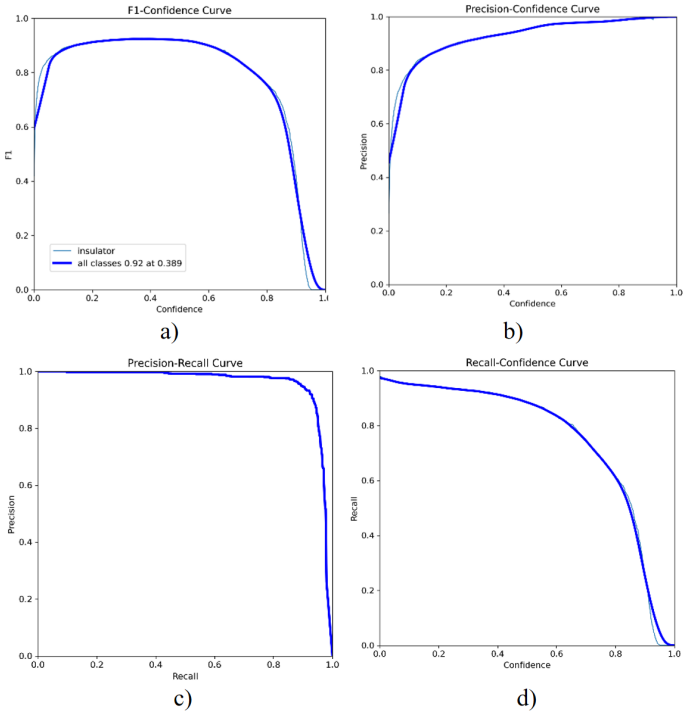


Figure 3. Performance of the model in phase 1 a) F1 confidence curve, b) precision curve, c) precision-recall and d) recall curve.

Table 1. Quantitative performance of the Phase 1 model.

Metrics	SE
mAP@0.5	0.9375 (93.75%)
mAP@0:75	0.7970 (79.70%)
mAP@0.5:0.95	0.7166 (71.66%)
Precision	0.9030 (90.30%)
Recall	0.9039 (90.39%)
F1-score	0.92 (92.00%)

3 Results and discussion

Given that the first phase of the proposed system is in charge of detection and localization of insulators, the training set from the MPID database contains 5019 images of the insulator, where the training set is divided by about 80%, while for validation and test it is divided by 10%. Additional expansion is done because during training using augmentation. Table 1 shows the results of the model's performance.

From Table 1, it can be seen that an IoU (Intersection over Union) of 0.5 for mAP shows a value of 93.75%, which is the reliability of the insulator finding data. A 90.30% precision shows that the insulator detection is accurate, while a 90.39% recall shows that the model finds insulators at 90.39% of the images. That the model is balanced is shown by the F1 parameter that shows that the model does not have a pronounced bias towards false positive or false negative errors. A more stringent metric is provided by the mAP parameter 0.5:0.95, which gives an average of precision at different overlap thresholds. Figure 2 shows a confusion matrix where it can be seen that the model classifies 94% of the insulators (True Positive) while only 6% (False Negative) is misclassified. If the analysis of the insulator were based on video, this 6% figure indicates that the probability that the insulator will remain undetected is low.

Figure 3 illustrates the performance of the model, i.e., the curve of the dependence of the metrics on the confidence threshold. Figure 3 (a) shows an F1

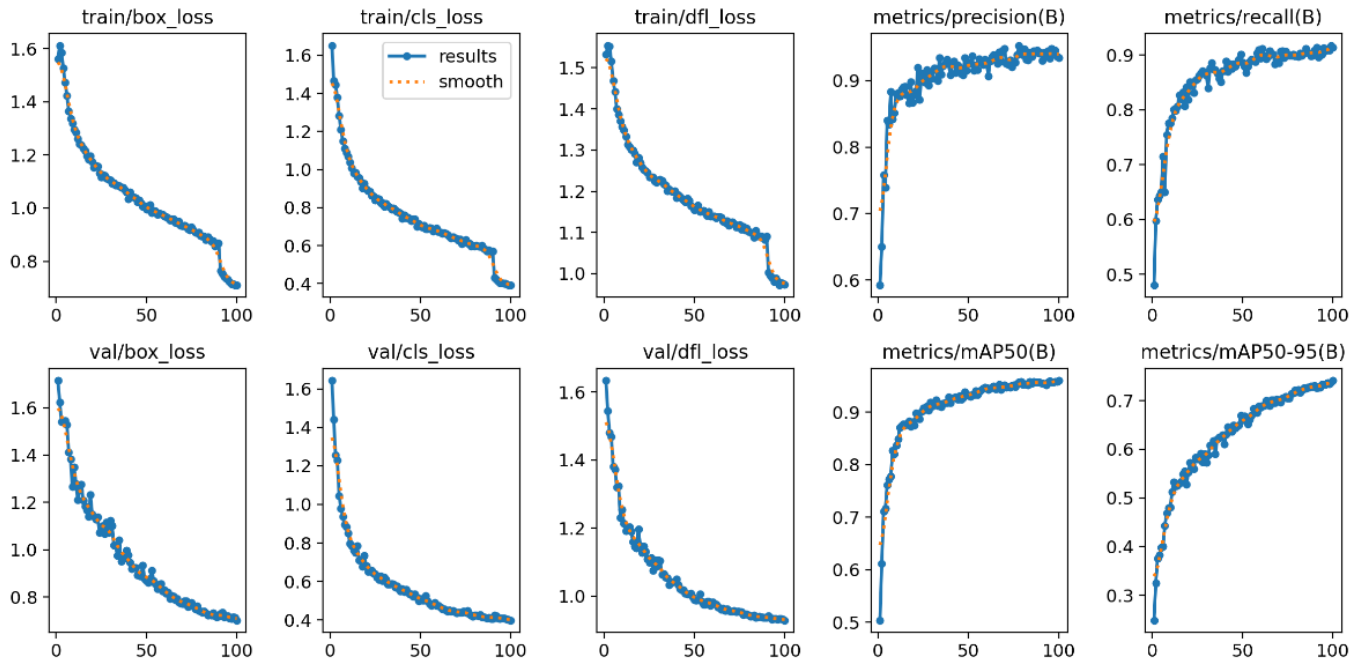


Figure 4. Loss and metric dynamics during Phase 1 training.

curve with a maximum value of 92% at a relatively low threshold of 0.389. This indicates that the model achieves an optimal balance between precision and responsiveness. In Figure 3 (b) you can see the precision confidence curve where you can see that at lower thresholds the accuracy remains around 85%, which is a suitable parameter for practical application. The Precision-Recall curve is shown in Figure 3 (c), while Figure 3 (d) shows the Recall-Confidence curve. It can be seen that the curve retains high accuracy values of over 95% even at low recall values of up to 0.8. These parameters indicate that the robustness of the model is high and that the possibility of detecting insulators is high even when the possibilities for prediction are low.

Figure 4 shows the dynamics during training, i.e., the loss curves for train and validation. Based on the loss functions, it can be seen that the model is not overfitted based on the minimal difference between train and validation loss. That the training is stable, showing a steady decline in functions without significant oscillations. When comparing the results with Table 1, it can be concluded that growth during the first 60-70 epochs stabilize thereafter.

Figure 5 shows an example of visual detection with a safety parameter. Figure 5 (a) shows an example of when the images are distance view, while (b) shows an example of when the images of the insulator are close. That is, it shows that the model detects insulators from different angles with a score of 0.8 to 0.9, however, an error can occur when the object is smaller than 50x50 pixels, as shown in Figure 5 (b) where the confidence score is 0.37. The stability of the prediction shows a confidence score of 0.9 in most images where the insulator is clearly shown.

The second phase is the classification using the Zenodo database. Figure 6 (a) shows the Accuracy curve, while b) shows the loss function, i.e., the loss curve of the function. As can be seen in Figure 6, the model shows a stable learning trend and validation that increases with each epoch. The ultimate validation accuracy is 96.5%, which is an indication that it is doing a good generalization, but potentially overfitting due to the uniformity of the data.

Figure 7 shows a confusion matrix where it can be seen that most classes are classified correctly, but also that errors occur between neighboring classes, especially between medium and high intensity.

Table 2 provides a classification report with performance metrics. It can be seen that the average values of Precision, Recall and F1-score are 0.93, 0.95 and 0.94 respectively. The best results are achieved by classes such as clean, excrement_med and soot_med, with F1 scores close to 0.96, which shows that the model very accurately recognizes and differentiates clearly defined and visually consistent examples.

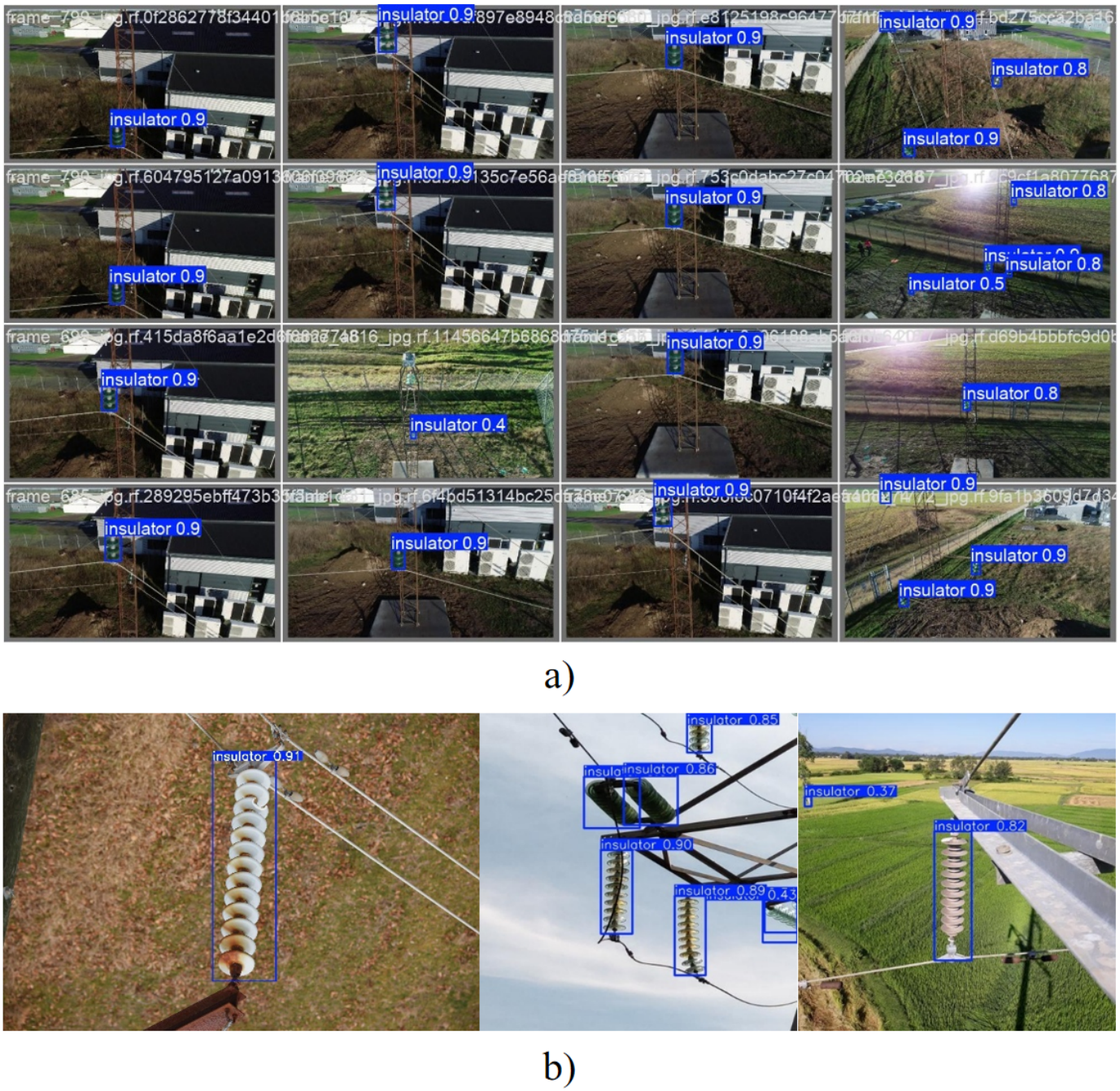


Figure 5. Example of visual detection of insulators a) distant and b) closer shots.

Figure 8 illustrates the visual diversity of the synthetic dataset used for initial training, showing representative samples for clean materials (glass, porcelain, polymer) and simulated pollution types (soot, salt, excrement). Figure 8 show example of insulators: a) clean glass, b) clean porcelain, c) clean polymer, d) soot, e) salt, f) excrement.

In the paper [24] they used generated synthetic insulators by training a model using VggNet and GoogleNet where they aged the accuracy of the classification of 98.99%, which also confirms the overestimation of the results and should only be seen as a proof of concept.

In order to overcome this problem in the proposed system, fine-tuning of this base with generated real images has been additionally done. During this process, the convolutional layers are frozen, and only the classification head of the network is trained. As shown in Figure 9 (a), the accuracy shows that the model reaches saturation, i.e. the maximum at one point in time, while the validation shows oscillatory growth and stabilizes only after 20 epochs. Such results indicate that the model has the ability to adapt. This phenomenon in learning transfer on smaller dataset

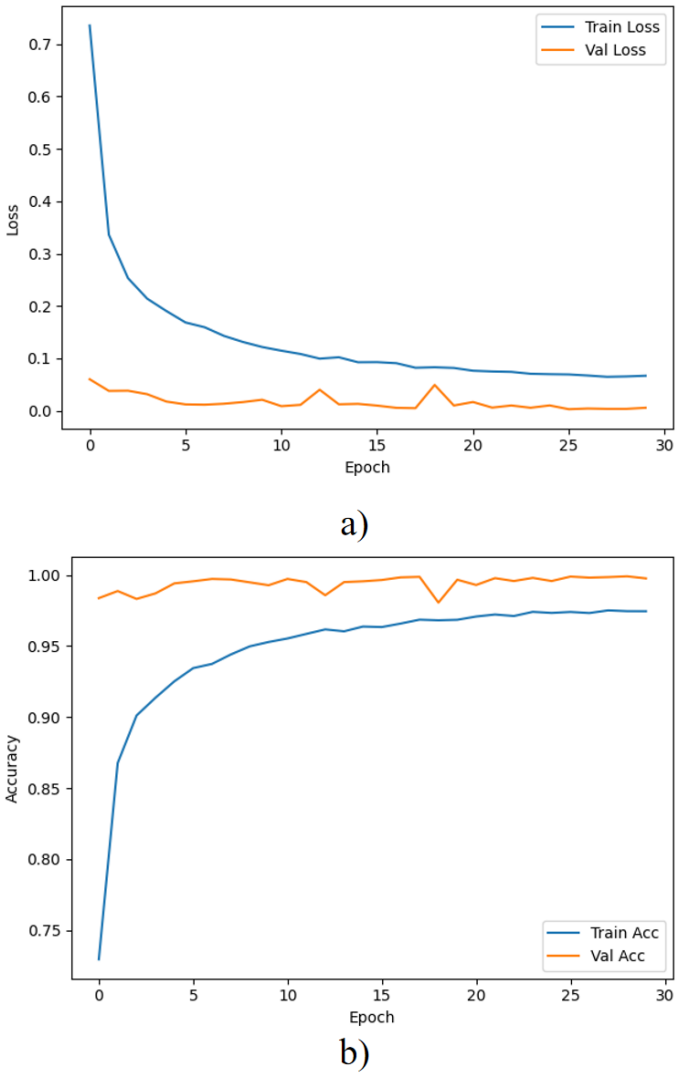


Figure 6. Performance of the model in phase 2 a) loss curve, b) accuracy curve.

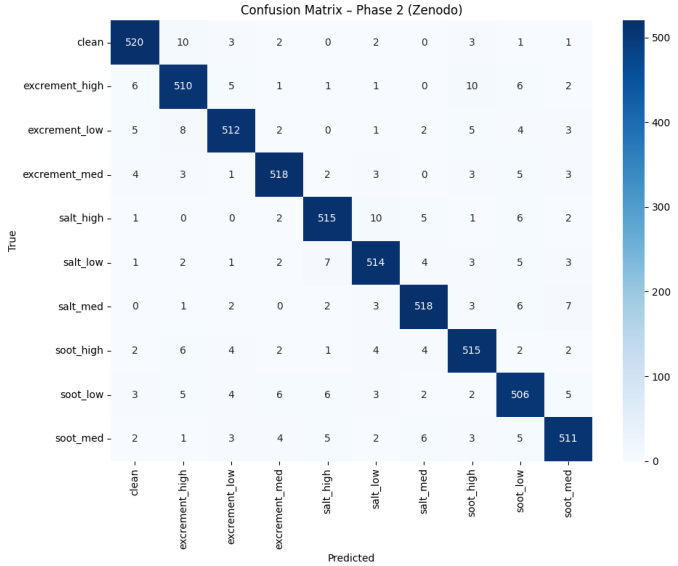


Figure 7. Confusion matrix for Phase 2 (Synthetic Zenodo dataset).

Table 2. Quantitative performance of the Phase 1 model.

Metrics	Precision	Recall	F1-score	Support
clean	0.95	0.96	0.96	542
excrement_high	0.91	0.94	0.92	542
excrement_low	0.92	0.94	0.93	542
excrement_med	0.96	0.96	0.96	542
salt_high	0.91	0.95	0.93	542
salt_low	0.93	0.95	0.94	542
salt_med	0.94	0.95	0.94	542
soot_high	0.94	0.95	0.94	542
soot_low	0.92	0.93	0.93	542
soot_med	0.95	0.94	0.95	542

is also known as domain adaptation. Oscillations on the validation curve show the nature of real samples that contain noise and variations, so the model is more difficult to generate compared to aesthetic data.

Figure 10 (a) shows a confusion matrix where you can see a detailed view of the classification performance. The results show that the matrix has a diagonal where classification errors in close classes such as medium and high levels of pollution are also clearly visible. To avoid further problems, the model has the ability to avoid classification by level of pollution and after training in only 4 classes. The confusion matrix for 4 classes and fine tuning is shown in Figure 10 (b). Quantitative metrics by class are shown in Figure 11 for the Precision, Recall, and F1 objective measures. Classes such as soot_med and clean achieve F1 above 0.85, which makes the approach very reliable for clear insulators. The very fact that it classifies clean insulators proves that the system has a real and practical application, especially if the dirt is cleaned in the same way, i.e. it does not require special fluids for different materials. Class excrement_high shows a slightly lower response, which can be explained by the high variability of visual appearance. However, this is solvable by the fact that the model can reduce the number of classes to 4. The average accuracy on real data is about 88% percent when using multiclass, so the random guess probability is only 10%. Due to errors in the classification of the level of soiling (low/med/high), a simplified model with only 4 classes (clean, excrement, salt, soot) was tested and simplified. This version achieves better generalization because it eliminates confusion between similar levels of pollution.

After the final model was trained, a set of 100 images simulating real conditions for insulator inspection was tested to evaluate performance. To validate the model's reliability, a test set of 100 images was picked to simulate a wide spectrum of field conditions.

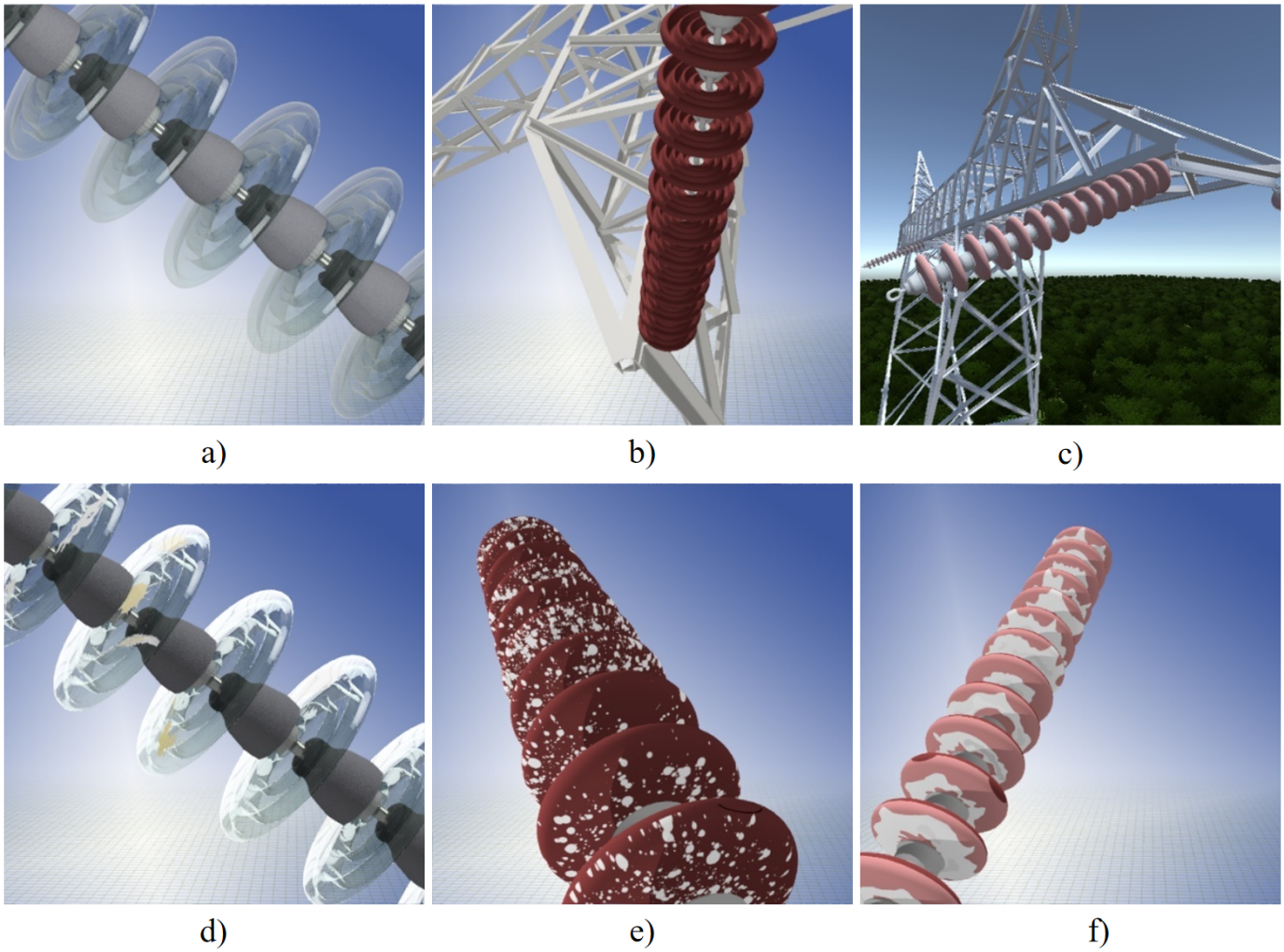


Figure 8. Example of a synthetic dataset for materials a) clean glass, b) clean porcelain, c) pure polymer, d) soot insulator, e) salt and f) excrement insulator.

This set was designed to be diverse, covering three insulator materials (porcelain, glass, and polymer) and varying environmental backgrounds (vegetation, sky, and transmission towers), as well as level of pollution. Due to this lack of large-scale authentic data, we manually curated these 100 images to act as a high-diversity validation set. Figure 12 shows the confusion matrix when detecting 100 images. In Figure 12, you can see where the approach was most often wrong, and that is when it comes to the transition from a medium to a higher level of pollution. The observed misclassification of blurred clean insulators as low-intensity pollution suggests a dependency on high-frequency features. Figure 13 shows the detection results. The distribution by class shows that there were 112 detections in total, while the other classes were more variously represented. The most common pollution classes are salt_med, soot_med and excrement_high which indicates a high sensitivity of the model to medium and high soiling. The CR data shows 84%, which means that at least one insulator has been successfully detected and classified, while the remaining 16% without detection can be attributed to poor image conditions. HRP of 24% shows high priority, that is, urgent "washing". This is a direct demonstration of the cost-effectiveness of such a system, as it allows operators to focus resources on only a quarter of the network or pillars that are critical.

The system's conservative detection behavior is evidenced by detecting insulators is caused by an ADPI that is close to 1 per image. Given that YOLO applies NMS, this data indicates that no problem of breakage has been observed. The MYL probability shows that the model's predictions are highly reliable in most cases. This is particularly important because YOLO functions in the system as the initial detection phase, the results of which are transferred to the next stages for further processing. A value over 0.7 confirms that the detections were stable and with high confidence, and this value is consistent with the

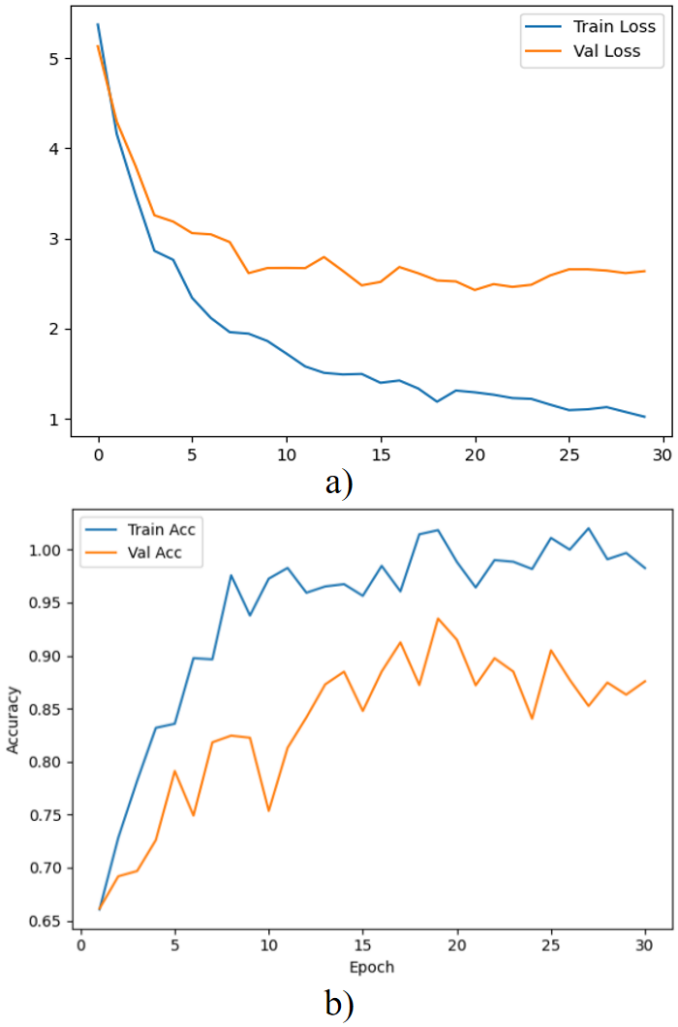


Figure 9. Performance of the model in phase 2 of fine-tuning a) loss curve, b) accuracy curve.

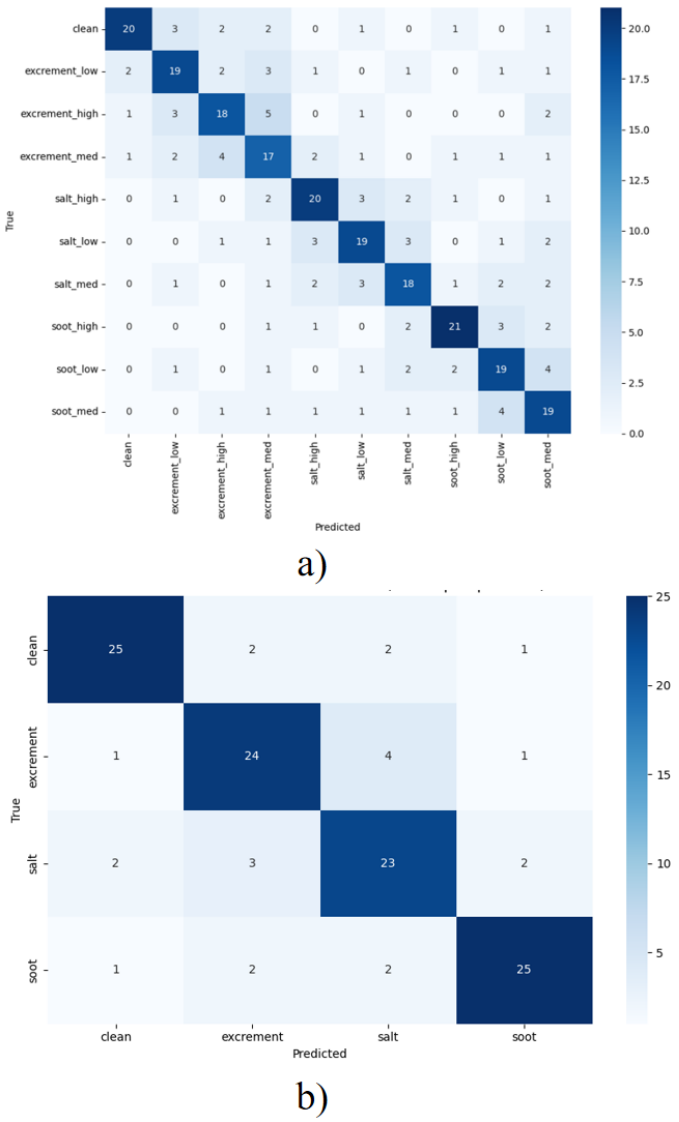


Figure 10. Confusion matrix for fine tuning in phase 2 for a) 10 classes and b) 4 classes.

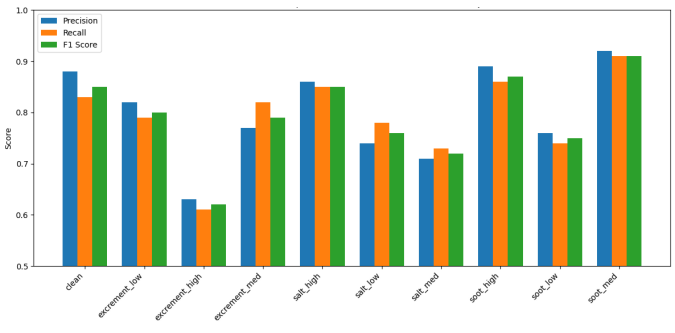


Figure 11. Quantitative performance by class (Precision, Recall, F1-score) for the 10-class model.

trained classification thresholds. Since the system is fine-tuning, an MCC value of 0.682 confirms the consistency of the system, i.e., that the system can detect more complex insulator pollution.

Figure 14 shows a visual example of detection on different types of insulators in different conditions and angles, as well as the level of pollution. In Figure 14 (a) we can see the detection when the insulator is completely clean and therefore, we have high values, which proves that the model is reliable. Figure 14 (b) is an example when the staining of excrement is of the medium level, but since the part of the staining is similar to that of salt, when the priority class with 10 classifications was used, then there is a value of 0.88, which showed on the confusion matrix, but also an MCC value of 0.68 at the level of the set. In other words, if there is no need for the level of pollution, the system will become much more reliable. Figure 14 (c) shows multiple detection, i.e., when there are insulators in the image that are the same or similar distance, and it can be seen that it successfully detected soot/dust pollution, which it categorized as high priority and medium, which is confirmed by the ADPI value of 1.28. Also, soot_high was detected with a score of 0.83,

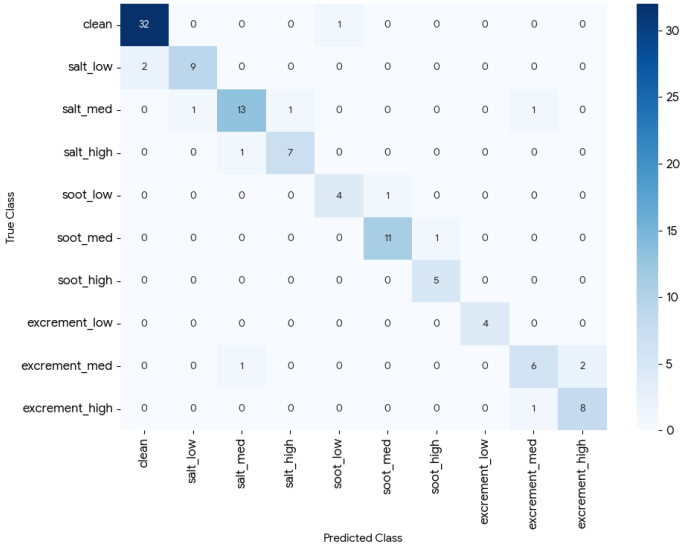


Figure 12. Confusion matrix for 100 tested images.

```
Number of images: 100
Total detections: 112
Images with at least one detection: 84
Coverage Rate (CR) = images_with_detection / total_images = 0.8400
Average Detections Per Image (ADPI) = 1.1200
Mean YOLO Confidence (MYC) = 0.7150
Mean Phase 3 Confidence (MCC) = 0.6820
High Priority Ratio (HPR) = images_with_high_priority / total_images = 0.2400

Class distribution (detections only):
    clean: 34
    salt_low: 10
    salt_med: 15
    salt_high: 8
    soot_low: 5
    soot_med: 12
    soot_high: 6
    excrement_low: 4
    excrement_med: 8
    excrement_high: 10
```

Figure 13. Results of the detection of the proposed approach.

while soot_med was detected with 0.75. This gradation in scores (where harder pollution is easier to detect) is a desirable feature for security systems. Figure 14 (d) is an example of when the soiling is salt and when the glass insulator. Reliability values are high for both classes at 90%, which shows what directly contributes to HPR (metric of 0.24), and is a key trigger for sending the drone for washing. However, when the images are of lower resolution, or there is a degradation such as blur (focus), the algorithm has an error factor, so in Figure 14 (e) it is shown when the approach correctly detects a clean insulator with a high score, however an insulator that is out of focus (blurred) recognizes it as a high level of dust. However, the system gave a low safety rating for last case.

The proposed approach achieves the highest precision, although the difference with VGG16 is relatively small (0.6%), the key advantage of EfficientNet-B0 lies in the drastically lower computational complexity. For safety-critical systems such as high-voltage lines, False Negative is more dangerous than False Positive.

Table 3. Comparison of the proposed approach with other architectures.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score
ResNet18	86.1	85.4	86.2	0.86
MobileNetV2	83.5	82.7	84.3	0.83
Vgg16	87.4	86.7	87.8	0.85
Proposed 10 class	**88.0**	**87.6**	**88.7**	**0.88**
Proposed 4 class	**91.0**	**90.5**	**91.3**	**0.91**

The proposed model with the highest Recall (88.7%) minimizes the risk of missed critical insulators. For example, in real applications, MobileNetV2 has a larger difference of +1.6%, which means too many false positives and therefore costs increase, while ResNet18 has a smaller difference of +0.8%, which means more false negatives. The security risk is even though Resnet is trained on 11.7M and only 2.2× more than MobileNetV2, but 26× less than VGG16. Although the 10-class model provides a detailed insight into the degree of pollution, analysis of the confusion matrix showed that most of the errors come from mixing different levels of intensity within the same type of pollution. For practical maintenance purposes, information about the type of pollution is often more important than a precise gradation of the layer thickness. As shown in Table 3, the results for the 4 classes show metrics of Accuracy 91.0%, Precision 90.5%, Recall 91.3% and F1-score 0.91 confirming that the system maintains a high level of accuracy.

4 Conclusion

This paper proposes a two-phase approach for automatic detection and classification of pollution on high-voltage insulators based on UAV images. First, a model was trained using the YOLO algorithm for precise localization of insulators, based on an MPID database of over 5000 insulators. After that, three types of pollution were analyzed, namely salt, soot and excrement. These pollutions were analyzed in the Zenodo database, which contains over 14,000 examples. The training was done using EfficientNet-B0 which did the classification of pollution and intensity (small, medium and high). Depending on the intensity of, a priority is created, which is achieved using the HPR parameter. In order to achieve greater accuracy, the model has been additionally trained with fine-tuning for each type and intensity of pollution.

First, in phase 1, the YOLO detector was used to detect the insulators, where an average accuracy of mAP@0.5 of 93.75% was achieved, while the F1 parameter was 92%, confirming the high reliability in the localization of the insulators, regardless of the complexity of

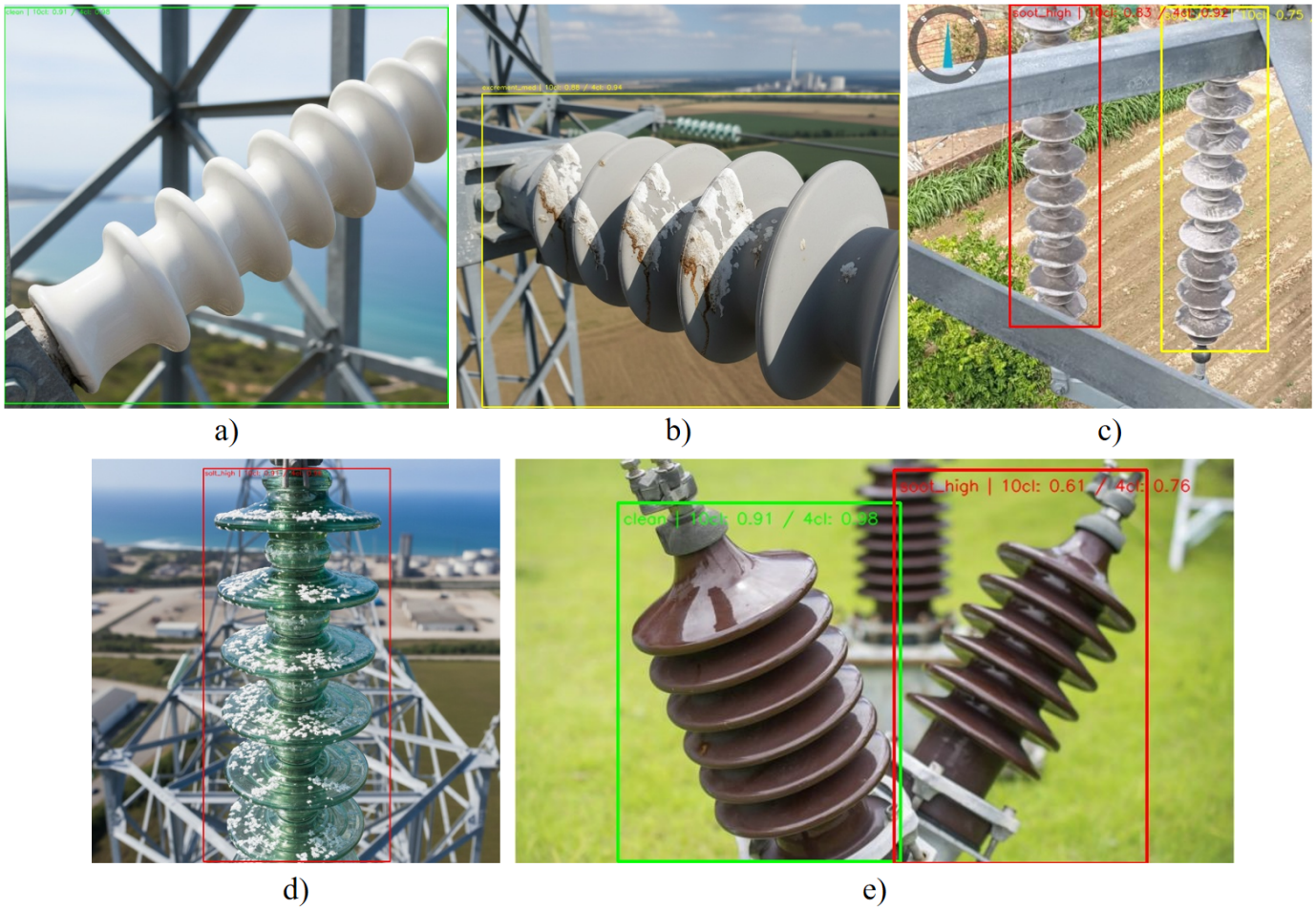


Figure 14. Visual example of detection with confidence values using 4 (4cl) and 10 (10cl) classes a) ideal clean insulator detection, b) medium level bird droppings detection, c) multiple detection, d) high priority detection and e) detection error analysis.

the background. This also indicates that the model achieves an optimal balance between accuracy and responsiveness. The confusion matrix and loss curves demonstrate the stability of detection and localization of the insulator, as well as model training.

In the second phase, EfficientNet was used to classify the type of pollution of small, medium and high intensity. In addition, fine-tuning has been done with the generated images to achieve greater accuracy. A classification accuracy of 88.0% has been achieved for a detailed analysis of 10 levels of pollution. In addition, an optimized model with 4 classes was developed that achieved an accuracy of 91.0%, confirming the flexibility of the system to adapt to different operational requirements. The confusion matrix and loss curves show that additional fine-tuning has further brought stability and precision to the system.

Metrics such as F1, precision, recall were analyzed in the training phase but also during the execution of the entire approach. The final report also provides the average number of detections per image, the mean YOLO confidence value for all detections, the mean confidence value of the pollution classifier.

The proposed model achieves the best balance of performance, with a Recall rate of 88.7%, minimizing the risk of missing critical defects (False Negatives), making it a safer choice compared to the ResNet18 and MobileNetV2 architectures.

With a CR of 84% and an HPR of 24%, the approach demonstrated the ability to efficiently filter large amounts of data and accurately identify priority points that require immediate intervention. The approach itself has the possibility of modularity, i.e., fine-tuning and additional extensions of the system, especially by adding new real images, which would increase the accuracy of detection and classification. Also, a potential real-world application would enable complete automation of the process where one drone would collect and analyze data, while the other would receive a prioritized report on the basis of which it

would perform cleaning tasks.

The current approach relies solely on visual (RGB) data. Future research will focus on the integration of thermal imaging (IR) and ultraviolet (UV) cameras. Such an analysis opens the possibility for additional research, especially in terms of pollution classifications by analyzing the structure of the insulator, but also by segmenting the pollution and switching to another spectral domain.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Mussina, D., Irmanova, A., Jamwal, P., & Bagheri, M. (2020). Multi-modal data fusion using deep neural network for condition monitoring of high voltage insulator. *IEEE Access, 8*, 184486-184496. [CrossRef]

[2] Chatzargyros, G., Papakonstantinou, A., Kotoula, V., Stimoniaris, D., & Tsiamitros, D. (2024). UAV Inspections of Power Transmission Networks with AI Technology: A Case Study of Lesvos Island in Greece. *Energies, 17*(14), 3518. [CrossRef]

[3] Sampedro, C., Rodriguez-Vazquez, J., Rodriguez-Ramos, A., Carrio, A., & Campoy, P. (2019). Deep Learning-Based System for Automatic Recognition and Diagnosis of Electrical Insulator Strings. *IEEE Access, 7*, 101283-101308. [CrossRef]

[4] Santos, T., Cunha, T., Dias, A., Moreira, A. P., & Almeida, J. (2024). UAV Visual and Thermographic Power Line Detection Using Deep Learning. *Sensors, 24*(17), 5678. [CrossRef]

[5] Tingyu, W., Xia, S., Jiaxing, L., & Yue, Z. (2024). A Deep Learning Based Detection Method for Insulator Defects in High Voltage Transmission Lines. *International Journal of Advanced Computer Science & Applications, 15*(10), 378-385. [CrossRef]

[6] Wang, T., Zhai, Y., Li, Y., Wang, W., Ye, G., & Jin, S. (2024). Insulator Defect Detection Based on ML-YOLOv5 Algorithm. *Sensors, 24*(1), 204. [CrossRef]

[7] Fahim, F., & Hasan, M. S. (2024). Enhancing the reliability of power grids: A YOLO based approach for insulator defect detection. *e-Prime - Advances in Electrical Engineering, Electronics and Energy, 9*, 100663. [CrossRef]

[8] Dang, Q., Shang, W., Luo, F., Lu, P., Lin, G., & Gui, X. (2024, August). Insulator Defect Detection Technology Based on Deep Learning. In *2024 4th International Conference on Energy Engineering and Power Systems (EEPS)* (pp. 941-947). IEEE. [CrossRef]

[9] Rong, S., He, L., Atici, S. F., & Cetin, A. E. (2025). Advanced YOLO-based Real-time Power Line Detection for Vegetation Management. *IEEE Transactions on Power Delivery, 40*(4), 2142-2153. [CrossRef]

[10] Gonçalves, R. S., De Oliveira, M., Rocioli, M., Souza, F., Gallo, C., Sudbrack, D., Trautmann, P., Clasen, B., & Homma, R. (2023). Drone–Robot to Clean Power Line Insulators. *Sensors, 23*(12), 5529. [CrossRef]

[11] Ferraz, H., Gonçalves, R. S., Moura, B. B., Sudbrack, D. E. T., Trautmann, P. V., Clasen, B. C., Homma, R. Z., & Bianchi, R. A. C. (2024). Automated classification of electrical network high-voltage tower insulator cleanliness using deep neural networks. *International Journal of Intelligent Robotics and Applications, 9*(3), 818-832. [CrossRef]

[12] Ergün, E. (2025). Artificial intelligence approaches for accurate assessment of insulator cleanliness in high-voltage electrical systems. *Electrical Engineering, 107*(3), 2969-2982. [CrossRef]

[13] Ferraz, H., Gonçalves, R., Moura, B., Sudbrack, D., Trautmann, P., Clasen, B., ... & Bianchi, R. A. (2024). Synthetic images datasets of clean and dirty string insulators used in high-voltage power lines. *Journal of the Brazilian Society of Mechanical Sciences and Engineering, 46*(11), 636. [CrossRef]

[14] Savadkoohi, E. M., Mirzaie, M., Seyyedbarzegar, S., Mohammadi, M., Khodsuz, M., Pashakolae, M. G., & Ghadikolaei, M. B. (2020). Experimental investigation on composite insulators AC flashover performance with fan-shaped non-uniform pollution under electro-thermal stress. *International Journal of Electrical Power & Energy Systems, 121*, 106142. [CrossRef]

[15] Luque-Vega, L. F., Castillo-Toledo, B., Loukianov, A., & Gonzalez-Jimenez, L. E. (2014). Power line inspection via an unmanned aerial system based on the quadrotor helicopter. In *Proceedings of the 2014 17th IEEE Mediterranean Electrotechnical Conference (MELECON)* (pp. 393-397). [CrossRef]

[16] Matikainen, L., Lehtomäki, M., Ahokas, E., Hyyppä, J., Karjalainen, M., Jaakkola, A., ... & Heinonen, T. (2016).

Remote sensing methods for power line corridor surveys. *ISPRS Journal of Photogrammetry and Remote sensing, 119*, 10-31. [CrossRef]

[17] Roboflow Universe. (2026). MPID dataset [Data set]. Roboflow. Retrieved from https://universe.roboflow.com/rt-at8om/mpid-twm1g

[18] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics (Version 8) [Computer software]. Retrieved from https://github.com/ultralytics/ultralytics

[19] Bianchi, R. A., Ferraz, H. F., Gonçalves, R. S., Moura, B., Sudbrack, D. E., Merini, A., ... & Homma, R. Z. (2024). A synthetic high-voltage power line insulator images dataset. *Data in Brief, 55*, 110688. [CrossRef]

[20] Gemini Team, Google. (2024). Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

[21] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) (pp. 779-788). [CrossRef]

[22] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

[23] Maksimović, V., Petrović, M., Savić, D., Jakšić, B., & Spalević, P. (2021). New Approach of Estimating Edge Detection Threshold and Application of Adaptive Detector Depending on Image Complexity. *Optik, 238*, 166476. [CrossRef]

[24] Liu, Y., Liu, D., Huang, X., & Li, C. (2023). Insulator defect detection with deep learning: A survey. *IET Generation, Transmission & Distribution, 17*(16), 3541–3558. [CrossRef]



Vladimir Maksimović was born in Serbia, in 1992. He received his Ph.D. degree in Electrical and Computer Engineering in 2022 from the Faculty of Technical Sciences, University of Priština in Kosovska Mitrovica. His main research interests include image processing, telecommunication, television and multimedia systems. He participated in several international Erasmus+ and national projects. He is an Assistant Professor at the Faculty of Technical Sciences University of Priština in Kosovska Mitrovica at the department Electronics and Telecommunications. (Email: vladimir.maksimovic@pr.ac.rs)



Branimir Jakšić was born in Serbia, in 1984. He received his Ph.D. degree in Electrical and Computer Engineering in 2015 from the Faculty of Electronics, University of Niš. His main research interests include telecommunications, information and communication technologies, and information systems. He participated in several international Erasmus+ and national projects. He is a Full Professor at the Faculty of Technical Sciences, University of Priština in Kosovska Mitrovica, in the field of Telecommunications and Information Systems. (Email: branimir.jaksic@pr.ac.rs)